

INTERDISCIPLINARY APPLIED MATHEMATICS

MATHEMATICAL BIOLOGY

Mathematical Population Genetics

I. Theoretical Introduction

Warren J. Ewens

Second Edition



Springer

Interdisciplinary Applied Mathematics

Volumes published are listed at the end of the book.

Interdisciplinary Applied Mathematics

Volume 27

Editors

S.S. Antman **J.E. Marsden**

L. Sirovich **S. Wiggins**

Geophysics and Planetary Sciences

Imaging, Vision, and Graphics

Mathematical Biology

L. Glass, **J.D. Murray**

Mechanics and Materials

R.V. Kohn

Systems and Control

S.S. Sastry, **P.S. Krishnaprasad**

Problems in engineering, computational science, and the physical and biological sciences are using increasingly sophisticated mathematical techniques. Thus, the bridge between the mathematical sciences and other disciplines is heavily traveled. The correspondingly increased dialog between the disciplines has led to the establishment of the series: *Interdisciplinary Applied Mathematics*.

The purpose of this series is to meet the current and future needs for the interaction between various science and technology areas on the one hand and mathematics on the other. This is done, firstly, by encouraging the ways that that mathematics may be applied in traditional areas, as well as point towards new and innovative areas of applications; and, secondly, by encouraging other scientific disciplines to engage in a dialog with mathematicians outlining their problems to both access new methods and suggest innovative developments within mathematics itself.

The series will consist of monographs and high-level texts from researchers working on the interplay between mathematics and other fields of science and technology.

Warren J. Ewens

Mathematical Population Genetics

I. Theoretical Introduction

Second Edition



Springer

Warren J. Ewens
Department of Biology
University of Pennsylvania
Philadelphia, PA 19104
USA
wewens@sas.upenn.edu

Editors

S.S. Antman
Department of Mathematics
and
Institute for Physical Science and Technology
University of Maryland
College Park, MD 20742
USA
ssa@math.umd.edu

J.E. Marsden
Control and Dynamical Systems
Mail Code 107-81
California Institute of Technology
Pasadena, CA 91125
USA
marsden@cds.caltech.edu

L. Sirovich
Division of Applied Mathematics
Brown University
Providence, RI 02912
USA
chico@camelot.mssm.edu

S. Wiggins
School of Mathematics
University of Bristol
Bristol BS8 1TW
UK
s.wiggins@bris.ac.uk

Mathematics Subject Classification (2000): 92-02, 92D10, 60J70

ISBN 978-1-4419-1898-7 ISBN 978-0-387-21822-9 (eBook)
DOI 10.1007/978-0-387-21822-9

© 2004 Springer Science+Business Media New York

Originally published by Springer-Verlag New York, Inc in 2004

Softcover reprint of the hardcover 2nd edition 2004

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Science+Business Media, LLC), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

For Bronnie and Dave

Preface

This is the first of a planned two-volume sequence discussing mathematical aspects of population genetics theory, with an emphasis on the evolutionary theory. This first volume is intended to discuss the more introductory aspects of the theory, with the second volume taking up more advanced and more recent aspects. Because of this, this first volume draws heavily on the first (1979) edition of this book, since the material in that edition may now be taken, to a large extent, as introductory to the contemporary theory. A second reason for drawing heavily on the 1979 edition is that many present-day students have asked for access to earlier material not now easily available. It is indeed remarkable how many results well-known in the 1970's, and appearing in the literature of the time, are rediscovered in the modern literature.

On the other hand, the subject has greatly expanded in scope and depth over the last twenty-five years. Many topics have been introduced during that time, or developed well beyond the level reached in the 1970's. No doubt the most important of these is the development of the theory of molecular population genetics. Introductory aspects of this theory molecular population genetics are taken up in the later chapters of this volume, but a far more extensive description of the molecular theory will be given in Volume II. As one example of this, the theory behind currently active haplotype mapping projects will be discussed. To this extent, Volume II will be largely data-based. It will thus also form connections between evolutionary genetics and currently active areas of problems of human genetics and bioinformatics. On the other hand, developments of the evolutionary theory itself will be considered also, taking up evolutionary questions relat-

ing to many species rather than evolutionary behavior within one species. Other evolutionary topics such as the game theoretical approach to evolution, the analysis of gene-environment interactions, gene conversion and the extended development of the concept of inclusive fitness, will also be discussed in Volume II.

Despite the emphasis on evolutionary population genetics in this volume, some material concerning human genetics, in particular those parts of the theory that are best discussed in evolutionary terms, has been included. One of the more pleasing developments over the last two decades has been a convergence of work in mathematical human genetics and mathematical evolutionary genetics, areas which in 1979 had very little overlap. A manifestation of this convergence is the recent volume on mathematical population genetics and human evolution by Donnelly and Tavaré (1997).

The aim of the 1979 edition, namely to focus on the purely mathematical aspects of population genetics theory, is retained in this book, even though it is recognized that this provides a narrow and distorted view on the subject of population genetics, and indeed of theoretical population genetics, as a whole. Thus, as in 1979, the book is intended as a complement to broader and more balanced accounts of population genetics generally. There are now many excellent books available devoted to this broader field, but these often do not attempt any depth of mathematical treatment, so that there is still a place for a narrowly focussed mathematical treatment.

Apart from this, there are now several excellent books on specific aspects of population genetics theory. Of these it is appropriate to mention that by Lynch and Walsh (1998) on quantitative traits, a topic not covered in this volume, Epperson (2003) on geographical genetics and books by Christiansen (2000) and Bürger (2000) on multilocus theory. All these books carry the theory beyond the introductory level aimed at in this volume.

One aim of the 1979 volume, not explicitly stated, was to induce mathematically-trained workers to enter the population genetics field. This aim is continued in this volume, and the mathematical beauty of many of the formulas in the molecular genetics chapters of this book should help in this endeavor.

The molecular nature of current data implies that statistical methods are used far more frequently than was the case in 1979, with the molecular data being used to test various hypotheses about the evolutionary process. For the statistical analyses discussed I have adopted the standard convention of employing upper-case letters to denote random variables and the corresponding lower-case letters to denote their observed values, except in cases where this seemed pedantic. This has also sometimes implied replacing Greek letters sometimes used in the literature for random variables by Roman letters. Probability distributions and density functions are written in lower case.

Despite the fact that the earlier chapters of this book are based heavily on the 1979 edition, the discussion does sometimes differ substantially from

that in the 1979 edition, especially where the 1979 viewpoint now seems to be misguided or out of date. As one example, the discussion of the Fundamental Theorem of Natural Selection is now quite different from that of the 1979 edition. The 1979 interpretation of the theorem, standard at the time, is now seen as incorrect and has been discarded. However I have no illusions about its ability to continue to exist as the textbook interpretation, offered to students, especially since the correct interpretation requires greater mathematical depth than does the textbook version.

Current theory in mathematical population genetics emphasizes retrospective analyses rather than the prospective analyses making up much of the classical theory. In particular, theory surrounding the Kingman coalescent process forms, quite appropriately, a significant part of current research. An introduction to this theory is given in Chapter 10, and a more extensive discussion will be given in Volume II. One of the aims of this book is to make connections between the prospective theory that much of the book considers with this retrospective theory. Apart from this, the classical prospective theory, considering properties of forward-going evolutionary processes, is still relevant to retrospective analyses. As one example of this, the theory surrounding the coalescent is often best developed by considering a process moving forward in time from a common ancestor to a sample of genes in the present generation, rather than by starting with the contemporary sample and moving backward in time to the common ancestor.

Despite the natural current emphasis on the retrospective theory, there are several reasons for discussing the prospective theory in some detail in this book. The Darwinian theory of evolution continues to be attacked by various interest groups, and these attacks are sometimes helped by incorrect statements about the prospective evolutionary theory made sometimes even by biologists. The many extraordinary statements made by Løvtrup (1987), for example, illustrate this.

Even professionals in population genetics contribute to this problem. Arguments against evolution as a Darwinian process have been based the concept of the substitutional genetic load, which I believe has been deleterious concept that should be dropped from into the theory. Substitutional load “theory”, as well as segregational load “theory”, is discussed, and I hope debunked, in Section 2.11. The “blind watchmaker” paradigm, periodically raised by outsiders to population genetics theory as refuting the Darwinian process and indeed evolution generally, is discussed, and I hope also debunked, in Section 1.6. On two more narrow points where those active in areas close to population genetics theory frequently abuse the theory, the correct as opposed to the textbook version of the Fundamental Theorem of Natural Selection, mentioned above, is described in Sections 2.9 and 7.4.5. A discussion of the much-misunderstood expression “effective population size”, often incorrectly used in with reference to the history of the human population, is given in Section 3.7.

The recent and welcome infusion of population genetics theory into a variety of disciplines associated with the evolutionary process has not been without some problems. Perhaps the most important of these is that it has led to an uncritical use of some formulas from the theory without due assessment of whether the formulas are appropriate to the situation at hand. All formulas in population genetics theory derive some model of the evolutionary process, and in some cases this model can be no more than a very rough approximation to reality. For this reason a new section has been added, in this volume, discussing the modeling process and what may reasonably be concluded from the models discussed in the population genetics literature.

On more technical matters, it has not always been possible to use the notation of various published papers whose results are described here, since in some cases the notation used in different papers for the same quantity differ, and in other cases different authors use the same symbol for different quantities. As in the 1979 version of this book, the notation is not consistent, so that the symbol " x_i " might variously mean the frequency of an allele in generation i , in subpopulation i , the frequency of the allele A_i , and so on. On a similar point, I have adopted American spelling but English punctuation conventions: The latter are more suited to a mathematical text.

It is a pleasure to acknowledge the inspiration I have received from my long-time colleagues Bob Griffiths and Geoff Watterson. It is also a pleasure to thank Peter Donnelly, John Kingman and Simon Tavaré for an equally close, albeit long-range, collegial association. I thank various colleagues for pointing out typographical errors in the 1979 edition of this book, and Alan Rogers for pointing out an error concerning exchangeable model calculations. Any errors observed in this volume will be gratefully received at wewens@sas.upenn.edu and an archive of these will be maintained at www.textbook-errata.org

Philadelphia, Pennsylvania, USA
October 2003

Warren J. Ewens

Contents

Preface	vii
Introduction	xvii
1 Historical Background	1
1.1 Biometricians, Saltationists and Mendelians	1
1.2 The Hardy–Weinberg Law	3
1.3 The Correlation Between Relatives	6
1.4 Evolution	11
1.4.1 The Deterministic Theory	11
1.4.2 Non-Random-Mating Populations	18
1.4.3 The Stochastic Theory	20
1.5 Evolved Genetic Phenomena	31
1.6 Modelling	35
1.7 Overall Evolutionary Theories	38
2 Technicalities and Generalizations	43
2.1 Introduction	43
2.2 Random Union of Gametes	44
2.3 Dioecious Populations	44
2.4 Multiple Alleles	49
2.5 Frequency-Dependent Selection	54
2.6 Fertility Selection	54
2.7 Continuous-Time Models	57

2.8	Non-Random-Mating Populations	62
2.9	The Fundamental Theorem of Natural Selection	64
2.10	Two Loci	67
2.11	Genetic Loads	78
2.12	Finite Markov Chains	86
3	Discrete Stochastic Models	92
3.1	Introduction	92
3.2	Wright–Fisher Model: Two Alleles	92
3.3	The Cannings (Exchangeable) Model: Two Alleles	99
3.4	Moran Models: Two Alleles	104
3.5	K-Allele Wright–Fisher Models	109
3.6	Infinitely Many Alleles Models	111
3.6.1	Introduction	111
3.6.2	The Wright–Fisher Infinitely Many Alleles Model	111
3.6.3	The Cannings Infinitely Many Alleles Model	117
3.6.4	The Moran Infinitely Many Alleles Model	117
3.7	The Effective Population Size	119
3.8	Frequency-Dependent Selection	129
3.9	Two Loci	129
4	Diffusion Theory	136
4.1	Introduction	136
4.2	The Forward and Backward Kolmogorov Equations	137
4.3	Fixation Probabilities	139
4.4	Absorption Time Properties	140
4.5	The Stationary Distribution	145
4.6	Conditional Processes	146
4.7	Diffusion Theory	148
4.8	Multi-dimensional Processes	151
4.9	Time Reversibility	153
4.10	Expectations of Functions of Diffusion Variables	153
5	Applications of Diffusion Theory	156
5.1	Introduction	156
5.2	No Selection or Mutation	158
5.3	Selection	165
5.4	Selection: Absorption Time Properties	167
5.5	One-Way Mutation	171
5.6	Two-Way Mutation	174
5.7	Diffusion Approximations and Boundary Conditions	176
5.8	Random Environments	181
5.9	Time-Reversal and Age Properties	188
5.10	Multi-Allele Diffusion Processes	192

6	Two Loci	201
6.1	Introduction	201
6.2	Evolutionary Properties of Mean Fitness	202
6.3	Equilibrium Points	208
6.4	Special Models	209
6.5	Modifier Theory	221
6.6	Two-Locus Diffusion Processes	227
6.7	Associative Overdominance and Hitchhiking	230
6.8	The Evolutionary Advantage of Recombination	235
6.9	Summary	239
7	Many Loci	241
7.1	Introduction	241
7.2	Notation	242
7.3	The Random Mating Case	243
7.3.1	Linkage Disequilibrium, Means and Variances	243
7.3.2	Recurrence Relations for Gametic Frequencies	245
7.3.3	Components of Variance	246
7.3.4	Particular Models	249
7.4	Non-Random Mating	254
7.4.1	Introduction	254
7.4.2	Notation and Theory	255
7.4.3	Marginal Fitnesses and Average Effects	256
7.4.4	Implications	258
7.4.5	The Fundamental Theorem of Natural Selection	259
7.4.6	Optimality Principles	261
7.5	The Correlation Between Relatives	266
7.6	Summary	274
8	Further Considerations	276
8.1	Introduction	276
8.2	What is Fitness?	276
8.3	Sex Ratio	277
8.4	Geographical Structure	278
8.5	Age Structure	282
8.6	Ecological Considerations	283
8.7	Sociobiology	285
9	Molecular Population Genetics: Introduction	288
9.1	Introduction	288
9.2	Technical Comments	290
9.3	Infinitely Many Alleles Models: Population Properties	292
9.3.1	The Wright–Fisher Model	292
9.3.2	The Moran Model	294
9.4	Infinitely Many Sites Models: Population Properties	297

9.4.1	Introduction	297
9.4.2	The Wright–Fisher Model	298
9.4.3	The Moran Model	300
9.5	Sample Properties of Infinitely Many Alleles Models	301
9.5.1	Introduction	301
9.5.2	The Wright–Fisher Model	301
9.5.3	The Moran Model	306
9.6	Sample Properties of Infinitely Many Sites Models	308
9.6.1	Introduction	308
9.6.2	The Wright–Fisher Model	308
9.6.3	The Moran Model	314
9.7	Relation Between Infinitely Many Alleles and Infinitely Many Sites Models	316
9.8	Genetic Variation Within and Between Populations	319
9.9	Age-Ordered Alleles: Frequencies and Ages	320
10	Looking Backward in Time: The Coalescent	328
10.1	Introduction	328
10.2	Competing Poisson and Geometric Processes	329
10.3	The Coalescent Process	330
10.4	The Coalescent and Its Relation to Evolutionary Genetic Models	331
10.5	Coalescent Calculations: Wright–Fisher Models	333
10.6	Coalescent Calculations: Exact Moran Model Results	338
10.7	General Comments	341
10.8	The Coalescent and Human Genetics	342
11	Looking Backward: Testing the Neutral Theory	346
11.1	Introduction	346
11.2	Testing in the Infinitely Many Alleles Models	349
11.2.1	Introduction	349
11.2.2	The Ewens and the Watterson Tests	349
11.2.3	Procedures Based on the Conditional Sample Frequency Spectrum	353
11.2.4	Age-Dependent Tests	354
11.3	Testing in the Infinitely Many Sites Models	355
11.3.1	Introduction	355
11.3.2	Estimators of θ	356
11.3.3	The Tajima Test	358
11.3.4	Other “Tajima-like” Testing Procedures	361
11.3.5	Testing for the Signature of a Selective Sweep	362
11.3.6	Combining Infinitely Many Alleles and Infinitely Many Sites Approaches	364
11.3.7	Data from Several Unlinked Loci	365
11.3.8	Data from Unlinked Sites	368

11.3.9	Tests Based on Historical Features	369
12	Looking Backward in Time:	
	Population and Species Comparisons	370
12.1	Introduction	370
12.1.1	The Reversibility Criterion	372
12.2	Various Evolutionary Models	373
12.2.1	The Jukes–Cantor Model	373
12.2.2	The Kimura Model and Its Generalizations	374
12.2.3	The Felsenstein Models	375
12.3	Some Implications	377
12.3.1	Introduction	377
12.3.2	The Jukes–Cantor Model	377
12.3.3	The Kimura Model	380
12.4	Statistical Procedures	381
	Appendix A: Eigenvalue Calculations	384
	Appendix B: Significance Levels for \hat{F}	385
	Appendix C: Means and Variances of \hat{F}	386
	References	387
	Author Index	409
	Subject Index	413

Introduction to the First Edition

Population genetics occupies a central place in a variety of important biological and social undertakings. It has for many years been crucial to an understanding of evolutionary processes, of plant and animal breeding programs, and of various diseases of particular importance to man. While increased research in these areas naturally leads to a greater understanding of them, it also shows, particularly with the mathematical theory of population genetics, that previous arguments have sometimes been misleading, important points have been glossed over, and our knowledge of the genetic behavior of populations is not as firm as might previously have been thought. This observation is all the more important because much recent controversy on developments within or connected to population genetics has sometimes relied on now outdated population genetics theory. In this connection one might mention sociobiology, the effects of genetic manipulation with recombinant DNA, nature-nurture and heritability studies, and the knowledge of the detailed constitution of genetic material and the consequent possibility of its artificial creation. The importance of these developments is immense, as is the need to base controversies on them on firm population genetic and other scientific knowledge.

Population genetics embraces observational, experimental and theoretical components. While population genetics theory is in large measure quantitative, the complexities of Nature ensure that nonmathematical reasoning eventually outruns the purely mathematical aspects of the theory, which are necessarily based on simplified models of biological behavior. Nevertheless, the purely mathematical aspects of population genetics theory comprise a very large area of applied mathematical research, and the

aim of this book is to give an account of this purely mathematical theory. Thus this book is not about population genetics theory, still less about population genetics itself. Indeed, the selection of material that must necessarily be made is biased towards that with the richest mathematical content, and this sometimes implies that topics of greater importance to population genetics generally are treated at shorter length than their real importance warrants. Given the number of books on population genetics and population genetics theory, I believe there is a place for an account of the purely mathematical theory, even if biased in this way. Despite this broad aim, the first chapter of this book is largely historical and considers more general questions on population genetics. This is so since I believe such a background is necessary even for a consideration of the purely mathematical theory.

The book has been aimed at the graduate or research level and should be supplemented by reading an introductory text. Perhaps the most useful for this purpose is C. C. Li's excellent *First Course in Population Genetics*. As indicated above, collateral reading in population genetics theory generally is also necessary to place the topics treated in this book in proper perspective.

What is the value of the mathematical side of population genetics theory? It may be argued that this merely makes quantitative arguments the general nature of which is already clear qualitatively. While in some measure this is true, there are many questions where common-sense qualitative arguments have led to quite incorrect conclusions on the genetic behavior of populations. This is true even for rather simple aspects of the theory and, of course, is increasingly true for more complex aspects and also aspects involving stochastic phenomena. This matter is discussed further in the concluding remarks of the book, to some extent in the light of examples of such questions treated in the preceding chapters.

The mathematical theory contributes in various degrees to the controversial areas mentioned in the opening paragraph. The theory of the correlation between relatives for a metrical trait, outlined in Chapters 7 and 8, is the key ingredient in heritability studies and in nature-nurture allocations. The small but growing mathematical theory of altruistic traits concerns perhaps the central question of sociobiology. Detailed knowledge of the nature of genetic material has already led to considerable quantitative theory, particularly in the nature of evolutionary processes: It is perhaps in this area that of those mentioned, mathematical population genetics theory will find its greatest application and from which, in turn, it will be most influenced so far as its nature and direction are concerned. The manipulation of genetic material now possible is perhaps, except in a negative sense, the area where mathematical theory is of least value. One population geneticist has claimed that the eventual goal of the study of evolution is to understand its processes quantitatively and thus be able to predict and control its course. The theory in this book, particularly that of Chapters 6 and 7, should indicate the difficulty of achieving the first

aim and the consequent great danger in an attempt to take control of evolutionary processes generally and in particular (as some enthusiasts would wish) of human evolution. The complexities of the genetic behavior of populations, as shown by the (still incomplete) mathematical theory, are far greater than our power to comprehend and control.

Various points concerning the presentation of this book should be mentioned. Aiming to concentrate on the mathematical theory, I have emphasized, particularly in Chapter 2, that such theory rests on models of biological reality which, no matter how simplistic, must be analyzed on their own without the injection of extraneous assumptions during the analysis. If such assumptions are brought in, and the assumptions injected contradict those implicit in the model, in principle any result, no matter how incorrect, can arise. Of course the conclusions reached from a model must be treated with caution, depending on the reality of the initial assumptions made, but this is a different matter from interfering with the analysis of a model in mid-stream. Several incorrect conclusions in population genetics have arisen from such *ad hoc* interference.

So far as terminology is concerned I have followed the standard usage of the subject, even when this is perhaps unsatisfactory. Two unfortunate expressions, "gene frequency" (instead of the more logical "allele frequency") and "additive genetic variance" (instead of, perhaps, "genetic variance") are entrenched in the literature, and I have used them here except on specific occasions when a more precise usage seemed necessary. The notation is not consistent throughout the book. Thus the symbol " x_i " might variously mean the frequency of an allele in generation i , in subpopulation i , the frequency of the allele A_i , and so on. Consistency would lead to cumbersome notation, and the context should always make clear, even if no explicit explanation is given, what any symbol stands for.

I have cited fewer rather than more references during this book, concentrating on those accounts that appear to be definitive, innovative, the most recent or in some other way important.

This book has benefited greatly from the advice and criticism of many friends and colleagues, of whom I should mention Marc Feldman, Walter Fitch, Bob Griffiths, Sam Karlin, Ray Littler, Tom Nagylaki, Eugene Seneta, Richard Spielman and Glenys Thomson. I must thank in particular Frank Norman for many patient hours spent explaining to me the intricacies of mathematical diffusion theory, John Gillespie for his constant advice on biological, evolutionary and mathematical questions, and above all Geoff Watterson for his most careful and detailed reading of drafts of this book and for much discussion and guidance on the topics it considers. Naturally I am responsible for all errors and obscurities in the final version.

Melbourne, Victoria, Australia, and
Philadelphia, Pennsylvania, USA
December 1976 to October 1978

Warren J. Ewens

1

Historical Background

1.1 Biometricians, Saltationists and Mendelians

Population genetics theory was initially developed, in the 1920's and 1930's, by Fisher, Haldane and Wright, and current theory still bears the imprint of the work of these three great masters. Indeed, so fundamental was their contribution that even today, it is difficult to move forward from the paradigms that they introduced. Such a move forward is, however, necessary, especially because of the availability of data from the human genome project and other genome projects, and the need to analyze these data using population genetic theory methods.

To make any such forward move, and to establish any new paradigm, will nevertheless require an understanding of the theory established by Fisher, Haldane and Wright, as well as an understanding of the historical context in which they found themselves. In short, their objective was to formulate an evolutionary paradigm based on the Mendelian hereditary mechanism. Perhaps the major difficulty in doing this arose from the divisions on evolutionary questions following the rediscovery of Mendelism in 1900. We therefore start by describing these divisions, which reinforced an already existing division among biologists about the nature of evolution.

The *Origin of Species* was published in 1859. Apart from the controversies it brought about on a nonscientific level, it set biologists at odds as to various aspects of the theory. That evolution had occurred was not, on the whole, questioned. What was more controversial was the claim that the agency bringing about evolution was natural selection, and, among se-

lectionists, there was disagreement about the the nature of a selectively induced evolutionary changes. Darwin adhered to the “gradualist” point of view, that changes in the nature of organisms in populations were gradual and incremental. Some of those who, in general, were his strongest supporters, for example T. H. Huxley and Francis Galton, were “saltationists”, believing that evolutionary changes most often occur in “jumps” of not inconsiderable magnitude. Two evolutionary schools of thought developed from these two points of view. Although any attempt to describe in brief terms the long and complex controversies that followed is bound to be incomplete, it is nevertheless possible to trace in general terms the threads of the arguments followed by members of both schools. A more detailed account of these matters is given by Provine, (1971).

Before doing so, it must be remembered that Mendel’s work, and hence the mechanism of heredity, was in effect unknown before 1900 and that in so far as a common view of heredity existed, it would have been that the characteristics of an individual are, or tend to be, a blending of the corresponding characteristics of his parents. It is, however, interesting to note that in a letter to Darwin in 1875 Galton came almost by pure reasoning to a proposition about the hereditary mechanism that was very close to the Mendelian one. Unfortunately his line of thought appears not to have been pursued: If it had been, the course of evolutionary thought during the next hundred years would have been very different. Details of Galton’s letter, and comments on it, are given by Olby (1965).

The blending hypothesis brought perhaps the most substantial scientific objection to Darwin’s theory. It is easy to see that with random mating, the variance in a population for any characteristic will, under the blending theory, decrease by a factor of one-half in each generation. Thus uniformity of characteristics would essentially be obtained after a few generations, so that eventually no variation would exist upon which natural selection could act. Since, of course, such uniformity is not observed, this argument is incomplete. But since variation of the degree observed could only occur by postulating further factors of strong effect which cause the characteristics of offspring to deviate from those of their parents, it cannot then be reasonably argued that selectively favored parents produce offspring who closely resemble them and who are thus themselves selectively favored. This argument was recognized by Darwin as a major obstacle to his theory of evolution through natural selection, and it is interesting to note that later versions of the *Origin* were, unfortunately, somewhat influenced by this argument.

Galton’s role in the controversy between the gradualists and the saltationists was somewhat ambiguous. On the one hand he was himself a believer in the saltation theory, and this no doubt influenced him in advancing in 1875 the hereditary theory referred to above. On the other hand, he pursued a close intellectual and personal relationship with Darwin and, through this, attempted to quantify the gradualist evolutionary process.

This led him to introduce the statistical concepts of correlation and regression, which became the main tools of a group of scientists, later known as biometricians, who were one of the inheritors of the gradualist Darwin theory. This group’s mathematical research in quantitative evolution began in the 1890s under the leadership of W.F.R. Weldon and Karl Pearson. At the same time the saltationists gained further adherents, notably William Bateson, and the struggle between the two groups became more intense as the century drew to a close.

The year 1900 saw the rediscovery of Mendelism. The particulate nature of this theory was of course appealing to the saltationists. Rather soon many biologists believed in a non-Darwinian process of evolution through mutational jumps – the view that “Mendelism had destroyed Darwinism” was not uncommon. On the other hand, the biometricians continued to believe in the Darwinian theory of gradualist evolution through natural selection and were thus, in the main, disinclined to believe in the Mendelian mechanism, or at least that this mechanism was of fundamental importance in evolution.

It would be pointless to follow in detail the sometimes bitter acrimony that then followed. Even the inspired arguments of Yule (1902), based on a mathematical analysis of the Mendelian system, that Mendelism and Darwinism could be reconciled, were largely ignored. And yet, paradoxically, Darwinism and Mendelism are not incompatible. Indeed, the former relies crucially on the latter, and further it would be difficult to conceive of a Mendelian system without some form of natural selection associated with it. To see why this should be so, it is now necessary to turn to the beginnings of the mathematical theory of population genetics.

1.2 The Hardy–Weinberg Law

We consider a random-mating monoecious population which is so large that genotype frequency changes may be treated as deterministic, and focus attention on a given gene locus at which two alleles may occur, namely A_1 and A_2 . Suppose that in any generation the proportions of the three genotypes A_1A_1 , A_1A_2 and A_2A_2 are X , $2Y$, and Z , respectively. Since random mating obtains, the frequency of matings of the type $A_1A_1 \times A_1A_1$ is X^2 , that of $A_1A_1 \times A_1A_2$ is $4XY$, and so on. We now consider the outcomes of each of these matings. If the very small probability of mutation is ignored, and if there are no fitness differentials between genotypes, elementary Mendelian rules indicate that the outcome of an $A_1A_1 \times A_1A_1$ mating must be A_1A_1 and that in an indefinitely large population, half the $A_1A_1 \times A_1A_2$ matings will produce A_1A_1 offspring, and the other half will produce A_1A_2 offspring, with similar results for the remaining matings.

It follows that since A_1A_1 offspring can be obtained only from $A_1A_1 \times A_1A_1$ matings (with overall frequency 1 for such matings), from $A_1A_1 \times A_1A_2$ matings (with overall frequency $\frac{1}{2}$ for such matings), and from $A_1A_2 \times A_1A_2$ matings (with frequency $\frac{1}{4}$ for such matings), and since the frequencies of these matings are X^2 , $4XY$, $4Y^2$, the frequency X' of A_1A_1 in the following generation is

$$X' = X^2 + \frac{1}{2}(4XY) + \frac{1}{4}(4Y^2) = (X + Y)^2. \quad (1.1)$$

Similar considerations give the frequencies $2Y'$ of A_1A_2 and Z' of A_2A_2 as

$$2Y' = \frac{1}{2}(4XY) + \frac{1}{2}(4Y^2) + 2XZ + \frac{1}{2}(4YZ) = 2(X + Y)(Y + Z), \quad (1.2)$$

$$Z' = \frac{1}{4}(4Y^2) + \frac{1}{2}(4YZ) + Z^2 = (Y + Z)^2. \quad (1.3)$$

The frequencies X'' , $2Y''$ and Z'' for the next generation are found by replacing X' , $2Y'$ and Z' , by X'' , $2Y''$ and Z'' and X , $2Y$ and Z by X' , $2Y'$ and Z' in (1.1)–(1.3). Thus, for example, using (1.1) and (1.2),

$$\begin{aligned} X'' &= (X' + Y')^2 \\ &= (X + Y)^2 \\ &= X', \end{aligned}$$

and similarly it is found that $Y'' = Y'$, $Z'' = Z'$. Thus, the genotype frequencies established by the second generation are maintained in the third generation and consequently in all subsequent generations. Frequencies having this property can be characterized as those satisfying the relation

$$(Y')^2 = X'Z'. \quad (1.4)$$

Clearly if this relation holds in the first generation, so that

$$Y^2 = XZ, \quad (1.5)$$

then not only would there be no change in genotypic frequencies between the second and subsequent generations, but also these frequencies would be the same as those in the first generation. Populations for which (1.5) is true are said to have genotypic frequencies in Hardy–Weinberg form.

We also observe that whereas there might be genotype frequency changes between generation 1 and generation 2, the frequency $x = X + Y$ of the allele A_1 does *not* change between these two generations. Nor of course does it change between any further generations. In accordance with common practice, we shall often use the expression “gene frequency”, and an expression such as “the frequency of the gene A_1 ”, rather than the “allele frequency” terminology employed above.

Since $X + 2Y + Z = 1$, only two of the frequencies X , $2Y$ and Z are independent. If, further, (1.5) holds, only one frequency is independent. Examination of the recurrence relations (1.1)–(1.3) shows that the most convenient quantity for independent consideration is the frequency x of the

allele A_1 . These conclusions may be summarized in the form of a theorem:

Theorem (Hardy–Weinberg). Under the assumptions stated, a population having genotypic frequencies X (of A_1A_1), $2Y$ (of A_1A_2) and Z (of A_2A_2) achieves, after one generation of random mating, stable genotypic frequencies x^2 , $2x(1-x)$, $(1-x)^2$ where $x = X + Y$ and $1-x = Y + Z$. If the initial frequencies X , $2Y$, Z are already of the form x^2 , $2x(1-x)$, $(1-x)^2$, then these frequencies are stable for all generations.

Numerical examples of this theorem were given by Castle (1903), who possibly (cf. Keeler (1968)) knew the theorem in full generality, by Yule (1906), and by Pearson (1904). The first published general proof was by Hardy (1908) and Weinberg (1908), and it is after these authors that the theorem has become known, normally as the “Hardy–Weinberg law”.

Why is this rather simple theorem, or as it is more frequently called “law”, so important? Unfortunately it is important for two different reasons, one purely technical, and concentration on the technical reason has sometimes tended to obscure its truly basic value. The technical point is that if, as we may reasonably assume in a random-mating population, equation (1.5) is true, the mathematical behavior of the population can be examined in terms of the single frequency x rather than in terms of the pair (X, Y) ; this is certainly a considerable convenience, but it is not fundamentally important. The really important part of the theorem lies in the stability behavior. If no external forces act, there is no intrinsic tendency for any variation present in the population, that is, variation caused by the existence of the three different genotypes, to disappear. This shows immediately that the major earlier criticism of Darwinism, namely the fact that variation decreases rapidly under the blending theory, does not apply with Mendelian inheritance. It is clear directly from the Hardy–Weinberg Law that under a Mendelian system of inheritance, variation tends to be maintained.

Of course, the action of selection itself often tends to destroy variation; this qualification is of some importance and we shall return to this point later and will find that the rate of loss of variation in any realistic Mendelian scheme involving selection is far less than the rate under any realistic blending scheme.

It is the “quantal” nature of the gene that leads to the stability behavior described by the Hardy–Weinberg law. It is thus interesting that the year in which the Mendelian theory was rediscovered, 1900, was the same year as the introduction of the quantum theory in physics. Both theories have been fundamental and crucial in their respective spheres. One can even claim that if there is intelligent life, that is, life that has evolved via natural selection, elsewhere in the universe, the heredity mechanism involved must be a quantal, maybe a Mendelian, one, since otherwise it is not clear how the variation necessary for evolution by natural selection can be maintained.

Thus the Hardy–Weinberg law shows that far from being incompatible, Darwinism and Mendelism are almost inseparable. It would be difficult to think of a hereditary process other than the quantal Mendelian scheme in which natural selection could act with such efficiency, while on the other hand fitness differentials between genotypes will normally lead to changes in gene frequencies and thus ultimately to evolution. We generalize the Hardy–Weinberg law later in this book to the case where more than two alleles are possible at the locus in question and also to the multilocus case. We shall also discuss extensions of it to non-random-mating populations. For the moment we shall be content with noting its historical significance.

It was thus beginning to become clear by the end of the first decade of the 20th century that a reconciliation between Darwinism and Mendelism was not only possible but indeed inevitable. In 1911 this was already apparent to a young student of mathematics who read, during that year, a paper on “Heredity” to the Cambridge University Eugenics Society, in which he stressed the necessity for this reconciliation. Such a reconciliation would carry with it a requirement to interpret, on Mendelian principles, the large bodies of data assembled by the biometricians on the correlations between relatives for various physical characteristics. Several years later R. A. Fisher, the young student in question, wrote a landmark paper (Fisher, (1918)) in population genetics in which this reconciliation was achieved.

Special cases of these correlations had been treated earlier by Pearson (1904) and Yule (1906), but Fisher’s (1918) work was the first one to consider the problem in a rather complete degree of generality. We therefore consider the approach he used, since several of the quantities which play a key role in his argument will appear subsequently to have considerable evolutionary importance.

1.3 The Correlation Between Relatives

Consider any character which is determined entirely by a locus A at which occur alleles A_1 and A_2 . Suppose that all A_1A_1 individuals have measurement m_{11} for this character, that all A_1A_2 individuals have measurement m_{12} , and that all A_2A_2 individuals have measurement m_{22} . For the moment we assume no environmental contributions: Once we know the genotype of any individual assume that we know the value of his measurement.

Suppose that random mating obtains with respect to this character and that the frequencies of A_1A_1 , A_1A_2 and A_2A_2 are in Hardy–Weinberg form x^2 , $2x(1-x)$ and $(1-x)^2$, respectively. Then the mean value \bar{m} of this measurement is given by

$$\bar{m} = x^2m_{11} + 2x(1-x)m_{12} + (1-x)^2m_{22},$$

and the variance σ^2 in the measurement is

$$\sigma^2 = x^2(m_{11} - \bar{m})^2 + 2x(1-x)(m_{12} - \bar{m})^2 + (1-x)^2(m_{22} - \bar{m})^2. \quad (1.6)$$

Table 1.1.

<i>G</i>	<i>M</i>	Son		
		A_1A_1	A_1A_2	A_2A_2
		m_{11}	m_{12}	m_{22}
Father	A_1A_1	x^3	$x^2(1-x)$	0
	A_1A_2	$x^2(1-x)$	$x(1-x)$	$x(1-x)^2$
	A_2A_2	0	$x(1-x)^2$	$(1-x)^3$

What is the covariance between father and son with respect to this measurement? Suppose first that the father is A_1A_1 . Then the son will be A_1A_1 if the mother transmits an A_1 gene to him, an event with probability x . Similarly the son will be A_1A_2 with probability $1-x$. The father himself will be A_1A_1 with probability x^2 . Continuing in this way it is possible to draw up a table of the probabilities of the various father-son combinations in genotype and hence in the character measured. Using G for genotype and M for measurement, we eventually find the values shown in Table 1.1.

The covariance between the measurement for the father and that for the son, assuming no change in the frequency of A_1 between the two generations, is thus

$$\begin{aligned} & x^3m_{11}^2 + 2x^2(1-x)m_{11}m_{12} + x(1-x)m_{12}^2 + 2x(1-x)^2m_{12}m_{22} \\ & \quad + (1-x)^3m_{22}^2 - \bar{m}^2 \\ & = x(1-x)(xm_{11} + (1-2x)m_{12} - (1-x)m_{22})^2. \end{aligned} \quad (1.7)$$

The correlation between the two measurements, found by dividing the covariance by the variance (since the variance for sons is the same as that for fathers), is then

$$x(1-x)\{xm_{11} + (1-2x)m_{12} - (1-x)m_{22}\}^2/\sigma^2. \quad (1.8)$$

It is useful to write this expression in a different form. If we define

$$\left. \begin{aligned} \sigma_A^2 &= 2x(1-x)\{xm_{11} + (1-2x)m_{12} - (1-x)m_{22}\}^2, \\ \sigma_D^2 &= x^2(1-x)^2\{2m_{12} - m_{11} - m_{22}\}^2, \end{aligned} \right\} \quad (1.9)$$

the expression (1.8) is clearly

$$\frac{1}{2} \frac{\sigma_A^2}{\sigma^2}. \quad (1.10)$$

Furthermore, it is simply a matter of algebra to show that

$$\sigma^2 = \sigma_A^2 + \sigma_D^2, \quad (1.11)$$

and in view of these relations it is of some interest to find interpretations for σ_A^2 and σ_D^2 .

In order to find an interpretation for σ_A^2 we consider what changes are made in the measurement in question if we replace an A_1 allele by an A_2 allele in some individual. The effect of doing this will, in general, depend on whether the replacement is made in an A_1A_1 individual or an A_1A_2 individual. The change is $m_{12} - m_{11}$ in the first case and $m_{22} - m_{12}$ in the second, and these will not generally be equal. We thus try to find some expression for this effect which in some sense is as close as possible to these two values, using the concept of a weighted least-squares fit.

Suppose we fit the measurements m_{11} , m_{12} and m_{22} as closely as possible, in the sense of weighted least squares, by values of the form $\bar{m} + 2\alpha_1$, $\bar{m} + \alpha_1 + \alpha_2$, $\bar{m} + 2\alpha_2$. Differentiation of the expression S , defined by

$$S = x^2(m_{11} - \bar{m} - 2\alpha_1)^2 + 2x(1-x)(m_{12} - \bar{m} - \alpha_1 - \alpha_2)^2 + (1-x)^2(m_{22} - \bar{m} - 2\alpha_2)^2$$

with respect to α_1 and α_2 with the derivative subsequently set to 0, gives eventually

$$\left. \begin{aligned} \alpha_1 &= x(m_{11} - \bar{m}) + (1-x)(m_{12} - \bar{m}), \\ \alpha_2 &= x(m_{12} - \bar{m}) + (1-x)(m_{22} - \bar{m}). \end{aligned} \right\} \quad (1.12)$$

as the best-fitting values. With this choice of α_1 and α_2 , the equation

$$x\alpha_1 + (1-x)\alpha_2 = 0 \quad (1.13)$$

is automatically satisfied. Often the minimization procedure is carried subject to the requirement that this equation holds, but since at the minimizing values this requirement is automatically satisfied, imposition of the requirement is not necessary. By contrast, when the above calculations are generalized to the case of many gene loci, a requirement of the form (1.13) will be needed.

We define the average effect of substituting A_2 for A_1 by

$$\alpha_2 - \alpha_1 = x(m_{12} - m_{11}) + (1-x)(m_{22} - m_{12}). \quad (1.14)$$

When more than two alleles are involved we shall find it more convenient to adopt a slightly different usage and to call α_1 the average effect of A_1 and α_2 the average effect of A_2 . The value (1.14) could have been found almost immediately by taking a weighted average of $m_{12} - m_{11}$ and $m_{22} - m_{12}$. The present approach, while less direct, does on the other hand yield further information. The minimum value of the expression S is easily seen to be

$$x^2(1-x)^2(2m_{12} - m_{11} - m_{22})^2, \quad (1.15)$$

and the difference between this and σ^2 , namely the sum of squares removed from σ^2 by fitting the parameters α_1 and α_2 , is

$$2x(1-x)\{xm_{11} + (1-2x)m_{12} - (1-x)m_{22}\}^2. \quad (1.16)$$

The expression (1.16) is identical to the quantity σ_A^2 defined in (1.9), while the residual sum of squares (1.15) is identical to the quantity σ_D^2 defined in (1.10). Because σ_A^2 can be derived in the way just outlined, it might reasonably be called the genic or allelic variance: It is that part of the total variance in the character which can be accounted for by the average effects of the alleles A_1 and A_2 , used in an additive fashion. A frequently used name for σ_A^2 is the “additive genetic variance” in the character measured, the word “genetic” meaning here “relating to genes”: This usage is perhaps unfortunate but because it is well established we follow it in this book. The residual variance σ_D^2 is called the dominance variance. Except for the trivial cases $x = 0$, $x = 1$, it is zero only if $m_{12} = \frac{1}{2}(m_{11} + m_{22})$, that is when there is no dominance in the measurement in question.

We may then express the result (1.10) as follows: Under the conditions assumed, the correlation between father and son in the measurement considered is half the ratio of the additive genetic variance to the total variance in the measurement. If we denote this ratio by ρ^2 , this result becomes

$$\text{corr}(\text{father, son}) = \frac{1}{2} \rho^2. \quad (1.17)$$

This correlation is always nonnegative, and will only take the value zero when $x = (m_{12} - m_{22}) / (2m_{12} - m_{11} - m_{12})$, a possibility that can arise only if m_{12} exceeds both m_{11} and m_{22} , or if m_{12} is less than both m_{11} and m_{22} . We emphasize strongly the fact that this correlation has been found by basing all calculations on the Mendelian nature of the hereditary process.

A table analogous to Table 1.1, considering in this case full sibs, shows that under the same assumptions made above,

$$\text{corr}(\text{full sibs}) = \frac{1}{2} \rho^2 + \frac{1}{4} \delta^2, \quad (1.18)$$

where $\delta^2 = \sigma_D^2 / \sigma^2$. Similar considerations, using tables of Mendelian associations rather more complex than those in Table 1.1, show that under the same assumptions,

$$\begin{aligned} \text{corr}(\text{uncle-nephew}) &= \frac{1}{4} \rho^2, & (1.19) \\ \text{corr}(\text{double first cousins}) &= \frac{1}{4} \rho^2 + \frac{1}{16} \delta^2, & (1.20) \end{aligned}$$

and so on.

Having obtained these results, Fisher (1918) then considered more complex situations, in particular cases where more than two alleles are possible at each locus, where characters are determined by the alleles at many loci, and where assortative mating obtains. We shall not pursue the complexities

associated with assortative mating: They are touched on briefly in Chapter 8. We also describe, in Chapter 7, a more efficient way of finding these correlations in the random-mating case. One generalization of these results is, however, straightforward. Fisher showed that for the one-locus multiple alleles case, the correlation formulae (1.17), (1.18), (1.19) and (1.20) remain unaltered provided that the additive and dominance variances are defined in the natural way through a generalization of the least-squares procedure just described. This is demonstrated in Section 2.4.

The analysis of the correlation for characters determined by many loci is far more complex than that for characters determined by one locus, since interactive effects must then be taken into account. In the case of a character which is correlated with fitness it is very hard to determine how important these interactive effects might be. If, however, the character is not correlated with fitness, we may reasonably assume (see Section 7.6) that

$$\text{freq}(A_i A_j B_k B_l \dots) = \text{freq}(A_i A_j) \times \text{freq}(B_k B_l) \times \dots \quad (1.21)$$

where A_1, A_2, \dots are the alleles possible at locus A partially determining this character, B_1, B_2, \dots are the alleles possible at a second locus B partially determining this character, and so on. Under random mating, equation (1.21) implies that the frequency of any chromosome, or *gamete*, can be written as the product of the frequencies of its constituent alleles. In this case the additive genetic variance τ^2 can be found, as we show later (Section 7.3.3), by simply summing the single-locus additive variances at the various individual loci (that is, in an obvious notation, $\tau^2 = \sum \sigma_A^2$), with a similar result for the total variance ($\omega^2 = \sum \sigma^2$): Our notation here is informal and is different from Fisher's. Thus assuming that (1.21) is true, the correlation in the character measured between father and son becomes

$$\text{corr}(\text{father}, \text{son}) = \frac{1}{2} \frac{\tau^2}{\omega^2}, \quad (1.22)$$

which is the natural generalization of (1.17). Similar values arise for the other relationships although, as will be observed in Chapter 7, the formulas for these other correlations often depend on the recombination structure between the loci determining the character. It is quite possible that while these results are true only when the character in question is not correlated with fitness, these values yield a satisfactory approximation even when there is some such correlation.

So far we have not taken any account of environmental variance. In practice it is difficult to do this, because of the unknown but presumably high environmental correlation for father and son, for brother and brother, and so on. Ignoring the possibilities of such environmental correlation, Fisher used formulae such as those above, in conjunction with observed correlations, to estimate the various components of variance in any character. We do not pursue the details of this here, and more will be said on this matter in

Chapter 8. It is sufficient to note at this stage that at least under simplified assumptions, the genetic component of the correlation between relatives is given in terms of some function of the additive and the dominance variances in the measurement of interest, and that the pattern of correlations predicted by the Mendelian mechanism agreed, for the data used by Fisher, reasonably well with those observed. As a result, Fisher had made a most significant beginning in reconciling biometry and Mendelism and for fusing these two into one discipline. From this point on population genetics, as the inheritor jointly of the Darwinian and the Mendelian theories, could start on a firm quantitative basis. Further, as we see in Section 1.4, the same variables used so effectively by Fisher in this reconciliation are, remarkably, central to the mathematical description of the evolutionary process.

1.4 Evolution

1.4.1 *The Deterministic Theory*

We turn now to the evolutionary consequences of Mendelism. The twin cornerstones of the Darwinian theory of evolution are variation and natural selection. Variation is provided, under a Mendelian system, ultimately by mutation: In all natural populations mutation provides a continual source of genetic variation. Since the different genotypes created by mutation will often have different fitnesses, that is will differ in viability, mating success, and fertility, natural selection will occur. Our task is to quantify this process, and we now outline the work done during the 1920s and 1930s in this direction. Such a quantification amounts to a scientific description of the Darwinian theory in Mendelian terms.

It is necessary, at least as a first step, to make a number of assumptions and approximations about the evolutionary process. Thus although mutation is essential for evolution, mutation rates are normally so small that for certain specific problems we may ignore mutational events. Further, although the fitness of an individual is determined in a complex way by his entire genetic make-up, and even then will often differ from one environment to another, we start by assuming as a first approximation that this fitness depends on his genotype at a single locus, or at least can be found by “summing” single locus contributions to fitness. It is also difficult to cope with that component of fitness which relates to fertility, and almost always special assumptions are made about this. More complete discussions of these problems will be given later in this book. If fitness relates solely to viability then much of the complexity is removed, and for convenience we make this assumption, at least for the moment.

Suppose then that the fitnesses and the frequencies of the three genotypes A_1A_1 , A_1A_2 , and A_2A_2 at a certain locus “A” are as given below:

	A_1A_1	A_1A_2	A_2A_2	
fitness	w_{11}	w_{12}	w_{22}	(1.23)
frequency	x^2	$2x(1-x)$	$(1-x)^2$	

We have written the frequencies of these genotypes in the Hardy–Weinberg form appropriate to random mating. (Non-random-mating populations are discussed in Section 1.4.2.) Now Hardy–Weinberg frequencies apply only at the moment conception, since from that time on differential viabilities alter genotype frequencies from the Hardy–Weinberg form. For this reason we will always, in this book, count frequencies in the population at the moment of conception of each generation.

Clearly the most interesting question to ask is: What is the behavior of the frequency x of the allele A_1 under natural selection? Since we take the fundamental units of the microevolutionary process to be the replacement in a population of an “inferior” allele by a “superior” allele, the answer to this question is essential to an understanding of the microevolutionary process.

This question was first attacked in certain specific cases by Norton (see Punnett, 1917), and later in much greater detail by Haldane (1924, 1926, 1927a, 1927b, 1930a, 1930b, 1932a) with a summary in Haldane (1932b). We consider here only the simplest of these cases. Before doing so, we observe that we are required to explain two seemingly contradictory phenomena. On the one hand we must explain the dynamic process of the substitution of one allele for another and, on the other hand, we must explain the observed existence of considerable, apparently stable, genetic polymorphism.

The first concern is to find the frequency x' of A_1 in the following generation. By considering the fitnesses of each individual and all possible matings, we find that

$$x' - x = \frac{x(1-x)\{w_{11}x + w_{12}(1-2x) - w_{22}(1-x)\}}{w_{11}x^2 + 2w_{12}x(1-x) + w_{22}(1-x)^2}. \quad (1.24)$$

Clearly continued iteration of the recurrence relation (1.24) yields the successive values taken by the frequency of A_1 . Unfortunately simple explicit expressions for these frequencies are not always available, and resort must be made to approximation.

Before discussing these approximations, we observe that x' depends on the ratios of the fitnesses w_{ij} rather than the absolute values, so that x' is unchanged if we multiply each w_{ij} by any convenient scaling constant. It is therefore possible to scale the w_{ij} in any way convenient to the analysis at hand. Different scalings are more convenient for different purposes. We indicate below two alternative scalings of the fitness values w_{ij} , and on different occasions either (1.25a), (1.25b), or (1.25c) will prove to be the

most useful. It should be emphasized that nothing is involved here other than convenience of notation.

Fitness Values			
A_1A_1	A_1A_2	A_2A_2	
w_{11}	w_{12}	w_{22}	(1.25a)
$1 + s$	$1 + sh$	1	(1.25b)
$1 - s_1$	1	$1 - s_2$	(1.25c)

We normally assume that except in extreme cases, perhaps involving lethality, the fitness differentials s , sh , s_1 and s_2 are small, perhaps of the order of 1%. In this case we ignore small-order terms in these parameters.

Using the fitness scheme (1.25b), the recurrence relation (1.24) may be replaced, to a sufficiently close approximation, by

$$x' - x = sx(1 - x)\{x + h(1 - 2x)\}. \quad (1.26)$$

If we measure time in units of one generation, this equation may be approximated, in turn, by

$$dx/dt = sx(1 - x)\{x + h(1 - 2x)\}. \quad (1.27)$$

If the time required for the frequency of A_1 to move from some value x_1 to some other value x_2 is denoted by $t(x_1, x_2)$, then clearly

$$t(x_1, x_2) = \int_{x_1}^{x_2} (sx(1 - x)\{x + h(1 - 2x)\})^{-1} dx. \quad (1.28)$$

Naturally this equation applies only in cases where, starting from x_1 , the frequency of A_1 will eventually reach x_2 .

While an explicit expression for $t(x_1, x_2)$ is possible, it is usually more convenient to use the expression (1.28) directly. Suppose first that $s > sh > 0$. Then it is clear from (1.27) that the frequency of A_1 steadily increases towards unity. However, as this frequency approaches unity, the time required for even small changes in it will be large, due to the small term $1 - x$ in the denominator of the integrand in (1.28). This behavior is even more marked in the case $h = 1$ (A_1 dominant to A_2 in fitness), for then the denominator in the integrand in (1.28) contains a multiplicative term $(1 - x)^2$. This very slow rate of increase is due to the fact that, once x is close to unity, the frequency of A_2A_2 , the genotype against which selection is operating, is extremely low. In the important particular case $h = \frac{1}{2}$, that is no dominance in fitness, (1.28) assumes the simple form

$$t(x_1, x_2) = \int_{x_1}^{x_2} \left\{ \frac{1}{2} sx(1 - x) \right\}^{-1} dx. \quad (1.29)$$

Table 1.2. Generations spent in various frequency ranges

h	Range					
	0.001–0.01	0.01–0.1	0.1–0.5	0.5–0.9	0.9–0.99	0.99–0.999
1/2	462	480	439	439	480	462
1	232	250	309	1,020	9,240	90,231

It is possible to evaluate the times required for any nominated changes in the frequency of A_1 from (1.28) and (1.29), and some representative values are given in Table 1.2.

The times shown in Table 1.2 support the conclusions just given and show that while selection acts so that variation is ultimately destroyed, the times required are usually very long, and are much longer than those required under any blending theory of inheritance. We may therefore often expect to observe considerable genetic polymorphism in populations even though they are subject to directional natural selection. We shall find several uses later for this table and its various generalizations. The papers by Haldane referred to above provide values analogous to those in Table 1.2 in increasingly complex conditions, for example inbreeding, the case of different sets of fitnesses in the two sexes. Clearly this procedure quantifies, at least approximately, the unit microevolutionary process of the replacement of an “inferior” allele by a “superior” allele.

It is clear that if $s < sh < 0$ a process parallel to the above, with A_2 steadily replacing A_1 , will occur. This process is a mirror image of the one just considered and needs no further comment.

An entirely different behavior arises when the fitness w_{12} of the heterozygote exceeds the fitnesses of both of the homozygotes. This case is most conveniently treated by using the fitness parameters (1.25c) with $s_1 > 0$, $s_2 > 0$. Here the recurrence relation (1.24) may be rewritten, to a sufficiently close approximation, as

$$x' - x = x(1 - x)\{s_2 - x(s_1 + s_2)\}. \quad (1.30)$$

It is clear that there will be no change in the frequency x of A_1 if x takes the particular value

$$x = x^* = \frac{s_2}{(s_1 + s_2)} = \frac{(w_{22} - w_{12})}{(w_{11} + w_{22} - 2w_{12})}. \quad (1.31)$$

Further, if $x < x^*$, then $x < x' < x^*$, while if $x > x^*$, then $x^* < x' < x$. Thus x^* is a point of stable equilibrium and, whatever its initial value, the frequency x of A_1 will steadily approach x^* . It is not difficult to see that if the heterozygote is the least fit genotype, so that $s_1 < 0$, $s_2 < 0$, then x^* is still an equilibrium point of the recurrence system (1.24), but in this case it is an unstable equilibrium and thus of little interest. In this case the frequency of A_1 will steadily decrease to zero if its initial value is less than x^* and will steadily increase to unity if initially greater than x^* .

The above considerations taken together show that a necessary and sufficient condition that there exist a stable equilibrium of the frequency of A_1 in the interval $(0, 1)$ is that the heterozygote have a larger fitness than both homozygotes. This most important fact was established by Fisher (1922), and gives one possible explanation for the occurrence of stable allelic frequencies in a population. Later we shall find a number of other possible explanations: For the moment we simply observe that under the Mendelian system we can explain the occurrence of both dynamic substitutional processes and static equilibrium configurations. Thus, by the 1920s the first major steps were already being taken to explain in Mendelian terms, and also to quantify, what are perhaps the two major properties of biological populations, namely their capacity to evolve and their capacity to maintain static variation over long periods.

We now consider the effect of mutation. Suppose that A_1 mutates to A_2 at a rate u and that A_2 mutates to A_1 at rate v . Then it is easy to see that if there is no selection,

$$x' = x(1 - u) + v(1 - x), \quad (1.32)$$

and that a stable equilibrium is reached when

$$x = x^* = \frac{v}{u + v}. \quad (1.33)$$

Suppose now that both selection and mutation occur. We have in mind mainly the case where selective differences are of order 10^{-2} while mutation rates are of order 10^{-5} or 10^{-6} . Consider first the case where heterozygote selective advantage exists so that under selection only, a stable equilibrium of the form (1.31) exists. It is clear under this assumption that if selection and mutation are now both taken into account there will exist a new stable equilibrium differing only trivially from that given by (1.31). We thus do not consider this case any further.

We next consider the case where A_1A_1 is the most fit genotype and A_2A_2 the least fit. Under the fitness scheme (1.25b), this assumption implies that $s > sh > 0$, and because mutation rates are assumed to be considerably smaller than fitness differentials, selective forces dominate mutation pressures for all but extreme frequencies of A_1 . Because of this there will exist a stable equilibrium point for the frequency of A_1 close to unity. More exactly we find, for this equilibrium point, the approximate formula

$$x = x^* = 1 - \frac{u}{s - sh} \quad (1.34)$$

for the equilibrium frequency of A_1 . If $s > 0$ and $h = 1$ (A_1 dominant to A_2), the corresponding formula is

$$x = x^* = 1 - \sqrt{u/s}. \quad (1.35)$$

Parallel formulas apply when $s < sh < 0$: Here we find, at equilibrium,

$$x = x^* = \frac{v}{|sh|}, \quad (1.36)$$

while when $s < 0$, $h = 1$,

$$x = x^* = \sqrt{|v/s|}. \quad (1.37)$$

All these formulas were arrived at during the 1920s. They imply a second way in which genetic variation may be maintained in a population, that is by “mutation-selection balance”. However, the frequency of one or other allele will be very small for any of the equilibria (1.34)–(1.37), although the frequency of the less frequent allele is less small where dominance is complete. Thus, when $s = 0.01$, $u = 10^{-6}$, the frequency of A_2 at equilibrium will be 0.01 when $h = 1$ (complete dominance) and 0.0002 when $h = \frac{1}{2}$ (no dominance).

We now consider further properties of mutation-selection equilibria such as (1.34), where the less frequent allele is quite rare and is maintained only by recurrent mutation from the favored allele. Under the fitness scheme (1.25b) the equilibrium mean fitness of the population would be $1 + s$ if the mutation rate were zero, since in this case A_1 would fix in the population. The occurrence of mutation causes the mean fitness to decrease somewhat from this value. So long as $h < 1$, this decrease is found, to a close approximation, to be $2u$. For $h = 1$ a somewhat different calculation, using (1.35), gives a decrease of u , and for values of h close to 1 a value closer to u than to $2u$ is found. In other words, the population suffers a decrease in mean fitness proportional to the mutation rate, but not to fitness differentials. Haldane (1937), who first obtained this result, made the assertion that this situation has been reached in present-day populations by evolutionary modification of the mutation rate, so that a small current decrease in mean fitness is traded off against an increase in genetic plasticity in the population suitable for possible future evolution. We term the loss in mean fitness the “mutational load” and later consider this and more general forms of genetic load in more detail.

We have observed earlier that the Mendelian system of heredity enables us to quantify, at least as a first approximation, the rate of allelic substitution in an evolutionary process. Is it possible to arrive at general principles, derived from the Mendelian system, which quantify the two main features of an evolutionary process through Darwinian natural selection, namely the requirement of variation for evolution to occur and second, the “improvement” brought about in a population through this evolution? In his Fundamental Theorem of Natural Selection (FTNS), Fisher (1930a, 1958) attempted to find such a principle. His presentation of this theorem was very obscure. The “conventional wisdom” version of this theorem, outlined below, is clearly not what he intended, but is nevertheless an interesting result. It is called here the “mean fitness increase theorem” (MFIT).

Consider a random-mating population where the fitness of any individual depends only on his genetic constitution at a single locus “ A ”. Suppose that two alleles, A_1 and A_2 , are possible at this locus and that the fitnesses of the three possible genotypes are as given in (1.25a). The population is assumed to reproduce in nonoverlapping generations, so that (1.24) is applicable. In any generation we may define the mean fitness \bar{w} of the population in that generation by

$$\bar{w} = w_{11}x^2 + 2w_{12}x(1-x) + w_{22}(1-x)^2, \quad (1.38)$$

where x is the frequency of A_1 in that generation. The frequency x' of A_1 in the following generation can be found from (1.24), and thus the mean fitness \bar{w}' in that generation can be computed as

$$\bar{w}' = w_{11}(x')^2 + 2w_{12}x'(1-x') + w_{22}(1-x')^2. \quad (1.39)$$

From this the change $\Delta\bar{w} = \bar{w}' - \bar{w}$ in mean fitness between these two generations is given exactly by

$$\begin{aligned} \Delta\bar{w} &= 2x(1-x)\{w_{11}x + w_{12}(1-2x) - w_{22}(1-x)\}^2 \\ &\quad \times \left\{w_{11}x^2 + \left(w_{12} + \frac{1}{2}w_{11} + \frac{1}{2}w_{22}\right)x(1-x) + w_{22}(1-x)^2\right\}\bar{w}^{-2}. \end{aligned} \quad (1.40)$$

Clearly $\Delta\bar{w}$ is nonnegative, so we may conclude that natural selection acts so as to increase, or at worst maintain, the mean fitness of the population. This is the first part of the MFIT, and in the very restricted case considered it provides a quantification in genetic terms of the Darwinian concept that an “improvement” in the population has been brought about by the action of natural selection.

We may also use (1.40) to quantify the second part of the Darwinian principle that variation, in our case genetic variation, is necessary for natural selection to operate. If the w_{ij} are all close to unity we may write, to a sufficiently close approximation,

$$\Delta\bar{w} \approx 2x(1-x)\{w_{11}x + w_{12}(1-2x) - w_{22}(1-x)\}^2. \quad (1.41)$$

The definition in (1.9) for the additive genetic variance in fitness then shows immediately that

$$\Delta\bar{w} \approx \sigma_A^2. \quad (1.42)$$

This approximation quantifies in genetic terms the second major element of the Darwinian theory, and correspondingly of the MFIT, namely that the rate of increase of mean fitness is essentially equal to the additive component of the genetic variance in fitness.

One might initially have thought that the total variance in fitness, namely

$$\sigma^2 = w_{11}^2x^2 + 2w_{12}^2x(1-x) + w_{22}^2(1-x)^2 - \bar{w}^2, \quad (1.43)$$

rather than the additive component of the variance, should be related to the increase in mean fitness. There are at least two arguments that show that this is not so. First, if the fitness values are of the form (1.25c) with

$s_1, s_2 > 0$, and if the population is at the equilibrium point (1.31), then the total variance in fitness will be positive and yet, because the population is at equilibrium, there will be no increase in mean fitness from one generation to the next. Second, and related to the first argument, the additive component of the genetic variance is that portion explained by “genes within genotypes” when these are freed, as far as is possible, from deviations due to dominance. Since, in the model we consider, changes in gene frequencies are the fundamental components of evolution, the rate of increase of mean fitness can be expected to be related to that component of the total genetic variance which is accounted for by the alleles themselves, that is the additive genetic, or genic, variance.

The MFIT is not the FTNS. The Fundamental Theorem in its full generality is deeper, more general and more complex than the MFIT. In particular, it applies in cases when mating is not at random and also when the fitness of any individual depends on his entire genomic make-up, not simply his genetic make-up at one single locus. In both these cases the MFIT breaks down. Because the FTNS is so general, we defer its exposition and proof to Chapter 2 (for the one locus case) and Chapter 7 (for the many locus case), where the machinery needed for it is developed.

As stated above, the MFIT does not hold as a theorem under non-random mating and when fitnesses depends on the genes at many loci. The fact that the MFIT does not hold when mating is not at random, implying non-Hardy-Weinberg frequencies in the parental generation, is immediately apparent. Suppose that the fitness of A_1A_1 is 1, that the fitness of A_1A_2 is 0.6 and the fitness of A_2A_2 is 1, and that some form of non-random mating has occurred so that in some parental generation, half the individuals in the population are A_1A_1 and half are A_2A_2 . Then the mean fitness is 1, and if mating is such that heterozygotes appear in the daughter generation, the mean fitness will decrease. Thus in this case the MFIT breaks down, as a mathematical theorem.

It is less immediately apparent that decreases in mean fitness can arise even under random mating if the fitness of any individual depends on the alleles at several loci. This case required a more complex analysis than that considered here, and is deferred to Chapter 7.

1.4.2 *Non-Random-Mating Populations*

Essentially all the theory above, and indeed the theory in most of this book, assumes a random-mating population. This reflects in part the theory in the literature as it now exists, and also a focus on animal populations. However the human population does not mate at random, and it is thus relevant to consider, at least briefly, some of the consequences to the theory when a population does not mate at random.

There are many forms of non-random mating, and here we consider one which brings out some of the salient features of this form of mating. Suppose

that the frequencies of the three genotypes in some parental generation are as given in (1.44).

$$\begin{array}{ccc} & A_1A_1 & A_1A_2 & A_2A_2 \\ \text{frequency} & X_{11} & 2X_{12} & X_{22} \end{array} \quad (1.44)$$

Suppose now that an individual mates specifically with an individual of the same genotype with probability f , and mates at random, possibly with an individual of the same genotype, with probability $1 - f$. By considering all possible matings, their frequencies and their genetic outputs, it is found that the genotype frequencies in the daughter generation are given as in (1.45).

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ f(X_{11} + \frac{1}{2}X_{12}) & fX_{12} & f(\frac{1}{2}X_{12} + X_{22}) \\ + (1 - f)x^2 & +(1 - f)x(1 - x) & +(1 - x)^2. \end{array} \quad (1.45)$$

Here $x = X_{11} + X_{12}$ is the frequency of A_1 in the parental generation.

The daughter generation values can be used for several purposes. First, they show that the frequency of the allele A_1 in daughter generation is the same as that in the parental generation. Thus this frequency remains constant throughout the evolutionary process. Second, they can be updated to find the various genotype frequencies in the following generation. Finally, by equating parental and daughter generation genotype frequencies we find the asymptotic ($t \rightarrow \infty$) values. This limiting process shows that the asymptotic heterozygote frequency H is given by

$$H = \frac{4(1 - f)x(1 - x)}{2 - f}, \quad (1.46)$$

while the two asymptotic homozygote frequencies are

$$(\bar{A}_1A_1) : x - \frac{1}{2}H, \quad (A_2A_2) : 1 - x - \frac{1}{2}H. \quad (1.47)$$

All these genotype frequencies are positive, and their values confirm that the asymptotic frequencies of A_1 and A_2 are at the original parental values x and $1 - x$ respectively. Thus in the sense that allelic frequencies are maintained, a central conclusion deriving from of the Hardy–Weinberg law concerning the preservation of genetic variation also holds for this non-random-mating population. One generation of random mating would immediately restore Hardy–Weinberg genotype frequencies. On the other hand, the variation that is maintained is to some extent cryptic, since the heterozygote frequency is less than that applying for a random-mating population with the same allelic frequencies.

Because of the preservation of variation, even though this is to some extent cryptic variation, in the non-random-mating case, we pay comparatively little attention to this case in this book, certainly less than is appropriate.

1.4.3 *The Stochastic Theory*

In this section we consider an aspect of evolutionary behavior which was considered at some length by Fisher, Haldane and Wright, namely the effect of the finite size of the population considered. This finiteness implies that changes in gene frequencies must be viewed as being part of a stochastic, rather than a deterministic, process. It is necessary, in order to arrive at a theoretical estimate of the importance of the stochastic factor, to set up a stochastic model which reasonably describes the behavior of a population in the stochastic case. Perhaps more than in any other part of the theory the choice of a model here is somewhat arbitrary, and we do not pretend that Nature necessarily follows at all closely the models we construct. (Modeling in population genetics is discussed further in Section 1.6.) Although they did not use the terminology of Markov chain theory, the methods used by Fisher and Wright are in fact those of this theory and its close relative, diffusion theory. A brief summary of parts of Markov chain theory is given in Section 2.12. We anticipate here some of the results given in that section, and present the conclusions of Fisher and Wright in the terminology of Markov chains.

We consider, as the simplest possible case, a diploid population of fixed size N . Suppose that the individuals in this population are monoecious, that no selective difference exist between the two alleles A_1 and A_2 possible at a certain locus "A," and that there is no mutation. There are $2N$ genes in the population in any generation, and it is sufficient to center our attention on the number X of A_1 genes. Clearly in any generation X takes one or other of the values $0, 1, \dots, 2N$, and we denote the value assumed by X in generation t by $X(t)$.

We must now assume some specific model which describes the way in which the genes in generation $t+1$ are derived from the genes in generation t . Clearly many reasonable models are possible and, for different purposes, different models might be preferable. We discuss various possible models later in this book: Naturally, biological reality should be the main criterion in our choice of model, but we shall also consider mathematical convenience in this choice. The model which we consider assumes that the genes in generation $t+1$ are derived by sampling with replacement from the genes of generation t . This means that the number $X(t+1)$ is a binomial random variable with index $2N$ and parameter $X(t)/2N$. More explicitly, given that $X(t) = i$, the probability p_{ij} that $X(t+1) = j$ is assumed to be given by

$$p_{ij} = \binom{2N}{j} (i/2N)^j \{1 - (i/2N)\}^{2N-j}, \quad i, j = 0, 1, 2, \dots, 2N. \quad (1.48)$$

While the model in this form was not written down explicitly by Fisher and Wright, it is clear that it was known to Fisher (1921), (1930a) and Wright (1931), who explicitly gave several formulas deriving from it. While

the model apparently originated with Fisher, we follow common practice of honoring both authors by calling it the Wright–Fisher model.

More precisely, we shall refer to the model (1.48) as the “simple” Wright–Fisher model, since it does not incorporate selection, mutation, population subdivision, two sexes or any other complicating feature. The purpose of introducing it is to allow an initial examination of the effects of stochastic variation in gene frequencies, without any further complicating features being involved. More complicated models, such as (1.58), (1.66), (3.68) and (3.72) that introduce factors such as selection, mutation and allow more than two alleles, but which share the binomial sampling characteristic of (1.48), will all be referred to generically as “Wright–Fisher” models.

We emphasize that all of these models are no more than crude approximations to biological reality. This fact is expanded upon in Sections 1.6 and 3.7. Later in this book we will introduce other models having properties different from those of Wright–Fisher models.

In the form of (1.48), it is clear that $X(\cdot)$ is a Markovian random variable with transition matrix $P = \{p_{ij}\}$, so that in principle the entire probability behavior of $X(\cdot)$ can be arrived at through knowledge of P and the initial value $X(0)$ of X . In practice, unfortunately, the matrix P does not lend itself readily to simple explicit answers to many of the questions we would like to ask, and we shall be forced, later, to consider alternative approaches to these questions.

On the other hand, (1.48) does enable us to make some comments more or less immediately. Perhaps the most important is that whatever the value $X(0)$, eventually $X(\cdot)$ will take either the value 0 or $2N$, and once this happens there will be no further change in the value of $X(\cdot)$. Genetically this corresponds, of course, to the fact that since the model (1.48) does not allow mutation, once the population is purely A_2A_2 or purely A_1A_1 , no variation exists, and no further evolution is possible at this locus. It was therefore natural for both Fisher and Wright to find, assuming the model of (1.48), the probability of eventual fixation of A_1 rather than A_2 , and perhaps more important, to attempt to find how much time might be expected to pass before fixation of one or other allele occurs. It is easy enough to see that the answer to the first question is $X(0)/2N$. This conclusion may be arrived at by a variety of methods, the one most appropriate to Markov chain theory being that the solution $\pi_j = j/(2N)$ satisfies (2.141) and its boundary conditions. Setting $j = X(0)$ leads to the required solution. A second way of arriving at the value $X(0)/2N$ is to note that $X(\cdot)/2N$ is a martingale, that is satisfies the “invariant expectation” formula

$$E\{X(t+1)/2N \mid X(t)\} = X(t)/2N, \quad (1.49)$$

and then use either martingale theory or informal arguments to arrive at the desired value. A third approach, more informal and yet from a genetical point of view perhaps more useful, is to observe that eventually every gene in the population is descended from one unique gene in generation zero.

The probability that such a gene is A_1 is simply the initial fraction of A_1 genes, namely $X(0)/2N$, and this must also be the fixation probability of A_1 .

It is far more difficult to assess the properties of the (random) time until fixation occurs. The most obvious quantity to evaluate is the mean time $\bar{t}\{X(0)\}$ taken until $X(\cdot)$ reaches 0 or $2N$, starting from $X(0)$. As it happens, no simple explicit formula for this mean time exists, although, as we see later, some simple approximations are available. Fisher and Wright, no doubt noting this difficulty, paid comparatively little attention to the mean fixation time, concentrating on an approach centering around the leading nonunit eigenvalue of P . It follows immediately from (1.48) that if we put $x(t) = X(t)/2N$,

$$E(x(t+1)\{1-x(t+1)\} | x(t)) = \{1 - (2N)^{-1}\}x(t)\{1-x(t)\}, \quad (1.50)$$

so that the expected value of the heterozygosity measure $2x(\cdot)\{1-x(\cdot)\}$ decreases by a factor of $1 - (2N)^{-1}$ each generation. It follows immediately that $1 - (2N)^{-1}$ is an eigenvalue of the matrix P , and the theory in Appendix A shows that it is the leading nonunit eigenvalue. We write the right and left eigenvectors corresponding to this eigenvalue as $\mathbf{r} = (r_0, r_1, r_2, \dots, r_{2N})$, and $\boldsymbol{\ell}' = (\ell_0, \ell_1, \ell_2, \dots, \ell_{2N})$ respectively. It follows from (1.50) that \mathbf{r}' is proportional to the vector

$$\{0, 2N-1, 2(2N-2), 3(2N-3), \dots, 2N-1, 0\}. \quad (1.51)$$

Unfortunately, no such simple formula exists for the left eigenvector $\boldsymbol{\ell}$. If we suppose that $\boldsymbol{\ell}$ and \mathbf{r} are normalized by the requirements

$$\sum_{k=1}^{2N-1} \ell_k = 1, \quad \sum_{k=0}^{2N} \ell_k r_k = 1, \quad (1.52)$$

then (2.140) shows that

$$\begin{aligned} p_{ij}(t) &= \text{Prob}\{X(t) = j | X(0) = i\} \\ &= r_i \ell_j \{1 - (2N)^{-1}\}^t + o\{1 - (2N)^{-1}\}^t \text{ for } t \text{ large.} \end{aligned} \quad (1.53)$$

Equations (1.50) and (1.53) jointly provide much interesting information. It is clear that especially in a large population, the mean heterozygosity of the population decreases extremely slowly with time as a result of the sampling drift implicit in the process under consideration. We conclude that although genetic variation must ultimately be lost under the model (1.48), the loss is usually very slow. This slow rate of loss may be thought of as a stochastic analogue of the “variation-preserving” property of infinite genetic populations shown by the Hardy–Weinberg law. It is appropriate to quote Fisher (1958, p. 95) on this conclusion: “No result could bring out more forcibly the contrast between the conservation of the variance in particulate inheritance, and its dissipation in inheritance confirming to the blending theory”. We shall generalize this conclusion later, taking into

account not only mutation but also complications brought about through variation in the population size through geographical factors, through the existence of two sexes, and so on.

What can be said about the distribution of $X(t)$ for large t , given $X(t) \neq 0, 2N$? Both Fisher (1958, pp. 90–96) and Wright (1931, pp. 111–116) paid considerable attention to this question. It is clear from (1.52) and (1.53) that

$$\lim_{t \rightarrow \infty} \text{Prob}\{X(t) = j \mid X(t) \neq 0, 2N\} = \ell_j, \quad j = 1, 2, \dots, 2N - 1. \quad (1.54)$$

Furthermore, both Fisher (1958, p. 94) and Wright (1931, p. 113) show that $\ell_j \approx (2N - 1)^{-1}$, so that the asymptotic distribution under consideration is essentially uniform. Although both Fisher and Wright devoted considerable attention to this distribution, and indeed to very accurate expressions for it, especially for very small and very large values of j , it is of far less importance than would appear from the extensive discussion that they devoted to it. The reason for this is that the complete spectral expression for $p_{ij}(t)$, of which (1.53) gives the leading terms and which was unknown to Fisher and Wright, shows that by the time this distribution becomes relevant it is almost certain that fixation or loss of A_1 will already have occurred. This observation, due to Kimura (1955a), will be taken up in more detail later. For the moment we use it to justify our passing over further discussion of this asymptotic distribution.

A more important question, also taken up by Fisher (1958, p. 96) and Wright (1931, p. 116), although in a rather different form than that used later in this book, is the following. Suppose that in an otherwise purely A_2A_2 population, a single new mutant A_1 gene arises. No further mutation occurs, so from this point on the model (1.48) applies. How much time will pass before the mutant is lost (probability $1 - (2N)^{-1}$) or fixed (probability $(2N)^{-1}$)? The mean number of generations \bar{t}_1 for one or other of these events may be written in the form

$$\bar{t}_1 = \sum_{j=1}^{2N-1} \bar{t}_{1,j}, \quad (1.55)$$

where $\bar{t}_{1,j}$ is the mean number of generations that the number of A_1 genes takes the value j before reaching either 0 or $2N$. Both Fisher and Wright found that

$$\bar{t}_{1,j} \approx 2j^{-1}, \quad j = 1, 2, \dots, 2N - 1, \quad (1.56)$$

so that using (1.55),

$$\bar{t}_1 \approx 2(\log(2N - 1) + \gamma), \quad (1.57)$$

where γ is Euler's constant 0.5772... This expression is the C^{-1} of Wright (1931, p. 117); Fisher (1958) found the extremely accurate expression $2(\log(2N - 1) + \gamma) + 0.200645 + O(N^{-1})$, which for large N is correct

to at least 5 decimal places, as well as expressions for $\bar{t}_{1,j}$ that are more accurate, for small j , than those provided by (1.56).

We derive the result (1.57) later (see (5.23)), using methods other than those employed by Fisher and Wright.

There is an ergodic equivalent to the expressions in (1.55) and (1.56) which is perhaps of more interest than (1.55) and (1.56) themselves, and which is indeed the route by which Fisher arrived at these formulas. Consider a sequence of independent loci, each initially purely " A_2A_2 ", and at which a unique mutation A_1 occurs in generation k in the k th member of the sequence. We may then ask how many such loci will be segregating for A_1 and A_2 after a long time has passed, and at how many of these loci will there be exactly j " A_1 " genes. It is clear that the mean values of these quantities are \bar{t}_1 and $\bar{t}_{1,j}$, respectively, and this gives us some idea, at least insofar as the model (1.48) is realistic, of how much genetic variation we may expect to see in any population at a given time. The question of the amount, and the nature, of the genetic variation that can be expected in a population at any given time will be taken up later at much greater length.

Wright (1931, p. 129) and Fisher (1958, p. 99) also considered the modifications to these results when selective differences exist. Again we do not pursue the details of their calculations since we arrive later at their results by other methods. Suppose we assume fitness values of the form (1.25b). Then it is reasonable to replace (1.48) by the model

$$p_{ij} = \binom{2N}{j} (\eta_i)^j (1 - \eta_i)^{2N-j}, \quad i, j = 0, 1, 2, \dots, 2N \quad (1.58)$$

where now

$$\eta_i = \frac{(1+s)i^2 + (1+sh)i(2N-i)}{(1+s)i^2 + 2(1+sh)i(2N-i) + (2N-i)^2}. \quad (1.59)$$

We may again ask what values \bar{t}_1 and $\bar{t}_{1,j}$ assume. This problem was attacked by Fisher and Wright only in the case $h = \frac{1}{2}$. We shall show later, for general values of h , that

$$\bar{t}_{1,j} \approx \frac{2 \int_0^1 \psi(y) dy}{2Nx(1-x)\psi(x) \int_0^1 \psi(y) dy}, \quad (1.60)$$

where $x = j/2N$ and

$$\psi(x) = \exp\{-2\alpha hx + (2h-1)\alpha x^2\}, \quad (1.61)$$

with α defined by $\alpha = 2Ns$. When there is no selection the value of α is 0, so that $\psi(x) = 1$, and the expression in (1.60) reduces to that in (1.56), as we would wish. For the zero dominance case, where $h = \frac{1}{2}$, the

approximation (1.55) reduces to

$$\bar{t}_{1,j} \approx \frac{2(1 - \exp\{-\alpha(1-x)\})}{2Nx(1-x)\{1 - \exp(-\alpha)\}}, \quad (1.62)$$

agreeing with the value given by Fisher (his $4an$ is our α). For $h \neq \frac{1}{2}$ the right-hand side in (1.60) cannot be evaluated explicitly, although clearly numerical approximation is possible. In all cases $\bar{t}_1 = \sum \bar{t}_{1,j}$.

Both Fisher and Wright used the approximation (1.62) to find the probability that a new mutant A_1 will eventually become fixed in the population. Their method, which is quite different from the one we consider later, is as follows. Suppose in (1.62) we put $x = 1 - \delta$ and consider small values of δ . Then (1.62) reduces in effect to

$$\frac{2\alpha}{2N\{1 - \exp(-\alpha)\}}, \quad (1.63)$$

which, as $\alpha \rightarrow 0$, approaches $2/2N$. We now argue that since the probability of fixation of A_1 for the neutral case ($\alpha = 0$) is known to be $(2N)^{-1}$, the probability of fixation in the case we are considering must be given by

$$\text{Prob}(A_1 \text{ fixes}) = \frac{s}{1 - \exp(-2Ns)}. \quad (1.64)$$

This is identical to the value given by Fisher (1958, p. 100) and Wright (1931, p. 133) upon setting our s equal to Fisher's $2a$ and Wright's $2s$.

Equation (1.64) influenced Fisher considerably. He was accustomed to think in terms of very large populations; thus he gave a table of values of \bar{t}_1 (see (1.55)) for values of N ranging from 10^6 to 10^{12} and wrote later of populations of size of a thousand million as though they were typical. The ratio of the right-hand side in (1.64) to the value $(2N)^{-1}$ applying for the case $s = 0$ is

$$\alpha/(1 - \exp(-\alpha)), \quad (1.65)$$

and for the values $\alpha = -4, 0$ and 4 this ratio takes the values $0.08, 1$ and 4 . Thus, as noted by Fisher, increasing α from -4 to $+4$ increases the probability of fixation of A_1 by a factor of about 50 . Thus in a population of size 10^9 , only a minute range of selective differences around zero lead effectively to the same fixation probability as for complete selective equivalence. As an alternative way of noting this, an increase in s from 0 to 10^{-6} increases the probability of fixation of A_1 by a factor of $2,000$ in a population of this size. These considerations strongly influenced Fisher in arriving at the view that selective differentials are of paramount importance in determining the genetic evolutionary behavior of populations, and that the randomness in the behavior of gene frequencies brought about by the finite nature of the population size in no way seriously undermines the Darwinian theory.

There were two reasons why Wright was less influenced than was Fisher by formulas such as (1.65). First, he was accustomed to think in terms

of population sizes far smaller than 10^9 . His view of the optimal circumstance under which evolution occurs, which we consider in more detail later, was rather different from Fisher's, and involves random changes in gene frequencies in populations of comparatively small size as one significant component. Second, Wright considered comparatively short-term behavior whereas Fisher was accustomed to focus on very long-term behavior, for which comparatively short-term stochastic effects are eventually dominated by the long-term effects of selective differences. We return to this comparison of emphases later.

A further problem of an essentially stochastic nature, considered almost exclusively by Wright (1930, pp. 133–134), concerns the stationary distribution of the frequency of A_1 when, in addition to the changes in frequencies brought about by selection and the random changes due to the finite nature of the population size, we allow mutation from A_1 to A_2 (at rate u) and from A_2 to A_1 (at rate v). In this case we may reasonably replace the transition probability (1.58) by

$$p_{ij} = \binom{2N}{j} (\eta_i^*)^j (1 - \eta_i^*)^{2N-j}, \quad (1.66)$$

where η_i^* is given by

$$\eta_i^* = (1 - u)\eta_i + (1 - \eta_i)v, \quad (1.67)$$

η_i being defined by (1.59). If we put $x = X(\cdot)/2N$, Wright showed in effect that the stationary distribution of x is of the form

$$f(x) = \text{const } x^{4Nv-1}(1-x)^{4Nu-1} \exp\{2\alpha h x - (2h-1)\alpha x^2\}, \quad (1.68)$$

the constant being chosen so that $\int_0^1 f(x)dx = 1$.

When the heterozygote is at a selective advantage it is perhaps better to use the fitness parameters (1.25c) to arrive at the equivalent formula

$$f(x) = \text{const } x^{4Nv-1}(1-x)^{4Nu-1} \exp\{2\alpha_2 x - (\alpha_1 + \alpha_2)x^2\}, \quad (1.69)$$

where $\alpha_i = 2Ns_i$. In these formulas the relative effects of the population size, the selective coefficient and the mutation rate on the form of the distribution can be ascertained. Thus if mutation rates are sufficiently small so that ($4Nu < 1, 4Nv < 1$), some accumulation of probability occurs near $x = 0$ and near $x = 1$. This does not, however, necessarily mean that most of the mass of the probability distribution is near these points, and it is quite possible that the most likely values for x are determined more by selection than by mutation.

As an example we consider the case $N = \frac{1}{4} \times 10^5$, $u = v = 5 \times 10^{-6}$ and, in the notation (1.25c), $s_1 = s_2 = 2 \times 10^{-3}$. Inserting these values in (1.69) we arrive at the stationary distribution

$$f(x) = Cx^{-1/2}(1-x)^{-1/2} \exp 200x(1-x) \quad (1.70)$$

for the frequency of A_1 . The constant C is again chosen so that $\int_0^1 f(x)dx = 1$. To compare the effects of mutation and selection we compare the integral of the density functions over two small sub-intervals, one near 0 and the other near $\frac{1}{2}$. Thus we find, for example, that the probability that the frequency x of A_1 is less than 0.0001 or greater than 0.9999 is approximately

$$2C \int_0^{0.0001} x^{-1/2} dx \approx 0.04 C, \quad (1.71)$$

while the probability that x is between 0.4999 and 0.5001 is approximately

$$0.0004 C \exp(50). \quad (1.72)$$

This is about 10^{22} times larger than the value given in (1.71), and indicates that in this case the selective forces have a far greater influence on the likely values that x will assume than have the mutation rates. Although this example has a high degree of symmetry implicit in it, a parallel result will hold for asymmetric cases where the selective coefficients and the mutation rates are of the same order of magnitude as those in this example. Thus if $u = 5 \times 10^{-6}$, $v = 10^{-5}$, $s_1 = 10^{-3}$, $s_2 = 2 \times 10^{-3}$, selection is again far more important than mutation in determining the likely values of x . In general this conclusion will hold so long as the selective differentials are at least 100 times larger than the mutation rate. If in the above example $s_1 = s_2 = 2 \times 10^{-4}$, the probability of a value of x less than 0.0001 or greater than 0.9999 is of the same order of magnitude as the probability of a value between 0.4999 and 0.5001, while if $s_1 = s_2 = 2 \times 10^{-5}$, the former probability is rather larger than the latter.

As a particularly important application of stochastic process theory, Fisher (1922), Haldane (1927b) and Wright (1931) all considered the specific problem of the probability of survival of a single new favorable mutant allele. This probability has already been computed, for the case of selection without dominance, in (1.64). A rather different approach, using the theory of branching processes, may be used to approximate this probability, and it is some interest to outline the elements of this method. To do this we follow the treatment of Fisher (1930a).

We consider a population with nonoverlapping generations, the various generations existing at a sequence of time points 0, 1, 2, 3, ..., and suppose X_n genes (or "individuals") at time n . Each of these X_n individuals gives rise to a number of offspring individuals and then dies. At time $n + 1$ each of these offspring in turn produces offspring, and so on. We suppose a given fixed distribution for the number of offspring for each individual and that the numbers of offspring to those individuals alive at any given time are independent. The values X_0, X_1, X_2, \dots , form a Markov chain: In this branching process Markov chain model no fixed upper limit can be set to the values of the X_i . We suppose that each individual leaves i offspring

with probability p_i , and introduce the generating function

$$p(z) = p_0 + p_1z + p_2z^2 + \cdots, \quad (1.73)$$

where z is a dummy variable. Clearly, if the mean and variance of the distribution $\{p_i\}$ are denoted μ and σ^2 , we have

$$p(1) = 1, \quad p'(1) = \mu, \quad p''(1) = \sigma^2 - \mu + \mu^2. \quad (1.74)$$

We assume that the branching process starts with one individual in generation 0 (that is, $X_0 = 1$). Then the generating function of the number of individuals in generation 1 is $p(z)$, and for generation 2 can be found in the following way. We have

$$\text{Prob}\{X_2 = i\} = \sum_j \text{Prob}\{X_2 = i \mid X_1 = j\} \times \text{Prob}\{X_1 = j\}. \quad (1.75)$$

Given that $X_1 = j$, the probability that $X_2 = i$ is evidently the coefficient of z^i in $\{p(z)\}^j$. Thus from (1.75)

$$\begin{aligned} \text{Prob}\{X_2 = i\} &= \text{coeff } z^i \text{ in } \sum_j \{p(z)\}^j p_j \\ &= \text{coeff } z^i \text{ in } p(p(z)). \end{aligned}$$

It follows that the generating function of the distribution of X_2 is $p(p(z))$ and, more generally, the generating function of the distribution of X_n is the n th functional iterate $p_n(z)$, defined by

$$p_n(z) = p(p_{n-1}(z)) = p_{n-1}(p(z)). \quad (1.76)$$

Fisher (1930a) was interested in three quantities. The first is the probability π_n that $X_n = 0$, the second is the limiting value of π_n as $n \rightarrow \infty$, and the third the conditional probability distribution of X_n for n large, given $X_n \neq 0$. By setting $z = 0$ in (1.71) we see immediately that π_n satisfies the functional relation

$$\pi_{n+1} = p(\pi_n), \quad n = 1, 2, 3, \dots, \quad (1.77)$$

with $\pi_0 = 0$. By letting $n \rightarrow \infty$ in (1.77), we see that limiting value π of π_n satisfies

$$\pi = p(\pi), \quad (1.78)$$

and it is not hard to show that the required value π is the smallest positive root of (1.78). Putting $\pi = 1 - \delta$ (δ small), a Taylor series expansion in (1.78) yields

$$1 - \delta \approx 1 - \delta p'(1) + \frac{1}{2} \delta^2 p''(1), \quad (1.79)$$

and if $\mu = 1 + \epsilon$ (ϵ small, positive), (1.74) and (1.79) show that

$$\delta \approx \frac{2\epsilon}{\sigma^2}. \quad (1.80)$$

We shall defer consideration of the conditional distribution of X_n ($X_n \neq 0$) for a moment and examine it only in a case of particular genetic interest.

We turn now to the application of these results in genetics, following the approach used by Fisher (1930a). Consider the case of a nonrecessive A_1 mutant gene introduced into a previously purely A_2A_2 population. Homozygotes A_1A_1 will not usually appear until the number of A_1 genes is comparatively large (of order \sqrt{N} , where N is the population size) and by this time the fate of the new mutant, that is whether it will die out or not, is usually in effect settled. Thus although it is clear that the assumptions made in the theory of branching processes are not exactly met for populations of fixed size, it should be possible using this theory to obtain rather close approximations to several quantities of evolutionary interest, increasing in accuracy as $N \rightarrow \infty$. If this is done, the expression "survival of a new mutant" is then taken to mean the increase in the frequency of a mutant to a point where the probability of loss of the mutant by accidents of sampling in anything other than a very long time may safely be ignored. We have in mind in particular either the fixation of the mutant in the population or the attainment of a quasi-stable equilibrium point determined, for example, by heterozygote selective advantage.

We are mainly interested in establishing results for populations of stable size, and by convention we do this by using the fitness scheme (1.25b), where the values are now taken as absolute fitnesses. We thus identify the unit fitness of the prevailing genotype A_2A_2 with stable population size. Assuming the model (1.58), we may reasonably suppose each mutant A_1 gene produces a random number of A_1 "offspring" according to the binomial distribution with index $2N$ and parameter $(1 + sh)/2N$. To a sufficient approximation we may replace this distribution by a Poisson distribution with parameter $1 + sh$. In this case the generating function (1.73) becomes

$$p(z) = \exp\{(z - 1)(1 + sh)\}, \quad (1.81)$$

and the approximation (1.80) yields

$$\delta \approx 2sh. \quad (1.82)$$

For $h = \frac{1}{2}$ this agrees with the value (1.64) found by diffusion methods, at least for values of N sufficiently large so that $\exp(-Ns)$ may safely be ignored. This confirms the view that the branching process approximation is most accurate for large N . Equation (1.77) becomes

$$\pi_{n+1} = \exp\{(\pi_n - 1)(1 + sh)\}, \quad (1.83)$$

an equation which may be iterated numerically to provide values of π_n for any value of n . This was done by Fisher for $s = 0$ and for $sh = 0.01$ and the numerical values found confirm the approximation (1.82), which Fisher did not use explicitly.

The case $s = 0$ is of particular interest. Here

$$\pi_{n+1} = \exp(\pi_n - 1). \quad (1.84)$$

Since $\pi_n \rightarrow 1$ as $n \rightarrow \infty$ it is interesting to attempt an approximate solution of (1.84) in the form

$$\pi_n \approx 1 - cn^{-1}.$$

Insertion of this trial value into (1.84) gives $c = 2$ and hence

$$\pi_n \approx 1 - 2n^{-1}. \quad (1.85)$$

This value was given by Fisher from inspection of the numerical iteration (1.84).

We turn finally to the conditional distribution of X_n , given $X_n \neq 0$ for the case $s = 0$. Here we merely outline Fisher's conclusion. It is clear that the unconditional mean of X_n is unity, and hence from (1.85) the conditional mean of X_n (given $X_n > 0$) is approximately $\frac{1}{2}n$. It is thus reasonable to consider the normalized variable $y_n = X_n/n$ which, given $X_n > 0$, we may hope will possess a limiting distribution as $n \rightarrow \infty$. By using generating function techniques, Fisher showed that the limiting ($n \rightarrow \infty$) distribution of y_n is

$$f(y) = 2 \exp(-2y), \quad y > 0, \quad (1.86)$$

so that in particular

$$\text{Prob}(X_n > kn) = \text{Prob}(y > k) \approx \exp(-2k). \quad (1.87)$$

In the case $sh > 0$, Fisher proved that the conditional distribution of $y = X_n/(1 + sh)^n$, given $X_n > 0$, is asymptotically

$$f(y) = 2sh \exp(-2shy), \quad y > 0. \quad (1.88)$$

Thus

$$\text{Prob}(X_n > X(1 + sh)^n \mid X_n = 0) = \text{Prob}(y > X \mid y > 0) \approx \exp(-2Xsh). \quad (1.89)$$

It should be emphasized that these conclusions, while they are arrived at by considering indefinitely large values of n , nevertheless apply only if the numbers of mutants involved is far less than the population size N , for it is only for such values that branching process approximations are legitimate. This is true particularly of equation (1.89).

What evolutionary conclusions can be drawn from these calculations? The first, and perhaps most important, is that while the survival probability (1.82) is small, it is nevertheless positive. Thus while the lines initiated by most favorable mutations will die out, and usually rather rapidly, the eventual survival of a favorable mutant is certain if mutation is recurrent. Thus taking the case $s = 0.01$, $h = \frac{1}{2}$, a mutation rate of 10^{-6} in a population of size 10^8 will produce 200 mutations per generation, and the probability that none of the mutational lines initiated in just the first generation survive is only $(0.99)^{200} \approx 0.14$. In a larger population, or with a larger mutation rate, this probability is diminished even further. It follows

that in large populations a favorable new mutant will begin to establish itself rather soon after mutation to it commences. We may then use equations such as (1.28) to consider how long various degrees of establishment will require. On the other hand, in small populations, and even more important with unique mutational events, the small individual probability of survival of the line initiated by a single mutant is a factor which must be incorporated into evolutionary considerations.

A second observation concerns the origin and potential selective advantage of a mutant which has spread to large numbers in a population. We may take as a numerical example a population of size 10^7 containing 10^5 A_1 genes. If these genes enjoy no selective advantage and arose from a single mutational event, (1.87) shows that the mutation most likely occurred at least 10^5 generations in the past. However, if the mutation to the allele in question is recurrent, the average time required for the current frequency 10^5 is rather less, while if the mutant possesses a selective advantage its present frequency can be explained by a comparatively rapid recent increase in numbers.

A final comment concerns populations whose sizes are not stationary. Any mutant in a population of uniformly increasing size will have its survival probability increased. We consider as an example a new mutant having selective advantage 0.01 arising in a population of 10^4 . Suppose now that the population doubles in size for eight generations and stabilizes at a size of 256×10^4 . If the doubling in population size were to continue indefinitely, the new mutant would have a probability π of loss satisfying the equation

$$\pi = \exp(2.02(\pi - 1)),$$

the solution of which is $\pi = 0.1978$. When doubling stops after eight generations the probability of loss of the mutant is rather greater than this, being approximately 0.3. A converse comment applies for mutants in decreasing populations. Thus populations that are increasing in size should exhibit some variety of forms compared to populations that have a stable size or are decreasing in size. The variety will perhaps diminish once stability of population size is reached, since some unfavorable mutants which increased in numbers because of the increase in population size will now die out. In practice, of course, any protracted increase in size must occur at a rather low rate, and thus this argument applies most to mutations whose selective advantage or disadvantage is rather small.

1.5 Evolved Genetic Phenomena

In the previous section we have asked the question: Assuming the Mendelian genetic scheme and given the numerical values of various genetic parameters, for example mutation rates, the degree of dominance, what conclusions can be drawn about evolutionary processes? Fisher, Wright and Haldane

also asked a converse question, namely: Given that evolution has occurred, what purely genetic characteristics can be explained as a result of this evolution? Perhaps the most interesting such questions concern mutation rates, dominance, linkage intensities, and the sex ratio, while on a broader level the existence of sexual dimorphism, a Mendelian phenomenon, and even the pervasiveness of the Mendelian scheme itself, can be considered. Here we limit attention to brief comments on the first four topics, again restricting attention to the work done in the pioneering period we are considering.

We have alluded already to the question of observed rates of mutation and the possibility that these are the results of evolutionary processes whereby the contrasting requirements of a low mutation rate, to preserve such favorable gene complexes as have been built up, and a high mutation rate, so that a large number of potentially or actually favorable new mutations will arise, are optimally balanced. It is difficult to quantify this argument, and no real attempt to do so was made during the time we are considering. Of course one must avoid the assumption that all presently observed genetic phenomena are at some sense at optimal values: it is certainly possible to argue that current mutation rates are partly the result of extrinsic factors having nothing to do with evolution, or at least that while they no doubt vary from locus to locus and time to time and are capable of some evolutionary modification, they are not presently at optimal evolutionary values.

We turn next to the question of dominance. Fisher argued that dominance is the outcome of an evolutionary process through an induced selection of modifier genes at loci other than the primary one under consideration. He was strongly influenced in this view by the observation that it is normally the prevailing wild-type allele that is dominant, so that in the course of its becoming the prevalent type it presumably acquired the dominance property. We consider the details of this argument in Section 6.5, and for the moment we only introduce the elements of the analysis.

We consider two alleles A_1 and A_2 at a locus and assume the fitness scheme of the form (1.25b). If A_1 mutates to A_2 at rate u we may suppose that the frequency of A_1 is at the mutation-selection equilibrium point (1.34). Suppose now that at a locus M , at which the allele M_2 was previously fixed, a mutant allele M_1 arises with the effect that those A_1A_2 individuals carrying the allele M_1 are altered in phenotypic expression towards that of the prevailing homozygote A_1A_1 . We assume that fitness is determined by the phenotype so that the fitness scheme takes the following form:

	A_1A_1	A_1A_2	A_2A_2	
M_1M_1	1	1	$1 - s$	
M_1M_2	1	$1 - sk$	$1 - s$	(1.90)
M_2M_2	1	$1 - sh$	$1 - s$	

Here $s > 0$ and $0 \leq k \leq h \leq 1$. Clearly M_1 is at an induced selective advantage to M_2 and will steadily increase in frequency to unity, bringing about dominance of A_1 over A_2 .

Several qualifications should be made about this argument. Perhaps the most important is that we have ignored any possible selective differences between M_1 and M_2 which might arise for reasons quite separate from dominance modification at the A locus. Clearly the rate of change in the frequency of M_1 though dominance modification is very small, since the selective superiority of M_1 over M_2 through this agency arises only in the comparatively rare heterozygotes A_1A_2 . It would require only a minute selective advantage of M_2 over M_1 for other reasons to overcome this. Wright (1929a,b) was strongly influenced by this argument in forming his doubts about Fisher's theory. Wright's view on evolution, which we shall examine more closely in Section 1.7, was centered around the assumption of an almost universal interactive effects of genes, so that the fate of any allele is determined by the net selective force acting on it, the direction of this force being normally determined by factors more important than dominance modification. Fisher, on the other hand, believed that the selective advantage due to dominance modification would ultimately be effective. We examine his argument in more detail in Section 6.5.

Wright (1934) put forward the more purely physiological view that dominance is a natural pristine characteristic, rather than an evolved characteristic, of an allele. We do not go into detail of this argument here. It is sufficient to note that the theory recognizes the role of genes in controlling the production of enzymes, which act as catalysts in physiological processes, and that one gene may well produce sufficient enzyme for a certain process so that no further effect is produced by a second gene. The reader is encouraged to read Fisher's and Wright's original papers on this matter, since in no other way than by reading them can the flavor of their long dispute on this matter, and its bearing on their respective evolutionary viewpoints, be appreciated.

We consider next the question of linkage modification. The circumstances under which Fisher envisioned the evolution of close linkage between two loci (see for example Fisher (1958, p. 116)) occur when, at two loci A and B , the allele A_1 is favored in the presence of B_1 while A_2 is favored in the presence of B_2 . This will imply that the double heterozygote A_1B_1/A_2B_2 will occur more frequently than the double heterozygote A_1B_2/A_2B_1 and that a recombination between A and B loci will break down the former in greater absolute numbers than they are formed by recombination from the latter. Thus a higher recombination fraction will lead to a greater breakdown of the "favored" gametes A_1B_1 and A_2B_2 and hence to a decrease in the mean fitness of the population. It is convenient to give an example of the form of fitness schemes envisioned for such a process. One set of

fitnesses having the desired characteristics is of the form

$$\begin{array}{cccc}
 & B_1B_1 & B_1B_2 & B_2B_2 \\
 A_1A_1 & 1 & 1 - a & 1 - 4a \\
 A_1A_2 & 1 - a & 1 & 1 - a \\
 A_2A_2 & 1 - 4a & 1 - a & 1
 \end{array} \tag{1.91}$$

This fitness scheme was introduced by Wright (1952) and considered in some detail by him for purposes other than that of present interest.

An analysis of the evolutionary behavior of cases where, as in (1.91), the fitness of any individual depends on his genetic constitution at more than one locus is more complicated than the single-locus analysis considered so far, and is discussed in some detail in Chapter 6. For the moment we simply present the result of this analysis as it applies to the model (1.91) and also, below, as it applies to the model (1.92).

The evolutionary behavior of a population for which the fitnesses are as given in (1.91) is not as simple as one might initially expect. It can be shown that for any value of the recombination fraction R between A and B loci ($0 < R \leq \frac{1}{2}$), there is an equilibrium point of gamete frequencies with all frequencies positive. However, this equilibrium is never stable. In other words, a fitness scheme of the form (1.91) cannot maintain a stable genetic polymorphism at either A or B locus and is thus of no use for considering the argument in question.

Another fitness scheme with fitnesses of the general desired form is

$$\begin{array}{cccc}
 & B_1B_1 & B_1B_2 & B_2B_2 \\
 A_1A_1 & 1 & 1 - a & 1 - 2a \\
 A_1A_2 & 1 - a & 1 + 2a & 1 - a \\
 A_2A_2 & 1 - 2a & 1 - a & 1
 \end{array} \tag{1.92}$$

This fitness scheme leads to an equilibrium point with

$$\begin{aligned}
 \text{freq}(A_1B_1) &= \text{freq}(A_2B_2) = c^* \\
 \text{freq}(A_1B_2) &= \text{freq}(A_2B_1) = \frac{1}{2} - c^*
 \end{aligned} \tag{1.93}$$

where c^* is the unique solution in $(\frac{1}{4}, \frac{1}{2})$ of the equation

$$12ac^3 - 8ac^2 + ac + R(1 + 2a)(c - \frac{1}{4}) = 0. \tag{1.94}$$

It is easy to verify geometrically that c^* increases as R decreases and that $c^* \rightarrow \frac{1}{2}$ as $R \rightarrow 0$.

We turn next to the equilibrium value of the mean fitness \bar{w} , considered as a function of c^* . This is

$$\bar{w} = 1 - 4ac^* + 12a(c^*)^2, \tag{1.95}$$

and since this is an increasing function of c^* for $\frac{1}{4} < c^* < \frac{1}{2}$, we conclude that the equilibrium value of \bar{w} is smallest when R is large and largest when R is small.

We show later that the equilibrium (1.95) is stable, at least for small R , and thus we have shown that for small R at least, the stable equilibrium mean fitness decreases as the recombination fraction between the loci increases.

Fisher now argued that if “different strains” have different recombination fractions, the strain with the smallest value will, because of its higher mean fitness, tend to replace the others, so that tight linkage will have evolved in the population. This argument, involving the new concept of interpopulational selection, will be considered further, with arguments not involving this form of selection, in Section 6.5.

The final characteristic we consider is the sex ratio. Fisher’s argument on this is curiously non-genetic in the sense that it could well have been made in pre-Mendelian times. The argument involves the introduction of the concept of “parental expenditure”, which does not initially appear to be a necessary, or indeed the most obviously appropriate, vehicle for explaining the sex ratio. The argument is that each offspring receives, while young, a certain expenditure on the part of its parents. Consider now a cohort of such offspring about to embark on reproduction. The males in this cohort will supply exactly half the ancestry of the descendants of this cohort, as will of course the females. Suppose now that the total parental expenditure on behalf of males is less than that of females. Then parents having the tendency to produce male offspring in excess will, for the same expenditure, tend to contribute disproportionately to the ancestry of subsequent generations. Since the same argument in reverse would apply if the expenditure on females were less, selection will tend to change the sex ratio to the point where an equal expenditure is made on female and male offspring. If now males suffer a heavier pre-adult mortality, then as compared to females more of this expenditure will take place for males who die early and do not participate in reproduction. It follows that the sex-ratio of males to females should exceed unity at birth but be lower than unity at the age of reproduction.

This argument leads to an evolutionary adjustment of the sex ratio. Whether the various assumptions implicit in it are valid is uncertain, and what appears to be a superior verbal argument, the consequences of which is that the sex ratio should be unity at the time of conception, is given in Crow and Kimura (1970, pp. 288–289). We examine an argument parallel to Fisher’s, but based more firmly on genetic concepts, in Chapter 8.

1.6 Modelling

Much of the discussion in previous sections concerns the analysis of some model. That is, some set of assumptions, usually incorporating mathematical formulas, is constructed attempting to describe the real-world process

or phenomenon being considered. The model is then analyzed by mathematical or other methods to find its properties, and the implications of these in the real world are then discussed. It is thus worthwhile to discuss, albeit briefly, the modeling process in mathematical evolutionary genetics.

The concept of mathematical modeling in biology was inherited from the very successful modeling process in physics. But the two respective natures of the modeling process in the two areas are quite different. In physics one aims at, and largely achieves, mathematical models that describe the real world very precisely, based for example on Newton's laws of dynamics. This allows, for example, the calculation of trajectories of space vehicles so that they arrive precisely at some desired location. No such precision is possible in evolutionary genetics. The biological world is too complex, and unpredictable phenomena ranging from mutations to large-scale ecological events are so prevalent, that no precise prediction of the course of evolution is possible. Nevertheless it is possible by using mathematical models to arrive at general principles that do lead to important evolutionary conclusions. The discussion above following the Hardy–Weinberg law is an example of this, and other examples will be given later in this book.

Even though mathematical models in evolutionary population genetics cannot hope to describe the real world with the precision that is often possible in physics, it is nevertheless important that any mathematical model used be well-defined and consistent, containing no internal contradictions. Further, no *ad hoc* assumptions, which can possibly contradict the implicit properties of the model, should be made during the course of the analysis of the implications of the model, since doing so can in principle lead to reaching any conclusion whatsoever.

More important, any mathematical model should aim at capturing the essential features of reality so that the conclusions drawn from it are useful. This was well known to the pioneers, who showed great skill in devising models that do this. Unfortunately, one aspect of the evolutionary process with which they were quite familiar was not sufficiently emphasized by them, and this has led to a recurring error by a succession of analysts, not usually geneticists, concerning the possibility of the evolution of the complex life forms that we see today by the Darwinian-Mendelian process. This error follows from an inappropriate model of the evolutionary process. The essence of the error can be seen from the following oversimplified example.

Suppose that we wish to attain some desired sequence of 19 letters, for example THEGREATWALLOFCHINA. Here we might think of the first letter, "T", as the first desired gene in a sequence, the second letter "H" as the second gene in the desired sequence, and so on. The incorrect model for an evolutionary process arriving at this sequence is as follows. Suppose that we randomly choose 19 letters. If they happen to form the desired sequence, we have evolved in one step to the desired sequence. However the probability of doing this is minute, being $(26)^{-19}$. In the much more likely event that we did not form the desired sequence, the first sequence is

entirely discarded and a new sequence of 19 letters is formed. We continue in this way until the desired sequence of genes happens to be reached, a procedure taking a mean of $(26)^{19}$ steps. Even at one step per second, this mean time is far longer than the time since the Big Bang.

But this is an incorrect model of evolution. Assuming that each of the letters, or genes, in the desired sequence is itself desirable, a more plausible model is that after the first random sequence has been chosen, any letter in this sequence that happened to match the corresponding letter in the desired sequence is retained. At the second step, a random choice is then made for those letters that did not match the desired sequence. Any letters obtained at this second step that match the corresponding letter in the desired sequence are retained, along with any that were retained at the first step. This process continues until the correct letter is obtained at all locations, a process taking on average only a few hundred steps.

While this second process is still a very crude representation of reality, it does model the genetic evolutionary process more appropriately than does the first process. A gene that is good for vision is not thrown out, but is retained, while a gene that is good for some other function evolves.

Clearly evolution is not aiming at some a priori target, but might arrive at the equally effective ATITANICCHINESEWALL instead of the sequence above. This does not affect the broad conclusion of the above argument.

It is a pity that a small proportion of scientists, often outside the field of genetics, regularly re-invent the incorrect modeling paradigm, since the negative views of the possibility of evolution that they form are then seized upon by creationists as support for their arguments. Fisher, Haldane and Wright all described the correct paradigm, or model, quite clearly, but unfortunately their message was not sufficiently absorbed into the theory, nor into scientific circles generally.

At a more minor level, there are other aspects of modeling theory that are often overlooked within the population genetics literature. Perhaps unfortunately, the simple Wright–Fisher model discussed at length in Section 1.4.3, has assumed a “gold standard” status, and serves as a reference distribution for several calculations in population genetics theory. This has arisen largely for historical reasons, and the fact that this is only one model among many, and is far less general and plausible than the Cannings model discussed in much detail later, is generally overlooked. We mention two examples where the fact that the Wright–Fisher model is no more than a reference model has been often overlooked, with unfortunate consequences.

First, the concept of the “effective population size”, discussed in more detail in Section 3.7, is defined with reference to the simple Wright–Fisher model (1.48). A certain model has effective population size N_e if some characteristic of the model has the same value as the corresponding characteristic for the simple Wright–Fisher model (1.48) whose actual size is N_e . Further, the comparison of several characteristics are possible, and this leads to different varieties of effective population size. Except in simple

cases, the concept is not directly related to the actual size of a population. For example, a population might have an actual size of 200 but, because of a distorted sex ratio, have an effective population size of only 25. This implies that some characteristic of the model describing this population, for example a leading eigenvalue, has the same numerical value as that of a Wright–Fisher model with a population size of 25. It would be more indicative of the meaning of the concept if the adjective “effective” were replaced by “in some given respect Wright–Fisher model equivalent”. Misinterpretations of effective population size calculations frequently follow from a misunderstanding of this fact. The concluding comments of Section 3.7 discuss this point at length.

Second, the fundamental genetic parameter θ will be introduced in Section 3.6 in the discussion of the Wright–Fisher model (3.72). For that model θ assumes the value $4Nu$, and the identification of θ and $4Nu$ is very common in the literature. However, for models other than Wright–Fisher models a different definition of θ is needed. This is particularly true of the exchangeable model of Cannings (1974) introduced in Section 3.3, which provides a most important generalization of Wright–Fisher models, and is also true of the Moran model introduced in Section 3.4. Much of the discussion in Chapter 9 refers to this point. The identification of θ with $4Nu$ arises in effect from an inappropriate assumption that the simple Wright–Fisher model (1.48) is the stochastic evolutionary model relevant to the situation at hand. The rather more general definition of θ as $4N_e u$ partly overcomes this problem, but does not do so entirely, since (as mentioned above) there are several distinct concepts of the effective population size N_e .

1.7 Overall Evolutionary Theories

We now outline the two contrasting views of evolution arrived at by Fisher and Wright.

Fisher’s view was focused on the long term. With this perspective, his evolutionary view in a way a simple one. He considered populations to be very large: The numerical values used in (Fisher (1930a)) for population size are often of order 10^9 or larger. Thus apart from the particular case of the probability of survival of an individual new mutant, stochastic effects are not regarded as being of central importance, and deterministic analyses are seen as being sufficient to describe the essence of evolutionary behavior. Even in the case of new mutants, where a stochastic analysis is unavoidable, we have seen that essentially deterministic behavior arises for recurrent mutations in large populations. Thus, once the genetic raw material has been furnished by mutation, natural selection is regarded as the sole important agency in shaping genetic evolution.

The nature of this selection is also seen as being rather straightforward. In the first place, since complexes of genes at various loci, even if harmonious, tend to be broken up ultimately by recombination, a stronger emphasis is placed on genes at single loci than that placed on gene complexes. This is not to deny the fact that as we have just seen, Fisher viewed interactive systems as being important. But, for example, so far as evolutionary processes are concerned, the effect of an interactive system such as (1.92) simply has the effect of yielding a selective advantage to M_1 over M_2 , and the primary emphasis is placed on this fact. This leads to the point of view (Fisher (1953)) that “it is often convenient to consider a natural population not so much as an aggregate of living individuals but as an aggregate of gene ratios”. Fisher would have regarded this view as an approximation, but one which is nevertheless sufficient to describe the main characteristics of evolution. This view pervades, directly or indirectly, his work not only in population genetics but also, interestingly enough, in the statistical theory of experimental design (see, for example, Fisher (1926, p. 511)), which was strongly influenced by, if indeed not suggested by, his research in genetics.

In population genetics a corollary of this view is that frequencies of gametes can be found, at least to a sufficient approximation, as the product of the frequencies of the constituent alleles. This approximation is implicit in his pioneer work in both quantitative and evolutionary genetics, except in special cases involving, for example, assortative mating. Thus, for example, in both fields the total additive genetic variance, a quantity of central importance, appears to be defined by him as the sum of the constituent one-locus marginal values (Fisher (1918, p. 405; 1958, p. 37)). We shall see later that while this is correct if indeed gamete frequencies can be so calculated, it is not generally so.

A further characteristic of Fisher’s evolutionary view, arising from the above considerations and the assumed very large sizes of populations, is that an allele having a net selective advantage, no matter how small, is destined for fixation, at least while the selective advantage persists. Thus, for example, one of his main objectives in putting forward his theory of the evolution of dominance through the natural selection of modifiers was to show that even a minute selective force would have evolutionary consequences. This was seen as being so even if the modifiers are subject to selective forces other than through dominance modification. Fisher’s reasoning on this point, (in particular Fisher (1934, pp. 372–373)) is not clear to this writer, who shares Wright’s (1934) doubts on its acceptability.

Against these views should be set the fact that Fisher’s Fundamental Theorem of Natural Selection, which we examine in detail in Sections 2.9 and 7.4.5, is a fully multilocus, indeed entire genome, result. In contradiction to the conventional wisdom view of it, the theorem does not assume random mating, whereas a high proportion of Wright’s mathematical work, discussed in more detail below, does make this assumption.

To summarize, Fisher's view on the nature of evolution involves large population sizes, an emphasis on the long term and on the main effects of single loci as contrasted with complexes of loci, and a steady and essentially deterministic increase in the frequency of each allele having a selective advantage, no matter how small, with regard to the various alternative alleles at its locus. Evolution can be viewed to a large extent on a locus-by-locus basis, and the net evolutionary pattern can be found by "adding" together such single-locus events.

Fisher's view has a grand simplicity to it. Is it, however, simplistic? The evolutionary theory reached by Wright (1931, 1956, 1960, 1965b, 1969b) appears, at least at first sight, to be more subtle. Wright arrived at his view of evolution by discussing in turn several modes of the way in which gene substitution by selection can occur. He first considered selection in a very large random-mating population in a stable environment. The rates of change of gene frequency can be assumed to follow, at least to a reasonable approximation, differential equations of the form (1.27). Successive substitutional processes depend on the occurrence of favorable new mutations, and these are seen as arising sufficiently rarely so that evolution in this manner takes place too slowly to be effective.

This led to a view of the circumstances most favorable to evolution that is more complex than Fisher's, and of a different nature. Wright proposed a three-phase process under which evolution could most easily occur. This view assumes that large populations are normally split up into semi-isolated subpopulations, or demes, each of which is comparatively small in size. Within each deme there exists a genotypic fitness surface, depending on the genetic constitution at many loci, and in conformity with the "increase in mean fitness" concept, gene frequencies tend to move so that local peaks in this surface are approached. The surface of mean fitness is assumed to be very complex with a multiplicity of local maxima, some higher than others. If a fully deterministic behavior obtains the system simply moves to the nearest selective peak and remains there. The importance of the comparatively small deme size is that such strict deterministic behavior does not occur: Random drift can move gene frequencies across a saddle and possibly under the control of a higher selective peak. Random changes in selective values can also perform the same function. In this way a succession of peaks can be reached, each one higher than the previous one. Interpopulational selection, arising from migration of individuals from demes which have higher selective peaks than have other demes, allows the favorable gene complex to spread ultimately throughout the entire population. The unit of selection here is the entire gene complex and not individual alleles. Indeed the latter are viewed as often having no absolute selective advantage, being perhaps favorable in some gene combinations but unfavorable in others.

A case where evolution can more easily take place under this mode compared to that of Fisher is that of two alleles, one at each of two loci, which

individually are deleterious but together are favorable. Calling the alleles in question A_1 and B_1 , one selective scheme where this might occur is the following:

$$\begin{array}{cccc}
 & B_1B_1 & B_1B_2 & B_2B_2 \\
 A_1A_1 & 1 + r & 1 + s & 1 - t \\
 A_1A_2 & 1 + s & 1 & 1 - u \\
 A_2A_2 & 1 - t & 1 - u & 1
 \end{array} \tag{1.96}$$

Here $r > s > 0$ and $t > u > 0$. Under a deterministic scheme the frequencies of A_1 and B_1 , if initially small, will be kept small (because of the selective disadvantage of $A_1A_2B_2B_2$ and $A_2A_2B_1B_2$ to $A_2A_2B_2B_2$). If however in one deme the frequencies of A_1 and B_1 can reach a sufficiently high value, the selective advantage of $A_1A_1B_1B_1$, and to a lesser extent of $A_1A_1B_1B_2$ and $A_1A_2B_1B_1$, will lead to fixation of A_1 and B_1 . In terms of the previous discussion, this implies passing across a saddle from a selective peak at frequency (A_1) = frequency (B_1) = 0 to a higher selective peak at frequency (A_1) = frequency (B_1) = 1. By migration the favored complex, involving the gamete A_1B_1 in high frequency, is now assumed to spread to all demes.

It will be clear that Wright's emphasis, at least compared to Fisher's, was on interactive genetic systems in which most characters are affected by the genes at many loci and most genes have pleiotropic effects, that is influence several characters. Fisher was of course fully aware of the importance of the interactive nature of genetic systems, as his work on the evolution of dominance shows. However, his view tended to the claim that in the very long term, the effects of single genes would be important. Wright's view was no doubt strongly influenced by his early experimental work on the coat color of guinea pigs, which revealed the importance of these interactive effects. From the very first (see, in particular, Wright (1935, 1952, 1969b)) his conceptual framework involved multilocus analysis and in particular an examination of the "optimum" model (1.91) and its various generalizations for more than two loci. We examine the model (1.91) in more detail in Chapter 6, and will find that Wright's analysis of this model is flawed. His analysis uses gene frequencies rather than the correct gametic frequencies, and a correct analysis using gametic frequencies shows that the equilibrium point of this model, which he investigated, is unstable and thus of no interest.

This leads to a further criticism of his work from a mathematical point of view. The only multilocus model that he analyzed mathematically is the model (1.91) discussed above. Thus despite his emphasis on multilocus fitness systems, he never analyzed one in an appropriate mathematical way. Further, the mean fitness increase theorem, a central feature of his fitness surface analysis, will be shown in Chapters 6 and 7 not to be correct as a mathematical theorem in the multilocus case.

Finally, just as we asked whether Fisher's view of evolution in Mendelian populations is too simplistic, it is equally reasonable to ask whether Wright's overall views, particularly those involving population subdivision with migration between partially isolated demes, are not too complex. His picture of evolution may well rely on an equipoise of migration rates, fitness differentials and deme sizes of an unrealistically finely-tuned nature. We shall examine this point later when assessing the role of these various factors, and of linkage, in evolution. It should however be mentioned that the facile criticism of Wright's evolutionary theory, that random drift is seen as an alternative to selection, has no basis in reality. Random drift is conceived of as acting merely as a trigger mechanism in the first phase of the process, changing gene frequencies within each deme before the more permanent and important changes brought about by selection.

The debate about whether Fisher's broad view or Wright's broad view is the more appropriate continues, pointlessly, to this day. Whatever differences Fisher and Wright may have had, they are dwarfed by their agreement on the need to formulate a new evolutionary theory based on Mendelian genetics and the essential identity of much of their (separate) calculations concerning this new evolutionary process. This implies that it is necessary to be familiar with their work, but also necessary to move forward from the paradigms established by these two giants. This volume is intended, on the one hand, to summarize mathematical aspects of this evolutionary theory as it was developed by Fisher, Haldane and Wright and their immediate successors, and on the other hand to introduce the molecular genetics-based contemporary theory. In the latter aim it is intended to form the basis of some of the material to be discussed in Volume II.

2

Technicalities and Generalizations

2.1 Introduction

This chapter is largely technical in nature. Its aim in part is to consider in more detail some of the theoretical points raised in Chapter 1, and in part to put these in a setting that allows a more detailed and up-to-date discussion of them in later chapters. A second aim is to introduce some further techniques not discussed in Chapter 1. Some rather straightforward generalizations of the theory are also made. Finally, the statement of the Fundamental Theorem of Natural Selection for one gene locus will be given and proved.

Population genetics models often make a number of simplifying assumptions, for example that random mating obtains, that fitnesses are fixed constants, that the population size is effectively infinite, and so on. In this chapter we consider what happens when some of these assumptions are relaxed or even dropped altogether. It is difficult enough to consider the effect of relaxing two or three of these assumptions simultaneously and quite impossible to consider the effect of relaxing them all. In the various sections of this chapter we therefore consider one or other generalization of the theory brought about by relaxing one or other of these assumptions, without attempting to assess the effect of simultaneous relaxation of two or more assumptions. Such an assessment must, at the moment, be largely nonquantitative.

2.2 Random Union of Gametes

In elementary textbooks the way in which the frequencies of the various genotypes in a daughter generation are derived from those in the parent generation is by means of a two-way table. All the various possible matings are listed, their frequencies and the relative frequencies with which they produce various offspring genotypes are noted, and thus the frequencies of the daughter generation genotypes are calculated. This procedure was outlined in Chapter 1 for the case of non-random-mating populations. It is far more efficient, however, for random-mating populations, to proceed in a different way. Restricting attention to autosomal loci, we observe that each individual transmits, for each locus, one gene to each of his/her offspring: The union of two such genes, one from each parent, defines at that locus the genotype of the offspring individual. Random mating of parents is equivalent to random union of genes. Thus, for example, using the notation of Section 1.2, since the frequency of A_1 in the parent generation is $X + Y$, the frequency of A_1A_1 in the daughter generation, being the probability that two genes drawn at random from the parent generation are both A_1 , is $(X + Y)^2$. This argument, and parallel arguments for the other genotypes, together give equations (1.1)–(1.3) immediately. Only minor extensions of the argument are needed for more complex cases such as sex-linked loci, multiple alleles, dioecious populations, and so on, and we use this form of argument below in developing the properties of these more complex models.

It was stated in Section 1.6 that explicit models should be set up before any mathematical analysis is attempted, so it is necessary to state more explicitly the model assumed in the above argument. It has been assumed that the population is monoecious, of effectively infinite size and that any daughter-generation individual is formed by the mating of two randomly chosen individuals of the parent generation. It is also assumed that there are no geographical effects, no mating success differentials, and so on. Perhaps most important, it is also assumed that distinct generations can be recognized, so that matings occur only between individuals of the same generation, and that these individuals do not participate in further mating once the daughter generation is formed. These assumptions imply that there is no population age structure. Later, models with assumptions that are more general than, and also rather different from, these will be introduced.

2.3 Dioecious Populations

In this section we drop the assumption that the population is monoecious and suppose instead that it is dioecious, that is admits two sexes. The other

assumptions of the previous section are maintained. We focus initially on the autosomal case, deferring the analysis of the sex-linked case to later.

Suppose first there is no selection, and that in a given generation the genotypic frequencies are as given in (2.1) below:

$$\begin{array}{rcccl}
 & A_1A_1 & A_1A_2 & A_2A_2 & \\
 \text{males:} & X_M & 2Y_M & Z_M & (2.1) \\
 \text{females:} & X_F & 2Y_F & Z_F &
 \end{array}$$

The argument of the random union of gametes, suitably modified to the dioecious case, shows that the frequency of A_1A_1 individuals among both males and females of the daughter generation is $(X_M + Y_M)(X_F + Y_F)$, with parallel formulas for A_1A_2 and A_2A_2 . This implies that after one further generation of random mating the frequencies in both sexes are in the Hardy-Weinberg form

$$\begin{array}{rcccl}
 A_1A_1 & A_1A_2 & A_2A_2 & & \\
 x^2 & 2x(1-x) & (1-x)^2 & & (2.2)
 \end{array}$$

where

$$x = \frac{1}{2}(X_M + X_F + Y_M + Y_F). \quad (2.3)$$

The frequencies of the three genotypes among males and among females now remain equal in all further generations. For this reason we often make the modeling simplification of ignoring the existence of two sexes, except of course in special cases, for example in discussing the sex ratio.

One case where the existence of two sexes has to be taken into account is that where genotype fitness values are different in males and females. Suppose then that viability selection exists, so that the relative fitnesses of the genotypes A_1A_1 , A_1A_2 and A_2A_2 in males are w_{11} , w_{12} , and w_{22} , with corresponding values v_{11} , v_{12} and v_{22} in females. We consider genotypic frequencies immediately after the formation of the zygotes of any generation, and suppose that in a given generation the males produce A_1 gametes with frequency x and A_2 gametes with frequency $1-x$. Let the corresponding frequencies for females be y and $1-y$. Then at the time of conception of the zygotes in the daughter generation the genotypic frequencies are, in both sexes,

$$\begin{array}{rcccl}
 A_1A_1 & A_1A_2 & A_2A_2 & & \\
 xy & x(1-y) + y(1-x) & (1-x)(1-y) & &
 \end{array}$$

By the age of maturity these frequencies will have been altered by differential viability to the relative values

$$\begin{array}{rcccl}
 A_1A_1 & A_1A_2 & A_2A_2 & & \\
 \text{males:} & w_{11}xy & w_{12}\{x(1-y) + y(1-x)\} & w_{22}(1-x)(1-y) & \\
 \text{females:} & v_{11}xy & v_{12}\{x(1-y) + y(1-x)\} & v_{22}(1-x)(1-y) &
 \end{array}$$

The frequencies x' and y' of A_1 gametes produced by males and females of the daughter generation are thus

$$x' = \frac{w_{11}xy + \frac{1}{2}w_{12}\{x(1-y) + y(1-x)\}}{w_{11}xy + w_{12}\{x(1-y) + y(1-x)\} + w_{22}(1-x)(1-y)}, \quad (2.4a)$$

$$y' = \frac{v_{11}xy + \frac{1}{2}v_{12}\{x(1-y) + y(1-x)\}}{v_{11}xy + v_{12}\{x(1-y) + y(1-x)\} + v_{22}(1-x)(1-y)}. \quad (2.4b)$$

These recurrence relations cannot in general be solved explicitly. It is nevertheless possible to arrive at certain important properties concerning their equilibrium points. It is clear that if selection favors the same allele in both males and females there will be no internal equilibrium, so the two cases of real interest are, first, that where different genes are favored in the two sexes, and second, that where overdominance is involved. Our analysis of these two cases follows that of Kidwell et al. (1977).

Suppose first there is no dominance in fitness for each sex and that selection acts in opposite directions in the two sexes. We thus write the fitnesses in the form

	A_1A_1	A_1A_2	A_2A_2
males	1	$1 - \frac{1}{2}s_m$	$1 - s_m$
females	$1 - s_f$	$1 - \frac{1}{2}s_f$	1

where $s_m, s_f > 0$. Solution of the equilibrium equations $x = x', y = y'$ gives, as the only possible equilibrium,

$$x = 1 - s_m^{-1} + \{(s_m s_f - s_m - s_f + 2)(2s_m s_f)^{-1}\}^{1/2},$$

$$y = 1 - s_f^{-1} + \{(s_m s_f - s_m - s_f + 2)(2s_m s_f)^{-1}\}^{1/2}.$$

This equilibrium will be admissible ($0 < x < 1, 0 < y < 1$) only if

$$\frac{s_m}{1 + s_m} < s_f < \frac{s_m}{1 - s_m} \quad (2.5a),$$

or, equivalently, if

$$\frac{s_f}{1 + s_f} < s_m < \frac{s_f}{1 - s_f}. \quad (2.5b)$$

When these conditions apply the equilibrium can be shown to be stable. We conclude that especially if s_m and s_f are small, additive selection acting in opposite directions in the two sexes will maintain a stable equilibrium only if the selective differences in the two sexes are fairly close.

Suppose now that dominance is introduced, so that the fitness scheme becomes

	A_1A_1	A_1A_2	A_2A_2
males	1	$1 - h_m s_m$	$1 - s_m$
females	$1 - s_f$	$1 - h_f s_f$	1

An interesting special case occurs when $h_f + h_m = 1$. Here the conditions (2.5) that there exist a single stable internal equilibrium point continue to apply. When $h_f + h_m < 1$ there will be at most one equilibrium point, and the conditions on s_m and s_f for this to occur are rather less stringent than (2.5). Thus, speaking roughly, for smaller h_f and h_m values, a larger range of s_m and s_f values will lead to an equilibrium point. When $h_f + h_m > 1$ it is possible that more than one internal equilibrium point can arise, but the conditions for this are not given here.

When directional selection obtains for one sex and overdominance in the other, one suspects that a stable polymorphic equilibrium is possible provided the directional selection is not too strong. We quantify this statement in a moment when considering conditions for a stable polymorphic equilibrium to exist.

It is of considerable interest to ask how effective the existence of different selective schemes in the two sexes is in maintaining genetic variation compared to the corresponding effect when identical selective schemes obtain in the two sexes. We attack this question quantitatively by considering the conditions for the existence of an internal polymorphism. For practical purposes we may suppose that such a polymorphism exists when the two equilibria $\text{freq}(A_1) = 0$ in males and females, $\text{freq}(A_2) = 0$ in males and females, are both unstable. If we linearize the recurrence relations (2.4) around $x = y = 0$ and around $x = y = 1$, we find that the condition for an internal polymorphism is that both the inequalities

$$(w_{12}/w_{22}) + (v_{12}/v_{22}) > 2, \quad (2.6a)$$

$$(w_{12}/w_{11}) + (v_{12}/v_{11}) > 2 \quad (2.6b)$$

should hold. These requirements are the natural extensions to the corresponding monoecious population requirement that the heterozygote be more fit than both homozygotes.

When A_1 is at a selective advantage in males (so that $w_{11} > w_{12} > w_{22}$) but overdominance applies in females (so that $v_{12} > v_{11}, v_{22}$), condition (2.6a) holds automatically. However, condition (2.6b) will hold only if the overdominance in females is sufficiently strong compared to the directional selection in males. Thus (2.6b) quantifies our earlier discussion of this point.

How stringent are the conditions given in (2.6)? Suppose we normalize so that $w_{12} = v_{12} = 1$. The conditions (2.6) then reduce to the requirements that the harmonic means of v_{11} and w_{11} , and that of v_{22} and w_{22} , should both be less than unity. Since harmonic means are less than arithmetic means, this is a less stringent requirement than that the arithmetic means both be less than unity. In other words, the existence of different selective parameters in the two sexes provides a stronger mechanism for maintaining genetic polymorphism than taking average selective values over the two sexes would suggest.

The above analysis concerns autosomal loci, and clearly a special analysis is needed in the sex-linked case. Taking the males as the heterogametic sex, the frequencies of the various genotypes in the sex-linked case can be written

male		female		
A_1	A_2	A_1A_1	A_1A_2	A_2A_2
x	$1 - x$	Y_{11}	$2Y_{12}$	Y_{22}

If there is no selection, the discussion outlined in the previous section shows that the frequencies in the following generation are

$$\begin{aligned} x' &= Y_{11} + Y_{12}, \\ Y'_{11} &= x(Y_{11} + Y_{12}), \\ 2Y'_{12} &= x(Y_{12} + Y_{22}) + (1 - x)(Y_{11} + Y_{12}), \\ Y'_{22} &= (1 - x)(Y_{12} + Y_{22}). \end{aligned}$$

In contrast to the autosomal case, one generation of random mating is not sufficient to yield equal frequencies of A_1 in the two sexes. Nor does one further generation of random mating produce female genotypic frequencies in Hardy–Weinberg form. On the other hand, since

$$x' - (Y'_{11} + Y'_{12}) = -\frac{1}{2}\{x - (Y_{11} + Y_{12})\},$$

the absolute value of the difference between male and female frequencies of A_1 is halved between successive generations. For practical purposes we may thus assume that after a short time, these frequencies are equal: If this is so, one further generation of random mating yields frequencies in the form

males		females		
A_1	A_2	A_1A_1	A_1A_2	A_2A_2
z	$(1 - z)$	z^2	$2z(1 - z)$	$(1 - z)^2$

where

$$z = \frac{1}{3}x + \frac{2}{3}(Y_{11} + Y_{12}).$$

When selection operates the behavior is clearly more complex, as is shown by Sprott (1957), Bennett (1957) and Cannings (1967, 1968). We do not go into details here, and in this book we give little attention, perhaps less than is deserved, to sex-linked genes, under the assumption that properties of autosomal loci are normally mirrored, perhaps with minor alterations, in the sex-linked case.

While, in both autosomal and the sex-linked cases, the evolutionary behavior of two-sex systems is slightly more complex than in the monoecious case, the important Mendelian properties of conservation of genetic variation and the suitability of the Mendelian system for evolutionary processes continue to apply.

2.4 Multiple Alleles

We turn now to the case of multiple alleles, considering only random-mating populations. Suppose that at an autosomal locus A , alleles A_1, A_2, \dots, A_k can occur. We consider a model identical to that of Section 2.2 and assume there is no selection. If the frequency of A_i in any generation is x_i , the concept of the random union of gametes shows that in the next generation the frequency of $A_i A_i$ will be x_i^2 and that of $A_i A_j$ ($i \neq j$) will be $2x_i x_j$. These frequencies are in generalized Hardy–Weinberg form and are maintained through future generations.

Suppose now that viability differentials exist and that the fitness of $A_i A_j$ is w_{ij} . It is clear that if we continue to count individuals at the moment of conception of each generation, the genotypic frequencies are in Hardy–Weinberg form at that time. The gene frequencies will normally change from one generation to another, and the appropriate recurrence relations are

$$x'_i = x_i \sum_j w_{ij} x_j / \bar{w}, \quad (2.7)$$

$$= x_i w_i / \bar{w}, \quad (2.8)$$

the sum (as with all sums in this section) being over $1, 2, \dots, k$. In this equation w_i , the “marginal fitness of the allele A_i ”, is defined as

$$w_i = \sum_j w_{ij} x_j. \quad (2.9)$$

In equations (2.7) and (2.8) the quantity \bar{w} , the mean fitness of the population, is defined by

$$\bar{w} = \sum_i x_i w_i = \sum_i \sum_j w_{ij} x_i x_j. \quad (2.10)$$

In view of the statement of the mean fitness increase theorem in Section 1.4, and the condition given there for the existence of a stable internal equilibrium point under the action of selection only, it is natural to ask whether the mean fitness increases from one generation to another in the multiple allele case, and to seek the conditions on the w_{ij} that ensure a stable internal equilibrium point (that is each $x_i > 0$) of gene frequencies.

The most efficient proof that mean fitness increases in the multiple allele case was given by Kingman (1961a) and is reproduced in detail here. The daughter generation mean fitness \bar{w}' is defined by $\bar{w}' = \sum \sum w_{ij} x'_i x'_j$, and we are required to prove that with this definition, $\bar{w}' - \bar{w} \geq 0$. Using (2.7),

we obtain

$$\begin{aligned}\bar{w}' &= \bar{w}^{-2} \left(\sum_i \sum_j w_{ij} (x_i w_i) (x_j w_j) \right) \\ &= \bar{w}^{-2} \left(\sum_i \sum_j \sum_m w_{ij} w_{im} x_i x_j x_m w_j \right).\end{aligned}$$

By interchanging the roles of j and m we also have

$$\bar{w}' = \bar{w}^{-2} \left(\sum_i \sum_j \sum_m w_{ij} w_{im} x_i x_j x_m w_m \right).$$

Thus by averaging, we find

$$\begin{aligned}\bar{w}' &= \frac{1}{2} \bar{w}^{-2} \left(\sum_i \sum_j \sum_m w_{ij} w_{im} x_i x_j x_m (w_j + w_m) \right) \\ &\geq \bar{w}^{-2} \left(\sum_i \sum_j \sum_m w_{ij} w_{im} (w_j w_m)^{1/2} x_i x_j x_m \right)\end{aligned}\tag{2.11}$$

$$\begin{aligned}&= \bar{w}^{-2} \sum_i x_i \left(\sum_j x_j w_{ij} (w_j)^{1/2} \right)^2 \\ &\geq \bar{w}^{-2} \left(\sum_i x_i \sum_j x_j w_{ij} (w_j)^{1/2} \right)^2\end{aligned}\tag{2.12}$$

$$\begin{aligned}&= \bar{w}^{-2} \left(\sum_j x_j (w_j)^{1/2} \sum_i x_i w_{ij} \right)^2 \\ &= \bar{w}^{-2} \left(\sum_j x_j (w_j)^{3/2} \right)^2 \\ &\geq \bar{w}^{-2} \left(\left\{ \sum_j (x_j w_j) \right\}^{3/2} \right)^2 \\ &= \bar{w}^{-2} \left(\sum_j x_j w_j \right)^3 \\ &= \bar{w}.\end{aligned}\tag{2.13}$$

In this sequence of steps the inequality (2.11) is justified by the inequality $\frac{1}{2}(a+b) \geq (ab)^{1/2}$ for positive quantities a and b , and the inequalities (2.12) and (2.13) are justified by the convexity property $\sum x_i a_i^n \geq \left(\sum x_i a_i \right)^n$ for nonnegative a_i and $n \geq 1$. If we assume each $x_i > 0$, this proof also shows that $\bar{w}' = \bar{w}$ if and only if $w_1 = w_2 = \dots = w_k$, and when this is so,

$$w_i = \bar{w}, \quad i = 1, 2, \dots, k.\tag{2.14}$$

This equation and (2.8) together imply that $x'_i = x_i$, so that the system is at an equilibrium point. We thus conclude that in the evolutionary system (2.7), the mean fitness always increases except when the system has reached an equilibrium point, where of course it remains unchanged. This conclusion also applies when some of the x_i are zero, although here of course (2.14) is true only for those values of i for which x_i is positive at the equilibrium point.

In view of the discussion in Chapter 1, it is natural to ask whether the change in mean fitness can be approximated by σ_A^2 , the additive genetic variance in fitness. The natural generalization of the procedure that led to (1.16) is to define σ_A^2 as the maximum sum of squares removed by $\alpha_1, \dots, \alpha_k$ in the expression S , defined by

$$S = \sum \sum x_i x_j (w_{ij} - \bar{w} - \alpha_i - \alpha_j)^2. \quad (2.15)$$

It is found that the values of the α_i that lead to the minimizing of S are

$$\alpha_i = w_i - \bar{w}, \quad i = 1, 2, \dots, k. \quad (2.16)$$

From this it follows, after some algebra, that

$$\sigma_A^2 = 2 \sum_i x_i (w_i - \bar{w})^2. \quad (2.17)$$

When $k = 2$ this reduces to the value given by (1.42).

We now wish to compare the expression in (2.17) with the mean fitness change $\bar{w}' - \bar{w}$, which we write as

$$\bar{w}' - \bar{w} = \bar{w}^{-2} \left[\sum \sum w_{ij} x_i x_j w_i w_j - \bar{w}^3 \right].$$

If $w_{ij} = \bar{w} + \delta_{ij}$, $w_i = \bar{w} + \delta_i$, where the δ_{ij} are assumed small, this becomes, on ignoring terms of order δ_{ij}^3 ,

$$\begin{aligned} \bar{w}' - \bar{w} &\approx \sum \sum \{ \delta_{ij} \delta_i + \delta_{ij} \delta_j + \delta_i \delta_j \} x_i x_j \\ &= 2 \sum_i x_i \delta_i^2 + \sum_i x_i \delta_i \sum_j x_j \delta_j \\ &= 2 \sum_i x_i \delta_i^2. \end{aligned} \quad (2.18)$$

This is identical to (2.17), and we conclude that for small fitness differentials the increase in mean fitness is very closely approximated by the additive genetic variance in fitness. Thus, under the assumptions made, in particular that of small fitness differentials, the MFIT holds for an arbitrary number of alleles at the locus. When fitness differentials are not small a rather different conclusion is found (Seneta (1973)).

Suppose that each x_i is positive. Then (2.17) shows that σ_A^2 is zero if and only if $w_1 = w_2 = \dots = w_k = \bar{w}$. If some of the x_i are zero, the additive genetic variance σ_A^2 is zero if (2.14) applies for those values of i for which x_i is positive. In both cases the discussion above shows that σ_A^2 is zero if and only if the system is at an equilibrium point. We see later that in multilocus systems the identification just reached for one locus, namely

$$\sigma_A^2 = 0 \Leftrightarrow \text{population in equilibrium} \quad (2.19)$$

no longer holds, although a restricted version of this conclusion can be found.

We consider now the evolution of a metrical character, not necessarily fitness, under the evolutionary system (2.7). Consider some character which for $A_i A_j$ individuals takes the measurement m_{ij} . The mean value \bar{m} of this character is given by $\bar{m} = \sum \sum x_i x_j m_{ij}$, and we wish to compute the change in this mean after one generation. To a first order of approximation,

$$\begin{aligned} \Delta \bar{m} &= 2 \sum \sum (\Delta x_i) x_j m_{ij} \\ &= 2 \sum (\Delta x_i) m_i \\ &= 2 \sum (\Delta x_i) (m_i - \bar{m}) \\ &\approx 2 \sum x_i (w_i - \bar{w}) (m_i - \bar{m}), \end{aligned} \tag{2.20}$$

where we have defined m_i , the marginal measurement for the allele A_i , by

$$m_i = \sum x_j m_{ij}. \tag{2.21}$$

A verbal description of this conclusion is that the change in the character is twice the covariance between marginal allelic values of the character itself and fitness. For further details, see Robertson (1966, 1968). When the character is fitness itself this conclusion reduces to that obtained in (2.18).

We turn now to the condition under which a stable equilibrium of gene frequencies exists. We first assume that each x_i is positive at the equilibrium. The equilibrium conditions (2.14) can be written

$$\begin{aligned} w_i - w_1 &= 0, \quad i = 2, 3, \dots, k, \\ x_1 + x_2 + \dots + x_k &= 1, \end{aligned} \tag{2.22}$$

and this is just a system of k linear equations in k unknowns. It thus possesses no solution, one solution or an infinity of solutions. The first and third cases arise only for special values of the w_{ij} , such as, for example, when all fitnesses are equal. In practice it is most interest to ignore these cases and suppose there is a unique solution of (2.22). Unfortunately this solution might be inadmissible, that is the condition $0 < x_i < 1$, $i = 1, \dots, k$, might not be met, and even if the equilibrium is admissible it need not be stable. Fortunately the stability criteria have been obtained (Kingman, (1961b)). A unique admissible solution to (2.22) will be stable if and only if the matrix $W = \{w_{ij}\}$ has exactly one positive eigenvalue and at least one negative eigenvalue. In this case the system moves, for any initial frequency point for which each x_i is positive, to this equilibrium. If the equilibrium (2.22) is not admissible or is unstable, the system (2.7) evolves in such a way that one or more alleles become eliminated. The behavior then becomes considerably more complicated, and in practice perhaps the best procedure is to note that the system always moves so that \bar{w} is maximized, so that finding the maximum value of \bar{w} subject to the constraints $0 \leq x_i \leq 1$, $\sum x_i = 1$, via the Kuhn–Tucker theory for quadratic programming, will

provide the stable equilibrium point. A result of Kingman (1961b) relevant to this is that if W has j positive eigenvalues, then at most $k - j + 1$ alleles will exist with positive frequencies at this equilibrium.

As the simplest possible example of this theory we consider the case where all homozygotes have fitness $1 - s$ ($0 < s < 1$), and all heterozygotes have fitness 1. Clearly there is an admissible equilibrium point at $x_i = k^{-1}$. This will be stable if the matrix

$$W = \begin{pmatrix} 1-s & 1 & 1 & \cdots & 1 \\ 1 & 1-s & 1 & \cdots & 1 \\ 1 & 1 & 1-s & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1-s \end{pmatrix}$$

has exactly one positive eigenvalue and at least one negative eigenvalue. But standard theory shows that the eigenvalues of this matrix are $k - s$, $-s$, \dots , $-s$, and thus the stability conditions are indeed met.

We turn finally to the correlation between relatives in the k -allele system, and take as an example the correlation between father and son. Suppose the father has genotype $A_i A_i$ (and thus measurement m_{ij}). The son will be $A_i A_j$ (and have measurement m_{ij}) with probability x_j , and since the frequency of $A_i A_i$ fathers is x_i^2 this will make a contribution to the covariance of

$$x_i^2 \sum x_j m_{ii} m_{ij} = x_i^2 m_i m_{ii}. \quad (2.23)$$

If the father is $A_i A_j$ (frequency $2x_i x_j$) the son will be $A_i A_i$ (probability $\frac{1}{2}x_i$) or $A_j A_j$ (probability $\frac{1}{2}x_j$), $A_i A_j$ (probability $\frac{1}{2}(x_i + x_j)$), $A_i A_\ell$ (probability $\frac{1}{2}x_\ell$) or $A_j A_\ell$ (probability $\frac{1}{2}x_\ell$). The contribution to the covariance corresponding to this case is

$$\begin{aligned} 2x_i x_j m_{ij} \left[\frac{1}{2}(x_1 m_{i1} + \cdots + x_k m_{ik}) + \frac{1}{2}(x_1 m_{j1} + \cdots + x_k m_{jk}) \right] \\ = x_i x_j m_{ij} (m_i + m_j). \end{aligned} \quad (2.24)$$

Adding (2.23) over all i and (2.24) over all i, j ($i < j$) we arrive at the covariance

$$\sum_i x_i^2 m_i m_{ii} + \sum_{i < j} \sum x_i x_j m_{ij} (m_i + m_j) - \bar{m}^2 = \sum_i x_i (m_i - \bar{m})^2.$$

This is just half the expression (2.17) (if we replace w_{ij} by the more general m_{ij}), and in this way we recover expression (1.10) for the correlation in the measurement between father and son, where now both variance terms have the more general k -allele interpretation. Identical conclusions apply for other relationships, and we conclude that the correlation formulas found in Chapter 1 are not affected by the number of alleles at the locus in question.

2.5 Frequency-Dependent Selection

In all of the above constant fitness values for each genotype have been assumed. It is likely in reality that many fitness values are not constant but depend on the number of individuals in the population, on the frequencies of the various alleles, or on both. In this short section we consider briefly some aspects of frequency-dependent selection. We assume the model of Section 2.2 with two alleles at the locus considered.

Using the fitness scheme (1.25a) we arrived at the equation

$$\Delta x = x(1-x)\{w_{11}x + w_{12}(1-2x) - w_{22}(1-x)\}/\bar{w},$$

and this equation continues to hold if the w_{ij} are functions of the allele frequency x . Clearly there are equilibria when $x = 0$, $x = 1$, or when

$$w_{11}x + w_{12}(1-2x) - w_{22}(1-x) = 0. \quad (2.25)$$

If the functions w_{ij} are sufficiently complex functions of x , (2.25) can have a number of solutions, several of which can be stable. There is little point in considering special cases. Further, \bar{w} need not be maximized at an equilibrium point of the system. (2.25) and the equation $d\bar{w}/dx = 0$ show that mean fitness will not be maximized at an equilibrium if, at that equilibrium,

$$x^2 dw_{11}/dx + 2x(1-x)dw_{12}/dx + (1-x)^2 dw_{22}/dx \neq 0.$$

Thus evolution can cause a steady decrease in mean fitness. In a classical example due to Wright (1948) it is supposed that the fitnesses of A_1A_1 , A_1A_2 , and A_2A_2 individuals are $1-s+t(1-x)$, 1, and $1+s-t(1-x)$ where $s, t > 0$. If $s < t$ there is a point of stable equilibrium where $x = x^* = 1-st^{-1}$, whereas the mean fitness is maximized at $\frac{1}{2}(\frac{1}{2} + x^*)$, halfway between x^* and $\frac{1}{2}$. Clearly, for suitable initial frequencies of A_1 , the mean fitness can steadily decrease during the course of evolution.

2.6 Fertility Selection

Until now we have assumed that selection operates through viability differentials. This assumption was made for mathematical convenience, and we now suppose that further selective differences between genotypes arise through differential fertility as well as through viability differences. The analysis now becomes more complex, since fertility relates to mating combinations rather than single genotypes. Our discussion assumes the natural generalizations of the model of Section 2.2 and closely follows the work of Bodmer (1965) and Kempthorne and Pollak (1970). We follow the natural generalization of (1.25a) and suppose that the viability of an A_iA_j genotype is w_{ij} ($i, j = 1, \dots, k$) (assumed the same in both sexes) and that the fertility of an $A_iA_j \times A_mA_n$ mating is f_{ijmn} . (We adopt some standard ordering convention such that A_iA_j is the male and A_mA_n the female.) It

is clear that male and female genotypic frequencies will be equal: Let X_{ij} be the frequency of A_iA_j just before the conception of a new generation. Those matings leading to A_iA_i offspring must be of the form $A_iA_j \times A_iA_m$ for some j and m . Consideration of the genotypic products of such matings shows that the frequency of A_iA_i at the birth of the next generation will be proportional to

$$\begin{aligned} Z_{ii} = & f_{iii}X_{ii}^2 + \frac{1}{2} \sum_{m \neq i} f_{iim}X_{ii}X_{im} + \frac{1}{2} \sum_{j \neq i} f_{ijj}X_{ij}X_{ji} \\ & + \frac{1}{4} \sum_{j \neq i} \sum_{m \neq i} f_{ijm}X_{ij}X_{im}. \end{aligned} \quad (2.26)$$

These A_iA_i individuals are now subject to viability selection between birth and the age of maturity, and it follows that the frequency X'_{ii} of A_iA_i just before the birth of the next following generation is given by

$$\mu X'_{ii} = w_{ii}Z_{ii}, \quad i = 1, 2, \dots, k, \quad (2.27a)$$

where μ is a normalizing constant to be discussed later. Similar considerations for A_iA_j individuals yield

$$\mu X'_{ij} = w_{ij}Z_{ij}, \quad i, j = 1, 2, \dots, k, i \neq j, \quad (2.27b)$$

where

$$\begin{aligned} Z_{ij} = & (f_{iij} + f_{jji})X_{ii}X_{jj} + \frac{1}{2} \sum_{m \neq j} f_{ijm}X_{ii}X_{jm} \\ & + \sum_{m \neq i} f_{imj}X_{im}X_{jj} + \frac{1}{4} \sum_{m \neq i} \sum_{n \neq j} f_{imn}X_{im}X_{jn}. \end{aligned}$$

The constant μ in (2.27a) and (2.27b) is now chosen so that $\sum \sum X'_{ij} = 1$.

These recurrence relations are far too complex to solve in general, and we make no attempt to do so. Questions concerning the existence and stability of equilibrium points of the system (2.27) have been discussed by Haderler and Liberman (1975), but we do not pursue them here. Some simplification is possible if it is supposed that the fertilities f_{ijmn} are of the multiplicative form

$$f_{ijmn} = a_{ij}b_{mn}, \quad (a_{ij} = a_{ji}, b_{mn} = b_{nm}). \quad (2.28)$$

Introducing the new variables

$$\begin{aligned} x_i = & (a_{ii}X_{ii} + \frac{1}{2} \sum_{j \neq i} a_{ij}X_{ij}) / \sum_{j \leq i} \sum_{j \leq i} a_{ij}X_{ij}, \\ y_i = & (b_{ii}X_{ii} + \frac{1}{2} \sum_{j \neq i} b_{ij}X_{ij}) / \sum_{j \leq i} \sum_{j \leq i} b_{ij}X_{ij}, \end{aligned} \quad (2.29)$$

the recurrence relations (2.27) become, for the multiplicative case,

$$\begin{aligned} \mu^* X'_{ii} &= w_{ii}x_iy_i, \\ \mu^* X'_{ij} &= w_{ij}(x_iy_j + x_jy_i), \quad i \neq j, \end{aligned} \quad (2.30)$$

where μ^* is a new normalizing constant ensuring that the sum of genotypic frequencies is unity. Use of (2.29) and (2.30) shows that

$$\begin{aligned} x'_i &= (a_{ii}w_{ii}x_iy_i + \frac{1}{2} \sum_{j \neq i} a_{ij}w_{ij}(x_iy_j + x_jy_i)) / \sum \sum a_{ij}w_{ij}x_iy_j, \\ y'_i &= (b_{ii}w_{ii}x_iy_i + \frac{1}{2} \sum_{j \neq i} b_{ij}w_{ij}(x_iy_i + x_jy_i)) / \sum \sum b_{ij}w_{ij}x_iy_j. \end{aligned} \quad (2.31)$$

These recurrence relations are identical in form to those in (2.4), and thus the latter system, once appropriate changes in fitnesses have been made to include the viability parameters, continue to apply. Some specific examples are given by Bodmer (1965). One question of particular interest is whether the mean fitness of the system increases with time. Unfortunately it is not at all evident that a natural definition for mean fitness exists in the fertility selection case. Using (2.30) and the analogy with previous recurrence systems, it would be reasonable to define mean fitness as

$$\sum_i w_{ii}x_iy_i + \sum_{i < j} w_{ij}(x_iy_j + x_jy_i). \quad (2.32)$$

With this definition, it is possible for mean fitness to decrease with time. Thus (Kempthorne and Pollak (1970)) if $k = 2$, $w_{11} = w_{12} = 1$, $w_{22} = 0.5$, $a_{11} = a_{12} = 1$, $a_{22} = 2$, $b_{11} = 0.25$, $b_{12} = b_{22} = 1$, $X_{11} = X_{22} = 0$, $X_{12} = 1$, then $x_i = y_i = 0.5$, and the mean fitness, as defined by (2.32), is 0.875. From (2.31), $x'_1 = x'_2 = \frac{1}{2}$, $y'_1 = 5/11$, $y'_2 = 6/11$ and using these values in (2.32) the daughter generation mean fitness is $19/22 \approx 0.864$. It is clear that this decrease is caused essentially because the genotype with highest fecundity has lowest viability.

Suppose now that in (2.28), it is assumed that $a_{ij} = b_{ij}$. Then immediately $x_i = y_i$ and that the birth of the new generation genotypic frequencies are in Hardy-Weinberg form. Further the recurrence relations (2.31) are of the form (2.7), and therefore the conclusions deriving from that system, including in particular the result that the mean fitness, defined now as $\sum \sum a_{ij}w_{ij}x_ix_j$, cannot decrease, continue to hold. The change in mean fitness again is approximately equal to the additive genetic variance when the latter is suitably defined so as to include both viability and fertility parameters.

Despite this, it is possible that (2.32) is not a natural definition of the mean fitness of the infant population. The classical definition is that the fitness of any genotype is proportional to half the number of offspring individuals (of whatever genotype) from individuals of the genotype in question, counting being performed at the same stage of the life cycle. We now attempt to find an algebraic definition of mean infant fitness along these lines.

Consider infants of genotype A_iA_j : These survive to adulthood with probability w_{ij} . An A_iA_j individual mating with an A_mA_n individual has

$a_{ij}a_{mn}$ offspring and crediting half of these to the A_iA_j individual and averaging over all A_mA_n , the A_iA_j individuals are credited with a proportionate amount

$$\frac{1}{2}w_{ij}a_{ij}\sum_m\sum_n x_mx_nw_{mn}a_{mn}/\bar{w} = w_{ij}a_{ij}\bar{m}/2\bar{w}$$

of offspring, where $m_{ij} = a_{ij}w_{ij}$ and $\bar{m} = \sum\sum x_ix_ja_{ij}w_{ij}$. The mean fitness of the infant population may then reasonably be defined as the weighted average of these quantities, or

$$\sum\sum x_ix_ja_{ij}\bar{m}/2\bar{w} = (\bar{m})^2/2\bar{w}. \quad (2.33)$$

In a parallel fashion the mean fitness of the adult population may be defined: Details are given by Kempthorne and Pollak (1970). Curiously neither the infant mean fitness, defined by (2.33), nor the adult mean fitness, must necessarily increase with time, decreases again possibly occurring when those genotypes with high fertility have low viability. We do not pursue this matter further and simply note the great complexity in general of fertility selection models. During most of the rest of this book selection will be taken to mean viability selection. This is no more than a reflection of the fact that, because the mathematics of viability fitness models is easier than that of fertility fitness models, more is known about viability selection models.

2.7 Continuous-Time Models

In all of this book so far it has been assumed that populations reproduce at discrete time points. There are certainly some real-world populations for which this is a reasonable assumption. On the other hand, it is sometimes more appropriate biologically, or simpler mathematically, to use continuous-time models in which births and deaths can take place at any instant. This normally leads to mathematical systems where changes in gene frequency are described by a differential equation or by differential equation systems. In this section we outline some of these mathematical models and discuss their properties, relying heavily on the definitive work of Nagylaki (1974c, 1976), Nagylaki and Crow (1974) and Kimura (1958).

Consider a locus "A" in a monoecious population and let this locus admit alleles A_1, \dots, A_k . At a given time let the number of A_iA_j individuals be n_{ij} , where we adopt an ordering notation such that the A_i gene has derived from the male parent. Define n_i by $n_i = \frac{1}{2}\sum(n_{ij} + n_{ji})$: Then $2n_i$ is the number of A_i genes in the population. If $N = \sum n_i$ is the population size we may write

$$x_i = n_i/N, \quad X_{ij} = n_{ij}/N \quad (2.34)$$

as the frequencies of A_i and the (ordered) genotype A_iA_j , respectively. Consider a continuous-time deterministic process of population change in which, if terms of order $(\delta t)^2$ are ignored throughout, $NX_{ij}d_{ij}\delta t$ individuals of genotype A_iA_j die in the time interval $(t, t + \delta t)$. Let $M\delta t$ be the number of matings during this time interval, $X_{im,nj}$ be the fraction of these matings which are of the (ordered) type $A_iA_m \times A_nA_j$, and $\tilde{a}_{im,nj}$ the number of offspring from such a mating. We introduce the standardized parameter $a_{im,nj} = M\tilde{a}_{im,nj}/N$, so that $NX_{im,nj}a_{im,nj}\delta t$ is the number of offspring from all (ordered) $A_iA_m \times A_nA_j$ matings in the time interval $(t, t + \delta t)$. Defining $n_{ij}(t)$ as the number of A_iA_j individuals in the population at time t and noting that A_iA_j individuals can arise from various ordered matings in various frequencies, we get

$$n_{ij}(t + \delta t) = n_{ij}(t) + \delta t \left(\sum_{m,n} NX_{im,nj}a_{im,nj} - d_{ij}n_{ij}(t) \right).$$

Letting $\delta t \rightarrow 0$ in the usual way, we obtain

$$\dot{n}_{ij} = \sum_{m,n} NX_{im,nj}a_{im,nj} - d_{ij}n_{ij}, \quad (2.35)$$

where the time derivative, here and below, is denoted by a superior dot. This equation and the verbal description leading to it form the basis of the model we shall consider.

It is convenient to define a birth-rate for A_iA_m individuals. Noting that the number of offspring (of whatever genotype) to such individuals acting as first partner in an $A_iA_m \times A_nA_j$ mating during $(t, t + \delta t)$ is $N \sum_{n,j} X_{im,nj}a_{im,nj}\delta t$ and that the number of A_iA_m individuals available to act as parents is n_{im} , it is reasonable for us to define the birth-rate b_{im} for such individuals by the equation

$$n_{im}b_{im} = N \sum_{n,j} X_{im,nj}a_{im,nj}. \quad (2.36)$$

From this, the fecundity b_i , mortality d_i , and ‘‘Malthusian parameter’’ m_i of the allele A_i are defined by

$$x_i b_i = \sum_j X_{ij} b_{ij}, \quad x_i d_i = \sum_j X_{ij} d_{ij}, \quad m_i = b_i - d_i. \quad (2.37)$$

The mean fecundity \bar{b} , mortality \bar{d} , and Malthusian parameter \bar{m} are then given by

$$\bar{b} = \sum x_i b_i, \quad \bar{d} = \sum x_i d_i, \quad \bar{m} = \bar{b} - \bar{d}. \quad (2.38)$$

Equations (2.35)–(2.38) jointly yield

$$\dot{N} = \bar{m}N, \quad (2.39)$$

$$\dot{X}_{ij} = \sum_{mn} X_{im,nj} a_{im,nj} - (d_{ij} + \bar{m}) X_{ij}, \quad (2.40)$$

and

$$\dot{x}_i = x_i(m_i - \bar{m}). \quad (2.41)$$

To make further progress it is necessary to make certain assumptions. We assume first that random mating obtains, so that

$$X_{im,nj} = X_{im} X_{nj} \quad (2.42)$$

and that $a_{im,nj}$ can be expressed in the additive form

$$a_{im,nj} = \beta_{im} + \beta_{nj} \quad (2.43)$$

for some set of parameters $\{\beta_{ij}\}$. Equation (2.43) is the natural analogue for continuous-time models of an equation like (2.28) for discrete-time models. Equations (2.37)–(2.43) then lead to

$$a_{im,nj} = \bar{b} + (b_{im} - \bar{b}) + (b_{nj} - \bar{b})$$

so that

$$\dot{X}_{ij} = x_i x_j (b_i + b_j - \bar{b}) - (d_{ij} + \bar{m}) X_{ij}. \quad (2.44)$$

Perhaps the most important question to ask is whether Hardy–Weinberg frequencies hold in this model. Defining $Q_{ij} = X_{ij} - x_i x_j$ as a measure of departure from Hardy–Weinberg, (2.41) and (2.44) yield

$$\dot{Q}_{ij} = x_i x_j (d_i + d_j - d_{ij} - \bar{d}) - (d_{ij} + \bar{m}) Q_{ij}. \quad (2.45)$$

Suppose that $d_i + d_j - d_{ij} - \bar{d} \neq 0$. Then even if Hardy–Weinberg frequencies obtain initially, (2.45) shows that they do not persist and do not hold at an equilibrium of the system (2.35). One particular consequence of this is that the rate of change of mean fitness is not necessarily approximately equal to the additive genetic variance in fitness. It is of some interest to determine the relationship between the two quantities, and we now do this in the simple special case where the quantities $a_{im,nj}$ and d_{ij} (which are functions of the $X_{im,nj}$ and of time) are adjusted so that the Malthusian parameter m_{ij} ($= b_{ij} - d_{ij}$) of the genotype $A_i A_j$ is constant in time.

To find the additive genetic variance we minimize the quantity S , defined by

$$S = \sum \sum X_{ij} (m_{ij} - \bar{m} - \alpha_i - \alpha_j)^2. \quad (2.46)$$

If Hardy–Weinberg frequencies do obtain, so that $X_{ij} = x_i x_j$, this would be done following the lines of the analysis in Section 2.4. To measure the effect of departure from Hardy–Weinberg frequencies we introduce the parameters θ_{ij} , defined by

$$X_{ij} = x_i x_j \theta_{ij}. \quad (2.47)$$

Clearly $\theta_{ij} \equiv 1$ implies that Hardy–Weinberg frequencies obtain. If we insert (2.47) into (2.46), we find that the minimization equations yield

$$x_i \alpha_i + \sum_j x_i x_j \theta_{ij} \alpha_j = \sum_j x_i x_j \theta_{ij} (m_{ij} - \bar{m}) \quad (2.48)$$

or

$$\alpha_i + \sum_j x_j \theta_{ij} \alpha_j = \sum_j x_j \theta_{ij} a_{ij} = a_i, \quad (2.49)$$

where we define

$$a_{ij} = m_{ij} - \bar{m}, \quad a_i = x_i^{-1} \sum_j X_{ij} a_{ij}. \quad (2.50)$$

Further, the additive genetic variance, being the sum of squares removed by this procedure, is

$$\sigma_A^2 = 2 \sum_i x_i a_i \alpha_i, \quad (2.51)$$

where a_i is defined explicitly by (2.50) and α_i implicitly by (2.49). In view of (2.41) this may also be written

$$\sigma_A^2 = 2 \sum_i \dot{x}_i \alpha_i. \quad (2.52)$$

We turn now to the rate of change of the mean fitness \bar{m} . By definition

$$\bar{m} = \sum_i \sum_j m_{ij} X_{ij}$$

and since under our assumptions the m_{ij} are constant,

$$\begin{aligned} \dot{\bar{m}} &= \sum_i \sum_j m_{ij} \dot{X}_{ij} \\ &= \sum_i \sum_j a_{ij} \dot{X}_{ij} \\ &= \sum_i \sum_j a_{ij} (\dot{x}_i x_j \theta_{ij} + x_i \dot{x}_j \theta_{ij} + x_i x_j \dot{\theta}_{ij}) \\ &= 2 \sum_i \sum_j a_{ij} \dot{x}_i x_j \theta_{ij} + \sum_i \sum_j a_{ij} x_i x_j \dot{\theta}_{ij} \quad (2.53) \\ &= 2 \sum_i \dot{x}_i \sum_j a_{ij} x_j \theta_{ij} + \sum_i \sum_j a_{ij} x_i x_j \dot{\theta}_{ij} \end{aligned}$$

$$\begin{aligned} &= 2 \sum_i \dot{x}_i (\alpha_i + \sum_j x_j \theta_{ij} \alpha_j) + \sum_i \sum_j a_{ij} x_i x_j \dot{\theta}_{ij} \\ &= \sigma_A^2 + 2 \sum_i \sum_j \dot{x}_i x_j \theta_{ij} + \sum_i \sum_j a_{ij} x_i x_j \dot{\theta}_{ij}. \quad (2.54) \end{aligned}$$

We wish to simplify the final two terms in (2.54). Now

$$x_j = \sum_i X_{ij} = \sum_i x_i x_j \theta_{ij}$$

so that

$$\sum_i x_i \theta_{ij} \equiv 1.$$

Differentiating with respect to t ,

$$\sum_i \dot{x}_i \theta_{ij} + \sum_i x_i \dot{\theta}_{ij} \equiv 0 \quad \text{for each } j.$$

Thus the second term in (2.54) can be written

$$-2 \sum \sum x_i x_j \alpha_j \dot{\theta}_{ij} = - \sum \sum x_i x_j (\alpha_i + \alpha_j) \dot{\theta}_{ij}.$$

The final two terms in (2.54) thus become

$$\sum \sum (a_{ij} - \alpha_i - \alpha_j) x_i x_j \dot{\theta}_{ij} = \sum \sum \delta_{ij} X_{ij} (\dot{\theta}_{ij} / \theta_{ij}),$$

where $\delta_{ij} = a_{ij} - \alpha_i - \alpha_j$ is a measure of nonadditivity in the Malthusian parameters m_{ij} . We conclude that

$$\dot{m} = \sigma_A^2 + \sum \sum X_{ij} \delta_{ij} d(\log \theta_{ij})/dt. \quad (2.55)$$

Thus the rate of increase of mean fitness is equal to the additive genetic variance in general only if Hardy–Weinberg frequencies hold (which, as we have seen in our model at least, they do not) or if the Malthusian parameter is additive ($m_{ij} = \alpha_i + \alpha_j$). A more general and more important conclusion, with m_{ij} no longer kept constant, is given by Kimura (1958).

How important then are departures from Hardy–Weinberg frequencies? In our model (2.45) shows that departures will be negligible after some time has passed if $d_i + d_j - d_{ij} - \bar{d} = 0$. But there is another circumstance under which departures will also be negligible. Suppose that the deviations $b_{ij} - \bar{b}$ and $d_{ij} - \bar{d}$ are all of order s , where s is a small parameter. Then Nagylaki (1976) has shown that the deviation Q_{ij} defined above changes in time (according to (2.45)) in such a way that after a small time period t_1 (an explicit formula for which is given by Nagylaki), Q_{ij} differs from zero only by a term of order s , even though at that time the gene frequencies themselves may be far from their equilibrium values. After time $2t$, the rate of change of Q_{ij} is of order s^2 . When this occurs a state of “quasi-Hardy–Weinberg” (QHW) is said to obtain. In this case departures from Hardy–Weinberg frequencies may be trivial, and as a consequence the mean fitness increase theorem should hold to an excellent approximation. More exactly, under the assumptions we have made, the term σ_A^2 in (2.55) is of order s^2 , and when QHW obtains the final term is of order s^3 . Thus the first term on the right-hand side will dominate the second, leading, as noted, to the essential accuracy of the theorem. The only exception to this rule occurs when the various frequencies are close to their respective equilibrium points: Since $\sigma_A^2 = 0$ at equilibrium, it is possible that near equilibrium σ_A^2 is smaller than the final term in (2.55). This is probably

of minor importance, and during the period of substantial change in gene frequencies the MFIT is effectively true.

2.8 Non-Random-Mating Populations

In this section and the next we consider properties of the discrete-time models considered above, focussing attention on the case where random mating is no longer assumed. In this section we consider calculations associated with the one-locus version of the mean fitness increase theorem (MFIT) and in the next on calculations associated with the Fundamental Theorem of Natural Selection (FTNS). In both sections we use a notation that generalizes readily to the multilocus extensions considered in later chapters.

Suppose that fitness depends on the genotype at one locus only, at which occur alleles A_1, A_2, \dots, A_k . Any form of mating is allowed, random or otherwise. We denote the frequency of the (ordered) genotype $A_u A_v$ at the time of conception of any generation of individuals by X_{uv} ($= X_{vu}$), so that the frequency x_u of the allele A_u is given by $x_u = \sum_v X_{uv}$.

We assume that the genotype $A_u A_v$ has (viability) fitness w_{uv} . The mean fitness \bar{w} of the population is then given by

$$\text{mean fitness} = \bar{w} = \sum_u \sum_v w_{uv} X_{uv}. \quad (2.56)$$

The additive genetic variance in fitness is found by the non-random-mating generalization of the procedure that led to the “random-mating” expression (2.17). That is, it is found by minimizing the function S , now defined more generally than in (2.15) as

$$S = \sum_u \sum_v X_{uv} (w_{uv} - \bar{w} - \alpha_u - \alpha_v)^2, \quad (2.57)$$

subject to the constraint

$$\sum_u x_u \alpha_u = 0. \quad (2.58)$$

The values of $\alpha_1, \alpha_2, \dots, \alpha_k$ found through this minimizing procedure, that is the *average effects* of the alleles A_1, A_2, \dots, A_k , are the implicit solutions of the equations

$$x_u \alpha_u + \sum_v X_{uv} \alpha_v = x_u a_u, \quad u = 1, 2, \dots, k, \quad (2.59)$$

where a_u , the *average excess* of the allele A_u , is given by

$$a_u = x_u^{-1} \sum_v X_{uv} (w_{uv} - \bar{w}). \quad (2.60)$$

Equation (2.59) shows that, under random mating, the average effect α_u of A_u and the average excess a_u of A_u are equal, since under random mating the second term on the left-hand side of (2.59) is 0. When mating is not random, α_u and a_u are, in general, different from each other.

Standard regression theory shows that the sum of squares removed by fitting the α_j values in (2.57), that is the additive genetic variance σ_A^2 , is given by

$$\sigma_A^2 = 2 \sum_u x_u a_u \alpha_u. \quad (2.61)$$

With the definition of a_u given in (2.60), the change Δx_u in the frequency of A_u between consecutive generations is

$$\Delta x_u = x_u a_u / \bar{w}, \quad (2.62)$$

so that an alternative expression for the additive genetic variance is

$$\sigma_A^2 = 2\bar{w} \sum_u \alpha_u \Delta x_u. \quad (2.63)$$

Similarly an alternative set of formulas implicitly defining the quantities $\{\alpha_u\}$ is

$$x_u \alpha_u + \sum_v X_{uv} \alpha_v = \bar{w} \Delta x_u, \quad u = 1, 2, \dots, k. \quad (2.64)$$

If we define D as a diagonal matrix whose u th term is x_u , P as a matrix whose (u, v) th term is X_{uv} , Δ as a vector of the Δx_u values and α as a vector of the α_u values, this equation can be written in matrix and vector form as

$$(D + P)\alpha = \bar{w}\Delta. \quad (2.65)$$

When this matrix form is used, the extension of the definition of the α_u to the multilocus case in Chapter 7 will be almost immediate.

An explicit solution of the equations in (2.59) for the α_u values is not in general possible. However in the two-allele case an explicit solution of is straightforward. For this case we get

$$\alpha_u = a_u x_1 x_2 / \{X_{11} X_{12} + 2X_{11} X_{22} + X_{12} X_{22}\}, \quad u = 1, 2. \quad (2.66)$$

Under random mating $X_{uv} = x_u x_v$, and this equation confirms that in this case a_u and α_u are equal. The equation also shows that under non-random mating, a_u and α_u have the same sign and are zero or nonzero together. In the two-allele case Fisher often described $\alpha_2 - \alpha_1$ as the average effect of replacing A_1 by A_2 , but in the k allele case, to which we now return, the definition of α_u simply as the average effect of A_u is rather more flexible.

We now consider the change in mean fitness from one generation to another. We write

$$w_{uv} = \bar{w} + \alpha_u + \alpha_v + \epsilon_{uv},$$

and with this definition, (2.58) implies that

$$\sum_u \sum_v X_{uv} \epsilon_{uv} = 0. \quad (2.67)$$

The frequency of A_u at the birth of any given generation is $\sum_v X_{uv}$, and in the next generation at birth it will be $\sum_j X_{uj} w_{uj} / \bar{w}$. Thus the change in mean fitness between consecutive generations becomes

$$\begin{aligned} \Delta \bar{w} &= \sum_u \sum_v X'_{uv} w_{uv} - \bar{w} \\ &= \sum_u \sum_v X'_{iuv} (\alpha_u + \alpha_v + \epsilon_{uv}) \\ &= 2 \sum_u \alpha_u x'_u + \sum_u \sum_v (X_{uv} + \Delta X_{uv}) \epsilon_{uv} \quad (2.68) \\ &= 2 \sum_u \alpha_u (\Delta x_u) + \sum_u \sum_v (\Delta X_{uv}) \epsilon_{uv} \quad (\text{from (2.59) and (2.67)}) \\ &= \sigma_A^2 / \bar{w} + \sum_u \sum_v (\Delta X_{uv}) \epsilon_{uv}, \quad (\text{from (2.61) and (2.62)}). \end{aligned}$$

If the second term on the right-hand side of this expression is small, the conclusion of the mean fitness increase theorem approximately applies.

2.9 The Fundamental Theorem of Natural Selection

We now turn to the Fundamental Theorem of Natural Selection (FTNS), considering first the discrete-time version, and later the continuous-time version, of this theorem.

Equation (2.58) shows that $\sum_u \sum_v X_{uv} (\alpha_u + \alpha_v) = 0$, and from this the mean fitness \bar{w} may be written in the form

$$\bar{w} = \sum_u \sum_v X_{uv} (\bar{w} + \alpha_u + \alpha_v). \quad (2.69)$$

In the FTNS, Fisher considered the change in mean fitness from one generation to another *only* through changes in the frequencies X_{uv} in the expression (2.69), with the quantities \bar{w} , α_u and α_v being kept constant. This is called the “partial change” in mean fitness, and we denote it by $\Delta_p(\bar{w})$. If X'_{uv} is the frequency of the (ordered) genotype $A_u A_v$ in the

daughter generation, this partial change $\Delta_p(\bar{w})$ is

$$\Delta_p(\bar{w}) = \sum_u \sum_v (X'_{uv} - X_{uv})(\bar{w} + \alpha_u + \alpha_v) \quad (2.70)$$

$$\begin{aligned} &= \sum_u \sum_v (X'_{uv} - X_{uv})(\alpha_u + \alpha_v) \\ &= 2 \sum_u \alpha_u \sum_j (X'_{uj} - X_{uj}) \\ &= 2 \sum_u \alpha_u \Delta x_u \end{aligned} \quad (2.71)$$

$$= \sigma_A^2 / \bar{w}. \quad (2.72)$$

The final step in this sequence comes from (2.63).

We call the interpretation of the FTNS in the above form the “Price” interpretation, since it was first given by Price (1972). This interpretation follows the spirit of the wording in Fisher (1930, 1958).

Thus the partial change in mean fitness is exactly equal to σ_A^2 / \bar{w} , and this is the one-locus statement of the FTNS. Thus, as asserted by Fisher (1930, 1958), the FTNS is an exact result, implying no approximations, and it applies to non-random-mating as well as random-mating populations, since no assumption about the mating scheme is made in the analysis. We extend the FTNS as an exact result in Chapter 7 to the case where fitness depends on an arbitrary number of loci, up to and including all those in the entire genome, under any form of mating, random or otherwise.

An alternative way of writing the FTNS in this interpretation is

$$\Delta_p(\bar{w}) = \sum_u \sum_v (\Delta X_{uv})(w_{uv})_\alpha = \sigma_A^2 / \bar{w}. \quad (2.73)$$

Here $(w_{uv})_\alpha = \bar{w} + \alpha_u + \alpha_v$ may be thought of as the best estimate of the fitness of the genotype $A_u A_v$ as predicted from the alleles in that genotype. In this form the Price interpretation bears an interesting similarity to a second interpretation to the FTNS, one which is closer in spirit to the wording in Fisher (1941), and which was developed by Lessard (1997). Lessard’s interpretation uses a concept of partial change different from, although mathematically equivalent to, that in the Price interpretation. In the Lessard interpretation the actual fitness w_{uv} of the genotype $A_u A_v$ is retained, but the change in genotype frequency is replaced by a “alleles derived” value. More explicitly, the statement of the theorem under this interpretation is that

$$\Delta_p(\bar{w}) = \sum_u \sum_v (\Delta X_{uv})_\alpha w_{uv} = \sigma_A^2 / \bar{w}, \quad (2.74)$$

where $(\Delta X_{uv})_\alpha$ is defined by

$$(\Delta X_{uv})_\alpha = \frac{(X_{uv})(\alpha_u + \alpha_v)}{\bar{w}}. \quad (2.75)$$

$(\Delta X_{uv})_\alpha$ is not the actual change in the frequency of the genotype $A_u A_v$ from one generation to another, but is thought of the change as predicted from the alleles A_u and A_v in that genotype. The similarity of the forms of the middle terms in (2.73) and (2.74), and the identity of the right-hand sides, together indicate the mathematical identity of the two concepts of partial change. The difference between the two concepts is in the interpretation: In the first interpretation the genes in a genotype may be thought of as assessing the genotype fitness, while in the second they may be thought of as assessing the change in the frequency of that genotype.

The background to Lessard's interpretation of the FTNS is as follows. Fisher (1941) discussed in some detail the circumstances under which the equation

$$\frac{\Delta X_{uu}}{X_{uu}} + \frac{\Delta X_{vv}}{X_{vv}} = 2 \frac{\Delta X_{uv}}{X_{uv}} \quad (2.76)$$

will hold for all u and v . If these equations do hold for all u and v , then $\Delta X_{uv}/X_{uv}$ can be expressed in the form

$$\frac{\Delta X_{uv}}{X_{uv}} = \beta_u + \beta_v, \quad (2.77)$$

for some set of constants $\beta_1, \beta_2, \dots, \beta_k$. From this,

$$X_{uv}(\beta_u + \beta_v) = \Delta X_{uv}. \quad (2.78)$$

Summation in this identity over all v gives

$$x_u \beta_u + \sum_v X_{uv} \beta_v = \Delta x_u \quad \text{for all } u. \quad (2.79)$$

Equation (2.64) then shows that we may take $\beta_v = \alpha_v/\bar{w}$ for all v , where α_v is the average effect of A_v . It follows from (2.77) that

$$\Delta X_{uv} = \frac{(X_{uv})(\alpha_u + \alpha_v)}{\bar{w}}. \quad (2.80)$$

Comparison of this equation with (2.75) shows that when all the equations of the form (2.76) hold, the actual change genotype frequency (2.80) is identical to the change as assessed by the alleles in the genotype. This implies that the *total* change in mean fitness is equal to the *partial* change defined in both equation (2.72) and equation (2.75).

However, equation (2.76) will hold only under very restrictive mating conditions. The random-mating case is perhaps the most important of these. Under random mating the equation $X_{uv}^2 = 4X_{uu}X_{vv}$ holds, so that $2 \log X_{uv} = \log 4 + \log X_{uu} + \log X_{vv}$. From this, $2\Delta \log X_{uv} = \Delta \log X_{uu} + \Delta \log X_{vv}$. If small-order terms are ignored, so that $\Delta \log x$ can be replaced by $(\Delta x)/x$, equation (2.76) then follows. More generally the conclusion still follows, to this level of approximation, if $X_{uv}^2 = \lambda X_{uu}X_{vv}$ for any fixed constant λ . Again ignoring small-order terms, it follows that the restriction $X_{uv}^2 = \lambda X_{uu}X_{vv}$ is required for the total change in mean

fitness to be predictable from parental generation genotype frequencies and fitnesses. The point of the FTNS is that random mating is *not* required for the theorem to hold, so that (2.76) does not necessarily hold. Then the total change in mean fitness is not predictable unless the mating scheme is known. Despite this, the FTNS holds whatever the mating scheme might be, and whether it be known or unknown.

It is straightforward to give also a continuous-time version of the FTNS. This shows that the continuous-time partial rate of change in mean fitness, defined as

$$\sum_u \sum_v \left(\frac{d}{dt} X_{uv} \right) (\bar{w} + \alpha_u + \alpha_v) \quad (2.81)$$

is exactly equal to the additive genetic variance. We do not provide the details since they closely follow those in the discrete-time case.

What biological relevance does the FTNS have? There are two points to raise here. First, the restrictive assumptions made in the theorem should be noted. Matters such as geographical dispersion, the existence of two sexes, stochastic changes in gene frequency in finite populations, and so on are ignored. On the other hand fertility selection is handled by Lessard and Castilloux's (1995) extension of the theorem to that case. Second, Fisher viewed the partial change in mean fitness as that change brought about by natural selection. It is not clear how this interpretation can be sustained, and it is possible that the MFIT, even though it is restricted to random-mating populations and, as we show in the following section, might not hold when fitness depends on a two-locus and more generally a multilocus genotype, nevertheless gives a greater biological insight into the evolutionary process than does the FTNS. Associated with this view is the approach, initiated by Nagylaki (1974c), which delimits the circumstances under which the MFIT is approximately true.

2.10 Two Loci

So far in this chapter we have assumed that the fitness of any individual depends on his genetic constitution at a single locus. This is of course only an initial simplification: We have already noted in Chapter 1 that for some questions, for example, the evolution of recombination rate, a more complicated theory is required. We now introduce briefly the case where fitness depends on the genetic constitution at two loci, deferring a more complete treatment to Chapter 6. Although such a "two-locus" theory may often be little more realistic than "single-locus" theory, it does allow at least two advances to be made. First, some assessment can be made of the accuracy of approximating two-locus behavior and measurements by combining two single-locus results. Second, no assessment of the evolutionary importance of linkage between loci can be made without at least a two-locus analysis.

For convenience we assume viability selection only, random mating and discrete nonoverlapping generations. Consider two loci “*A*” and “*B*” at which occur alleles A_1, A_2 and B_1, B_2 , respectively, and let the recombination fraction between the loci be R ($0 < R \leq 0.5$). (When $R = 0$ the two loci in effect become one locus, the theory of which has already been considered. This is why we impose the assumption $R > 0$.) It is convenient conceptually to suppose that these loci are on the same chromosome: The unlinked case ($R = 0.5$) may be treated by imagining the distance along the chromosome between the two loci to be so long that the recombination fraction between them is 0.5. We then use the words gamete and chromosome interchangeably in what follows.

It is possible to write down recurrence relations connecting the (ten) zygotic frequencies (of $A_1B_1/A_1B_1, A_1B_2/A_1B_1, \dots, A_2B_2/A_2B_2$). These relations show that a simpler set of recurrence relations can be found for the frequencies of the four gametes A_1B_1, A_1B_2, A_2B_1 and A_2B_2 , called here gametes 1, 2, 3, 4, respectively. This simplification arises through the concept of the random union of gametes and is parallel to treating gene frequencies rather than genotypic frequencies at a single locus.

We consider first the case where there is no selection. The gametes forming the zygotes of any generation may be thought of as being drawn randomly from a pool containing gametes of type 1–4 in certain proportions. These gametes will not necessarily be passed on to the next generation of gametes in the same proportions since, for example, there will be a decrease in the frequency of A_1B_1 gametes through recombination in A_1B_1/A_2B_2 individuals which might not be exactly counterbalanced by an increase through recombination in A_1B_2/A_2B_1 individuals. If the frequency of gamete i is denoted c_i ($i = 1, \dots, 4$), these arguments and some straightforward calculations show that the frequencies c'_i in the next generation are given by

$$\begin{aligned} c'_1 &= c_1 + R(c_2c_3 - c_1c_4), \\ c'_2 &= c_2 - R(c_2c_3 - c_1c_4), \\ c'_3 &= c_3 - R(c_2c_3 - c_1c_4), \\ c'_4 &= c_4 + R(c_2c_3 - c_1c_4), \end{aligned} \tag{2.82}$$

or more economically as

$$c'_i = c_i + \eta_i R(c_2c_3 - c_1c_4), \tag{2.83}$$

where

$$\eta_1 = \eta_4 = 1, \quad \eta_2 = \eta_3 = -1. \tag{2.84}$$

Several conclusions can be drawn immediately from these equations. First, since $c'_1 + c'_2 = c_1 + c_2$ and $c'_1 + c'_3 = c_1 + c_3$, there is no change in the frequencies of A_1 and B_1 . This confirms, fortunately, the one-locus analysis

of Chapter 1. Second, elementary algebra shows that

$$c'_1c'_4 - c'_2c'_3 = (1 - R)(c_1c_4 - c_2c_3), \quad (2.85)$$

so that since $R > 0$,

$$c_1(t)c_4(t) - c_2(t)c_3(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty. \quad (2.86)$$

It follows that under the assumptions we have made, in particular that of no selection, we may reasonably assume that the equation

$$c_1c_4 - c_2c_3 = 0 \quad (2.87)$$

holds if the population has evolved for some time. It is important to establish what this equation means in genetical terms. Algebraic manipulation shows that (2.87) is equivalent to

$$\text{freq}(A_iB_j) = \text{freq}(A_i) \times \text{freq}(B_j) \quad (2.88)$$

for all possible pairs i, j . When (2.88), or equivalently (2.87), holds, the population is said to be in a state of *linkage equilibrium* with respect to these loci. The quantity $c_1c_4 - c_2c_3$, which we denote by D , is often called the “coefficient of linkage disequilibrium”. As we see below, this can be a rather misleading expression for the quantity $c_1c_4 - c_2c_3$, which we would prefer to call the “coefficient of association”. An alternative expression for D , sometimes more useful than $c_1c_4 - c_2c_3$, is

$$D = c_1 - \text{freq. } A_1 \times \text{freq. } B_1. \quad (2.89)$$

We turn now to the case where selective differences between genotypes exist. In the previous chapter we used a fitness display such as that in (1.92), which focusses attention on the genotypes at each of the two loci. For theoretical purposes, however, it is usually more convenient to adopt a notation focussed around the two gametes making up each individual. This is so since, as (2.82) shows, gametic frequencies are the most natural vehicle for studying evolutionary behavior in two-locus systems under random mating. We thus adopt the fitness scheme shown in (2.90) below:

	A_1B_1	A_1B_2	A_2B_1	A_2B_2	
A_1B_1	w_{11}	w_{12}	w_{13}	w_{14}	
A_1B_2	w_{21}	w_{22}	w_{23}	w_{24}	(2.90)
A_2B_1	w_{31}	w_{32}	w_{33}	w_{34}	
A_2B_2	w_{41}	w_{42}	w_{43}	w_{44}	

In the notation of this fitness scheme the fitness of zygotes made up of gametes i and j is written as w_{ij} (which we assume equal to w_{ji}). If coupling and repulsion double heterozygotes have the same fitness, then also $w_{23} = w_{14}$. We make this assumption throughout. If, for specific purposes, we wish to adopt a fitness display emphasizing single-locus genotypes, (2.90)

becomes

$$\begin{array}{cccc}
 & B_1B_1 & B_1B_2 & B_2B_2 \\
 A_1A_1 & w_{11} & w_{12} & w_{22} \\
 A_1A_2 & w_{13} & w_{14} & w_{24} \\
 A_2A_2 & w_{33} & w_{34} & w_{44}
 \end{array} \tag{2.91}$$

The marginal fitness w_i of gamete i is defined by

$$w_i = \sum_j c_j w_{ij}, \tag{2.92}$$

and the mean fitness \bar{w} of the population then becomes

$$\bar{w} = \sum_i \sum_j c_i c_j w_{ij} = \sum_i c_i w_i. \tag{2.93}$$

Consideration of all possible matings, their frequencies, and their genetic outputs, as well as the fitnesses of the various genotypes, shows that the gametic frequencies c'_i in the following generation are given by

$$c'_i = \bar{w}^{-1} (c_i w_i + \eta_i R w_{14} (c_2 c_3 - c_1 c_4)), \quad i = 1, 2, 3, 4. \tag{2.94}$$

Here η_i is defined in (2.84). If the w_{ij} are all equal, these recurrence relations reduce to (2.83). These important equations are due in this form to Lewontin and Kojima (1960), but they were essentially derived earlier, for a continuous-time model, by Kimura (1956b). Our present aim is to discuss some of the more immediate consequences of these equations.

First, the mean fitness, as defined in (2.93), is similar in form to the definition (2.10) with $k = 4$. It follows from the discussion in Section 2.4 that if we assume that mean fitness is maximized at a unique internal ($c_i > 0$) point, then at this point $w_i = \bar{w}$, where now w_i and \bar{w} defined by (2.92) and (2.93). What is the connection between this maximization point and the equilibrium points of the system (2.94)? The equations $c'_i = c_i$ show that the system (2.94) is in equilibrium when

$$\bar{w} = w_i + c_i^{-1} \eta_i R w_{14} (c_2 c_3 - c_1 c_4), \quad i = 1 \dots 4. \tag{2.95}$$

Unless linkage equilibrium holds at the equilibrium point, this point cannot be a point of maximum fitness. We show later that linkage equilibrium holds at equilibrium only in special cases, so that mean fitness can decrease in the system (2.94). The MFIT cannot then be true in general in two-locus selection systems. By contrast, we shall show in Section 7.4.5 that the FTNS does hold with a multilocus fitness scheme, and thus in particular with a two-locus fitness scheme.

We now demonstrate the possible decrease in mean fitness by a numerical example. Suppose, using the notation (2.91), that the fitness scheme is

	B_1B_1	B_1B_2	B_2B_2	
A_1A_1	1.000	1.024	1.021	(2.96)
A_1A_2	1.025	1.066	1.026	
A_2A_2	1.018	1.019	1.007	

and let $R = \frac{1}{2}$, so that A and B loci unlinked. If initially

$$c_1 = 0.168, \quad c_2 = 0.362, \quad c_3 = 0.292, \quad c_4 = 0.178, \quad (2.97)$$

the population mean fitness is 1.033106. The mean fitness now *decreases* for about 14 generations and after that steadily increases, reaching a value of 1.031212 at the equilibrium point

$$c_1 = 0.24136, \quad c_2 = 0.28164, \quad c_3 = 0.22192, \quad c_4 = 0.25508. \quad (2.98)$$

The net effect of the evolution of the population from the starting point (2.97) to the equilibrium point (2.98) is to decrease mean fitness by 0.001894. At this equilibrium point the value of $D = c_1c_4 - c_2c_3$ is -0.000935 .

Apart from the fact that mean fitness can decrease, the above analysis demonstrates two further points. The first is that the coefficient of linkage disequilibrium can be nonzero at an equilibrium point of the evolutionary system, even though the two loci upon which fitness depends are unlinked. This is why we prefer the term “coefficient of association” for the quantity $c_1c_4 - c_2c_3$, rather than the term “coefficient of linkage disequilibrium”.

The second point to observe is that the location of the equilibrium point or points of (2.94) will depend on the recombination fraction R between the loci in those cases where linkage equilibrium does not obtain at equilibrium. Thus various values of R can be considered and the equilibrium mean fitnesses computed for each. When $R = 0$ the “equilibrium” equation (2.95) and the “maximization” equation $\bar{w} = w_i$ ($i = 1, \dots, 4$) agree, so that if each $c_i > 0$ at equilibrium, the value of R for which the greatest equilibrium mean fitness is achieved is for $R = 0$. This conclusion remains true if some of the c_i are zero at equilibrium but strangely, as we see later, it is not necessarily true that equilibrium mean fitness is a monotonically decreasing function of R . To the extent that equilibrium mean fitness is maximized for extremely tight linkage, the argument of Fisher given in Chapter 1 concerning the evolution of tight linkage between epistatic loci is justified. This argument can be made only when $D \neq 0$ at equilibrium for all R values: If $D = 0$ at equilibrium for all R the equilibrium mean fitness is independent of R .

The third topic we treat, at rather greater length, concerns the additive genetic variance in fitness. We are particularly interested in the relationship between this and the two marginal single-locus values, and we begin by

defining the latter. Using the fitness scheme (2.91), we may define the marginal fitnesses of the various single-locus genotypes as follows:

Genotype	Frequency	Marginal Fitness
A_1A_1	$(c_1 + c_2)^2$	$(w_{11}c_1^2 + 2w_{12}c_1c_2 + w_{22}c_2^2)/(c_1 + c_2)^2 = u_{11}$
A_1A_2	$2(c_1 + c_2) \times (c_3 + c_4)$	$(w_{13}c_1c_3 + w_{14}c_1c_4 + w_{14}c_2c_3 + w_{24}c_2c_4)/(c_1 + c_2)(c_3 + c_4) = u_{12}$
A_2A_2	$(c_3 + c_4)^2$	$(w_{33}c_3^2 + 2w_{34}c_3c_4 + w_{44}c_4^2)/(c_3 + c_4)^2 = u_{22}$
B_1B_1	$(c_1 + c_3)^2$	$(w_{11}c_1^2 + 2w_{13}c_1c_3 + w_{33}c_3^2)/(c_1 + c_3)^2 = v_{11}$
B_1B_2	$2(c_1 + c_3) \times (c_2 + c_4)$	$(w_{12}c_1c_2 + w_{14}c_1c_4 + w_{14}c_2c_3 + w_{34}c_3c_4)/(c_1 + c_3)(c_2 + c_4) = v_{12}$
B_2B_2	$(c_2 + c_4)^2$	$(w_{22}c_2^2 + 2w_{24}c_2c_4 + w_{44}c_4^2)/(c_2 + c_4)^2 = v_{22}$

(2.96)

From (1.42), the marginal additive genetic variance at the A locus may be defined as

$$2(c_1 + c_2)(c_3 + c_4)G_A^2, \quad (2.97)$$

where

$$G_A = u_{11}(c_1 + c_2) + u_{12}(1 - 2c_1 - 2c_2) - u_{22}(c_3 + c_4). \quad (2.98)$$

Similarly the marginal additive genetic variance at the B locus is

$$2(c_1 + c_3)(c_2 + c_4)G_B^2, \quad (2.99)$$

where

$$G_B = v_{11}(c_1 + c_3) + v_{12}(1 - 2c_1 - 2c_3) - v_{22}(c_2 + c_4). \quad (2.100)$$

We now find the two-locus additive genetic variance. To do this we assign additive parameters α_{11} and α_{12} to A_1 and A_2 and α_{21} and α_{22} to B_1 and B_2 , and then minimize the expression

$$S = c_1^2(w_{11} - \bar{w} - 2\alpha_{11} - 2\alpha_{21})^2 + 2c_1c_2(w_{12} - \bar{w} - 2\alpha_{11} - \alpha_{21} - \alpha_{22})^2 + \cdots + c_4^2(w_{44} - \bar{w} - 2\alpha_{12} - 2\alpha_{22})^2$$

with respect to the α_{ij} . Now that two loci are involved in the minimization it is appropriate to add constraints on the α_{ij} , since, for example, adding some constant to each α_{1x} and subtracting the same constant from each α_{2x} does not change the value of S . Such a change would, however, affect the definitions of marginal additive genetic variances. The natural constraints to impose are those which arise automatically in the one-locus case as given in (2.58). In the two-locus case these are

$$(c_1 + c_2)\alpha_{11} + (c_3 + c_4)\alpha_{12} = 0, \quad (c_1 + c_3)\alpha_{21} + (c_2 + c_4)\alpha_{22} = 0, \quad (2.101)$$

and the minimization is carried out subject to these constraints. Details of this procedure are given by Kojima and Kelleher (1961) and Kimura (1965)

and are not pursued here. It is found that the additive genetic variance can be written as

$$2\{(c_1 + c_2)(c_3 + c_4)H_A^2 + 2H_AH_BD + (c_1 + c_3)(c_2 + c_4)H_B^2\}, \quad (2.102)$$

where H_A and H_B are the solutions of the equations

$$\begin{aligned} H_A + \{(c_1 + c_2)(c_3 + c_4)\}^{-1}DH_B &= G_A, \\ H_B + \{(c_1 + c_3)(c_2 + c_4)\}^{-1}DH_A &= G_B, \end{aligned} \quad (2.103)$$

G_A and G_B being given by (2.98) and (2.100).

Several interesting conclusions follow from these equations. Perhaps the most important is that if $D = 0$ (that is, linkage equilibrium between the two loci) then $H_A = G_A$, $H_B = G_B$, and the true two-locus additive genetic variance is the sum of the two single-locus marginal values. When $D \neq 0$ this is no longer true, and there is no simple relationship between this sum and the true two-locus additive genetic variance value. This is an important conclusion since it seems to be widely assumed in the classical literature (see for example Fisher (1918, p. 405), (1958, p. 37) and Wright (1969, p. 439)) that in a multilocus system the true additive genetic variance can be found by simply summing single-locus marginal values. Since we have shown above that changes in mean fitness can be negative in two-locus systems, and thus cannot be equal to any form of genetic variance, it follows that

$$\Delta\bar{w}, \quad \sigma_A^2(\text{two-locus}), \quad \sum \sigma_A^2(\text{single-locus marginals}) \quad (2.104)$$

have in general no clear and obvious connection with each other. This conclusion is generalized in Section 7.3.3.

These conclusions may also be associated with properties of changes in gene frequency. Equations (2.97), (2.99), and (2.102) show that

$$\sigma_A^2(\text{two-locus}) - \sum \sigma_A^2(\text{single-locus marginals}) = 2D(G_AH_B + H_AG_B), \quad (2.105)$$

and if D is small this may be approximated by $-4DG_AG_B$. Since

$$\Delta(\text{frequency } A_1) = (c_1 + c_2)(c_3 + c_4)G_A/\bar{w},$$

with a corresponding expression for $\Delta(\text{frequency } B)$, it is found, if terms of order D^2 are ignored, that the left-hand side in (2.105) may be written

$$\frac{-4D\bar{w}^2\Delta(\text{frequency } A_1)\Delta(\text{frequency } B_1)}{(c_1 + c_2)(c_3 + c_4)(c_1 + c_3)(c_2 + c_4)}.$$

This gives an interesting relationship between the various additive genetic variances, the linkage disequilibrium, and the gene frequency changes in a two-locus system. If in a certain generation $\Delta(\text{frequency } A_1) = 0$, then to the order of accuracy we use the equation $G_A = 0$ holds, and the total additive genetic variance is simply the marginal B locus value. However, this is true only as an approximation and, more precisely, whenever there

is linkage disequilibrium between A and B loci there is a small perturbation from the A locus to the total additive variance, even though gene frequencies are not changing at that locus.

We expect the additive genetic variance to be of importance in discussing the correlation between relatives. Before exploring this, we recall that gene frequencies alone are not sufficient to describe the evolution of two-locus systems, so that it is reasonable to argue that the additive genetic variance, which fundamentally involves gene frequencies, is not the appropriate component of variance for evolutionary considerations. We thus consider a variance defined by gamete frequencies which, since gamete frequencies do describe the evolutionary behavior, might be thought to be of greater evolutionary significance than the additive genetic variance.

The marginal fitnesses w_i of the four gametes have been defined in (2.92). The total chromosomal, or gametic, variance in fitness, denoted σ_G^2 , may be defined by

$$\sigma_G^2 = 2 \sum_{i=1}^4 (w_i - \bar{w})^2 c_i, \quad (2.106)$$

the factor 2 being inserted because there are two gametes per zygote. Suppose now we attempt to fit the marginal gametic fitnesses by additive components depending on the genes on each gamete. This is done by minimizing

$$\begin{aligned} & c_1(w_1 - \bar{w} - \alpha_{11} - \alpha_{21})^2 + c_2(w_2 - \bar{w} - \alpha_{11} - \alpha_{22})^2 \\ & + c_3(w_3 - \bar{w} - \alpha_{12} - \alpha_{21})^2 + c_4(w_4 - \bar{w} - \alpha_{12} - \alpha_{22})^2 \end{aligned}$$

with respect to α_{11} , α_{12} , α_{21} and α_{22} , subject to the constraints in (2.101). The sum of squares so removed may be described as being due to the additive effects of genes within gametes, and for short may be called the additive gametic variance. It is found (see Kimura, (1965)) that this is identical to the additive genetic variance (2.102) and thus the latter, perhaps unexpectedly, is of use in evolutionary and other considerations. This conclusion is generalized in Section 7.3.3. The total gametic variance in (2.106) has three degrees of freedom, of which the additive component of it has two. The remaining degree of freedom is taken up by the epistatic gametic variance σ_{EG}^2 , which is

$$\sigma_{EG}^2 = 2(w_1 - w_2 - w_3 + w_4)^2 / (c_1^{-1} + c_2^{-1} + c_3^{-1} + c_4^{-1}). \quad (2.107)$$

This is zero if and only if an additive genetic fitness scheme exactly fits the marginal gametic fitnesses.

We turn now to the correlation between relatives, restricting attention to the case where (2.88) holds, that is that the two loci are in linkage equilibrium. This assumption was also made by Fisher (1918). We consider both linked and unlinked loci: Fisher's 1918 analysis is concerned only with

the unlinked case. Our treatment is based on Cockerham (1954, 1956) and Kempthorne (1954).

We first isolate various components of the total variance of the character measured. Suppose that the measurements for the various genotypes are

$$\begin{array}{cccc}
 & B_1B_1 & B_1B_2 & B_2B_2 \\
 A_1A_1 & m_{11} & m_{12} & m_{13} \\
 A_1A_2 & m_{21} & m_{22} & m_{23} \\
 A_2A_2 & m_{31} & m_{32} & m_{33}
 \end{array} \quad (2.108)$$

We form these measurements into a single vector $\mathbf{m} = (m_{11}, m_{12}, \dots, m_{33})'$. If the frequency of A_1 is x and of B_1 is y , then since linkage equilibrium is assumed, the frequency of $A_1A_1B_1B_1$ is x^2y^2 , of $A_1A_1B_1B_2$ is $2x^2y(1-y)$ and so on. It is convenient to write these frequencies as the entries in a diagonal matrix F , so that

$$F = \begin{pmatrix} x^2y^2 & & & \\ & 2x^2y(1-y) & & 0 \\ & & \ddots & \\ & 0 & & (1-x)^2(1-y)^2 \end{pmatrix}. \quad (2.109)$$

Evidently the mean value \bar{m} in the measurement is given by

$$\bar{m} = x^2y^2m_{11} + 2x^2y(1-y)m_{12} + \dots + (1-x)^2(1-y)^2m_{33}. \quad (2.110)$$

Further, adopting the notation of (2.96), the marginal means of A_1A_1 , A_1A_2 and A_2A_2 are

$$\begin{aligned}
 u_{11} &= y^2m_{11} + 2y(1-y)m_{12} + (1-y)^2m_{13}, \\
 u_{12} &= y^2m_{21} + 2y(1-y)m_{22} + (1-y)^2m_{23}, \\
 u_{22} &= y^2m_{31} + 2y(1-y)m_{32} + (1-y)^2m_{33}.
 \end{aligned} \quad (2.111)$$

Similarly the marginal means at the B locus are

$$\begin{aligned}
 v_{11} &= x^2m_{11} + 2x(1-x)m_{21} + (1-x)^2m_{31}, \\
 v_{12} &= x^2m_{12} + 2x(1-x)m_{22} + (1-x)^2m_{32}, \\
 v_{22} &= x^2m_{13} + 2x(1-x)m_{23} + (1-x)^2m_{33}.
 \end{aligned} \quad (2.112)$$

Finally the total variance σ^2 in the character measured is

$$\sigma^2 = x^2y^2m_{11}^2 + \dots + (1-x)^2(1-y)^2m_{33}^2 - \bar{m}^2 = \mathbf{m}'F\mathbf{m} - \bar{m}^2. \quad (2.113)$$

This total variance has eight degrees of freedom, and our aim is to break it down into the sum of eight components, each having one degree of freedom and each being of genetical significance. These components will measure two additive variances, one at each of the two loci, two dominance variances, one at each of the two loci and the four interaction variances.

Suppose a matrix T exists such that $TFT' = I$ (or equivalently $(T')^{-1}F^{-1}T^{-1} = I$), where I is the unit 9×9 matrix, and define a vector

\mathbf{z} by $\mathbf{z} = TF\mathbf{m}$. Then

$$\begin{aligned}\mathbf{m}'F\mathbf{m} &= \mathbf{z}'(T')^{-1}F^{-1}FF^{-1}T^{-1}\mathbf{z} \\ &= \mathbf{z}'\mathbf{z} \\ &= z_1^2 + z_2^2 + \cdots + z_9^2.\end{aligned}\quad (2.114)$$

If the last row in T can be chosen to be $(1, 1, \dots, 1)$, then $z_9 = \bar{m}$ and

$$\sigma^2 = z_1^2 + z_2^2 + \cdots + z_8^2. \quad (2.115)$$

The equation $TFT' = I$ reduces to the requirement

$$x^2y^2t_{i1}t_{j1} + 2x^2y(1-y)t_{i2}t_{j2} + \cdots + (1-x)^2(1-y)^2t_{i9}t_{j9} = \delta_{ij}, \quad (2.116)$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. The choice $t_{91} = t_{92} = \dots = t_{99} = 1$ does satisfy (2.116) with $i = j = 9$. Thus σ^2 can indeed be broken down into the sum (2.115), where

$$z_i = x^2y^2t_{i1}m_{11} + 2x^2y(1-y)t_{i2}m_{12} + \cdots + (1-x)^2(1-y)^2t_{i9}m_{33}, \quad (2.117)$$

provided that the t_{ij} satisfy (2.116) and the further requirement

$$x^2y^2t_{i1} + 2x^2y(1-y)t_{i2} + \cdots + (1-x)^2(1-y)^2t_{i9} = 0, \quad i = 1 \dots, 8. \quad (2.118)$$

Apart from these purely mathematical requirements we wish to choose that t_{ij} so that the z_i have the genetical interpretations described above.

Suppose z_1^2 and z_2^2 are to represent the additive and dominance variance components of the character from the A locus. Recalling equations (1.9) and using the marginal fitness values (2.111), we would like to have

$$\begin{aligned}z_1^2 &= 2x(1-x)\{xu_{11} + (1-2x)u_{12} - (1-x)u_{22}\}^2, \\ z_2^2 &= x^2(1-x)^2\{2u_{12} - u_{11} - u_{22}\}^2.\end{aligned}\quad (2.119)$$

Such a representation is in fact possible if, in (2.117), we choose

$$\begin{aligned}t_{11} &= t_{12} = t_{13} = x^{-1}\{2x(1-x)\}^{1/2}, \\ t_{14} &= t_{15} = t_{16} = (1-2x)\{2x(1-x)\}^{-1/2}, \\ t_{17} &= t_{18} = t_{19} = -(1-x)^{-1}\{2x(1-x)\}^{1/2},\end{aligned}\quad (2.120)$$

and

$$\begin{aligned}t_{21} &= t_{22} = t_{23} = -x^{-1}(1-x), \\ t_{24} &= t_{25} = t_{26} = 1, \\ t_{27} &= t_{28} = t_{29} = (1-x)^{-1}x.\end{aligned}\quad (2.121)$$

These choices do satisfy the requirements (2.116) and (2.118), and thus our desired representation (2.119) is allowable. A parallel procedure gives additive and dominance variance components at the B locus as

$$z_3^2 = 2y(1-y)\{yv_{11} + (1-2y)v_{12} - (1-y)v_{22}\}^2$$

and

$$z_4^2 = y^2(1-y)^2\{2v_{12} - v_{11} - v_{22}\}^2.$$

Once more, with the choice of the t_{ij} implicit in these definitions, the orthogonality conditions are met. If z_5^2 is to represent the additive-by-additive component of the total variance it would be natural to choose $t_{5i} = t_{1i} \times t_{3i}$, and the remaining three interactive components would naturally be chosen by similar multiplications. If this is done it is found that all the orthogonality conditions are met, and this also implies that the representation (2.115) is completed. We do not go into details here and note only that the various components can be expressed as

$$\begin{aligned} (\text{add} \times \text{add}): z_5^2 &= 4xy(1-x)(1-y)\{xye_{11} + x(1-y)e_{12} + (1-x)ye_{21} \\ &\quad + (1-x)(1-y)e_{22}\}^2, \tag{2.122} \\ (\text{add} \times \text{dom}): z_6^2 &= 2x(1-x)y^2(1-y)^2\{x(e_{11} - e_{12}) + (1-x)(e_{21} - e_{22})\}^2, \\ (\text{dom} \times \text{add}): z_7^2 &= 2x^2(1-x)^2y(1-y)\{y(e_{11} - e_{21}) + (1-y)(e_{12} - e_{22})\}^2, \\ (\text{dom} \times \text{dom}): z_8^2 &= x^2y^2(1-x)^2(1-y)^2\{e_{11} - e_{12} - e_{21} + e_{22}\}^2, \end{aligned}$$

where

$$\begin{aligned} e_{11} &= m_{11} - m_{12} - m_{21} + m_{22}, \\ e_{12} &= m_{12} - m_{13} - m_{22} + m_{23}, \\ e_{21} &= m_{21} - m_{22} - m_{31} + m_{32}, \\ e_{22} &= m_{22} - m_{23} - m_{32} + m_{33}. \end{aligned}$$

These expressions, given more generally to include the effect of inbreeding, were derived by Cockerham (1954). It is sometimes convenient to write

$$\sigma_A^2 = z_1^2 + z_3^2, \quad \sigma_D^2 = z_2^2 + z_4^2, \quad \sigma_{AA}^2 = z_5^2, \quad \sigma_{AD}^2 = z_6^2 + z_7^2, \quad \sigma_{DD}^2 = z_8^2,$$

so that

$$\sigma^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2. \tag{2.123}$$

A slightly shorter representation collects the final three terms as a single term σ_E^2 (epistatic variance), but for our purposes this is not useful, since the final three terms in (2.123) are involved differently in the correlation between relatives, and are therefore best kept separate.

Consider now the father-son and the full sib correlations in the measurement. It is possible to write down all 81 father-son genotypic combinations and, using a table extending Table 1.1, arrive at a father-son covariance. By doing this and a parallel procedure for full sibs, it is found that if the A and B loci are unlinked,

$$\text{corr}(\text{father-son}) = \left(\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_{AA}^2\right)/\sigma^2, \tag{2.124a}$$

$$\text{corr}(\text{full sibs}) = \left(\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \frac{1}{4}\sigma_{AA}^2 + \frac{1}{8}\sigma_{AD}^2 + \frac{1}{16}\sigma_{DD}^2\right)/\sigma^2. \tag{2.124b}$$

Cockerham (1956) demonstrated that, when the two loci are linked, the former expression remains unchanged but that the latter must be replaced by

$$\begin{aligned} \text{corr}(\text{full sibs}) = & \left\{ \frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \frac{1}{8}(3 - 4R + 4R^2)\sigma_{AA}^2 \right. \\ & + \frac{1}{4}(1 - 2R + 2R^2)\sigma_{AD}^2 \\ & \left. + \frac{1}{4}(1 - 2R + 2R^2)^2\sigma_{DD}^2 \right\} / \sigma^2. \end{aligned} \quad (2.125)$$

The effect of linkage is always to increase the full sib correlation compared to the value for the unlinked case. We derive these formulas later in Chapter 7 as particular cases of correlations where the trait in question depends on an arbitrary number of loci, using a more efficient approach.

The analysis in this section has assumed a discrete-time model, and it is expected that qualitatively similar conclusions would hold for a continuous model. One possible complication for such models does, however, occur. In the discrete models the frequency of any genotype is found immediately from the frequencies of the gametes making up this genotype, so that, for example,

$$\text{freq}(A_1 A_1 B_1 B_2) = 2 \text{freq}(A_1 B_1) \text{freq}(A_1 B_2). \quad (2.126)$$

In the continuous-time model of Nagylaki and Crow (1974) the existence of linkage disequilibrium between the two loci implies that “Hardy–Weinberg” equations such as (2.126) are no longer true. This is of some interest since many theoretical analyses of continuous-time two-locus models have assumed the truth of equations like (2.126). However, Nagylaki (1976) has shown that when fitness differentials are small a state of “quasi-Hardy–Weinberg” soon emerges when genotypic frequencies can, to a very close approximation, be found from the constituent gametic frequencies.

2.11 Genetic Loads

A genetic load is said to arise if the population mean fitness is less than that of some optimal value which in some idealized sense it could take. The two forms of genetic load that have caused considerable controversy in the literature are the substitutional load and the segregational load. In both cases the load ℓ is defined by

$$\ell = (w_{\max} - \bar{w}) / \bar{w}, \quad (2.127)$$

where w_{\max} is the fitness of the most fit genotype and \bar{w} is the mean fitness. If we normalize fitnesses so that the mean fitness is 1, we replace (2.127) by

$$\ell \approx w_{\max} - 1. \quad (2.128)$$

Our aim in this section is to analyze the formal calculations for both forms of load. These formal calculations have remained implicit rather than explicit in the analyses of proponents of genetic loads as calculated by the formula (2.128). Before doing this we briefly review the historical context.

The load concept was introduced by Haldane (1957, 1961) in the substitutional case. As a result of his load calculations, Haldane placed a quite conservative limit on the rate at which favorable new alleles at different loci, arising perhaps by mutations or perhaps by an environmental change rendering a previously unfavorable allele favorable, could spread throughout a population. Specifically, he came to the conclusion that as a result of what became known as the substitutional load (his “cost of natural selection”), substitutional processes at different loci could not start more frequently than about 300 generations apart.

As we observe below, a load in effect refers to a variance in fitness, not to a mean fitness. The essence of the substitutional load argument is that if many selectively driven substitutional processes are occurring in some population at any given time, then there will exist a substantial variance in fitness of these individuals in the population of interest at that time. Individuals carrying the favored allele at all the loci substituting will then have a very high fitness, that is will be required to produce an extremely large number of offspring. This is in effect the substitutional load placed on the population.

The load concept was subsequently extended to define a segregational load, the motivation being the observation, in the 1960’s, that there exists considerable genetic variation in natural populations. The segregational load argument claimed that under a selective explanation for the variation, perhaps because of heterozygote advantage at many of the loci exhibiting genetic variation, the most fit individuals in the population would again have a very high fitness and thus would be required to produce an extremely large number of offspring. This led to comments such as that of Dobzhansky (1970, page 220), that “higher vertebrates and man do not possess enough ‘load space’ to maintain more than a few balanced polymorphisms,” leading to the view (page 224) that selection favoring heterozygotes “cannot explain the polymorphisms observed in man.” At about the same time, segregational load arguments and subsequently substitutional load arguments were used by Kimura (1968) to support his neutral theory of evolution. The aim of this section is to show that the (implicit) arguments of Haldane, Dobzhansky and Kimura are all unjustified.

The segregational and substitutional genetic load “problem” arises when segregation occurs or substitutions take place at many loci simultaneously. The implicit assumption made in load calculations by proponents of the load concept is that multilocus fitnesses are obtained by first constructing single locus fitnesses and then multiplying these over the loci segregating or substituting. We initially make this (surely unrealistic) assumption so as

to follow load calculations and arguments, but later discuss more realistic fitness models.

We start with a discussion of the segregational load. This load exists because of segregation at a number of loci arising from heterozygote selective advantage at each locus. For simplicity we assume two alleles segregating at each locus and with a fitness scheme where, at each locus, each homozygote has fitness $1 - \frac{1}{2}s$ and the heterozygote has fitness $1 + \frac{1}{2}s$. Thus with the multiplicative assumption, and with two loci segregating, the two-locus fitness scheme (2.91) would be

	B_1B_1	B_1B_2	B_2B_2
A_1A_1	$(1 - \frac{1}{2}s)^2$	$(1 - \frac{1}{2}s)(1 + \frac{1}{2}s)$	$(1 - \frac{1}{2}s)^2$
A_1A_2	$(1 - \frac{1}{2}s)(1 + \frac{1}{2}s)$	$(1 + \frac{1}{2}s)^2$	$(1 - \frac{1}{2}s)(1 + \frac{1}{2}s)$
A_2A_2	$(1 - \frac{1}{2}s)^2$	$(1 - \frac{1}{2}s)(1 + \frac{1}{2}s)$	$(1 - \frac{1}{2}s)^2$

With many loci segregating the multilocus fitness scheme is the natural generalization of the two-locus fitness scheme above. We emphasize again that this model is discussed here since this is the model implicitly assumed in load calculations.

The equilibrium properties of this model are not straightforward. We shall see later (see (6.33)) that when the recombination fraction R between A and B loci is sufficiently large, the stable equilibrium frequencies of A_1 , A_2 , B_1 and B_2 are all $1/2$, and the mean fitness is 1, as a straightforward multiplication of single-locus values would suggest. However, when R is sufficiently small the picture is more complicated and the population mean fitness exceeds 1 at the stable equilibrium point of the system. We defer consideration of this case until later and assume for the moment the “loose linkage” case.

More generally, for m sufficiently loosely linked loci and a multiplicative fitness model generalizing the two-locus scheme above, the equilibrium frequencies of all alleles at all loci are $1/2$. Any individual is a heterozygote at j of these m loci with probability

$$\binom{m}{j} \left(\frac{1}{2}\right)^m,$$

so that the equilibrium population mean fitness is

$$\sum_{j=0}^m \binom{m}{j} \left(\frac{1}{2}\right)^m \left(1 + \frac{1}{2}s\right)^j \left(1 - \frac{1}{2}s\right)^{m-j} = 1. \quad (2.129)$$

An individual heterozygous at all loci has fitness $(1 + \frac{1}{2}s)^m$, and a formal application of the definition (2.128) implies that the segregational load is

$$\left(1 + \frac{1}{2}s\right)^m - 1 \approx e^{sm/2} - 1. \quad (2.130)$$

This can be substantial for large values of m , and this the formal calculation directly leads to the segregational load “problem”.

We return to this calculation below, and turn next to the substitutional load. We consider first the substitution process at one single gene locus, and initially, to follow formal substitutional load calculations, we do *not* scale fitnesses to make the mean population fitness equal to 1.

Suppose that at the locus of interest, fitnesses of the form (1.25b) apply, with $s > 0$. It is convenient, and does not materially affect the substance of the argument, to assume that $h = 0.5$. Then because of natural selection, the frequency of the allele A_1 will steadily increase in the population. When the frequency of A_1 is x the population mean fitness is $1 + sx$, and the load as defined by (2.128), is $s(1 - x)$. The overall substitutional load L for the entire substitution process is defined as the sum of this quantity during the process when x increases from a small value x_1 (at time t_1) to a value x_2 close to unity (at time t_2). Thus

$$\begin{aligned} L &= \sum s(1 - x) \\ &\approx \int_{t_1}^{t_2} s(1 - x) dt \\ &= 2 \int_{x_1}^{x_2} x^{-1} dx \quad \text{from (1.27)} \\ &= 2 \log(x_2/x_1). \end{aligned}$$

Since x_2 is close to 1, this differs only trivially from $-2 \log x_1$. Unfortunately the value chosen for x_1 will depend to a large extent on the view one takes of the most likely form of genetic evolution, and the discussion in Section 1.7 becomes relevant to the argument. A value often chosen for evolutionary load arguments is $x_1 = 0.0001$, and this gives $L = 18.4$. When $h \neq 0.5$ the load as calculated using this form of calculation usually exceeds 18.4, and for operational purposes the “representative value” $L = 30$ is generally used in the load argument. We therefore adopt this value also.

What does this calculation mean for the offspring requirement of the individuals in any given generation? Suppose that all selection is through viability differences and the number of reproducing adults in each generation remains constant at N . A considerable proportion of the depletion in population numbers between birth and the age of reproduction is non-genetic. Taking only the genetic component, and supposing there is no depletion through genetic deaths of the optimal genotype A_1A_1 , a straightforward calculation shows that when the frequency of A_1 is x , there must be $N(1+s)/(1+sx)$ individuals at birth, so that after differential viabilities operate there are N individuals at the age of maturity. Thus the average individual is required to leave approximately $1 + s(1 - x)$ offspring after

non-genetic deaths are taken into account, so that there will be $Ns(1-x)$ “genetic deaths” in each generation associated with the evolutionary process. Summed over the entire process this gives NL individuals in all. If each substitutional process takes T generations, this implies an average of NL/T such “deaths” in each generation.

Consider now a sequence of loci at which substitutions start regularly n generations apart. For convenience it is assumed that the same fitness parameters apply for all these loci as for the single locus discussed above. As in the segregational load argument, it is implicitly assumed in load arguments that fitnesses are multiplicative over loci, so initially we make this assumption also. As with the segregational load, the substitutional load relates to the fitness, or offspring requirement, of an individual of the most genotype. In this case this is an individual with the superior genotype “ A_1A_1 ” at each locus undergoing substitution.

At any one time there will be T/n substitutions in progress and thus a total of $(NL/T)(T/n) = NL/n$ “selective deaths” per generation. From this it is found that the offspring requirement of the most fit individual, assuming the multiplicative model of fitness with and with linkage equilibrium always holding between loci, is

$$(1 + L/T)^{T/n} \approx \exp(L/n) \approx \exp(30/n) \quad (2.131)$$

if we take the “representative value” 30 for L as discussed above.

The value $n = 300$ reached by Haldane (1957), as described above, arises from the fact that with this value of n , the expression in (2.131) is about 1.1, conforming to his view that an “excess reproductive requirement” of 10% is the maximum that can be expected, at least in mammals.

Kimura and Ohta (1971a) estimated that in the evolutionary history of mammals approximately six substitutions have been completed per generation in any evolutionary line. This implies that $n = 1/6$, 1800 times smaller than the Haldane “limiting” value, or equivalently implying substitutions occurring at 1800 times the upper rate as calculated by Haldane. Insertion of the value $n = 1/6$ in (2.131) leads to a substitutional load of $e^{180} \approx 10^{78}$. This form of calculation was a major factor in the development of the neutral theory, since it was argued (Kimura (1968)) that the amount of genetic substitution estimated to have taken place in evolution, in particular in mammalian evolution, could not be explained by selective processes because of a claimed unbearable substitutional genetic load that selective substitutions would imply. Thus (Kimura and Ohta, 1971a) claimed that

“to carry out mutant substitution at the above rate, each parent must leave $e^{180} \approx 10^{78}$ offspring for only one of the offspring to survive. This was the main reason why random fixation of selectively neutral mutants was first proposed by one of us as the main factor in molecular evolution.”

Because of calculations and claims of this type, it is clearly necessary to discuss the assumptions, both explicit and implicit, in formal load calculations.

We start with the expression in (2.131), and observe that this expression refers not to the offspring requirement of every individual, as is implied in the above quotation, but to the requirement of an individual of the maximum possible fitness when the population mean fitness is now scaled to 1. It is therefore appropriate to focus on this individual and on his fitness.

Our calculations show that the fitness e^{180} is arrived at by assuming that fitnesses are multiplicative over loci. This is a quite unreasonable assumption, and the large offspring requirement of the most fit individual is a direct consequence of it. It is certainly true that in nature substantial epistasis occurs, and if this is so there will be a considerable reduction to the load from that calculated formally by using marginal fitnesses and multiplicativity, as discussed below. The unreasonableness of the multiplicative assumption was stressed long ago, in particular by Wright (1930).

The second, and more important, problem concerns the very existence of an individual of the optimal multilocus genotype. It is extremely unlikely that such an individual ever exists. To simplify the argument we continue to consider the multiplicative case discussed above. It can be shown that with the individual locus fitness values $1+s, 1+s/2, 1$ for " A_1A_1, A_1A_2 and A_2A_2 ", as is assumed above, and with $s = 0.01$, $n = 1/6$, initial frequency = 0.0001, final frequency = 0.9999, there will be 22,080 loci substituting at any one time. The various favored alleles at each of these 22,080 loci will take a variety of frequencies in $(0, 1)$, and in particular at those loci where the substitution has only recently started, the frequency of the favored allele will be quite low. By calculating the means of the frequencies x_1, x_2, \dots of the favored allele at the various loci substituting, using (1.28), it is found that the probability that an individual taken at random is of this optimal genotype is on the order of $10^{-23,200}$. This value is so extremely small that a theory basing its numerical computations on the offspring requirement of such an individual must demand reconsideration. This point also was stressed by Wright (1977, p. 481).

What is needed is a calculation of the fitness of the individuals who might reasonably be expected to occur in the population of interest. Here the finite size of any population is an important factor in the calculations. Some progress on amending load calculations for this purpose may be made by using the statistics of extreme values in a population of given finite size (Kimura (1969), Ewens (1970)). It is convenient, for purposes of illustration only, to maintain the multiplicativity assumption here so as to discuss the point at issue. The starting point is to find the variance of the distribution of the fitness of an individual taken at random from the population, if the population mean fitness is scaled to unity. In the case considered above this variance is s/n (Ewens (1970), Crow and Kimura (1970, p. 252)). For $s = 0.01$, $n = 1/6$, this is a variance of 0.06, so that the standard deviation

in fitness is approximately 0.245. The rather low value for this standard derivation arises because it is most unlikely that any individual will have a genetic constitution which differs markedly, in terms of the number of favored genes carried, from the average.

If s is extremely small we may suppose, to a first approximation, that the distribution of fitness is a normal distribution. The statistical theory of extreme values (see Pearson and Hartley (1958, Table 28)) shows that, for example in a population of size 10^5 , the most fit individual that is likely to occur will have a fitness approximately four standard deviations in excess of the mean. In the present case this implies a fitness of $1 + 4(0.245) = 1.98$. On average, then, the most fit individual that is likely to exist in the population is required to produce only about two offspring in order to effect the gene substitutions observed. This is clearly an easily achievable goal.

A parallel argument holds for the segregational load as calculated in (2.130). The segregational load is clearly the excess over the mean of the offspring of the most fit individual, in the segregation load case the multiple heterozygote. The probability that an individual chosen at random in the population is of this genotype is $(1/2)^m$, and when m is large it is extremely unlikely that any individual in a population even of size several million has this genotype. As with the substitutional load, it is more reasonable to consider the fitness of the most fit individual likely to arise in the population. This is done as follows.

The mean fitness of the population is calculated in (2.129). The variance in fitness then found as

$$\sum_{j=0}^m \binom{m}{j} \left(\frac{1}{2}\right)^m \left(1 + \frac{1}{2}s\right)^{2j} \left(1 - \frac{1}{2}s\right)^{2(m-j)} - 1. \quad (2.132)$$

This expression reduces to

$$\left(1 + \frac{1}{4}s^2\right)^m - 1 \approx e^{ms^2/4} - 1. \quad (2.133)$$

For the case $m = 10,000$, $s = 0.01$ this is about 0.28. A fitness four standard deviations above the mean is only just in excess of 3, and arguing as above for the substitutional load, this clearly is an achievable fitness for the most fit individual likely to arise in a population of size 10^5 .

The essence of the argument, in both the substitutional load and the segregational load cases, is that in a finite population only a minute proportion of all theoretically possible genotypes are realized, and that those that are realized are not normally very "extreme". In particular the fitness of the most fit existing genotype is not extreme, and in the substitution case, substitutions at the required rate can easily be achieved through each individual's producing as many offspring as this most fit existing genotype, with consequent differential viability effecting the required substitutions.

There are many further arguments that make the substitutional load calculations leading to the value e^{180} of dubious value. First, it has been

assumed in all the calculations that selection arises entirely through viability differences. To the extent that fertility selection occurs, the offspring requirement is correspondingly lowered, in the sense that the calculation of the offspring requirement of the most fit individual is not a calculation of any relevance to the average individual.

Second, it has been assumed so far that fitnesses are fixed constants, and are not, for example, frequency-dependent. It is possible to devise frequency-dependent selection schemes for which there is no segregational load at a stable equilibrium. Thus in the fitness scheme

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ 1 + a(1 - 2x) & 1 & 1 - a(1 - 2x) \end{array} \quad (2.134)$$

where x is the frequency of A_1 and a is a small parameter, the point $x = 0.5$ is a point of stable equilibrium, and at this point all genotypes have equal fitness and there is no genetic load. On the other hand, it is unlikely that frequency-dependent fitnesses can reduce the substitutional load to zero, since with a change in gene frequencies due to selection, some selective differentials are necessary and hence some load. Little information is available on the extent to which frequency-dependent selection can reduce substitutional load.

We now consider the effects of linkage disequilibrium, and later of epistasis and linkage disequilibrium jointly, on load calculations. Stationary points of an evolutionary system exhibiting linkage disequilibrium generally have a higher mean fitness than points where linkage equilibrium holds at stationarity, and thus have a lower genetic load than that at linkage equilibrium equilibria. This is particularly so when the selective system implies epistasis. However, even in the simple multiplicative case, where we can say there is no multiplicative epistasis, the stable equilibrium points of the evolutionary system can display linkage disequilibrium and thus a decreased segregational load. For example, the calculations of Franklin and Lewontin (1970) show that in the case of 36 equally spaced linked loci, a multiplicative fitness scheme generalizing the two-locus multiplicative fitness scheme above with $s = 0.1$, and with recombination fraction 0.0025 between adjacent loci, the load when calculated from (2.130) is about 5, but when calculated using the actual population mean fitness is about 1.6. The smaller load arises from the linkage disequilibrium arising for this model. This point has also in effect been made by Lewontin (1974, pp. 289–290) in the context in discussing the effect of linkage disequilibrium on mean fitness.

Next, the joint effects of epistasis and linkage disequilibrium can decrease the segregational load substantially. Thus, for example, numerical computation shows that with the epistatic scheme (2.96) and with $R = 0.001$, there is a stable equilibrium set of gametic frequencies at

$$c_1 = 0.013, \quad c_2 = 0.469, \quad c_3 = 0.503, \quad c_4 = 0.015. \quad (2.135)$$

At this point the population mean fitness is 1.0417 and thus the genetic load as defined by (2.128) is 0.0233.

Suppose now that marginal fitness values for this case are found from (2.111), and the load calculated according to (2.127) using these marginal values and the marginal genotypic frequencies. The loads so calculated are 0.0212 for the A locus and 0.0210 for the B locus. The sum of these is almost twice that of the true load: For $R = 0$ it would be exactly twice. Evidently for general fitness schemes involving tight linkage and epistasis, the procedure leading to the load calculation of e^{180} , namely the calculation of a multilocus segregational load through an amalgamation of single-locus segregational load calculations, can lead to serious errors.

If we take into account, then, the unreasonable multiplicative fitness requirement implicit in load calculations, the unreasonable concentration on the fitness requirement of essentially impossible genotypes, the possibility of very substantial linkage disequilibria, the possibility of frequency-dependent fitnesses and a variety of other ecological and evolutionary arguments concerning the real nature of selective processes, it appears that there is no reason for load arguments to imply very conservative bounds on the number of loci that can undergo simultaneous selective substitution processes, no "load space" argument limiting the number of balanced polymorphisms arising at any one time in a population, and no load theory support for the neutral theory of evolution.

2.12 Finite Markov Chains

Some of the arguments presented later in this book use the theory of finite Markov chains, and in this section a brief and informal introduction to the theory of these is presented.

Consider a discrete random variable X which at time points $0, 1, 2, 3, \dots$ takes one or other of the values $0, 1, 2, \dots, M$. We shall say that X , or the system, is in state E_i if X takes the value i . Suppose that at some time t , the random variable X is in state E_i . Then if the probability p_{ij} that at time $t + 1$, the random variable is in state E_j is independent of t and also of the states occupied by X at times $t - 1, t - 2, \dots$, the variable X is said to be Markovian, and its probability laws follow those of a finite Markov chain. If the initial probability (at $t = 0$) that X is in E_i is a_i then the probability that X is in the state $E_i, E_j, E_k, E_l, E_m \dots$ at times $0, 1, 2, 3, 4 \dots$ is $a_i p_{ij} p_{jk} p_{kl} p_{lm} \dots$.

Complications to Markov chain theory arise if periodicities occur, for example, if X can return to E_i only at the time points $t_1, 2t_1, 3t_1, \dots$ for some integer $t_1 > 1$. Further minor complications arise if the states E_0, E_1, \dots, E_M can be broken down into noncommunicating subsets. To avoid unnecessary complications, which never in any event arise in genetical

applications, we suppose that no periodicities exist and that, apart from the possibility of a small number of absorbing states, (E_i is absorbing if $p_{ii} = 1$), no breakdown into noncommunicating subsets occur.

It is convenient to collect the p_{ij} into a matrix $P = \{p_{ij}\}$, so that

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0M} \\ p_{10} & p_{11} & \cdots & p_{1M} \\ \vdots & & & \\ p_{M0} & p_{M1} & \cdots & p_{MM} \end{pmatrix}. \quad (2.136)$$

The probability $p_{ij}^{(2)}$ that X is in E_j at time $t + 2$, given it is in E_i at time t , is evidently

$$p_{ij}^{(2)} = \sum_k p_{ik} p_{kj}.$$

Since the right-hand side is the (i, j) th element in the matrix P^2 , and if we write $P^{(2)} = \{p_{ij}^{(2)}\}$, then

$$P^{(t)} = P^t \quad (2.137)$$

for $t = 2$. More generally (2.137) is true for any positive integer t . In all cases we consider, P^t can be written in the spectral form

$$P^t = \lambda_0^t r_0 \ell'_0 + \lambda_1^t r_1 \ell'_1 + \cdots + \lambda_M^t r_M \ell'_M \quad (2.138)$$

where $\lambda_0, \lambda_1, \dots, \lambda_M$ ($|\lambda_0| \geq |\lambda_1| \geq \cdots \geq |\lambda_M|$) are the eigenvalues of P and (ℓ_0, \dots, ℓ_M) and (r_0, \dots, r_M) , normalized so that

$$\ell'_i \mathbf{r}_i = \sum_{j=0}^M \ell_{ij} r_{ij} = 1, \quad (2.139)$$

are the corresponding left and right eigenvectors, respectively. Suppose E_0 and E_M are absorbing states and that no other states are absorbing. Then $\lambda_0 = \lambda_1 = 1$ and if $|\lambda_2| > |\lambda_3|$ and $i, j = 1, 2, \dots, M - 1$,

$$p_{ij}^{(t)} = r_{2i} \ell_{2j} \lambda_2^t + o(\lambda_2^t) \quad (2.140)$$

for large t . Thus the leading nonunit eigenvalue λ_2 plays an important role in determining the rate at which absorption into either E_0 and E_M occurs.

Let π_j be the probability that eventually E_M (rather than E_0) is entered, given initially that X is in E_i . By considering values of X at consecutive time points it is seen that the π_i satisfy

$$\pi_i = \sum_{j=0}^M p_{ij} \pi_j, \quad \pi_0 = 0, \quad \pi_M = 1. \quad (2.141)$$

For the genetic model (1.48) (with $M = 2N$) the solution of (2.141) was $\pi_i = i/M$. The mean times \bar{t}_i until absorption into E_0 or E_M occurs, given

that X is in E_i , similarly satisfy

$$\bar{t}_i = \sum_{j=0}^M p_{ij} \bar{t}_j + 1, \quad \bar{t}_0 = \bar{t}_M = 0. \quad (2.142)$$

Starting with X in E_i the members of the set of mean times $\{\bar{t}_{ij}\}$ that X is in E_j before absorption into either E_0 or E_M satisfy the equations

$$\bar{t}_{ij} = \sum_{k=0}^M p_{ik} \bar{t}_{kj} + \delta_{ij}, \quad \bar{t}_{0j} = \bar{t}_{Mj} = 0, \quad (2.143)$$

where $\delta_{ij} = 1$ and $i = j$ and $\delta_{ij} = 0$ otherwise. Further,

$$\bar{t}_{ij} = \sum_{n=0}^{\infty} p_{ij}^{(n)}, \quad \bar{t}_i = \sum_{j=1}^{M-1} \bar{t}_{ij}. \quad (2.144)$$

An expression can also be found for the variance σ_i^2 of the time before absorption, given initially X in E_i , namely

$$\sigma_i^2 = 2 \sum_{j=1}^{M-1} \bar{t}_{ij} \bar{t}_j - \bar{t}_i - (\bar{t}_i)^2. \quad (2.145)$$

It is possible to derive the general form of the distribution of the time that X is in E_j if initially in E_i . Suppose that, starting in E_i , the probability that X ever enters E_j is α_{ij} and that once in E_j , the probability that X ever returns to E_j is r_j . Then the probability that E_j is occupied exactly n times before absorption takes place at E_0 or E_m is

$$\left. \begin{array}{ll} 1 - \alpha_{ij} & \text{for } n = 0 \\ \alpha_{ij}(r_j)^{n-1}(1 - r_j) & \text{for } n \geq 1. \end{array} \right\} \quad (2.146)$$

This is clearly a modified geometric distribution. The mean is thus

$$\begin{aligned} \bar{t}_{ij} &= \alpha_{ij}(1 - r_j) \sum_{n=1}^{\infty} nr_j^{n-1} \\ &= \alpha_{ij}/(1 - r_j) \end{aligned} \quad (2.147)$$

and the variance is

$$\begin{aligned} \sigma_{ij}^2 &= \alpha_{ij}(1 - r_j) \sum_{n=1}^{\infty} n^2(r_j)^{n-1} - \bar{t}_{ij}^2 \\ &= \bar{t}_{ij} \{1 - \bar{t}_{ij} + 2r_j/(1 - r_j)\}. \end{aligned} \quad (2.148)$$

It is possible to find an expression for r_j and hence to calculate (2.148) but we do not enter into details here.

Consider now only those cases for which E_M is the absorbing state eventually entered. Writing X_t for the value of X at time t , we get

$$\begin{aligned} p_{ij}^* &= \text{Prob}\{X_{t+1} \text{ in } E_j \mid X_t \text{ in } E_i, E_M \text{ eventually entered}\} \\ &= \text{Prob}\{X_{t+1} \text{ in } E_j \text{ and } E_M \text{ eventually entered} \mid X_t \text{ in } E_i\} \\ &\quad \div \text{Prob}\{E_M \text{ eventually entered} \mid X_t \text{ in } E_i\} \\ &= p_{ij}\pi_j/\pi_i, \quad (i, j = 1, 2, \dots, M), \end{aligned} \quad (2.149)$$

using conditional probability arguments and the Markovian nature of X . Let \tilde{P} be the matrix derived from P by omitting the first row and first column and let

$$V = \begin{pmatrix} \pi_1 & & & \\ & \pi_2 & & \\ & & \ddots & 0 \\ 0 & & & \pi_M \end{pmatrix}. \quad (2.150)$$

Then if $P^* = \{p_{ij}^*\}$, (2.149) shows that

$$P^* = V^{-1}\tilde{P}V. \quad (2.151)$$

Standard theory shows that the eigenvalues of P^* are identical to those of P (with one unit eigenvalue omitted) and that if $\ell'(\mathbf{r})$ is any left (right) eigenvector of \tilde{P} , then the corresponding left and right eigenvector of P^* are $\ell'V$ and $V^{-1}\mathbf{r}$. Further, if $P^{*(n)}$ is the matrix of conditional n step transition probabilities,

$$P^{*(n)} = (P^*)^n = V^{-1}\tilde{P}^nV$$

so that

$$p_{ij}^{*(n)} = p_{ij}^{(n)}\pi_j/\pi_i, \quad (2.152)$$

a conclusion that can be reached directly as with (2.149). If \bar{t}_{ij}^* is the conditional mean time spent in E_j , given initially X in E_i , then

$$\begin{aligned} \bar{t}_{ij}^* &= \sum_{n=0}^{\infty} p_{ij}^{*(n)} \\ &= (\pi_j/\pi_i) \sum_{n=0}^{\infty} p_{ij}^{(n)} \\ &= \bar{t}_{ij}\pi_j/\pi_i. \end{aligned} \quad (2.153)$$

If there is only one absorbing state interest centers solely on properties of the time until the state is entered. Taking E_0 as the only absorbing state and E_i as the initial state, the mean time t_i until absorption satisfies (2.142) with the single boundary condition $\bar{t}_0 = 0$, and the mean number of visits to E_j satisfies (2.143) with the single condition $\bar{t}_{0j} = 0$.

If there are no absorbing states P will have a single eigenvalue and all other eigenvalues will be strictly less than unity in absolute value. Equation (2.138) then shows that

$$\lim_{t \rightarrow \infty} P^t = \mathbf{r}_0 \ell'_0 \tag{2.154}$$

and since r_0 is of the form $(1, 1, 1, \dots, 1)'$,

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \ell_{0j} \quad \text{for all } i. \tag{2.155}$$

Using a slightly different notation we may summarize this by saying

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \phi_j, \tag{2.156}$$

where $\phi' = (\phi_0, \phi_1, \dots, \phi_M)$ is the unique solution of the two equations

$$\phi' = \phi' P, \quad \sum_{j=0}^M \phi_j = 1. \tag{2.157}$$

The vector α is called the stationary distribution of the process and in genetical applications exists only if fixation of any allele is impossible (e.g. if all alleles mutate at positive rates).

If the matrix P is a continuant (so that $p_{ij} = 0$ if $|i - j| > 1$) explicit formulas can be found for most of these quantities. We write $p_{i,i+1} = \lambda_i$ and $p_{i,i-1} = \mu_i$ in conformity with standard notation in this case. If E_0 and E_M are both absorbing states the probability π_i in (2.141) becomes, explicitly,

$$\pi_i = \sum_{k=0}^{i-1} \rho_k / \sum_{k=0}^{M-1} \rho_k, \tag{2.158}$$

where

$$\rho_0 = 1, \quad \rho_k = \frac{\mu_1 \mu_2 \mu_3 \cdots \mu_k}{\lambda_1 \lambda_2 \cdots \lambda_k}.$$

Further

$$\begin{aligned} \bar{t}_{ij} &= \frac{(1 - \pi_i) \sum_{k=0}^{j-1} \rho_k}{\rho_{j-1} \mu_j}, \quad (j = 1, \dots, i), \\ \bar{t}_{ij} &= \frac{\pi_i \sum_{k=j}^{M-1} \rho_k}{\rho_j \lambda_j}, \quad (j = 1 + 1, \dots, M - 1). \end{aligned} \tag{2.159}$$

Equations (2.144) and (2.153) then yield $\bar{t}_i, \bar{t}_{ij}^*$ and \bar{t}_i^* immediately. When there is only one absorbing state (2.144) still holds, but now \bar{t}_{ij} is defined

by

$$\bar{t}_{ij} = \begin{cases} \mu_j^{-1} \left\{ 1 + \frac{\lambda_{j-1}}{\mu_{j-1}} + \frac{\lambda_{j-1}\lambda_{j-2}}{\mu_{j-1}\mu_{j-2}} + \cdots + \frac{\lambda_{j-1}\lambda_{j-2}\cdots\lambda_1}{\mu_{j-1}\mu_{j-2}\cdots\mu_1} \right\} \\ \quad (j = 1, 2, \dots, i) \\ \bar{t}_{ii} \left(\frac{\lambda_i\lambda_{i+1}\cdots\lambda_{j-1}}{\mu_{i+1}\mu_{i+2}\cdots\mu_j} \right) \quad (j = i + 1, \dots, M) \end{cases} \quad (2.160)$$

if E_0 is the absorbing state and by

$$\bar{t}_{ij} = \begin{cases} \frac{\mu_{j+1}\mu_{j+2}\cdots\mu_i}{\lambda_j\lambda_{j+1}\cdots\lambda_{i-1}} \bar{t}_{ii} \quad (j = 0, 1, \dots, i - 1) \\ \lambda_i^{-1} \left\{ 1 + \frac{\mu_{i+1}}{\lambda_{i+1}} + \frac{\mu_{i+1}\mu_{i+1}}{\lambda_{i+1}\lambda_{i+2}} + \cdots + \frac{\mu_{i+1}\mu_{i+2}\cdots\mu_{M-2}}{\lambda_{i+1}\lambda_{i+2}\cdots\lambda_{M-1}} \right\} \\ \quad (j = i, i + 1, \dots, M - 1) \end{cases} \quad (2.161)$$

if E_M is the absorbing state. In this case of course there can be no further concept of a conditional mean absorption time.

Finally, when there are no absorbing states, the stationary distribution ϕ is defined by

$$\phi_i = \phi_0 \frac{\lambda_0\lambda_1 \cdots \lambda_{i-1}}{\mu_1\mu_2 \cdots \mu_i}, \quad (2.162)$$

where ϕ_0 is chosen so that $\sum \alpha_i = 1$.

Various further results are possible for continuant Markov chain models, an accessible summary being given in Kemeny and Snell (1960). We shall draw on the formulas given above on a number of occasions throughout this book.

We conclude our discussion of finite Markov chains by introducing the concept of time reversibility. Consider a Markov chain admitting a stationary distribution $\{\phi_0, \phi_1, \dots, \phi_M\}$. Then we define the process to be reversible if, at stationarity,

$$\text{Prob}\{X_t, X_{t+1}, \dots, X_{t+n}\} = \text{Prob}\{X_t, X_{t-1}, \dots, X_{t-n}\} \quad (2.163)$$

for every t and n . A necessary and sufficient condition for this is that the stationary state has been reached and that the equation

$$\phi_i p_{ij} = \phi_j p_{ji} \quad (2.164)$$

hold for all i, j . Certain classes of Markov chains are always reversible. For example, if the transition matrix is a continuant, (2.162) and (2.163) jointly show that the Markov chain at stationarity is reversible. Certain other chains, in particular several having genetical relevance, are reversible: we shall consider these later when discussing the uses to which the concept of reversibility can be put.

3

Discrete Stochastic Models

3.1 Introduction

In the last section of the previous chapter some elementary finite Markov chain theory was introduced. In this chapter we apply this theory to various Markov chain models which arise in genetics. We shall find that the complexities of these models are such that not all questions of genetical interest can in practice be answered by using Markov chain theory, and in the next two chapters we shall introduce diffusion theory to arrive at a more complete, although approximate, description of the properties of Markov chain models of interest in genetics.

3.2 Wright–Fisher Model: Two Alleles

In Chapter 1 we were led to the Wright–Fisher model (1.48) as a simple approximate representation of the stochastic behavior of gene frequencies in an idealized finite population. Our first aim is to discuss some of the properties of this model in the light of the theory of Section 2.12. We have already noted that in the model (1.48), the number X of A_1 genes is a Markovian random variable with two absorbing states, $X = 0$ and $X = 2N$. Further, the probability that eventually $X = 2N$, given that initially $X = i$, is simply $i/2N$.

We now ask whether the theory of Section 2.12 gives us further information on the behavior of X before an absorbing state is reached. The most

interesting quantities are the mean time \bar{t}_i until absorption, given initially $X = i$, and the mean number of times \bar{t}_{ij} that X takes the value j before absorption. While in principle these expressions can be found from (2.142) and (2.143), in practice solution of these equations seems extremely difficult for this model, and simple expressions for these mean times have not yet been found. It is indeed likely that no simple expressions exist for them. On the other hand, it is possible to find a simple approximation for \bar{t}_i by the following line of argument.

In (2.142) we put $M = 2N$, $i/M = x$, $j/M = x + \delta x$, and $\bar{t}_i = \bar{t}(x)$. We suppose $\bar{t}(x)$ is a twice differentiable function of a continuous variable x . Then (2.142) can be written

$$\bar{t}(x) = \sum \text{Prob}\{x \rightarrow x + \delta x\} \bar{t}(x + \delta x) + 1 \quad (3.1)$$

$$= \mathbf{E}\{\bar{t}(x + \delta x)\} + 1 \quad (3.2)$$

$$\approx \bar{t}(x) + \mathbf{E}(\delta x)\{\bar{t}(x)\}' + \frac{1}{2}\mathbf{E}(\delta x)^2\{\bar{t}(x)\}'' + 1, \quad (3.3)$$

where all expectations are conditional on x and in (3.3) only the first three terms in an infinite Taylor series have been retained. Since from (1.48)

$$\mathbf{E}(\delta x) = 0, \quad \mathbf{E}(\delta x)^2 = (2N)^{-1}x(1-x),$$

(3.3) gives

$$x(1-x)\{\bar{t}(x)\}'' \approx -4N. \quad (3.4)$$

The solution of this equation, subject to the boundary conditions $\bar{t}(0) = \bar{t}(1) = 0$, is

$$\bar{t}(p) \approx -4N\{p \log p + (1-p) \log(1-p)\}, \quad (3.5)$$

where $p = i/2N$ is the initial frequency of A_1 . We shall see later that this is the so-called diffusion approximation to the mean absorption time, although we have here not made any reference to diffusion processes.

In the case $i = 1$, so that $p = (2N)^{-1}$, the value appropriate if A_1 is a unique new mutation in an otherwise purely A_2A_2 population, (3.5) reduces to

$$\bar{t}\{(2N)^{-1}\} \approx 2 + 2 \log 2N \text{ generations}, \quad (3.6)$$

while when $p = \frac{1}{2}$,

$$\bar{t}\{\frac{1}{2}\} \approx 2.8N \text{ generations}. \quad (3.7)$$

This very long mean time, for equal initial frequencies, is of course intimately connected with the fact that the leading nonunit eigenvalue of the transition matrix in (1.48) is very close to unity.

Suppose now the condition is made that A_1 eventually fixes. The possible values for X are $1, 2, 3, \dots, 2N$ and (2.149) shows that the conditional

transition probability p_{ij}^* is

$$\begin{aligned} p_{ij}^* &= \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(\frac{2N-i}{2N}\right)^{2N-j} \frac{j}{i} \\ &= \binom{2N-1}{j-1} \left(\frac{i}{2N}\right)^{j-1} \left(\frac{2N-i}{2N}\right)^{2N-j}. \end{aligned} \quad (3.8)$$

An intuitive explanation for the form of p_{ij}^* is that under the condition that A_1 fixes, at least one A_1 gene must be produced in each generation. Then p_{ij}^* is the probability that the remaining $2N-1$ gene transmissions produce exactly $j-1$ A_1 genes. An argument parallel to that leading to (3.4) gives

$$(1-x)\{\bar{t}^*(x)\}' + \frac{1}{2}x(1-x)\{\bar{t}^*(x)\}'' = -2N \quad (3.9)$$

for the conditional mean time $\bar{t}^*(x)$ to fixation, given a current frequency of x . The solution of (3.9), subject to $\bar{t}^*(1) = 0$ and the requirement

$$\lim_{x \rightarrow 0} \bar{t}^*(x) \text{ is finite,} \quad (3.10)$$

and assuming initially $x = p$, is

$$\bar{t}^*(p) = -4Np^{-1}(1-p) \log(1-p). \quad (3.11)$$

We observe from this that

$$\bar{t}^*\{(2N)^{-1}\} \approx 4N - 2 \text{ generations,} \quad (3.12)$$

$$\bar{t}^*\left\{\frac{1}{2}\right\} \approx 2.8N \text{ generations,} \quad (3.13)$$

$$\bar{t}^*\{1 - (2N)^{-1}\} \approx 2 \log 2N \text{ generations.} \quad (3.14)$$

The approximation (3.13) is to be expected from (3.7), since by symmetry, when the initial frequency of A_1 is $\frac{1}{2}$, the conditioning should have no effect on the mean fixation time. On the other hand, (3.12) and (3.14) provide new information, and show that while when the initial frequency of A_1 is $(2N)^{-1}$ it is very unlikely that fixation of A_1 will occur, in the small fraction of cases when fixation of A_1 does occur, an extremely long fixation time may be expected. Further conclusions will be given later when we consider the diffusion approximation to the Wright–Fisher model (1.48).

As noted in Chapter 1, the initial analysis of the model (1.48) by Fisher and Wright paid particular attention to the leading eigenvalue of the transition matrix, regarded as a measure of the rate at which one or other allele is lost from the population. Although, as we see below, the eigenvalues are of less use than expressions like (3.5) and (3.11) for this purpose, they are nevertheless of some interest, so we now write down the formulas for these eigenvalues.

Since the matrix defined by the p_{ij} in (1.48) is the transition matrix of a Markov chain, it follows that one eigenvalue of the matrix is automatically

1. Denoting this eigenvalue by λ_0 , the remaining eigenvalues, first derived by Feller (1951), are

$$\lambda_j = (2N)(2N-1)\cdots(2N-j+1)/(2N)^j, \quad j = 1, 2, \dots, 2N. \quad (3.15)$$

This confirms the values $\lambda_1 = 1$ and $\lambda_2 = 1 - (2N)^{-1}$ found earlier by other methods. We derive the eigenvalues in (3.15) in Section 3.3 as particular cases of an important model of Cannings (1974) which generalizes the Wright–Fisher model.

Although considerable attention has been paid to the leading nonunit eigenvalue λ_2 and, to a lesser extent, to the complete set (3.15), it is possible to argue that these eigenvalues are of limited usefulness. First, (2.151) shows that the eigenvalues in the conditional process, where eventual fixation of a specified allele is assumed, are the same as those in the unconditional process. On the other hand, the mean fixation time values are quite different in the two cases, as (3.5) and (3.11) show, and thus are not adequately described by knowledge of the eigenvalues alone. Second, we shall show later that at least in the model (1.48), by the time that the term defined by the leading nonunit eigenvalue in the spectral expansion (2.138) dominates the remaining terms, it is very likely that loss or fixation of A_1 will already have occurred.

Suppose now that A_1 mutates to A_2 at rate u but that there is no mutation from A_2 to A_1 . It is then reasonable to replace the model (1.48) by

$$p_{ij} = \binom{2N}{j} (\psi_i)^j (1 - \psi_i)^{2N-j} \quad (3.16)$$

where $\psi_i = i(1-u)/2N$. Here eventual loss of A_1 is certain, and interest centers on properties of the time until A_1 is lost, either using eigenvalues or mean time properties. For the moment we consider mean time properties and note that an argument parallel to that leading to (3.4) shows that to a first approximation, the mean time $\bar{t}(x)$, given a current frequency x , satisfies

$$-4Nux\{\bar{t}(x)\}' + x(1-x)\{\bar{t}(x)\}'' = -4N. \quad (3.17)$$

If initially $x = p$, the solution of this equation, subject to the requirements $\bar{t}(0) = 0$,

$$\lim_{x \rightarrow 1} \bar{t}(x) \text{ is finite,}$$

is

$$\bar{t}(p) = \int_0^1 t(x, p) dx \text{ generations,} \quad (3.18)$$

where for $\theta \neq 1$,

$$\left. \begin{aligned} t(x, p) &= 4Nx^{-1}(1-\theta)^{-1}\{(1-x)^{\theta-1} - 1\}, & 0 \leq x \leq p, \\ t(x, p) &= 4Nx^{-1}(1-\theta)^{-1}(1-x)^{\theta-1}\{1 - (1-p)^{1-\theta}\}, & p \leq x \leq 1, \end{aligned} \right\} \quad (3.19)$$

and $\theta = 4Nu$. The corresponding formulas for the case $\theta = 1$ are found from (3.19) by standard limiting processes.

It may be shown (Griffiths, 2003) that with the definition of $t(x, p)$ in (3.19), $\bar{t}(p)$ may be written as

$$\bar{t}(p) = \sum_{j=1}^{\infty} \frac{4N}{j(j-1+\theta)} \left(1 - (1-p)^j\right). \quad (3.20)$$

The function $t(x, p)$ in (3.19) is more informative than it initially appears since, as we see later, $t(x, p)\delta x$ provides an excellent approximation to the mean number of generations for which the frequency of A_1 takes a value in $(x, x + \delta x)$ before reaching zero.

There are two interesting special cases of (3.20). First, when $\theta = 2$,

$$\left. \begin{aligned} t(x, p) &= 4N, & 0 \leq x \leq p, \\ t(x, p) &= 4Nx^{-1}(1-x)\{(1-p)^{-1} - 1\}, & p \leq x \leq 1, \end{aligned} \right\} \quad (3.21)$$

and from this,

$$\bar{t}(p) = \frac{-4Np \log p}{1-p}, \quad (3.22)$$

a conclusion that can also be found directly from (3.20). Second, when $p = 1$, (3.20) gives immediately

$$\bar{t}(1) = \sum_{j=1}^{\infty} \frac{4N}{j(j-1+\theta)}. \quad (3.23)$$

We shall return to these two cases later, when discussing the expressions in (9.102) and (9.95).

Suppose next that A_2 also mutates to A_1 at rate v . It is now reasonable to define ψ_i in (3.16) by

$$\psi_i = \{i(1-u) + (2N-i)v\}/2N. \quad (3.24)$$

There now exists a stationary distribution $\phi' = (\phi_0, \phi_1, \dots, \phi_{2N})$ for the number of A_1 genes, given in principle by (2.157). The exact form of this distribution is complex, and we consider later an approximation to it. On the other hand, certain properties of this distribution can be extracted from (3.16) and (3.24). The stationary distribution satisfies the equation $\phi' = \phi'P$, where P is defined by (3.16) and (3.24), so that if ξ is a vector with i th element i ($i = 0, 1, 2, \dots, 2N$) and μ is the mean of the stationary distribution,

$$\mu = \phi' \xi = \phi' P \xi.$$

The i th ($i = 0, 1, 2, \dots, 2N$) component of $P\xi$ is

$$\sum_j j \binom{2N}{j} \psi_i^j (1 - \psi_i)^{2N-j}$$

and from the standard formula for the mean of the binomial distribution, this is $2N\psi_i$ or

$$i(1 - u) + (2N - i)v.$$

Thus,

$$\begin{aligned} \phi' P\xi &= \sum \{i(1 - u) + (2N - i)v\} \alpha_i \\ &= \mu(1 - u) + v(2N - \mu). \end{aligned}$$

It follows that

$$\mu = (1 - u)\mu + v(2N - \mu)$$

or

$$\mu = 2Nv/(u + v). \quad (3.25)$$

In view of the deterministic stationary frequency (1.33), this value is not surprising. Similar arguments show that the variance σ^2 of the stationary distribution is

$$\sigma^2 = 4N^2 uv / \{(u + v)^2 (4Nu + 4Nv + 1)\} + \text{smaller order terms.} \quad (3.26)$$

Further moments can also be found, but we do not pursue the details.

The above values are sufficient to answer a question of some interest in population genetics, namely “what is the probability of two genes drawn together at random are of the same allelic type?” If the frequency of A_1 is x and terms of order N^{-1} are ignored, this probability is $x^2 + (1 - x)^2$. The required value is the expected value of this over the stationary distribution, namely

$$E\{x^2 + (1 - x)^2\} = 1 - 2E(x) + 2E(x^2).$$

If $u = v$, $4Nu = \theta$, (3.25) and (3.26) together show that this is

$$\text{Prob (two genes of same allelic type)} \approx (1 + \theta)/(1 + 2\theta). \quad (3.27)$$

This probability can be arrived at in another way, which we now consider since it is useful for purposes of generalization. Let the required probability be F and note that this is the same in two consecutive stationary generations. Two genes drawn at random in any generation will have a common parent gene with probability $(2N)^{-1}$, or different parent genes with probability $1 - (2N)^{-1}$, which will be of the same allelic type with probability F . The probability that neither of the genes drawn is a mutant, or that both are, is $u^2 + (1 - u)^2$, while the probability that precisely one is a mutant is

$2u(1-u)$. It follows that

$$F = \{u^2 + (1-u)^2\} \left\{ \frac{1}{2N} + F \left(1 - \frac{1}{2N}\right) \right\} \\ + 2u(1-u)(1-F) \left(1 - \frac{1}{2N}\right).$$

Thus exactly

$$F = \frac{1 + 2u(1-u)(2N-2)}{1 + 4u(1-u)(2N-1)},$$

and approximately

$$F = (1 + \theta)/(1 + 2\theta), \quad (3.28)$$

in agreement with (3.27). A third approach (see (5.71)) yields the same answer.

Suppose now that selection exists and that the genotypes A_1A_1 , A_1A_2 , and A_2A_2 have fitnesses given by (1.25a). In view of (1.24) a reasonable stochastic model is found by assuming that the transition matrix for the number of A_1 individuals is (3.16), where now

$$\psi_i = \bar{w}^{-1} (\{w_{11}x^2 + w_{12}x(1-x)\}(1-u) \\ + \{w_{12}x(1-x) + w_{22}(1-x)^2\}v), \quad (3.29)$$

where $x = i/2N$ and \bar{w} is defined by (1.38). The qualitative properties of this model are clear: When $u = v = 0$, one or other absorbing state, $X = 0$, $X = 2N$, is eventually reached. When $u > 0$, $v = 0$, A_1 is eventually lost from the population, and when $u, v > 0$ there will exist a stationary distribution for the number of A_1 genes. Essentially no quantitative results concerning this behavior are known, and the best that can be done is to consider approximations. We do this in Chapter 5 by using diffusion theory, and for the moment foreshadow this approach by deriving an approximate formula for absorption probabilities when $u = v = 0$.

We suppose that $w_{11} = 1 + s$, $w_{12} = 1 + sh$ and $w_{21} = 1$, where s is of order N^{-1} . Put $\alpha = 2Ns$ and, in (2.141), write $i = 2Nx$, $j = 2N(x + \delta x)$. Then this equation may be written

$$\pi(x) = \sum \text{Prob}(x \rightarrow x + \delta x) \pi(x + \delta x) \\ \approx \sum \text{Prob}(x \rightarrow x + \delta x) \left\{ \pi(x) + \delta x \pi'(x) + \frac{1}{2} (\delta x)^2 \pi''(x) \right\} \\ = \pi(x) + \text{E}(\delta x) \pi'(x) + \frac{1}{2} \text{E}(\delta x)^2 \pi''(x).$$

Under the assumptions we have made,

$$\text{E}(\delta x) = (2N)^{-1} \alpha x(1-x) \{x + h(1-2x)\} + O(N^{-2}), \\ \text{E}(\delta x)^2 = (2N)^{-1} x(1-x) + O(N^{-2}).$$

Thus to the order of approximation we use, these calculations give

$$2\alpha \{x + h(1-2x)\} \pi'(x) + \pi''(x) = 0.$$

The solution of this equation, subject to the obvious boundary conditions $\pi(0) = 0$, $\pi(1) = 1$, is

$$\pi(x) = \int_0^x \psi(y) dy / \int_0^1 \psi(y) dy, \quad (3.30)$$

where

$$\psi(y) = \exp(-\alpha y\{2h + y(1 - 2h)\}).$$

In the particular case $h = \frac{1}{2}$, for which the heterozygote is intermediate in fitness between the two homozygotes, this reduces to

$$\pi(x) = \{1 - \exp(-\alpha x)\} / \{1 - \exp(-\alpha)\}. \quad (3.31)$$

It is of some interest to use this approximate formula to get some idea of the effect of the selective differences on the probability of fixation of A_1 . Suppose for example that $N = 10^5$, $s = 10^{-4}$, and $x = 0.5$. Then $\alpha = 20$ and, from (3.31), $\pi(0.5) = 0.999955$. By contrast, for $s = 0$ we have $\pi(0.5) = 0.5$. Evidently the rather small selective advantage 0.0001, which is no doubt too small to be observed in laboratory experiments, is nevertheless large enough in evolutionary terms to have a significant effect on the fixation probability. Clearly this occurs because, while selection might have only a minor effect in any generation, the number of generations until fixation occurs is so very large that the cumulative effect of selection is considerable. We consider this problem at greater length later when more general models are considered and when a more powerful theory is available to handle them.

3.3 The Cannings (Exchangeable) Model: Two Alleles

An important generalization of the Wright–Fisher form of model was introduced by Cannings (1974). We consider a “population” of genes of fixed size $2N$, reproducing at time points $t = 0, 1, 2, 3, \dots$. The stochastic rule determining the population structure at time $t+1$ is quite general, provided that any subset of genes at time t has the same distribution of “descendant” genes at time $t+1$ as any other subset of the same size. Thus, if the i th gene leaves y_i descendant genes we require only that $y_1 + \dots + y_{2N} = 2N$ and that the distribution of (y_i, y_j, \dots, y_k) be independent of i, j, \dots, k . In particular all genes must have the same offspring probability distribution. This distribution must have mean 1 and we denote the variance of this distribution by σ^2 . This interpretation of σ^2 is used throughout this book when Cannings models are considered. In some Cannings models a gene

present at time t can also be present at time $t + 1$, and is then counted as one of its own descendants. An example of this is discussed later.

The Wright–Fisher model (1.48) is a particular case of the Cannings model, since in the model (1.48) $(y_1, y_2, \dots, y_{2N})$ have a symmetric multinomial distribution. However the Cannings model is more general and realistic than the Wright–Fisher model.

Our first calculation concerning the Cannings model relates to eigenvalues. Let the genes be divided into two allelic classes, A_1 and A_2 , and let X_t be the number of A_1 genes at time t . Then we have

Theorem 3.1 (Cannings (1974)). If

$$p_{ij} = \text{Prob}\{X_{t+1} = j \mid X_t = i\}, \quad i, j = 0, 1, 2, \dots, 2N,$$

then the eigenvalues of the matrix $\{p_{ij}\}$ are

$$\lambda_0 = 1, \quad \lambda_j = E(y_1 y_2 \cdots y_j), \quad j = 1, 2, \dots, 2N. \quad (3.32)$$

Since we use this theorem, or generalizations of it, several times below we reproduce here a proof of it, following Cannings (1974).

Proof. Let $P = \{p_{ij}\}$. Suppose that a nonsingular matrix Z and an upper triangular matrix A can be found such that $PZ = ZA$. Since this equation implies $P = ZAZ^{-1}$, the eigenvalues of P are identical to those of A which, because of the special nature of A , are its diagonal elements. Consider now the nonsingular matrix Z , defined by

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1^2 & 1^3 & \cdots & 1^{2N} \\ 1 & 2 & 2^2 & 2^3 & \cdots & 2^{2N} \\ 1 & 3 & 3^2 & 3^3 & \cdots & 3^{2N} \\ \vdots & & & & & \\ 1 & 2N & (2N)^2 & (2N)^3 & \cdots & (2N)^{2N} \end{pmatrix}.$$

With this definition of Z the (i, j) th element of PZ is

$$\sum_k p_{ik} k^j,$$

which can be written

$$E[\{X(t+1)\}^j \mid X(t) = i].$$

Similarly the (i, j) th element of ZA is of the form

$$\sum_{k=0}^j a_{kj} i^k,$$

which may be written as

$$a_{jj} i^{[j]} + \text{terms in } i^{j-1}, \quad j^{i-2}, \dots, i^1 i^0.$$

Here $i^{[j]} = i(i-1)(i-2)\cdots(i-j+1)$. It follows from this that if we can write

$$\mathbb{E}[\{X(t+1)\}^j \mid X(t) = i] = a_{jj}i^{[j]} + \text{terms in } i^{j-1}, i^{j-2}, \dots \quad (3.33)$$

then the a_{jj} ($j = 0, 1, 2, \dots, 2N$) are the eigenvalues of P . In the Cannings model,

$$\begin{aligned} \mathbb{E}[\{X(t+1)\}^j \mid X(t) = i] &= \mathbb{E}\{y_1 + y_2 + \cdots + y_i\}^j \\ &= i\mathbb{E}\{y_1^j\} + \cdots + i^{[j]}\mathbb{E}(y_1 y_2 \cdots y_j), \end{aligned}$$

and it follows that a representation of the form (3.33) is indeed possible for this model, with

$$a_{jj} = \mathbb{E}(y_1 y_2 \cdots y_j), \quad j = 0, 1, 2, \dots, 2N.$$

This completes the proof of the theorem. Cannings also asserted that except in the trivial case $y_j \equiv 1$, the eigenvalues obey the inequalities

$$1 = \lambda_0 = \lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_k = \lambda_{k+1} = \cdots = \lambda_{2N} = 0$$

for some k . However, Gladstien (1978) demonstrated that this is not quite true, and that all that can be asserted is that

$$1 = \lambda_0 = \lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_k = \lambda_{k+1} = \cdots = \lambda_{2N}.$$

It was noted above that in the simple Wright–Fisher model (1.48), any set y_1, y_2, \dots, y_j has a multinomial distribution with index $2N$ and common parameter $(2N)^{-1}$. This implies that if we write

$$\frac{(2N)!}{y_1! y_2! \cdots y_j! (2N - y_1 - \cdots - y_j)!} = \binom{2N}{\mathbf{y}},$$

the eigenvalue $\lambda_j, j = 1, 2, \dots, 2N$ is given by

$$\begin{aligned} \lambda_j &= \sum \cdots \sum y_1 y_2 \cdots y_j \binom{2N}{\mathbf{y}} \left(\frac{1}{2N}\right)^{\sum y_i} \left(1 - \frac{j}{2N}\right)^{2N - \sum y_i} \\ &= (2N)(2N-1)\cdots(2N-j+1)/(2N)^j. \end{aligned} \quad (3.34)$$

This confirms the values given in (3.15), found originally by other methods.

Theorem 3.1 shows that for the Cannings model, the leading nonunit eigenvalue is $\lambda_2 = \mathbb{E}(y_1 y_2)$ where, as defined before Theorem 3.1, y_i is the number of descendent genes of the i th gene in the population. Now $\sum y_j \equiv 2N$, so that the variance of $(\sum y_j)$ is 0. Then by symmetry,

$$2N \operatorname{var}(y_i) + 2N(2N-1) \operatorname{covar}(y_i, y_j) = 0.$$

This implies that

$$\operatorname{covar}(y_i, y_j) = -\sigma^2/(2N-1), \quad (3.35)$$

where $\sigma^2 = \text{var}(y_i)$. Immediately then,

$$\begin{aligned}\lambda_2 &= \text{E}(y_1 y_2) \\ &= \text{covar}(y_1, y_2) + \text{E}(y_1)\text{E}(y_2) \\ &= 1 - \sigma^2/(2N - 1).\end{aligned}\tag{3.36}$$

To confirm this formula we observe that in the Wright–Fisher model, y_i has a binomial distribution with index $2N$ and parameter $(2N)^{-1}$. Thus

$$\lambda_2 = 1 - \{1 - (2N)^{-1}\}/(2N - 1) = 1 - (2N)^{-1},$$

agreeing with the “ $j = 2$ ” case in the expression in (3.34).

Other properties of the Cannings model follow easily. For example, it is clear by symmetry that the probability of eventual fixation of any allele in such a model must be its initial frequency. Further, suppose that there are $X(t)$ A_1 genes in the Cannings model at time t , and write $X(t) = i$ for convenience. If we relabel genes so that the first i genes are A_1 ,

$$\begin{aligned}\text{var}\{X(t+1) \mid X(t)\} &= \text{var}(y_1 + \cdots + y_i) \\ &= i\sigma^2 + i(i-1)\text{covar}(y_1, y_2) \\ &= i(2N - i)\sigma^2/(2N - 1),\end{aligned}\tag{3.37}$$

from (3.35). If $x(t) = X(t)/2N$, it follows that

$$\text{var}\{x(t+1) \mid x(t)\} = x(t)\{1 - x(t)\}\sigma^2/(2N - 1).\tag{3.38}$$

To find the eigenvalues of the matrices defined by (3.16) and (3.24) we use a second theorem due to Cannings (1974). Suppose that if mutation does not exist, the conditions for Theorem 3.1 hold. Now assume that A_1 mutates to A_2 at rate u , with reverse mutation at rate v . Write $x_i = y_i + z_i$, where $y_i = 1$ or 0 depending on whether or not the i th gene at time t continues to exist at time $t + 1$. Thus, $y_i = 0$ in the model (3.16), but we are considering now more general conditions than those specified by this equation. The variable z_i is the number of offspring genes from the i th gene at time t . If this gene is of type A_1 , define z_{i1} as the (random) number of its A_1 (that is, nonmutated) offspring: z_{i1} has a distribution which depends on z_i . Similarly if the i th gene is of type A_2 let z_{i2} be the random number of its A_1 (that is mutant) offspring. Then we have

Theorem 3.2 (Cannings (1974)). The eigenvalues of the matrix P describing the stochastic behavior of the number of A_1 genes are

$$\begin{aligned}\lambda_0 &= 1, \quad \lambda_j = \sum \text{Prob}(z_1, \dots, z_j) \left\{ \text{E} \prod_{i=1}^j (y_i + z_{i1} - z_{i2} \mid z_1, \dots, z_j) \right\}, \\ &\quad (j = 1, 2, \dots, 2N).\end{aligned}\tag{3.39}$$

The proof of this theorem is omitted here. In the model defined by (3.16) and (3.24), $y_i \equiv 0$ and $z_1 \dots z_j$ have a multinomial distribution with

index $2N$ and common parameter $(2N)^{-1}$. Further, given z_i , z_{i1} and z_{i2} have binomial distributions with respective parameters $1 - u$ and v . Thus

$$E(z_{i1} - z_{i2} \mid z_i) = (1 - u - v)z_i$$

and

$$\begin{aligned} \lambda_j &= \sum \text{Prob}(z_1, \dots, z_j)(1 - u - v)^j z_1 \cdots z_j \\ &= (1 - u - v)^j E(z_1 \cdots z_j) \\ &= (1 - u - v)^j \{2N(2N - 1) \cdots (2N - j + 1)/(2N)^j\}, \quad j = 1, 2, \dots, 2N. \end{aligned} \quad (3.40)$$

The conclusion of (3.34) has been used in reaching this formula. The leading nonunit eigenvalue λ_1 is $1 - u - v$ and is thus independent of N . This is extremely close to unity and suggests a very slow rate of approach to stationarity in this model. The eigenvalues (3.40) apply also in the one-way mutation model, for which we simply put $v = 0$ in (3.40).

The conditional branching process model is a particular case of the Cannings model. In this model it is supposed that each gene produces k offspring with probability f_k ($k = 0, 1, 2, 3, \dots$), with the numbers of offspring from different parents being assumed independent. If $f(s) = \sum f_i s^i$, the generating function of the distribution of the total number of offspring genes is $[f(s)]^{2N}$. We now make the condition that the total number of such offspring is $2N$. If at time t there were i A_1 genes, the probability p_{ij} that at time $t + 1$ there will be j A_1 genes is

$$p_{ij} = \frac{\text{coeff } t^j s^{2N} \text{ in } [f(ts)]^i [f(s)]^{2N-i}}{\text{coeff } s^{2N} \text{ in } [f(s)]^{2N}}. \quad (3.41)$$

Transition probabilities of this form were introduced by Moran and Waterson (1959), who used them to find explicit expressions for the leading nonunit eigenvalue in dioecious populations with various family structures. Extensions to this theory were given by Feldman (1966).

Karlin and McGregor (1965) have analyzed the conditional branching process model in detail. They show in particular that the eigenvalues of the matrix $\{p_{ij}\}$ are

$$\lambda_0 = \lambda_1 = 1, \quad \lambda_k = \frac{\text{coeff } s^{2N-k} \text{ in } [f(s)]^{2N-k} [f'(s)]^k}{\text{coeff } s^{2N} \text{ in } [f(s)]^{2N}}, \quad k = 2, 3, \dots, 2N. \quad (3.42)$$

These must agree with the values found in (3.32), since a conditional branching process is a Cannings model. We check that this agreement holds for the eigenvalue λ_2 . It is clear from (3.41) that

$$\sum_j p_{ij} t^j = \frac{\text{coeff } s^{2N} \text{ in } [f(ts)]^i [f(s)]^{2N-i}}{\text{coeff } s^{2N} \text{ in } [f(s)]^{2N}}.$$

Differentiating twice with respect to t and putting $t = 1$,

$$\sum_j j(j-1)p_{ij} = \lambda_2 i(i-1) + \eta_2 i, \quad (3.43)$$

where λ_2 is defined by (3.42) and η_2 is some constant independent of i and j . Now $\sum j p_{ij} = i$ by symmetry, and $\sum j(j-1)p_{1j} = \sigma^2$, where σ^2 is defined after (3.35). Thus putting $i = 1$ in (3.43) we get $\eta_2 = \sigma^2$ and then putting $i = 2$, we get

$$\sum_j j(j-1)p_{2j} = 2\lambda_2 + 2\sigma^2,$$

so that

$$\sum_j j^2 p_{2j} = 2\lambda_2 + 2\sigma^2 + 2. \quad (3.44)$$

But the left-hand side in (3.44) is $E(y_1 + y_2)^2$, where y_i is the random number of offspring genes left by parental gene i . It follows that

$$2 + 2\sigma^2 + 2E(y_1 y_2) = 2\lambda_2 + 2\sigma^2 + 2$$

or

$$\lambda_2 = E(y_1 y_2),$$

as required. Parallel calculations can be made for the remaining eigenvalues, but we do not pursue the details here.

3.4 Moran Models: Two Alleles

The conclusions reached so far depend on the assumption that the appropriate model to describe the stochastic behavior of the number of A_1 genes is one or other form of the model (3.16). Different conclusions are reached for models other than these, and we consider now a model due to Moran (1958) for which this is so. Moran's model has the additional advantage of allowing explicit expressions for many quantities of evolutionary interest, although, strictly, it applies only for haploid populations.

Consider then a haploid population in which, at time points $t = 1, 2, 3, \dots$, an individual is chosen at random to reproduce. After reproduction has occurred, an individual is chosen to die (possibly the reproducing individual but not the new offspring individual). This model is an example of birth and death models, studied extensively in the stochastic process literature. As is discussed later, the model can be generalized by allowing mutation and selection, the latter being introduced by weighting the probability that an individual of a specific genotype is chosen either to give birth or to die.

We consider first the simplest case where there is no selection or mutation. Suppose the population consists of $2N$ haploid individuals (we use this notation to allow direct comparison with the diploid case), each of which is either A_1 or A_2 . Suppose also that at time t , the number of A_1 individuals is i . Then at time $t + 1$ there will be $i - 1$ A_1 individuals if an A_2 is chosen to give birth and an A_1 individual is chosen to die. The probability of this, under our assumptions, is

$$p_{i,i-1} = i(2N - i)/(2N)^2. \quad (3.45)$$

Similar reasoning shows that

$$p_{i,i+1} = i(2N - i)/(2N)^2, \quad (3.46)$$

$$p_{i,i} = \{i^2 + (2N - i)^2\}/(2N)^2. \quad (3.47)$$

The matrix defined by these transition probabilities is a continuant, so that much of the theory of Section 2.12 can be applied to it. In the notation of that section,

$$\lambda_i = \mu_i = i(2N - i)/(2N)^2, \quad \rho_i = 1, \quad i = 0, 1, 2, \dots, 2N. \quad (3.48)$$

It follows that the probability π_i of fixation of A_1 , given currently i A_1 individuals, is

$$\pi_i = i/2N, \quad (3.49)$$

and that using the notation of Section 2.12,

$$\begin{aligned} \bar{t}_{ij} &= 2N(2N - i)/(2N - j), \quad j = 1, 2, \dots, i, \\ \bar{t}_{ij} &= 2Ni/j, \quad j = i + 1, \dots, 2N - 1. \end{aligned} \quad (3.50)$$

Thus immediately

$$\bar{t}_i = 2N(2N - i) \sum_{j=1}^i (2N - j)^{-1} + 2Ni \sum_{j=i+1}^{2N-1} j^{-1}, \quad (3.51)$$

$$\begin{aligned} \bar{t}_{ij}^* &= 2N(2N - i)j/\{i(2N - j)\}, \quad j = 1, 2, \dots, i, \\ \bar{t}_{ij}^* &= 2N, \quad j = i + 1, \dots, 2N - 1, \end{aligned} \quad (3.52)$$

$$\bar{t}_i^* = 2N(2N - i)i^{-1} \sum_{j=1}^i j(2N - j)^{-1} + 2N(2N - i - 1). \quad (3.53)$$

An interesting example of these formulas arises in the case $i = 1$, corresponding to a unique A_1 mutant in an otherwise purely A_2 population. Here $\bar{t}_{1j}^* = 2N$ for all j , so that given that the mutant is eventually fixed, the number of A_1 genes takes, on average, each of the values $1, 2, \dots, 2N - 1$ a total of $2N$ times. The conditional mean fixation time is given by

$$\bar{t}_1^* = 2N(2N - 1) \quad (3.54)$$

birth and death events. The variance of the conditional absorption time can also be written down but we do not do so here.

The eigenvalues of the matrix (3.30) can be found by using Theorem 3.1. Take any collection of j genes and note that the probability that one of these is chosen to reproduce is $j/2N$, with the same probability that one is chosen to die. For this model a gene can be (and indeed usually is) one of its own “descendants”. Using the notation of Theorem 3.1, the product $y_1 y_2 \cdots y_j$ can take only three values:

- 0 if one of these genes is chosen to die and the gene so chosen is not chosen to reproduce,
- 2 if one of the genes is chosen to reproduce and none is chosen to die,
- 1 otherwise.

Thus $\lambda_0 = 1$ and

$$\begin{aligned} \lambda_j &= E(y_1 y_2 \cdots y_j) \\ &= 0\{j(2N - 1)/(2N)^2\} + 2j(2N - j)/(2N)^2 + 1 - j(4N - j - 1)/(2N)^2 \\ &= 1 - j(j - 1)/(2N)^2, \quad j = 1, 2, \dots, 2N. \end{aligned} \tag{3.55}$$

Various expressions for the corresponding eigenvectors, first found by Waterson (1961) using Chebychev polynomials, and later by Gladstien (1978) using other methods, have been given. We are particularly interested in the largest nonunit eigenvalue and its associated eigenvectors. The required eigenvalue is

$$\lambda_2 = 1 - 2/(2N)^2, \tag{3.56}$$

and elementary calculations show that the corresponding right eigenvector \mathbf{r} and left eigenvector ℓ' are given by

$$\begin{aligned} \mathbf{r} &= (0, 1(2N - 1), 2(2N - 2), \dots, i(2N - i), \dots, 1(2N - 1), 0)' \\ \ell' &= (-\frac{1}{2}(2N - 1), 1, 1, 1, \dots, 1, -\frac{1}{2}(2N - 1)). \end{aligned}$$

Thus the asymptotic distribution of the number X_t of A_1 genes for large t , given $X_t \neq 0$, $2N$, is uniform over the values $\{1, 2, 3, \dots, (2N - 1)\}$. The fact that λ_2 is very close to unity agrees with the very large mean absorption times (3.51) for intermediate values of i .

If mutation from A_1 to A_2 is allowed (at rate u), with no reverse mutation, A_1 must eventually become lost, and interest centers on properties of the time for this to occur. The model is now amended to

$$\begin{aligned} p_{i,i-1} &= \{i(2N - i) + ui^2\}/(2N)^2 = \mu_i \\ p_{i,i+1} &= i(2N - i)(1 - u)/(2N)^2 = \lambda_i \\ p_{i,i} &= 1 - p_{i,i-1} - p_{i,i+1}. \end{aligned}$$

Equation (2.160) can now be used to find \bar{t}_{ij} and thus \bar{t}_i . We do not present explicit expressions since it will be more useful, later, to proceed via ap-

proximations. If mutation from A_2 to A_1 (at rate v) is also allowed, the model becomes

$$\begin{aligned} p_{i,i-1} &= \{i(2N-i)(1-v) + ui^2\}/(2N)^2 = \mu_i \\ p_{i,i+1} &= \{i(2N-i)(1-u) + v(2N-i)^2\}/(2N)^2 = \lambda_i \\ p_{i,i} &= 1 - p_{i,i-1} - p_{i,i+1}. \end{aligned} \quad (3.57)$$

Here a stationary distribution arises for the number of A_1 genes in the population, and the typical probability ϕ_j in this distribution ϕ is found, from (2.162), to be

$$\phi_j = \phi_0 \frac{(2N)! \Gamma\{j+A\} \Gamma\{B-j\}}{j!(2N-j)! \Gamma\{A\} \Gamma\{B\}}. \quad (3.58)$$

Here $\Gamma\{\cdot\}$ is the well-known gamma function, $A = 2Nv/(1-u-v)$, $B = 2N(1-v)/(1-u-v)$, $C = 2Nu/(1-u-v)$, $D = 2N/(1-u-v)$ and $\alpha_0 = \Gamma\{B\} \Gamma\{A+C\} / [\Gamma\{D\} \Gamma\{C\}]$. Although these expressions are exact they are rather unwieldy, and we consider below a simple approximation to ϕ_j .

The Markov chain defined by (3.57), having a stationary distribution and a continuant transition matrix, is automatically reversible, as shown by the closing remarks in Chapter 2. This is not necessarily true for other genetical models: It can be shown, for example, that the Wright–Fisher Markov chain defined jointly by (3.16) and (3.24) is not reversible. What does reversibility mean in genetical terms? All the theory we have considered so far is *prospective*, that is, given the current state of a Markov chain, probability statements are made about its future behavior. Recent developments in population genetics theory often concern the *retrospective* behavior: The present state is observed, and questions are asked about the evolution leading to this state. For reversible processes these two aspects have many properties in common, and information about the prospective behavior normally yields almost immediately useful information about the retrospective behavior. We shall see later how the identity of prospective and retrospective probabilities can be used to advantage in discussing various evolutionary questions.

The eigenvalues of (3.57) can be found by applying Theorem 3.2. Here $y_i = 1$ unless the i th gene has been chosen to die, in which case $y_i = 0$. Similarly z_i , z_{i1} and z_{i2} are zero unless the i th gene has been chosen to reproduce. It is found after some calculation that $\lambda_0 = 1$ and

$$\lambda_j = 1 - \frac{j(u+v)}{(2N)} - \frac{j(j-1)(1-u-v)}{(2N)^2}, \quad j = 1, \dots, 2N. \quad (3.59)$$

These eigenvalues apply also in the case $v = 0$. The leading nonunit eigenvalue is $1 - (u+v)/(2N)$, and since $2N$ time units in the process we consider may be thought to correspond to one generation in the Wright–Fisher model, this agrees closely with the value $1 - u - v$ found in (3.40) in that model.

We now obtain approximations for several of the above quantities. It is evident from (3.51) that

$$\bar{t}(p) \approx -(2N)^2 \{p \log p + (1 - p) \log(1 - p)\}, \quad (3.60)$$

where $p = i/2N$. The similarity between this formula and (3.5) is interesting. A factor of $2N$ may be allowed in comparing the two to convert from birth and death events to generations. There remains a further factor of 2 to explain, and we show later why this factor exists.

Consider next the expression (3.58). Put $x = j/(2N)$, $u = \alpha/(2N)$, $v = \beta/(2N)$ and let j and $2N$ increase indefinitely with x , α and β fixed. Using the Stirling approximation $\Gamma\{y+a\}/\Gamma\{y\} \sim y^a$ for large y , moderate a , the stationary probability ϕ_j in (3.58) becomes, approximately,

$$\phi_j \sim (2N)^{-1} \frac{\Gamma\{\alpha + \beta\}}{\Gamma\{\alpha\}\Gamma\{\beta\}} x^{\beta-1} (1-x)^{\alpha-1}, \quad (3.61)$$

at least for values of x not extremely close to 0 or 1. Clearly this approximation expression is far simpler than the exact value (3.58). The values for \bar{t}_{ij} may be calculated from (2.160) and (3.57), and from these the value of \bar{t}_i . This is

$$\begin{aligned} \bar{t}_i \approx (2N)^2 (1 - \theta)^{-1} & \left(\int_0^p x^{-1} \{(1-x)^{\theta-1} - 1\} dx \right. \\ & \left. + \int_p^1 x^{-1} (1-x)^{\theta-1} \{1 - (1-p)^{1-\theta}\} dx \right) \end{aligned} \quad (3.62)$$

birth and death events, where $p = i/(2N)$, $x = j/(2N)$ and θ is defined for the diffusion approximation to this Moran model as $2Nu$. In the particular case $p = (2N)^{-1}$ this is, to a close approximation,

$$\bar{t}_i \approx 2N \left(1 + \int_{(2N)^{-1}}^p x^{-1} (1-x)^{\theta-1} dx \right) \quad (3.63)$$

birth and death events. When $\theta = 1$ the form of \bar{t}_i may be found by application of L'Hospital's rule.

Selection can be incorporated in this model by assuming differential birth rates or differential death rates. The two approaches give similar results so we consider here only the case where death rates differ. To do this we suppose that if at any time there are i A_1 genes in the population the probability that the next individual chosen to die is A_1 is

$$\mu_1 i / \{\mu_1 i + \mu_2 (2N - i)\}. \quad (3.64)$$

If $\mu_1 = \mu_2$ there is no selection while if $\mu_1 < \mu_2$ the allele A_1 has a selective advantage over A_2 . It follows that the transition matrix for the number of

A_1 individuals has elements

$$\begin{aligned} p_{i,i-1} &= \mu_1 i(2N-i)/[2N\{\mu_1 i + \mu_2(2N-i)\}], \\ p_{i,i+1} &= \mu_2 i(2N-i)/[2N\{\mu_1 i + \mu_2(2N-i)\}], \\ p_{i,i} &= 1 - p_{i,i-1} - p_{i,i+1}. \end{aligned} \quad (3.65)$$

The matrix defined by (3.65) is a continuant, and the theory of Section 2.12 applies. In the notation of that section,

$$\rho_0 = 1, \quad \rho_k = (\mu_1/\mu_2)^k,$$

and the probability π_i of eventual fixation of A_1 , given an initial number of i A_1 individuals, is

$$\pi_i = \{1 - (\mu_1/\mu_2)^i\} / \{1 - (\mu_1/\mu_2)^{2N}\}. \quad (3.66)$$

If now $\mu_1/\mu_2 = 1 - \frac{1}{2}s$, where s is small and positive, A_1 has a slight selective advantage over A_2 and (3.66) can be approximated by

$$\pi(x) \approx \{1 - \exp(-\frac{1}{2}\alpha x)\} / \{1 - \exp(-\frac{1}{2}\alpha)\}, \quad (3.67)$$

where $x = i/2N$ and $\alpha = 2Ns$. This formula differs from (3.31) by a factor of 2 in the exponents. This is not because the selective differences differ by a factor of 2, since indeed they do not, but from a more deep-rooted difference between the two models which we examine later.

It is possible to use (3.65) in conjunction with the continuant formulas of Section 2.12 to get expressions for mean absorption times, conditional mean absorption times, and so on. We do not do this here since the formulas become very unwieldy and uninformative, and since also we later consider simple approximations for these quantities. It may finally be remarked that no formula is known for the eigenvalues of the matrix defined by (3.65).

3.5 K-Allele Wright–Fisher Models

The models considered so far can easily be extended to allow K different alleles at the locus in question, where K is an arbitrary positive integer. In this case the population configuration at any time can be described by a vector (X_1, X_2, \dots, X_K) , where X_i is the number of genes of allelic type A_i . If we assume, as is usual, that $X_1 + X_2 + \dots + X_K = 2N$, only $K - 1$ elements in the above vector are independent. It is however convenient to retain all elements in the vector. The most interesting cases of these models arise when there is no mutation and a generalization of the Cannings model determines the evolution of the population. In this case any allele A_i can be treated on its own, all other alleles being classed simply as non- A_i , and much of the theory of the preceding sections can be applied. One problem for which the preceding theory is inadequate is to find the mean time until loss of the first allele lost, the mean time until loss of the second allele lost,

and so on. This more complex problem and various associated problems is discussed in Section 5.10.

We consider in detail only the K -allele generalization of the model Wright–Fisher (1.48), namely

$$\begin{aligned} & \text{Prob}\{Y_i \text{ genes of allele } A_i \text{ at time } t+1 \mid X_i \text{ genes of allele} \\ & \quad i \text{ at time } t, \quad i = 1, 2, \dots, K\} \\ &= \frac{(2N)!}{Y_1! Y_2! \dots Y_K!} \psi_1^{Y_1} \psi_2^{Y_2} \dots \psi_K^{Y_K} \end{aligned} \quad (3.68)$$

where $\psi_i = X_i/(2N)$. In this case the model (3.68) is in effect a Cannings model and the theory for the Cannings model given above, or straightforward generalizations of it, can be used. The eigenvalues of the matrix defined by (3.68) are precisely the values in (3.34), where now λ_j has multiplicity $(K+j-2)!/\{(K-2)!/j!\}$, ($j = 2, 3, \dots, 2N$). The eigenvalue $\lambda_0 = 1$ has total multiplicity K . These eigenvalues have the interesting interpretation (Littler (1975)) that

$$\text{Prob}\{\text{at least } j \text{ allelic types remain present at time } t\} \sim \text{const } \lambda_j^t. \quad (3.69)$$

Expressions for the mean times between losses of alleles are given explicitly later (see (5.122) and (5.123)), where it will be shown that the eigenvalue expression (3.69) does not give useful information about these mean times.

When mutation exists between all alleles there will exist a multi-dimensional stationary distribution of allelic numbers. The means, variances and covariances in this distribution can be found by procedures analogous to those leading to (3.25) and (3.26). We consider in detail only the case where mutation is symmetric: In this case the probability that any gene mutates is assumed to be u , and given that a gene of allelic type A_i has mutated, the probability that the new mutant is of type A_j is $(K-1)^{-1}$, ($j \neq i$). By symmetry, the mean number of genes of allelic type A_i alleles in the stationary distribution must be $2N/K$. However, it sometimes occurs that this is not a likely value for the actual number of genes of any allelic type to arise, and we see this best by finding the probability F that two genes taken at random from the population are of the same allelic type. Generalizing the argument that led to (3.28) we find, ignoring terms of order u^2 , that

$$F = ((2N)^{-1} + \{1 - (2N)^{-1}\}F)(1 - 2u) + (1 - (2N)^{-1})(1 - F)(2u/(K-1)).$$

If we write $\theta = 4Nu$, this gives

$$F \approx (K-1 + \theta)/(K-1 + K\theta). \quad (3.70)$$

This expression agrees with that in (3.28) for $K = 2$, and letting $K \rightarrow \infty$ we find

$$F \approx (1 + \theta)^{-1}. \quad (3.71)$$

This formula demonstrates a theme that will recur later. If θ is small then $F \approx 1$. This implies that it is very likely that one or other allele appears with high frequency in the population, with the remaining alleles having negligible frequency, despite the fact that all alleles are selectively equivalent. The imbalance arises because of stochastic effects, and is quite different from that predicted by considering the mean allele frequencies only.

The eigenvalues of the matrix defined by the symmetric mutation model are the values (3.34) if λ_i is multiplied by $\{1 - uK(K - 1)^{-1}\}^i$. The multiplicity of λ_i is $(i + K - 2)!/\{i!(K - 1)!\}$.

In view of the comments concerning the Cannings model made in Section 3.7 it is plausible that (3.70) and (3.71) hold with θ defined by $\theta = 4Nu/\sigma^2$. There is also a K -allele Moran model which allows various exact formulas, but for this model interest centers more on the infinitely many alleles case, to which we now turn.

3.6 Infinitely Many Alleles Models

3.6.1 Introduction

In this section we consider three so-called “infinitely many alleles” models, namely the Wright–Fisher model, the Cannings and the Moran model.

The discussion of the Wright–Fisher model is more extensive than that for the remaining models. This is not because it is more important than the other two: Indeed, the Wright–Fisher model is a particular case of the more general, and more plausible, Cannings model. The extensive discussion of the Wright–Fisher model arises for two reasons. The first of these is that calculations for this model are comparatively straightforward, and the second is that results for this model can be taken over almost directly for the Cannings model, with an appropriate change in the definition of the parameter θ arising in many of the formulas found.

Results for the Wright–Fisher and the Cannings infinitely many alleles models are usually diffusion approximations. By contrast, the infinitely many alleles Moran model allows many exact calculations.

In Chapter 9 we discuss why infinitely many alleles models are of interest and will develop some of their properties at greater length than is done in this section.

3.6.2 The Wright–Fisher Infinitely Many Alleles Model

The Wright–Fisher infinitely many alleles model follows the generic binomial sampling characteristic of all Wright–Fisher models. Mutation is intrinsic to the model, but the nature of the new mutants is different from anything assumed so far, the key difference being that all mutant genes are assumed to be of a new allelic type, not currently or previously seen in the

population. This implies that if the mutation rate is u , and if in generation t there are X_i genes of allelic type A_i ($i = 1, 2, 3, \dots$), then the probability that in generation $t + 1$ there will be Y_i genes of allelic type A_i , together with Y_0 new mutant genes, all of different novel allelic types, is

$$\text{Prob}\{Y_0, Y_1, Y_2, \dots \mid X_1, X_2, \dots\} = \frac{(2N)!}{\prod Y_i!} \prod \pi_i^{Y_i}, \quad (3.72)$$

where $\pi_0 = u$ and $\pi_i = X_i(1 - u)/(2N)$, $i = 1, 2, 3, \dots$

This model differs fundamentally from previous mutation models (which allow reverse mutation) in that since each allele will sooner or later be lost from the population, there can exist no nontrivial stationary distribution for the frequency of any allele. Nevertheless we are interested in stationary behavior, and it is thus important to consider what concepts of stationarity exist for this model. To do this we consider delabeled configurations of the form $\{a, b, c, \dots\}$, where such a configuration implies that there exist a genes of one allelic type, b genes of another allelic type, and so on. The specific allelic types involved are not of interest. The possible configurations can be written down as $\{2N\}$, $\{2N - 1, 1\}$, $\{2N - 2, 2\}$, $\{2N - 2, 1, 1\}$, \dots , $\{1, 1, 1, \dots, 1\}$ in dictionary order: The number of such configurations is $p(2N)$, the number of partitions of $2N$ into positive integers. For small values of N values of $p(2N)$ are given by Abramowitz and Stegun (1965, Table 24.5), who provide also asymptotic values for large N . It is clear that (3.72) implies certain transition probabilities from one configuration to another. Although these probabilities are extremely complex and the Markov chain of configurations has an extremely large number of states, nevertheless standard theory shows that there exists a stationary distribution of configurations, some of the characteristics of which we now explore.

We consider first the probability that two genes drawn at random are of the same allelic type. For this to occur neither gene can be a mutant and, further, both must be descended from the same parent gene (probability $(2N)^{-1}$) or different parent genes which were of the same allelic type. Writing $F_2^{(t)}$ for the desired probability in generation t , we get

$$F_2^{(t+1)} = (1 - u)^2 \left\{ (2N)^{-1} + \{1 - (2N)^{-1}\} F_2^{(t)} \right\}. \quad (3.73)$$

At equilibrium, $F_2^{(t+1)} = F_2^{(t)} = F_2$ and thus

$$F_2 = \{1 - 2N + 2N(1 - u)^{-2}\}^{-1} \sim (1 + \theta)^{-1}, \quad (3.74)$$

where, as is standard for Wright–Fisher models, $\theta = 4Nu$. This is identical to the limiting ($K \rightarrow \infty$) value in (3.71), a fact that we return to later.

Consider next the probability $F_3^{(t+1)}$ that three genes drawn at random in generation $t + 1$ are of the same allelic type. These three genes will all be descendants of the same gene in generation t , (probability $(2N)^{-2}$),

of two genes (probability $3(2N-1)((2N)^{-2})$) or of three different genes (probability $(2N-1)(2N-2)((2N)^{-2})$). Further, none of the genes can be a mutant, and it follows that

$$F_3^{(t+1)} = (1-u)^3(2N)^{-2}(1+3(2N-1)F_2^{(t)} + (2n-1)(2N-2)F_3^{(t)}). \quad (3.75)$$

At equilibrium $F_3^{(t+1)} = F_3^{(t)} = F_3$, and rearrangement in (3.75) yields

$$F_3 \approx 2(2+\theta)^{-1}F_2 \approx 2!/[(1+\theta)(2+\theta)]. \quad (3.76)$$

Continuing in this way we find

$$F_i^{(t+1)} = (1-u)^i[(2N-1)(2N-2)\cdots(2N-i+1)(2N)^{1-i}F_i^{(t)} + \text{terms in } F_{i-1}^{(t)}, \dots, F_2^{(t)}] \quad (3.77)$$

and that for small values of i ,

$$F_i \approx (i-1)!/[(1+\theta)(2+\theta)\cdots(i-1+\theta)]. \quad (3.78)$$

We can also interpret F_i as the probability that a sample of i genes contains only one allelic type, or, in other words, that the sample configuration is $\{i\}$. This conclusion may be used to find the probability of the sample configuration $\{i-1, 1\}$. The probability that in a sample of i genes, the first $i-1$ genes are of one allelic type while the last gene is of a new allele type is $F_{i-1} - F_i$. The probability we require is, for $i \geq 3$, just i times this, or

$$\text{Prob}\{i-1, 1\} = i\{F_{i-1} - F_i\} \approx i(i-2)\theta/[(1+\theta)(2+\theta)\cdots(i-1+\theta)]. \quad (3.79)$$

For $i = 2$ the required probability is

$$\text{Prob}\{1, 1\} \approx \theta/(1+\theta). \quad (3.80)$$

The probabilities of other configurations can be built up in a similar way. We illustrate this by considering the probability $F_{2,2}^{(t+1)}$ that, of four genes drawn at random in generation $t+1$, two are of one allelic type and two of another. Clearly none of the genes can be a mutant, and furthermore they will be descended from four different parent genes of configuration $\{2, 2\}$, from three different parent genes of configuration $\{2, 1\}$, the singleton being transmitted twice, or from two different parent genes, both transmitted twice. Considering the probabilities of the various events, we find

$$F_{2,2}^{(t+1)} = (1-u)^4(2N)^{-3}((2N-1)(2N-2)(2N-3)F_{2,2}^{(t)} + 2(2N-1)(2N-2)F_{2,1}^{(t)} + 3(2N-1)F_{1,1}^{(t)}). \quad (3.81)$$

Retaining only higher-order terms and letting $t \rightarrow \infty$, we obtain

$$F_{2,2} \approx (3+\theta)^{-1}F_{2,1} = 3\theta/((1+\theta)(2+\theta)(3+\theta)). \quad (3.82)$$

Continuing in this way we find (Ewens (1972), Karlin and McGregor (1972)) an approximating partition probability formula for a sample of

n of genes, where it is assumed that $n \ll N$. This formula can be presented in various ways. Perhaps the most useful formula arises if we define $\mathbf{A} = (A_1, A_2, \dots, A_n)$ as the vector of the (random) numbers of allelic types each of which is represented by exactly j genes in the sample. With this definition,

$$\text{Prob}(\mathbf{A} = \mathbf{a}) = \frac{n! \theta^{\sum a_j}}{1^{a_1} 2^{a_2} \dots n^{a_n} a_1! a_2! \dots a_n! S_n(\theta)}. \quad (3.83)$$

Here $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $S_n(\theta)$ is defined as $\theta(\theta+1)(\theta+2) \dots (\theta+n-1)$.

It is necessary that $\sum j A_j = \sum j a_j = n$, and it is convenient to denote $\sum A_j$, the (random) number of different allelic types seen in the sample, by K , and $\sum_j a_j$, the corresponding observed number in a given sample, by k . By suitable summation in (3.83) the probability distribution of the random variable K may be found as

$$\text{Prob}(K = k) = |S_n^k| \theta^k / S_n(\theta), \quad (3.84)$$

where $|S_n^k|$ is the coefficient of θ^k in $S_n(\theta)$. Thus $|S_n^k|$ is the absolute value of a Stirling number of the first kind (see Abramowitz and Stegun, (1965)).

From (3.84), the mean of K is

$$E(K) = \frac{\theta}{\theta} + \frac{\theta}{\theta+1} + \frac{\theta}{\theta+2} + \dots + \frac{\theta}{\theta+n-1}, \quad (3.85)$$

the variance of K is

$$\text{var}(K) = \theta \sum_{j=1}^{n-1} \frac{j}{(\theta+j)^2}, \quad (3.86)$$

and the probability that $K = 1$ is

$$\frac{(n-1)!}{(\theta+1)(\theta+2) \dots (\theta+n-1)}. \quad (3.87)$$

A formula equivalent to (3.83) is the following. Suppose that in the sample we observe k different allelic types. We label these in some arbitrary order as types $1, 2, \dots, k$. Then the probability that $K = k$ and also that with the types labelling in the manner chosen, there are n_1, n_2, \dots, n_k genes respectively observed in the sample of these various types, is

$$\frac{n! \theta^k}{k! n_1 n_2 \dots n_k S_n(\theta)}. \quad (3.88)$$

These various formulas lead to interesting questions of inference, which we take up in detail in Sections 9.5 and 11.2.

Equation (3.73) can be rewritten in the form

$$F_2^{(t+1)} - F_2^{(\infty)} = (1-u)^2 \{1 - (2N)^{-1}\} \{F_2^{(t)} - F_2^{(\infty)}\}, \quad (3.89)$$

and this implies that $(1-u)^2 \{1 - (2N)^{-1}\}$ is an eigenvalue of the Markov chain configuration process discussed above. A similar argument using

(3.75) shows that a second eigenvalue is $(1-u)^3\{1-(2N)^{-1}\}\{1-2(2N)^{-1}\}$. Equations (3.77) and (3.81) suggest that $(1-u)^4\{1-(2N)^{-1}\}\{1-2(2N)^{-1}\}\{1-3(2N)^{-1}\}$ is an eigenvalue of multiplicity 2. It is found more generally that

$$\lambda_i = (1-u)^i\{1-(2N)^{-1}\}\{1-2(2N)^{-1}\} \cdots \{1-(i-1)(2N)^{-1}\} \quad (3.90)$$

is an eigenvalue of the configuration process matrix and that its multiplicity is $p(i) - p(i-1)$, where $p(i)$ is the partition number given above. This provides a complete listing of all the eigenvalues. For details see Ewens and Kirby (1975).

We consider next the mean number of alleles existing in the population at any time. Any specific allele A_m will be introduced into the population with frequency $(2N)^{-1}$, and after a random number of generations will leave it, never to return. The frequency of A_m is a Markovian random variable with transition matrix in (3.16), with ψ_i defined immediately below (3.16). There will exist a mean time that $E(T)$ that A_m remains in the population. The mean number of new alleles to be formed each generation is $2Nu$, and the mean number to be lost each generation through mutation and random drift is $E(K)/E(T)$, where $E(K)$ is the mean number of alleles existing in each generation. It follows, by balancing the number of alleles gained each generation with the number lost, that at stationarity,

$$E(K) = 2NuE(T). \quad (3.91)$$

An approximation to $E(T)$ is found by putting $p = (2N)^{-1}$ in (3.19). This gives, to a close approximation,

$$E(K) \approx \theta + \int_{(2N)^{-1}}^1 \theta x^{-1}(1-x)^{\theta-1} dx. \quad (3.92)$$

A more detailed approximation is possible. If $E(K(x_1, x_2))$ is the mean number of alleles present in the population with frequency in any interval (x_1, x_2) ($(2N)^{-1} \leq x_1 < x_2 \leq 1$), then

$$E(K(x_1, x_2)) \approx \int_{x_1}^{x_2} \theta x^{-1}(1-x)^{\theta-1} dx. \quad (3.93)$$

This equation can be used to confirm (3.85). An allele whose population frequency is x is observed in a sample of size n with probability $1 - (1-x)^n$. From this and (3.93) it follows that the mean number of different alleles observed in a sample of size n is approximately

$$\int_0^1 \{1 - (1-x)^n\} \theta x^{-1}(1-x)^{\theta-1} dx, \quad (3.94)$$

and the value of this expression is equal to that given in (3.85). The function

$$\phi(x) = \theta x^{-1}(1-x)^{\theta-1} \quad (3.95)$$

is called the “frequency spectrum” of the process considered. Ignoring small-order terms, it has the (equivalent) interpretations that the mean number of alleles in the population whose frequency is in $(x, x + \delta x)$, and also the probability that there exists an allele in the population whose frequency is in this range, is, for small δx , equal to $\theta x^{-1}(1-x)^{\theta-1}\delta x$.

The frequency spectrum can be used to arrive at further results reached more laboriously by discrete distribution methods. Thus, for example,

$$\begin{aligned} & \text{Prob}\{\text{only one allele observed in a sample of } n \text{ genes}\} \\ & \approx \theta \int_0^1 x^n \{x^{-1}(1-x)^{\theta-1}\} dx \\ & = (n-1)! / ((1+\theta)(2+\theta) \cdots (n-1+\theta)) \end{aligned}$$

and this agrees with the expression in (3.78) with the notational change of n to i . More complex formulas such as (3.83) can be re-derived using multivariate frequency spectra, but we do not pursue the details.

The form of the frequency spectrum also shows that when θ is small, the most likely situation to arise at any time is that where one allele has a high frequency and the remaining alleles are all at a low frequency. This occurs for two reasons. The first of these is historical: Different alleles enter the population at different times, and an “older” allele has had more time to reach a high frequency than a “younger” allele. Second, imbalances in allelic frequencies arise through stochastic fluctuations, as in the K -allele model as discussed below (3.71). This imbalance agrees qualitatively with that found in the K -allele model of Section 3.5. We shall later find a number of uses for frequency spectra, all arising through their definitions in equations of the form (3.93).

Although the theory is by no means clear, it is plausible that to a first approximation, all the results given in this section continue to apply in more complicated Wright–Fisher models, involving perhaps two sexes or geographical structure, if the parameter θ is defined as

$$\theta = 4N_e u, \quad (3.96)$$

where N_e is one or other version of the effective population size (see Section 3.7).

Various generalizations of the selectively neutral Wright–Fisher infinitely many alleles model are possible. One generalization supposes that all heterozygotes have the same fitness $1 + s$ ($s > 0$) and all homozygotes have fitness 1. An extreme example arises for self-sterility alleles where homozygotes cannot appear. For this case we put $s = \infty$. For selective models the simple symmetry arguments, which lead to (3.91) no longer apply, and

a more complex analysis is necessary. We consider this analysis further in Chapter 5. A second generalization supposes that alleles fall into two classes, with individuals having two “favored” alleles having fitness $1 + 2s$, that those having only one favored allele having fitness $1 + s$, and that those with no favored allele having fitness 1. This model also is considered in more detail in Chapter 5.

3.6.3 *The Cannings Infinitely Many Alleles Model*

The reproductive mechanism in the nonoverlapping generations Cannings infinitely many alleles model follows that of the general principles of the Cannings two-allele model of Section 3.3. That is, the model allows any reproductive scheme consistent with the exchangeability and symmetry properties of the two-allele model. The mean number of offspring genes from any “parental” gene is 1, and the variance of the number of offspring genes is σ^2 , necessarily the same for each parental gene. The model follows the mutation mechanism of the Wright–Fisher infinitely many alleles model described above, in that all mutant offspring genes are assumed to be of novel allelic types.

Many of the results of the Wright–Fisher infinitely many alleles model apply for the Cannings model, at least to a close approximation, provided that the parameter θ , arising in many formulas in Section 3.6.2, is replaced by θ/σ^2 , as justified by the discussion leading to (3.111) below. We therefore use these Wright–Fisher formulae, with this change of definition, to apply for the Cannings model.

3.6.4 *The Moran Infinitely Many Alleles Model*

The Moran infinitely many alleles model is the natural extension to the infinitely many alleles case of the Moran two alleles model considered in Section 3.4. Haploid individuals, which we may identify with genes, are created and lost through a birth and death process, as in the two-alleles case, but in the infinitely many alleles model it is assumed that an offspring gene is a mutant with probability u and that any new mutant is of an entirely novel allelic type, not currently or previously existing in the population.

The stochastic behavior of the frequency of any allelic type in the population is then governed by (3.57), implying that there can be no concept of stationarity of the frequency of any nominated allelic type. On the other hand, as with the Wright–Fisher and Cannings models, there will exist a concept of the stationary distribution of allelic configurations. The possible configurations of the process are the same as those for those models, but for the Moran model an exact population probability can be given for each configuration. Suppose that β_j ($j = 1, 2, \dots, 2N$) is the number of allelic types with exactly j representative genes in the population, so that $\sum j\beta_j = 2N$. The quantity β_j is the population analogue of the sample

number α_j in (3.83). The exact stationary distribution of the population configuration process is (Trajstman, (1974))

$$\text{Prob}(\beta_1, \beta_2, \dots, \beta_{2N}) = \frac{(2N)! \theta^{\sum \beta_j}}{1^{\beta_1} 2^{\beta_2} \dots (2N)^{\beta_{2N}} \beta_1! \beta_2! \dots \beta_{2N}! S_{2N}(\theta)}. \quad (3.97)$$

Here $S_j(\cdot)$ is defined below (3.83) and θ is defined for this model by

$$\theta = 2Nu/(1 - u). \quad (3.98)$$

This is a different definition of often θ than that applying for the Wright–Fisher model, the difference arising because of the effective population size applying for the Moran model.

The expression (3.97) is of exactly the same form as (3.83), with n replaced by $2N$ and α_j by β_j . Thus several of the calculations arising from (3.83) are exact for the Moran population process. For example, the distribution of the number K_{2N} of allelic types in the population is given exactly by (3.84), with n replaced by $2N$. Thus, immediately from (3.84), the monomorphism probability that $K_{2N} = 1$ is, exactly,

$$P_{\text{mono}} = \frac{(2N - 1)!}{(1 + \theta)(2 + \theta) \dots (2N - 1 + \theta)}. \quad (3.99)$$

The mean of K_{2N} is given by (3.85), with in both cases n replaced by $2N$ and θ defined by (3.98), and the variance K_{2N} is

$$\text{var}(K_{2N}) = \theta \sum_{j=1}^{2N-1} \frac{j}{(\theta + j)^2}. \quad (3.100)$$

A further exact result for the Moran model concerns its exact frequency spectrum, for which (3.95) gives the diffusion approximation in the Wright–Fisher model.

To find this we consider first the “two-allele” model (3.57). In the infinitely many alleles case we think of A_1 as a new arisen allele formed by mutation and A_2 as all other alleles. (2.160) can be used to find the mean number $\mu(T)$ of birth and death events before its certain loss from the population. This is

$$\mu(T) = (2N + \theta) \sum_{j=1}^{2N} j^{-1} \left(\binom{2N}{j} \binom{2N + \theta - 1}{j}^{-1} \right), \quad (3.101)$$

In the case $\theta = 2$, this is about $2N \log(2N)$ birth and death events, or about $\log(2N)$ “generations”. The corresponding approximation for the Wright–Fisher model, found from (3.19), is also $\log(2N)$ generations, but this formal equality is misleading because of the different definitions of θ in the two cases.

The expression (3.101) has the further interpretation that its typical term is the mean number of birth and death events for which there are exactly j

copies of the allele in question before its loss from the population. The form of ergodic argument that led to (3.92) shows that at stationarity, the mean of the number K_{2N} of different allelic types represented in the population is $u\mu(T)$, which is

$$\theta \sum_{j=1}^{2N} j^{-1} \left(\binom{2N}{j} \binom{2N+\theta-1}{j}^{-1} \right), \quad (3.102)$$

where here and throughout we use the standard gamma function definition

$$\binom{M}{m} = \frac{\Gamma(M+1)}{m!\Gamma(M-m+1)} = \frac{M(M-1)\cdots(M-m+1)}{m!}$$

for non-integer M . The expression (3.102) simplifies to

$$\frac{\theta}{\theta} + \frac{\theta}{\theta+1} + \frac{\theta}{\theta+2} + \cdots + \frac{\theta}{\theta+2N-1}.$$

This is identical to the expression given in (3.85), with n replaced by $2N$, as we know it must be. However the expression (3.102) provides the further information that the typical j th term gives the stationary mean number of alleles arising with j representing genes in the population at any time. In other words, the exact frequency spectrum for the Moran model is

$$\theta j^{-1} \left(\binom{2N}{j} \binom{2N+\theta-1}{j}^{-1} \right), \quad j = 1, 2, \dots, 2N. \quad (3.103)$$

A standard asymptotic formula for the gamma function for large N shows the parallel between this exact expression with the diffusion theory frequency spectrum (3.95).

Many further exact results for the Moran model are available. Many of these relate to “time” and “age” properties, and will be discussed at length in Chapter 9, where “time” and “age” questions are of central interest.

3.7 The Effective Population Size

While the Wright–Fisher model (1.48) is less plausible than several other available models as a description of biological reality, it has, perhaps for historical reasons, assumed a central place in population genetics theory. We have already noted three properties of this model:

- (i) its maximum nonunit eigenvalue = $1 - (2N)^{-1}$,
- (ii) the probability that two genes taken at random are descendants of the same parent gene = $(2N)^{-1}$,
- (iii) $\text{var}\{x(t+1) \mid x(t)\} = x(t)\{1 - x(t)\}/(2N)$, where $x(t)$ is the fraction of A_1 genes in generation t .

In view of these properties it is perhaps natural, if the Wright–Fisher model (1.48) is to be used as a standard, to define the effective population size in diploid models that are more complicated and realistic than (1.48) in the following way:

$$N_e^{(e)} = \text{eigenvalue effective population size} = \frac{1}{2}(1 - \lambda_{\max})^{-1}, \quad (3.104)$$

$$N_e^{(i)} = \text{inbreeding effective population size} = (2\pi_2)^{-1}, \quad (3.105)$$

$$N_e^{(v)} = \text{variance effective population size} \\ = \frac{x(t)\{1 - x(t)\}}{2 \operatorname{var}\{x(t+1) \mid x(t)\}}. \quad (3.106)$$

Here λ_{\max} is the largest nonunit eigenvalue of the transition matrix of the model considered and π_2 is the probability, in this model, that two genes taken at random in any generation are descendants of the same parent gene. Similarly, $\operatorname{var}\{x(t+1)\}$ is the conditional variance of the frequency of A_1 in generation $t+1$ in the more complicated model, given the value of this frequency in generation t .

A fourth concept of effective population size, namely the mutation effective size, is also possible (Ewens (1989)) but we do not consider this concept here.

Our aim is to compute the three effective population sizes defined above for two classes of models that generalize the simple Wright–Fisher model (1.48). The first class is the Cannings model considered in Section 3.3 and the second comprises Wright–Fisher models that incorporate complicating features such as two sexes, geographical subdivision, fluctuating population sizes, and so on.

We consider first the Cannings model, and limit attention for the moment to those versions of the model where generations do not overlap. Equations (3.36) and (3.104) show immediately that for these models, the eigenvalue effective population size $N_e^{(e)}$ is given by

$$N_e^{(e)} = (N - \frac{1}{2})/\sigma^2, \quad (3.107)$$

where, as in Section 3.3, σ^2 is the variance in the number of offspring genes from any given gene. Equations (3.38) and (3.106) show that the variance effective population size $N_e^{(v)}$ is given by

$$N_e^{(v)} = (N - \frac{1}{2})/\sigma^2. \quad (3.108)$$

A value for $N_e^{(i)}$ can be found in the following way. Suppose that the i th gene in generation t leaves m_i offspring genes in generation $t+1$, ($\sum m_i = 2N$). Then the probability, given m_1, \dots, m_{2N} , that two genes drawn at random in generation $t+1$ are descendants of the same gene is

$$\sum_{i=1}^{2N} m_i(m_i - 1) / \{2N(2N - 1)\}. \quad (3.109)$$

The probability π_2 in (3.105) is the expectation of this quantity. Now m_i has mean unity and variance σ^2 , so that on taking expectations, $\pi_2 = \sigma^2/(2N - 1)$. From this,

$$N_e^{(i)} = (N - \frac{1}{2})/\sigma^2. \quad (3.110)$$

It follows from these various equations that for the Cannings model, all three effective population sizes are equal.

One application of this conclusion is the following. If leading terms only are retained, all three definitions of the effective population size in the Cannings model are N/σ^2 . From the remarks surrounding (3.96), it is plausible that the various Wright–Fisher infinitely many alleles model results given in Section 3.6 apply for the nonoverlapping generation Cannings model if θ is defined wherever it occurs by $4N_e u$. That is, to a close approximation, we define θ for the Cannings model by

$$\theta = 4N_e u/\sigma^2. \quad (3.111)$$

This definition is used for the calculations in Chapters 9 and 10, where the Cannings model plays an important role.

The above definitions of the effective population size are not appropriate for models such as (3.30) where generations overlap. If we write N_e for any one of the effective population sizes defined in (3.104)–(3.106), it seems reasonable for such models to define the effective population size as $N_e k/(2N)$, where k is the number of individuals to die each time unit. Since $k = 2N$ for models where generations do not overlap, this leaves (3.104)–(3.106) unchanged for such models. For the Moran model (3.30), where $k = 1$, this convention yields

$$N_e^{(e)} = N_e^{(i)} = N_e^{(v)} = \frac{1}{2}N. \quad (3.112)$$

However, in contrast to our approach for the Cannings model, we do not use this observation to use Wright–Fisher diffusion approximation results from Section 3.6 for the Moran infinitely many alleles model, since exact calculations are available for that model, as described in Section 9.3. Our interest in (3.112) arises for another reason, namely that it shows that the effective population size in the Moran model is half that in the Wright–Fisher model. We now discuss the reason for this.

Arguments parallel to those leading to (3.5) show that if two alleles A_1 and A_2 are allowed in the population, the mean time until fixation of one or other allele in the Cannings model is

$$\bar{t}(p) \approx -(4N - 2)\{p \log p + (1 - p) \log(1 - p)\}/\sigma^2, \quad (3.113)$$

where p is the initial frequency of A_1 and σ^2 is defined above. This formula explains the factor of 2 discussed after equation (3.60). In the Wright–Fisher model $\sigma^2 \approx 1$ while in the Moran model $\sigma^2 \approx 2/(2N)$. Setting aside the factor $2N$ as explained by the conversion from generations to birth and death events, it is clear that the crucial factor is the difference in the

variance in offspring distribution. It is also this factor which leads to the difference between (3.31) and (3.67) and that between other similar pairs of formulas.

So far we have ignored the diploid nature of most organisms of interest, and we now consider a definition of effective population size for the diploid case. We do this here for a Cannings model. An inbreeding effective population number is sometimes defined where attention is focussed on the diploid nature of the organisms in the population. This number will be denoted $N_e^{(id)}$, and is defined as the reciprocal of the probability that two genes taken at random in generation $t + 1$ are descended from the same individual in generation t . This is tantamount, in the Cannings model, to selecting two genes at random in generation t and asking whether the two genes drawn at random in generation $t + 1$ are both descended from one or other or both of these. In the notation of (3.109) the probability of this event can be written as the expected value of

$$\sum_{i=1}^N (m_i + m_{N+i})(m_i + m_{N+i} - 1) / \{2N(2N - 1)\}. \quad (3.114)$$

It is not hard to see this leads to

$$N_e^{(id)} = \frac{4N - 2}{\sigma_d^2 + 2}, \quad (3.115)$$

where σ_d^2 is the variance of the number of offspring genes from each (diploid) individual. It is therefore necessary to extend the definition of a Cannings model to the diploid case. We define a diploid Cannings model as one for which the concept of exchangeability given in Section 3.2 relates to monoecious diploid individuals. We also assume that the gene transmitted by any individual to any offspring is equally likely to be each of the two genes in that individual, is independent of the gene(s) transmitted by this individual to any other offspring, and is also independent of the genes transmitted by any other individual. With these conventions it can be shown that

$$\sigma^2 = \frac{\sigma_d^2 + 2}{4}, \quad (3.116)$$

where σ^2 is the Cannings model gene ‘‘offspring number’’ variance, and from this it follows that the expressions in (3.110) and (3.115) are identical.

We turn next to the second class of models where a definition of effective population size is useful, namely those Wright–Fisher models which attempt to incorporate biological complexity more than does the simple Wright–Fisher model (1.48).

The first model considered allows for the existence of two sexes. Suppose in any generation there are N_1 diploid males and N_2 diploid females, with $N_1 + N_2 = N$. The model assumes that the genetic make-up of each individual in the daughter generation is found by drawing one gene at random, with replacement, from the male pool of genes, and similarly one gene with

replacement from the female pool. If $X_1(t)$ represents the number of A_1 genes among males in generation t and $X_2(t)$ the corresponding number among females, then $X_1(t+1)$ can be represented in the form

$$X_1(t+1) = i(t+1) + j(t+1), \quad (3.117)$$

where $i(t+1)$ has a binomial distribution with parameter $X_1(t)/(2N_1)$ and index N_1 , and $j(t+1)$ has a binomial distribution with parameter $X_2(t)/(2N_2)$ and index N_1 . A similar remark applies to $X_2(t+1)$, where now the index is N_2 rather than N_1 . Evidently the pair $\{X_1(t), X_2(t)\}$ is Markovian, and there will exist a transition matrix whose leading nonunit eigenvalue we require to find so that we can calculate $N_e^{(e)}$.

To do this we use the theory of Appendix A. It is necessary to find some function $Y(X_1, X_2)$ which is zero in the absorbing states of the system, positive otherwise, and for which

$$E\{Y\{X_1(t+1), X_2(t+1)\} \mid X_1(t), X_2(t)\} = \lambda Y(t) \quad (3.118)$$

for some constant λ . Such a function always exists, but some trial and error is usually necessary to find it. In the present case it is found, after much labor, that a suitable function is

$$Y(X_1, X_2) = \frac{1}{2}C\{X_1(2N_1 - X_1)(2N_1)^{-2} + X_2(2N_2 - X_2)(2N_2)^{-2}\} \\ + \{1 - (X_1 - N_1)(X_2 - N_2)N_1^{-1}N_2^{-1}\}, \quad (3.119)$$

where

$$C = \frac{1}{2}\{1 + (1 - 2N_1^{-1} - 2N_2^{-1})^{1/2}\}.$$

With this definition the eigenvalue λ becomes

$$\lambda = \frac{1}{2}[1 - (4N_1)^{-1} - (4N_2)^{-1} + \{1 + N^2(4N_1N_2)^{-2}\}^{1/2}], \quad (3.120)$$

or approximately

$$\lambda \approx 1 - (N_1 + N_2)(8N_1N_2)^{-1}. \quad (3.121)$$

From this result and (3.104) it follows that to a close approximation,

$$N_e^{(e)} = 4N_1N_2N^{-1}. \quad (3.122)$$

If $N_1 = N_2 (= \frac{1}{2}N)$, then $N_e^{(e)} \approx N$, as we might expect, while if N_1 is very small and N_2 is large, $N_e^{(e)} \approx 4N_1$. This latter value is sometimes of use in certain animal breeding programs.

The inbreeding population size is found much more readily. Two genes taken at random in any generation will have identical parent genes if both are descended from the same "male" gene or both from the same "female" gene. The probability of identical parentage is thus

$$\pi_2 = \frac{1}{2} \frac{N-1}{2N-1} \{(2N_1)^{-1} + (2N_2)^{-1}\},$$

and from this it follows that

$$N_e^{(i)} = (2\pi_2)^{-1} \approx 4N_1N_2N^{-1}. \quad (3.123)$$

The variance effective population size cannot be found so readily, and indeed strictly it is impossible to use (3.106) to find such a quantity, since an equation of this form does not exist in the two-sex case we consider. The fraction of A_1 genes is not a Markovian variable and in particular, using the notation of (3.106), the variance of $x(t+1)$ cannot be given in terms of $x(t)$ alone. This indicates a real deficiency in this mode of definition of effective population size. On the other hand, we shall see in the next chapter that sometimes a “quasi-Markovian” variable exists in terms of which a generalized expression for the variance effective population size may be defined. In the present case the weighted fraction of A_1 genes, defined as

$$x(t) = X_1(t)/(4N_1) + X_2(t)/(4N_2)$$

has the required quasi-Markovian properties, and

$$\text{var}\{x(t+1) | x(t)\} = x(t)\{1-x(t)\}N(8N_1N_2)^{-1} + O(N_1^{-2}, N_2^{-2}).$$

From this a generalized variance effective population size may be defined, in conjunction with (3.106), as

$$N_e^{(v)} = 4N_1N_2N^{-1}. \quad (3.124)$$

Thus for this model, $N_e^{(e)} \approx N_e^{(i)} \approx N_e^{(v)}$, although strict equality does not hold for any of these relations.

We return now to the case of a monoecious population and consider complications due to geographical structure. A simplified model for this situation which, despite its obvious biological unreality, is useful in revealing the effect of population subdivision, has been given by Moran (1962). It is supposed that the total population, of size $N(H+1)$, is subdivided into $H+1$ sub-populations each of size N , and that in each generation K genes chosen at random migrate from subpopulation i to subpopulation j for all i, j ($i \neq j$). Suppose that in subpopulation i there are $X_i(t)$ A_1 genes in generation t . There is no single Markovian variable describing the behavior of the total population, but the quantities $X_i(t)$ are jointly Markovian, and to find $N_e^{(e)}$ it is necessary to find some function $Y(t) = Y\{X_1(t), \dots, X_{H+1}(t)\}$ obeying the requirements of Appendix A. It is found, after some trial and error, that a suitable function $Y(t)$ is

$$\begin{aligned} Y(t) = & [A - D + \{(A - D)^2 + 4BC\}^{1/2}] \sum_i X_i(t) \{2N - X_i(t)\} \\ & + 2B \sum_{i \neq j} X_i(t) \{2N - X_j(t)\}, \end{aligned} \quad (3.125)$$

where

$$\begin{aligned} A &= (4N^2 + H^2K^2 + K^2H - 2N - 4NKH)/4N^2, \\ B &= (4KN - K^2H - K^2)/(4N^2), \\ C &= (4HKN - K^2H^2 - K^2H)/(4N^2), \\ D &= (4N^2 + HK^2 + K^2 - 4HK)/(4N^2). \end{aligned}$$

With this definition of $Y(t)$, the eigenvalue λ satisfying

$$E\{Y(t+1) \mid X_1(t), \dots, X_{H+1}(t)\} = \lambda Y(t)$$

is

$$\lambda = \frac{1}{2}(A + D + \{(A - D)^2 + 4BC\}^{1/2}). \quad (3.126)$$

If small-order terms are ignored, this yields eventually

$$N_e^{(e)} \approx N(H+1)\{1 + (2K(H+1))\}^{-1} \quad (3.127)$$

for large H and K . This equation is in fact accurate to within 10% even for $H = K = 1$, and it thus reveals that population subdivision leads to only a slight increase in the eigenvalue effective population size compared to the value $N(H+1)$ obtaining with no subdivision.

The inbreeding effective population size $N_e^{(i)}$ can be found most efficiently by noting that it is independent of K , since the act of migration is irrelevant to the computation of its numerical value. Thus immediately from (3.110)

$$N_e^{(i)} = \{N(H+1) - \frac{1}{2}\}/\{1 - (2N)^{-1}\}, \quad (3.128)$$

since each gene produces a number of offspring according to a binomial distribution with index $2N$ and parameter $(2N)^{-1}$. This value clearly differs only trivially from the true population size $N(H+1)$ and, for small H and K , it differs slightly from $N_e^{(e)}$.

Because of these two results, one may be tempted to ignore geographical sub-division in modeling evolutionary population genetic processes.

The computation of $N_e^{(v)}$ is beset with substantial difficulties since there exists no scalar Markovian variable for the model. Indeed, unless migration rates are of a large order of magnitude, there is not even a "quasi-Markovian" variable. Because of this no satisfactory value for $N_e^{(v)}$ has yet been put forward for the geographical structure case.

We consider finally a population whose size assumes cyclically the sequence of values $N_1, N_2, N_3, \dots, N_k, N_1, N_2, \dots$. There is no unique value of $N_e^{(e)}$, $N_e^{(i)}$ or $N_e^{(v)}$ in this case, and it is convenient to extend our previous definition to cover k consecutive generations of the process. If the population size in generation $t+k$ is N_i , it is easy to see that if $X(t)$ is the number of A_1 genes in generation t , and in each generation reproduction

occurs according to the model (1.48),

$$E[X(t+k)\{2N_i - X(t+k)\} | X(t)] = X(t)\{2N_i - X(t)\} \prod_{i=1}^k \{1 - (2N_i)^{-1}\}.$$

Defining now $N_e^{(e)}$ by the equation

$$\{1 - (2N_e^{(e)})^{-1}\}^k = \prod_{i=1}^k \{1 - (2N_i)^{-1}\},$$

it is clear that if k is small and the N_i large,

$$N_e^{(e)} \approx k\{N_1^{-1} + \dots + N_k^{-1}\}^{-1}. \quad (3.129)$$

Thus the eigenvalue effective population size is effectively the harmonic mean of the various population sizes taken during the k -generation cycle. A parallel formula holds for $N_e^{(i)}$, although here it is easier to work through the probability $Q(t+k)$ that two genes in generation $t+k$ do *not* have the same ancestor in generation t . Clearly

$$Q(t+k) = \{1 - (2N_{i-1})^{-1}\}Q(t+k-1),$$

and iteration over k generations gives

$$Q(t+k) = \prod_{i=1}^k \{1 - (2N_i)^{-1}\}Q(t).$$

Elementary calculations now show that $N_e^{(i)}$ is also essentially equal to the harmonic mean of the various population sizes. Again, if $x(t)$ is the fraction of A_1 genes in generation t ,

$$\text{var}\{x(t+k) | x(t)\} = \frac{1}{2}k\{N_1^{-1} + N_2^{-1} + \dots + N_k^{-1}\}x(t)\{1-x(t)\} + O(N_i^{-2}).$$

This shows that to a suitable approximation, $N_e^{(v)}$ is also the harmonic mean of the various population sizes.

This conclusion has been generalized by Karlin (1968). Karlin assumed that in any generation the population size takes one or other of the values $N_1, N_2, N_3, \dots, N_m$ according to Markov chain rules, so that there exists a probability q_{ij} of a transition from a population of size N_i to a population of size N_j . The cyclic case just considered arises if $q_{i,i+1} = 1$ for $i = 1, 2, \dots, m-1$, $q_{m1} = 1$. The leading nonunit eigenvalue, and hence $N_e^{(e)}$, depends on the transition matrix $\{q_{ij}\}$ as well as the particular form for $f(s)$ assumed in (3.41). Explicit effective population size values are hard to achieve in general, but in all cases for which expressions can be found, $N_e^{(e)}$ is close to the weighted harmonic mean of the possible population sizes. Thus if $\{f_k\}$ is assumed to be a Poisson distribution, leading to a generalized binomial transition probability extending (1.48), and if $q_{ij} = q_j$

for all i , the leading nonunit eigenvalue is

$$1 - \sum q_j (2N_j)^{-1},$$

so that

$$N_e^{(e)} = \left[\sum q_j N_j^{-1} \right]^{-1}.$$

So far we have ignored the effects of age structure. The definition and calculation of the variance effective population size in the age-structure case is given by Pollak (2000).

We conclude with some general observations. First, we have only considered one source of complexity at a time in computing effective population sizes. While the theory is no doubt very complex, it is reasonable to hope that the effective population size in, for example, a geographically subdivided population admitting two sexes would be given by a natural composition of the effective sizes for the subdivision and the two-sex cases respectively.

Second, many papers and several textbooks use what appears to be a definition of $N_e^{(i)}$ defined by the outcome of a given experiment or by a given field observation. Thus, for example, in the (diploid) formula (3.115) the symbol σ_d^2 is sometimes replaced by V , where V is defined by

$$V = \sum (n_i - 2)^2 / N,$$

where n_i is the (random) number of genes produced by the i th individual. While such a definition might be of use in a retrospective analysis of a given experiment, it is not allowable for theoretical purposes since V is a random variable and thus can take quite different values for two different populations that have identical properties and thus must have the same value of $N_e^{(i)}$.

Third, all three effective population sizes suffer some defects. Thus, $N_e^{(e)}$ and $N_e^{(v)}$ are defined assuming two alleles at the locus in question, $N_e^{(i)}$ is not defined in terms of allelic type and is thus possibly superior to $N_e^{(e)}$ and $N_e^{(v)}$ as a pure measure of population structure although, as we note in a moment, each expression has its special interpretation and usefulness. Further $N_e^{(i)}$ is not of much value in characterizing various properties of the geographically structured case.

Fourth, although the three effective sizes are often nearly equal in the examples considered above, they can in other cases differ substantially. This occurs particularly in populations with nonconstant size. An extreme example arises when a single heterozygote in generation t gives rise to a very large number of offspring in generation $t + 1$. Here both $N_e^{(v)}$ and $N_e^{(e)}$ are very large, but $N_e^{(i)}$ is unity. Thus $N_e^{(v)}$ and $N_e^{(e)}$ tend to be defined in terms of the future evolution of the population, whereas $N_e^{(i)}$ is concerned

more with its past. In some circumstances one effective size is of most interest and for other circumstances, another.

Fifth, we have not defined an effective population size for continuous-time models. There is no reason to believe that these differ significantly from those given above. Some specific formulae are given by Felsenstein (1971), Hill (1972) and Kimura and Crow (1972). Nor have we considered the complications that can occur when fertility parameters are inherited (see, for example, Nei, (1966)), or in a variety of other situations.

Sixth, problems arise with the definition of an effective population size in cases such as the human population when the population size has steadily increased. None of the definitions of the effective population size given above handles this situation in a satisfactory way. Given the current focus on the evolution of the human population, this is particularly unfortunate.

Seventh, we recall what is perhaps our main motive in defining an effective population size, namely to consider whether various complex population structures can lead to a significantly increased importance for random drift compared to its importance in the model (1.48). Our conclusions show that this occurs when there is extremely large variance in offspring number, when the population size is cyclic and the smallest size the population assumes during the cycle is very small and when, in a dioecious population, the number of breeding individuals in one sex is very small. In all other cases, particularly in the case of geographically subdivided populations, there appears to be little significant scope for random drift beyond that applying for the model (1.48). We discuss the consequences of some of these observations later in this book.

Finally, the concept of the effective population size is widely misused in the literature, especially in areas outside of, but associated with, evolutionary genetics. In particular, its connection to the simple Wright–Fisher model (1.48) seems to be widely unknown. The effective population size of some population is no more than the size of a simple Wright–Fisher model population having some characteristic in common with the population of interest. The effective sizes for two different characteristics (for example, variance, inbreeding) might very well differ, so that the purpose for using the effective population size concept is relevant. The value of the concept is that calculations applying for the simple Wright–Fisher model, especially diffusion theory calculations, can sometimes, and for some specific purpose, be used to provide results for the population of interest, replacing N wherever it appears by the appropriate N_e . (Even this claim is essentially a heuristic one, and the theory for using Wright–Fisher model formulas for non-Wright–Fisher models, with this substitution, is not well developed.) In any event, since the model (1.48) provides at best a rough approximation to reality, the implication that however it is calculated, the effective size bears any necessary similarity to an actual population size is without any foundation.

3.8 Frequency-Dependent Selection

There is no requirement that the fitness values w_{ij} in the model defined by (3.16) and (3.29) should be fixed constants, although so far we have assumed that they are. The analysis we have carried out, in particular the derivation of (3.30), continues to be valid when the w_{ij} , (or s and h in (3.30)) are functions of x . This can lead to some interesting consequences. Thus for small t the fitness scheme

$$w_{11} = 1 + t\left(1 - \frac{1}{2}x\right), \quad x_{12} = 1, \quad w_{22} = 1 + \frac{1}{2}tx,$$

is equivalent to (1.25b) if we put $h = \frac{1}{2}x/(1-x)$. Using this value in (3.30) gives $\pi(x) = x$. Thus survival probabilities for this frequency-dependent fitness scheme are the same as those obtaining when there are no fitness differentials.

3.9 Two Loci

In this section we consider different two two-locus Markov chain analogues of the one-locus model (1.48). While a good deal of progress on the problems considered in this section is possible using diffusion theory (Ohta and Kimura (1969a, b), Littler (1973)), we defer consideration of a diffusion analysis to Section 6.6, and consider here only results found from Markov chain theory.

The first of the two-locus models that we consider is the “random union of zygotes” model of Kimura (1963), Watterson (1970), Serant and Villard (1972) and Littler (1973), and the second is the “random union of gametes” model of Karlin and McGregor (1968) and Hill and Robertson (1966, 1968). For convenience we call these here the RUZ and RUG models, respectively. A general theory of Weir and Cockerham (1974) yields many results for both models. We follow as far as possible the notation of Section 2.10 in discussing these models.

The RUZ model is defined as follows. Suppose a population of fixed size N contains, in generation t , $X_{ij}(t)$ individuals whose genotype is made up of gametic types i and j ($i \leq j$) (for a definition of a gamete of type i , see Section 2.10). Let α_{ijk} be the probability that a gamete produced by such an individual is of type k . When $(i, j) = (1, 4)$ or $(2, 3)$, these probabilities will involve the recombination fraction R . Then

$$c_k(t) = \sum_{i \leq j} X_{ij}(t) \alpha_{ijk} N^{-1}$$

is the probability that a gamete chosen at random forming generation $t+1$ is of type k . It is now assumed that the N individuals in generation $t+1$ have their gametes determined by $2N$ independent trials, with the probability of a gamete of type k on any trial being $c_k(t)$.

The values $X_{ij}(t+1)$ are thus determined from the $X_{ij}(t)$ only through the quantities $c_k(t)$. It follows that the random vector (c_1, c_2, c_3, c_4) evolves as a Markov chain. The transition matrix of this chain can be found from that of the X_{ij} , and is perhaps best written down in terms of the joint moment-generating function

$$E(e^{\sum \theta_j c_j(t+1)} | c_i(t)) = \left(\sum_{i=1}^4 \sum_{j=1}^4 c_i(t) c_j(t) e^{\sum_{k=1}^4 \alpha_{ijk} \theta_k / N} \right)^N. \quad (3.130)$$

This equation was given by Watterson (1970) and requires the biologically reasonable definition $\alpha_{ijk} = \alpha_{jik}$.

Before examining the consequences of (3.130) we introduce the RUG model. Here we ignore the zygote stage in our formation of the model and simply assume, following (2.84), that if generation t produces $n_i(t)$ gametes of type i ($i = 1, \dots, 4$), the probability that generation $t+1$ produces $n_i(t+1)$ gametes of type i ($i = 1, \dots, 4$) is

$$\frac{(2N)!}{\prod_{i=1}^4 n_i(t+1)!} \prod_{i=1}^4 \psi_i^{n_i(t+1)}, \quad (3.131)$$

where

$$\psi_i = c_i(t) + \eta_i R \{c_1(t)c_4(t) - c_2(t)c_3(t)\}, \quad c_i(t) = n_i(t)/(2N),$$

and the η_i have been defined in (2.84). The models defined by (3.130) and (3.131) are not equivalent in general, but are so in the limiting case $R = 0$.

The qualitative behaviors of both models are identical. All four gametes will segregate in the population, with the possibility that one gamete is temporarily absent not excluded, until after a random time whose distribution is determined by R , N and the initial gamete frequencies, one or other allele is lost from the population. From this time on segregation continues at one locus only and the one-locus model (1.48) applies. Eventually one or other allele at this locus is lost, and the population consists entirely of one of the four gametic types. Clearly questions of particular interest concern the time that all four gametes exist in the population, the probability that a nominated gamete is eventually fixed and, because linkage disequilibrium is of major interest in two-locus systems, the transient behavior of the coefficient of linkage disequilibrium, defined in generation t by

$$D(t) = c_1(t)c_4(t) - c_2(t)c_3(t). \quad (3.132)$$

We now consider three quantities in order to obtain further information about the properties of both models. The first is the eigenvalue μ for which, for large t ,

$$\text{Prob}(\text{segregation continues at both loci}) \sim c\mu^t. \quad (3.133)$$

The second is the probability of ultimate fixation of gamete type i , and the third is the mean $E\{D(t)\}$ and mean square $E\{D(t)\}^2$ of the coefficient of linkage disequilibrium at time t .

To find the eigenvalue μ defined by (3.132) we follow the approach of Watterson (1970, 1972). We consider the variable $D(t)$, defined by (3.132), as well as the variables $S(t)$ and $Z(t)$, defined by

$$S(t) = c_1(t)c_4(t) + c_2(t)c_3(t),$$

$$Z(t) = \{c_1(t) + c_2(t)\}\{c_1(t) + c_3(t)\}\{c_2(t) + c_4(t)\}\{c_3(t) + c_4(t)\}.$$

It is then found that conditional on $c_1(t)$, $c_2(t)$, $c_3(t)$, $c_4(t)$,

$$E\{S(t+1)\} = a_{11}S(t) + a_{12}\{D(t)\}^2 + a_{13}Z(t),$$

$$E\{D(t+1)\}^2 = a_{21}S(t) + a_{22}\{D(t)\}^2 + a_{23}Z(t),$$

$$E\{Z(t+1)\} = a_{31}S(t) + a_{23}\{D(t)\}^2 + a_{23}Z(t).$$

Here the a_{ij} are constants whose values depend only on N and R . It is clear that given the initial values $c_i(0)$,

$$E \begin{pmatrix} S(t) \\ \{D(t)\}^2 \\ Z(t) \end{pmatrix} = A^t \begin{pmatrix} S(0) \\ \{D(0)\}^2 \\ Z(0) \end{pmatrix}, \quad (3.134)$$

where $A = \{a_{ij}\}$. Since $S(t)$ is positive if and only if all four alleles continue to segregate in generation t , a generalization of the method of Appendix A shows that the leading nonunit eigenvalue of the transition matrix implied by (3.134) is the leading eigenvalue of A . This may be calculated either algebraically or numerically. Watterson found that the largest eigenvalue μ can be written in the form

$$\mu = 1 - (2N)^{-1} - y(2N)^{-1},$$

where y is the solution of a certain cubic equation (see Watterson (1970, 1972b), Littler (1973)).

The most useful discussion of the properties of μ is through numerical examples, but some limiting cases are of special interest. Thus $R = 0$ implies $y = 0$ and hence

$$\mu = 1 - (2N)^{-1}.$$

This is to be expected since for $R = 0$ the model is equivalent to a one-locus model with four alleles for which this eigenvalue has already been established above. Perhaps of more importance for consideration of properties of two-locus systems is to consider the behavior of μ for R moderate and N large. Here Watterson (1970, 1972) found that

$$\mu = \{1 - (2N)^{-1}\}^2 + O(N^{-3}). \quad (3.135)$$

The value of μ is very close to $\{1 - (2N)^{-1}\}^2$ when R is not extremely small and N greater than about 50. Some numerical values showing this, taken

Table 3.1. Values of the eigenvalue μ (see text for definition) in both RUG and RUZ models

		R			
		0.01	0.10	0.20	0.50
N	10	0.941	0.910	0.905	0.903
	25	0.972	0.961	0.961	0.960
	50	0.984	0.980	0.980	0.980

from Littler (1973), are given in Table 3.1. The reason why the eigenvalue μ takes the form shown in (3.135) is obvious enough. When N and R are not both small the two loci behave almost independently, so that the probability that segregation continues at both loci is close to the square of the probability that segregation continues at any one locus. As $R \rightarrow 0$ the two segregation behaviors become more dependent.

A parallel evaluation of μ for the RUG model has been made by Littler (1973). Here Littler sets up equations of the form

$$\begin{aligned} E\{[D(t+1)]^2 \mid c_i(t)\} &= b_{11}\{D(t)\}^2 + b_{12}I(t) + b_{13}Z(t), \\ E[I(t+1) \mid c_i(t)] &= b_{21}\{D(t)\}^2 + b_{22}I(t) + b_{23}Z(t), \\ E[Z(t+1) \mid c_i(t)] &= b_{31}\{D(t)\}^2 + b_{32}I(t) + b_{33}Z(t). \end{aligned}$$

where

$$I(t) = \{c_1(t)c_4(t) - c_2(t)c_3(t)\}\{1 - 2c_1(t) - 2c_2(t)\}\{1 - 2c_1 - 2c_3(t)\},$$

$Z(t)$ is as defined above and the b_{ij} are constants depending only on N and R . This gives, conditional on the values $c_i(0)$,

$$E \begin{pmatrix} \{D(t)\}^2 \\ I(t) \\ Z(t) \end{pmatrix} = B^t \begin{pmatrix} \{D(0)\}^2 \\ I(0) \\ Z(0) \end{pmatrix}, \tag{3.136}$$

where $B = \{b_{ij}\}$. This is analogous to (3.134), and the leading eigenvalue μ for this model is the leading eigenvalue of the matrix B .

As for the RUZ model, μ decreases as a function of R from $1 - (2N)^{-1}$ at $R = 0$, but is always greater than $\{1 - (2N)^{-1}\}^2$ for $R \leq 0.5$ (Karlin and McGregor (1968)). For the combinations of N and R values listed in Table 3.1, the numerical values given apply to the order of accuracy shown for the RUG model also, although the formulas for the eigenvalues in the two stochastic models are different. For values of N and R not listed the agreement between the two is not quite so close. Nevertheless, the general discussion given for the values of μ in the RUZ model also apply here for the RUG model.

We now turn to probabilities of fixation for the various gametes. These were found by Kimura (1963) for the RUZ model and by Karlin and McGregor (1968) for the RUG model. Suppose we can find func-

tions $\phi_i\{c_1(t), c_2(t), c_3(t), c_4(t)\}$, which we abbreviate to $\phi_i(t)$, having the property that

$$E\{\phi_i(t+1) \mid c_1(t), c_2(t), c_3(t), c_4(t)\} = \phi_i(t), \quad i = 1, \dots, 4, \quad (3.137)$$

where $\phi_i(\infty) = 1$ if gamete i eventually fixes, $\phi_i(\infty) = 0$ otherwise. Then by iteration in (3.137),

$$E\{\phi_i(\infty) \mid c_1(0), c_2(0), c_3(0), c_4(0)\} = \phi_i(0).$$

The left-hand side is the probability that gamete i fixes. We conclude that if we can find functions $\phi_i(t)$ satisfying (3.137), gamete fixation probabilities are given by the values of $\phi_i(0)$. Functions $\phi_i(t)$ satisfying (3.137) always exist, and can usually be found after some trial and error.

One complication arises with this procedure. The RUZ model concerns zygotes rather than gametes and the initial composition of the population then relates to zygote frequencies rather than gamete frequencies. Nevertheless the essence of the above procedure still applies. For the RUZ model it is found that functions $\phi_i(t)$ satisfying (3.137) are

$$\phi_i(t) = c_i(t) + \frac{\eta_i 2NR D(t)}{2NR + 1}$$

while for the RUG model,

$$\phi_i(t) = c_i(t) + \frac{\eta_i 2NR D(t)}{2NR + 1 - R}.$$

It follows immediately for the RUG model that the probability of fixation of gametes of type i is

$$c_i(0) + \frac{\eta_i 2NR D(0)}{2NR + 1 - R}. \quad (3.138)$$

Matters are slightly more complicated for the RUZ model since we must give probabilities in terms of initial zygotic frequencies. It is found (Watterson (1970)) that the required value is

$$c_i^* + \frac{\eta_i 2NR D^* + \Delta (2N)^{-1}}{2NR + 1},$$

where c_i^* is the frequency of the gamete i among the zygotes of the initial generation, $D^* = c_1^* c_4^* - c_2^* c_3^*$ and $\Delta = \{X_{14}(0) - X_{23}(0)\}(2N)^{-1}$.

In both cases, whenever R is fixed and moderate and N is large, the fixation probability in gamete i is approximately $c_i + \eta_i D$, where c_i stands for c_i^* in the RUZ model and $c_i(0)$ in the RUG model. But this is just the initial value of the product of the frequencies of the two alleles making up the i th gamete. Thus, to a close approximation, the probability in this case that any gamete fixes is simply the product of the probabilities that the two corresponding alleles fix. This arises because the segregation processes at the two loci are effectively independent. When R is small this is no longer

true, and the association between the two loci must be taken into account when computing fixation probabilities.

We turn finally to the behavior of the linkage disequilibrium function $D(t)$. A considerable part of two-locus theory relates to this quantity, so it is important to discuss its behavior in detail for the two models under consideration.

Consider first the model (3.130). It is easy to show for this model that

$$E\{D(t+1) \mid D(t)\} = \{1 - (2N)^{-1} - R\}D(t), \quad (3.139)$$

and hence the mean value of $D(t)$ decreases to zero geometrically fast with t at rate $1 - (2N)^{-1} - R$. Unless R is close to zero this is quite a rapid rate. A parallel remark applies for the model (3.131), where we find

$$E\{D(t+1) \mid D(t)\} = \{1 - (2N)^{-1}\}\{1 - R\}D(t).$$

The convergence to zero of $E\{D(t)\}$ is only slightly slower for this model than that for the model in (3.130).

More detailed information about the behavior of $D(t)$ will depend on knowledge of the variance of $D(t)$, or equivalently, since the above expressions easily yield the mean of $D(t)$, on the expected mean square $E\{D(t)\}^2$. Suppose that the initial value of D is zero. Since $D = 0$ once fixation of one or other allele occurs, we might expect that in this case the variance of D will increase from zero as t increases and, after achieving a maximum for intermediate values of t , decrease again to zero. If initially D is nonzero we might perhaps expect the variance of D monotonically to decrease to zero.

Although the behavior of the variance of D is by no means simple, these expectations are in essence confirmed for the RUG model by Hill and Robertson (1968) and for both models by Littler (1973). Equations (3.134) and (3.136) show that $E\{D(t)\}^2$ can be written in the form

$$E\{D(t)\}^2 = \alpha_1\mu_1^t + \alpha_2\theta_1^t + \alpha_3\theta_2^t,$$

for the RUZ model and

$$E\{D(t)\}^2 = \beta_1\mu_2^t + \beta_2\rho_1^t + \beta_3\rho_2^t,$$

for the RUG model. Here the α_i , β_i and θ_i are constants depending on N and R and μ_1 , μ_2 are the leading eigenvalues of A and B , respectively. While for large t the behavior of $E\{D(t)\}^2$ in both cases is determined largely by the maximum eigenvalue μ , the behavior for small t is quite complicated and all eigenvalues and eigenvectors are needed to describe it. To gain more information about the transient behavior of $D(t)$, Littler (1973) investigated the behavior of $E(D(t))^2$ as a function of t . When D is initially zero, the variance of D increases to a maximum value of order 0.01 and then decreases. The maximum is reached sooner and is slightly greater for small values of N . For large t , the variance of D is smallest for small values of N . When D is not zero initially, $E\{D(t)\}^2$ decreases with t , and

for large t is minimized for small values of N . The eigenvalue μ determining the ultimate behavior of $E\{D(t)\}^2$ is much closer to unity than the value $1 - (2N)^{-1} - R$ implied by (3.139) for the ultimate behavior of $E\{D(t)\}$. Because of this it has sometimes been asserted that observed values of D can differ significantly from the mean. In view of the above results this conclusion may not be drawn, and it is clear that the eigenvalues on their own do not give a complete picture of the true behavior of $D(t)$.

In the RUG model, $E\{D(t)\}^2$ can be found by a procedure similar to that used in the RUZ model. Numerical examples have been given by Hill and Robertson (1968). They found that if all gametes have initial frequency 0.25 and $NR = 1$, $\text{var}\{D(t)\}$ reaches a maximum of 0.006 at about N generations, while when $NR = 4$ a maximum of 0.003 is reached after about N generations. In general the larger NR , the smaller the maximum and the sooner it is reached. For large populations and unlinked loci, $\text{var}\{D(t)\}$ is always extremely small, indicating that by random effects only, it is unlikely that D will ever assume a large value.

All the above has assumed that there is no mutation, so that eventual fixation of one or other gamete is certain. If mutation exists at positive rate from A_i to A_j ($i \neq j$) and B_i to B_j ($i \neq j$) there will exist a stationary distribution of gamete frequencies and thus a stationary distribution of D . Since we are interested in the extent of likely variation of D from zero, we consider now the stationary mean value of D^2 in this mutation case, following the analysis of Ohta and Kimura (1969b).

Suppose that the mutation rates are u_1 (from A_1 to A_2), v_1 (from A_2 to A_1), u_2 (from B_1 to B_2) and v_2 (from B_2 to B_1). We consider the three quantities $Z(t)$, $I(t)$ and $\{D(t)\}^2$ introduced above. Ohta and Kimura set up a recurrence relation similar to that given above for the case without mutation, where the coefficients b_{ij} now include mutation terms. By letting $t \rightarrow \infty$, nondegenerate limits are found for the expectations of these quantities and in particular for $E\{D(\infty)\}^2$. We defer giving an explicit formula here since a slightly simpler formula will be given in Section 6.6, based upon diffusion theory.

4

Diffusion Theory

4.1 Introduction

In the previous chapter we encountered some difficulty in deriving explicit formulas for several quantities of evolutionary interest, particularly when the population behavior was described by the Wright–Fisher model (1.48) or any of its generalizations. Even for models such as (3.30), where explicit formulas can often be found, the effects of the genetic parameters are sometimes obscured by the complexities of the expressions that arise. For both these reasons, it would be most useful to us if we could find approximate formulae for these quantities by reasonably accurate expressions which are not only comparatively simple, but which also display explicitly the effects of the various genetic parameters involved. Fortunately there exists a general approach which very often does all this for us, namely in approximating the discrete process by a continuous-time continuous-space diffusion process.

A substantial and mathematically deep theory of diffusion processes exists. We outline those aspects of this theory that are of use to us in Section 4.7. Our approach to diffusion processes does not, however, proceed through this theory, being often rather intuitive and avoiding theoretical niceties. This is in part because for us the fundamental process is always a discrete one, usually a finite Markov chain, for which some of these niceties are irrelevant, and in part because the mathematical depth of formal diffusion theory is inappropriate to the level of this book. We shall in particular

assume without question the existence of a unique diffusion process having certain properties that we require.

Diffusion theory has a long and honorable place in population genetics theory, going back to Fisher (1922). In this chapter we consider the elements of the theory divorced from specific genetical applications, and in Chapter 5 the theory developed here will be applied to a variety of genetical models.

4.2 The Forward and Backward Kolmogorov Equations

We consider a discrete Markov chain with state space $\{0, 1, 2, \dots, M\}$, transition matrix $P = \{p_{ij}\}$ and initial value k for the random variable whose properties are described by this Markov chain: This notation will be used throughout this chapter. For convenience we write $p_{ki}^{(t)}$ as $f(i; k, t)$, so that

$$f(j; k, t + 1) = \sum_i f(i; k, t)p_{ij}. \quad (4.1)$$

We re-scale the space axis by a factor M^{-1} and consider the new variables

$$x = iM^{-1}, \quad x + \delta x = jM^{-1}, \quad (4.2)$$

and write $p = kM^{-1}$. In all applications of interest to us, $E(\delta x|x) = O(M^{-\gamma})$ and $\text{var}(\delta x|x) = O(M^{-\gamma})$, where $\gamma = 1$ or 2 ; now change the time scale so that possible changes in the random variable can occur at time points $\delta t, 2\delta t, 3\delta t, \dots$, where $\delta t = M^{-\gamma}$. The re-scaled process is of course essentially identical to the original process and in particular is still a discrete process. Nevertheless we feel that as $M \rightarrow \infty$ the process converges in some way to a continuous-time continuous-space diffusion process, and our aim is to identify this diffusion process and to discover some of its properties.

Suppose that in the discrete process the moments of the change δx , given the current value x at time t , satisfy the equations

$$E(\delta x) = a(x)\delta t + o(\delta t), \quad (4.3)$$

$$\text{var}(\delta x) = b(x)\delta t + o(\delta t), \quad (4.4)$$

$$E(|\delta x|^3) = o(\delta t). \quad (4.5)$$

Here $a(x)$ and $b(x)$ are assumed to be functions of x but not of t . We write (4.1) in the form

$$f(x + \delta x; p, t + \delta t) = \int f(x; p, t)f(x + \delta x; x, \delta t) dx,$$

where here and below all integrals have terminals 0 and 1. We now formally expand on both sides as Taylor series in δt and δx . Using (4.3) - (4.5) and

retaining leading terms only, we eventually arrive at the equation

$$\frac{\partial f(x; t)}{\partial t} = -\frac{\partial}{\partial x}\{a(x)f(x; t)\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\{b(x)f(x; t)\}. \quad (4.6)$$

This is the forward Kolmogorov (Fokker–Planck or diffusion) equation and is of fundamental importance in the theory of population genetics. This formal procedure can be justified by the mathematical theory referred to briefly in Section 4.7.

Since small $\delta t \rightarrow 0$ corresponds to large M , we now assume that there exists a diffusion process on $[0, 1]$ that satisfies (4.3)–(4.5) and possesses a density function $f(x; t)$ which satisfies (4.6). We expect this process to approximate the original discrete process in the sense that for $0 < g < h < 1$,

$$\int_g^h f(x; t) dx \quad (4.7)$$

provides a good approximation to the probability that the original unscaled discrete random variable is between Mg and Mh at time $M\gamma t$.

In the procedure leading to (4.6), little mention was been made of the initial value p of the diffusion variable, and p does not appear explicitly in (4.6). However, the function $f(x; t)$ should be written more fully $f(x; p, t)$, since the solution of the equation depends on the value of p . There is, however, a second equation that makes a more explicit and indeed fundamental use of the value of p . If we consider instead of the time points $(0, t, t + \delta t)$ the new time points $(0, \delta t, t + \delta t)$, we arrive at the equation

$$f(x; p, t + \delta t) = \int g(\delta p; p)f(x; p + \delta p, t)d(\delta p). \quad (4.8)$$

Here δp is the change in the value of the random variable in the time interval $(0, \delta t)$ and $g(\delta p; p)$, its probability density function. Expanding the integrand as above and retaining leading terms, we arrive at the equation

$$\frac{\partial f(x; p, t)}{\partial t} = a(p)\frac{\partial f(x; p, t)}{\partial p} + \frac{1}{2}b(p)\frac{\partial^2 f(x; p, t)}{\partial p^2}. \quad (4.9)$$

This is the backward Kolmogorov equation, which for several purposes is more useful than the forward equation (4.6).

Some care must be exercised in the interpretation of (4.9). As stated above, the density function $f(x; p, t)$ depends on p , and all that is claimed is that as a function of p , this density function satisfies (4.9). The statement sometimes made that (4.9) implies a time reversal and that p is a random variable with x fixed is incorrect: The random variable in (4.9) is the current gene frequency x .

An explicit solution of (4.6), or of (4.9), can sometimes be achieved, as we see in the next chapter. The solution is usually of the eigenfunction

expansion form

$$f(x; p, t) = \sum_{i=1}^{\infty} g_i(x, p) \exp(-\lambda_i t), \quad (4.10)$$

where the λ_i ($0 \leq \lambda_1 < \lambda_2 < \lambda_3 \dots$) are eigenvalue constants and the $g_i(x, p)$, the associated eigenfunctions. This form of solution is clearly analogous to (2.138), a parallel we examine in more detail in particular cases. Remarkably, a considerable amount of information concerning the diffusion process (4.6) can be found without computing the explicit solution (4.10), as we now see.

4.3 Fixation Probabilities

In this and the next three sections we assume without question the existence of a diffusion process on $[0, 1]$ satisfying (4.3)–(4.5) and admitting a density function satisfying (4.6) and (4.9).

An equation parallel to (4.9) can be found by replacing $f(x; p, t)$ by $F(x; p, t)$ throughout, where

$$F(x; p, t) = \int_0^x f(y; p, t) dy, \quad (4.11)$$

so that

$$\frac{\partial F(x; p, t)}{\partial t} = a(p) \frac{\partial F(x; p, t)}{\partial p} + \frac{1}{2} b(p) \frac{\partial^2 F(x; p, t)}{\partial p^2}. \quad (4.12)$$

Suppose now that both $x = 0$ and $x = 1$ are absorbing states of the diffusion process. From (4.12) we arrive at the equation

$$\frac{\partial P_0(p; t)}{\partial t} = a(p) \frac{\partial P_0(p; t)}{\partial p} + \frac{1}{2} b(p) \frac{\partial^2 P_0(p; t)}{\partial p^2}, \quad (4.13)$$

where $P_0(p; t)$ is the probability that absorption has occurred at $x = 0$ at or before time t . The same equation holds for the probability $P_1(p; t)$ that absorption has occurred at $x = 1$ at or before time t . Although $P_0(p; t)$ and $P_1(p; t)$ obey the same equation, their values differ due to different boundary conditions. By letting $t \rightarrow \infty$, the probability $P_0(p)$ that absorption ever occurs at $x = 0$ satisfies the equation

$$0 = a(p) \frac{dP_0(p)}{dp} + \frac{1}{2} b(p) \frac{d^2 P_0(p)}{dp^2}. \quad (4.14)$$

Since $P_0(p)$ clearly satisfies the boundary conditions $P_0(0) = 1$, $P_0(1) = 0$, it is straightforward to solve (4.14) explicitly to get

$$P_0(p) = \int_p^1 \psi(y) dy / \int_0^1 \psi(y) dy, \quad (4.15)$$

where

$$\psi(y) = \exp(-2 \int_0^y \{a(z)/b(z)\} dz). \quad (4.16)$$

Similarly the probability $P_1(p)$ that absorption eventually occurs at $x = 1$ is found to be

$$P_1(p) = \int_0^p \psi(y) dy / \int_0^1 \psi(y) dy. \quad (4.17)$$

We have already found these formulas as approximations to the values in a finite Markov chain in (3.30) and (3.31), where a different notation was used, and without reference to diffusion processes. Although we have carried out a scaling of the time axis in passing from the original Markov chain to the diffusion process, there is no need to re-scale the values (4.15) and (4.17) when using them as approximations in the Markov chain. This is no longer true for questions concerning the time until absorption, as we now see.

4.4 Absorption Time Properties

We start by assuming that both $x = 0$ and $x = 1$ are absorbing barriers and consider the mean time until one or other boundary is reached in the diffusion process of interest. Equation (4.13) and the corresponding equation for $x = 1$ show that if $\phi(t; p)$ is the density function of the time t until absorption occurs, then $\phi(t; p)$ satisfies the equation

$$\frac{\partial \phi(t; p)}{\partial t} = a(p) \frac{\partial \phi(t; p)}{\partial p} + \frac{1}{2} b(p) \frac{\partial^2 \phi(t; p)}{\partial p^2}. \quad (4.18)$$

Then

$$\begin{aligned}
 -1 &= -\int_0^{\infty} \phi(t; p) dt \\
 &= -[t\phi(t; p)]_0^{\infty} + \int_0^{\infty} t \frac{\partial \phi}{\partial t} dt \\
 &= 0 + \int_0^{\infty} t \left\{ a(p) \frac{\partial \phi}{\partial p} + \frac{1}{2} b(p) \frac{\partial^2 \phi}{\partial p^2} \right\} dt
 \end{aligned}$$

so that

$$-1 = a(p) \frac{d\bar{t}(p)}{dp} + \frac{1}{2} b(p) \frac{d^2 \bar{t}(p)}{dp^2}, \quad (4.19)$$

providing an interchange in the order of integration and differentiation is justified, that the mean fixation time is finite, and that $t\phi(t; p) \rightarrow 0$ as $t \rightarrow \infty$. Here

$$\bar{t}(p) = \int_0^{\infty} t\phi(t; p) dt \quad (4.20)$$

is the mean time until one or other absorbing boundary is reached, given the initial frequency p . The solution of (4.19), subject to the boundary conditions $\bar{t}(0) = \bar{t}(1) = 0$, is best expressed in the form

$$\bar{t}(p) = \int_0^1 t(x; p) dx, \quad (4.21)$$

where

$$t(x; p) = 2P_0(p)[b(x)\psi(x)]^{-1} \int_0^x \psi(y) dy, \quad 0 \leq x \leq p, \quad (4.22)$$

$$t(x; p) = 2P_1(p)[b(x)\psi(x)]^{-1} \int_x^1 \psi(y) dy, \quad p \leq x \leq 1. \quad (4.23)$$

For the original Markov chain we approximate the mean absorption time by

$$M^{\gamma} \bar{t}(p). \quad (4.24)$$

The representation (4.21) suggests a more detailed examination of the function $t(x; p)$. This function has the interpretation that

$$\int_{x_1}^{x_2} t(x; p) dx \quad (4.25)$$

is the mean time in the diffusion process that the random variable spends in the interval (x_1, x_2) before absorption. Correspondingly, we approximate the mean number of times in the Markov chain that the discrete random variable takes the value $j (= Mx)$ before absorption by

$$\bar{t}_{k,j} \approx M^{\gamma-1} t(x; p). \quad (4.26)$$

The representation (4.25) allows further conclusions to be drawn. Let $g(x)$ be any well-behaved function of x and consider the integral $I_g(p)$ of this function over the time until absorption occurs. This integral is a random variable, since its value will depend on the actual path traced out by the diffusion variable, and its mean value from (4.25) is clearly

$$E(I_g(p)) = \int_0^1 g(x) t(x; p) dx. \quad (4.27)$$

In a similar way, if $\sum g(x)$ is the sum of the function $g(x)$ in the discrete process,

$$E(\sum g(x)) \approx M^\gamma \int_0^1 g(x) t(x; p) dx. \quad (4.28)$$

There is an alternative way of deriving (4.27) akin to the derivation of the backward (4.9). We note that the integral of $g(p)$ over $(0, \delta t)$ is approximately $g(p)\delta t$, so that writing $E[I_g(p)]$ as $\mu(p)$ for convenience,

$$\begin{aligned} \mu(p) &\approx g(p)\delta t + E \left[\int_0^1 g(x) t(x; p + \delta p) dx \right] \\ &\approx g(p)\delta t + \mu(p) + a(p)\delta t \frac{d\mu}{dp} + \frac{1}{2} b(p)\delta t \frac{d^2\mu}{dp^2} + o(\delta t). \end{aligned}$$

Dividing by δt and letting $\delta t \rightarrow 0$, we get

$$a(p) \frac{d\mu(p)}{dp} + \frac{1}{2} b(p) \frac{d^2\mu(p)}{dp^2} = -g(p). \quad (4.29)$$

This equation generalizes (4.19) and may be solved, subject to the boundary conditions $\mu(0) = \mu(1) = 0$, to derive (4.27).

It is possible to derive higher moments of the absorption time and more generally of $I_g(p)$. For the absorption time we have

$$\begin{aligned}
 -2\bar{t}(p) &= -2 \int_0^\infty t\phi(t; p) dt \\
 &= -[t^2\phi(t; p)]_0^\infty + \int_0^\infty t^2 \frac{\partial\phi}{\partial t} dt \\
 &= \int_0^\infty \left\{ a(p) \frac{\partial t^2\phi(t; p)}{\partial p} + \frac{1}{2}b(p) \frac{\partial^2 t^2\phi(t; p)}{\partial p^2} \right\} dt \\
 &= a(p) \frac{dS(p)}{dp} + \frac{1}{2}b(p) \frac{d^2S(p)}{dp^2}, \tag{4.30}
 \end{aligned}$$

where $S(p)$ is the second moment of the absorption time. In this procedure we have formally interchanged the order of integration and differentiation.

Equation (4.30) can be solved for $S(p)$, subject to the boundary conditions $S(0) = S(1) = 0$, and hence a formula for the variance of the absorption time can be found. Clearly this procedure can be generalized to find any moment of the absorption time, but the formulas become complicated and we present here only an expression for the variance $\sigma^2(p)$. This is

$$\begin{aligned}
 \sigma^2(p) &= 4 \left[P_1(p) \int_p^1 \psi(x) \int_x^1 \xi(y) dy dx \right. \\
 &\quad \left. - P_0(p) \int_0^p \psi(x) \int_x^p \xi(y) dy dx \right] - [\bar{t}(p)]^2 \tag{4.31}
 \end{aligned}$$

where $\psi(x)$ has been defined in (4.16) and

$$\xi(x) = [b(x)\psi(x)]^{-1} \bar{t}(x). \tag{4.32}$$

It is also possible to find higher moments of the random variable $I_g(p)$. This has been done by Nagylaki (1974a), and we here only outline the method. Denoting the n th moment of this variable by $\mu^{(n)}(p)$, the successive moments satisfy the recurrence relation

$$-(n+1)g(p)\mu^{(n)}(p) = a(p) \frac{d\mu^{(n+1)}(p)}{dp} + \frac{1}{2}b(p) \frac{d^2\mu^{(n+1)}(p)}{dp^2}, \tag{4.33}$$

and the boundary condition $\mu^{(n)}(0) = \mu^{(n)}(1) = 0$. For $n = 1$ this generalizes (4.30), and higher moments may be found from (4.33) by iteration. In particular by choosing $g(x) = 1$, ($0 < x < 1$), $g(x) = 0$, otherwise, higher moments of the absorption time can be found from (4.33).

We now use the diffusion process to find an approximation for the distribution of the sojourn time in any state of the Markov chain. Equation (2.146) shows that the distribution of the sojourn time depends on two parameters (there denoted by α_{ij} and r_j), and (2.147) shows how these parameters are related to the mean of the sojourn time. Suppose first we wish to approximate the distribution of the sojourn time at j , where $j > k$, k being the initial value. The parameter α_{kj} is the probability that state $\{j\}$ is ever reached and equation (4.17) is readily adapted to give

$$\alpha_{kj} \approx \int_0^p \psi(y) dy / \int_0^x \psi(y) dy, \quad (4.34)$$

where $k = pM$, $j = xM$, and the drift and diffusion coefficients $a(y)$ and $b(y)$, needed to calculate $\psi(y)$, are associated with the diffusion process approximating the Markov chain. We approximate r_j by using (2.147) and (4.26) to find

$$a_{kj}/(1 - r_j) \approx M^{\gamma-1} \bar{t}(x; p). \quad (4.35)$$

Combining (4.34) and (4.35), the sojourn time distributions is given by (2.146) where α_{kj} is approximated by (4.34) and r_j by

$$r_j \approx 1 - \frac{1}{2} M^{1-\gamma} \left(b(x)\psi(x) / \left(\int_0^x \psi(y) dy \int_0^1 \psi(y) dy \right) \right). \quad (4.36)$$

When $j < k$ (2.147) continues to hold, but here we approximate α_{kj} by

$$\alpha_{kj} \approx \int_p^1 \psi(x) dx / \int_x^1 \psi(x) dx. \quad (4.37)$$

The approximation (4.36) remains unchanged.

All of the above formulas require modification when there is only one absorbing state. We do not go into details here and only state the conclusions. If $\{0\}$ is the only absorbing state (4.21) continues to hold, but $t(x; p)$ must be redefined as

$$t(x; p) = 2(b(x)\psi(x))^{-1} \int_0^x \psi(y) dy, \quad 0 \leq x \leq p, \quad (4.38)$$

$$t(x; p) = 2(b(x)\psi(x))^{-1} \int_0^p \psi(y) dy, \quad p \leq x \leq 1. \quad (4.39)$$

Similarly, when $\{1\}$ is the only absorbing state, we have

$$t(x; p) = 2(b(x)\psi(x))^{-1} \int_p^1 \psi(y) dy, \quad 0 \leq x \leq p, \quad (4.40)$$

$$t(x; p) = 2(b(x)\psi(x))^{-1} \int_x^1 \psi(y) dy, \quad p \leq x \leq 1. \quad (4.41)$$

In both cases (4.24)–(4.29) hold, except that revised boundary conditions are needed for (4.29) to produce the solution (4.27), where now $t(x; p)$ is given either by (4.38) and (4.39) or by (4.40) and (4.41).

4.5 The Stationary Distribution

We have assumed above that in the Markov chain we are interested in there has existed at least one absorbing state. In several cases of interest there are no absorbing states, and there exists a stationary distribution $\{\phi_j\}$ satisfying (2.156) and given implicitly as the solution to (2.157). Since an explicit expression for this distribution has not been found in many examples of genetic interest, we aim in this section to approximate this distribution by finding the stationary distribution of the approximating diffusion process. It will turn out that this leads to a very simple form for this approximating distribution in which the effects of the general parameters are clearly displayed.

Our starting point is the forward Kolmogorov equation in (4.6). If we integrate throughout formally with respect to x , there results eventually

$$\frac{\partial}{\partial t}[1 - F(x; t)] = a(x)f(x; t) - \frac{1}{2} \frac{\partial\{b(x)f(x; t)\}}{\partial x}. \quad (4.42)$$

Here $F(x; t)$ is the distribution function

$$F(x; t) = \int_0^x f(y; t) dy. \quad (4.43)$$

This formal derivation suggests that the right-hand side in (4.42) is the rate of flow of probability (from left to right) across the point x at time t . This interpretation can be verified, and we thus call the right-hand side in (4.42) the probability flux of the diffusion process. If a stationary distribution $f(x)$ exists this probability flux will be zero if $f(x; t)$ is replaced by $f(x)$, so that the stationary distribution satisfies the equation

$$-a(x)f(x) + \frac{1}{2} \frac{d\{b(x)f(x)\}}{dx} = 0. \quad (4.44)$$

Integration shows that the solution of this equation is

$$f(x) = \text{const}[b(x)]^{-1} \exp\left(2 \int^x a(y)/b(y) dy\right), \quad (4.45)$$

where the constant is allocated so that

$$\int_0^1 f(x) dx = 1. \quad (4.46)$$

So far as the original Markov chain is concerned, our interpretation is that the diffusion approximation to the stationary probability that the random variable in the Markov chain lies in $[Mx_1, Mx_2]$ is given by

$$\text{Prob}\{Mx_1 \leq X \leq Mx_2\} \approx \int_{x_1}^{x_2} f(x) dx. \quad (4.47)$$

This approximation turns out to be satisfactory except when $x_1 \approx 0$ or $x_2 \approx 1$, in which case special arguments, which we shall consider later, are needed.

4.6 Conditional Processes

In this section we consider diffusion processes where 0 and 1 are both absorbing barriers. It is often of interest to single out those diffusions for which a nominated absorbing barrier is eventually reached, and we do this by the theory of conditional processes. For definiteness we assume the barrier in question is $x = 1$, although we shall also give some formulas applying when it is $x = 0$.

Since there can be no stationary distribution for such conditional processes, and since also there is no interest in fixation probabilities, interest centers almost entirely on properties of the time until fixation. Regarding the diffusion as an approximation to a Markov chain, it is clear from (2.153) that the sojourn time function (4.22) and (4.23) should be replaced by

$$t^*(x; p) = t(x; p)P_1(x)/P_1(p). \quad (4.48)$$

This gives

$$t^*(x; p) = 2P_0(p)P_1(x)[P_1(b)b(x)\psi(x)]^{-1} \int_0^x \psi(y) dy, \quad 0 \leq x \leq p, \quad (4.49)$$

$$t^*(x; p) = 2P_1(x)[b(x)\psi(x)]^{-1} \int_x^1 \psi(y) dy, \quad p \leq x \leq 1. \quad (4.50)$$

We consistently use the asterisk notation (*) to denote functions computed conditional on eventual absorption at $x = 1$ and, below, the double asterisk notation (**) when conditioning on eventual absorption at $x = 0$. Thus conditional on eventual absorption at $x = 0$, the sojourn time function is, by arguments parallel to those just given,

$$t^{**}(x; p) = 2P_0(p)[b(x)\psi(x)]^{-1} \int_0^x \psi(y) dy, \quad 0 \leq x \leq p, \quad (4.51)$$

$$t^{**}(x; p) = 2P_0(x)P_1(p)[P_0(p)b(x)\psi(x)]^{-1} \int_x^1 \psi(y) dy, \quad p \leq x \leq 1 \quad (4.52)$$

Equation (2.152) suggests an even stronger result than these, namely that the conditional density functions $f^*(x; p, t)$ and $f^{**}(x; p, t)$ of the diffusion variable at time t satisfy

$$f^*(x; p, t) = f(x; p, t)P_1(x)/P_1(p), \quad (4.53)$$

$$f^{**}(x; p, t) = f(x; p, t)P_0(x)/P_0(p). \quad (4.54)$$

It is clear that (4.49)–(4.52) can be used immediately to find the conditional mean times before absorption. These were originally found by Kimura and Ohta (1969) by a method other than that just outlined. We now indicate a third way in which these conditional mean times can be derived, namely by finding the conditional process analogues to the Kolmogorov equations (4.6) and (4.9). To do this we must find the conditional process drift and diffusion coefficients analogous to those defined by (4.3) and (4.4). Let A be the event that absorption eventually occurs at $x = 1$ and $p^*(x \rightarrow x + \delta x)$ be the conditional probability density, given A , of a transition from x to $x + \delta x$ in time δt . Then

$$\begin{aligned} p^*(x \rightarrow x + \delta x) &= p(x \rightarrow x + \delta x \text{ and } A)/\text{Prob}(A) \\ &= p(x \rightarrow x + \delta x)P_1(x + \delta x)/P_1(x) \\ &\approx p(x \rightarrow x + \delta x)[1 + \delta x P_1'(x)/P_1(x)], \end{aligned}$$

where we use the dash notation (') to refer to differentiation with respect to x . Then the conditional process drift coefficient $a^*(x)$ is found from

$$\begin{aligned} a^*(x)\delta t &= \int (\delta x)p^*(x \rightarrow x + \delta x)d(\delta x) \\ &\approx \int (\delta x)p(x \rightarrow x + \delta x)[1 + (\delta x)P_1'(x)/P_1(x)]d(\delta x) \\ &= \{a(x) + b(x)P_1'(x)/P_1(x)\}\delta t. \end{aligned}$$

Thus it follows that

$$a^*(x) = a(x) + b(x)P_1'(x)/P_1(x). \quad (4.55)$$

It is found similarly that

$$b^*(x) = b(x). \quad (4.56)$$

In the case of the Wright–Fisher model, with no selection or mutation, so that $a(x) = 0$, $b(x) = x(1 - x)$, $P_1(x) = x$, (4.55) gives

$$a^*(x) = 1 - x. \quad (4.57)$$

This value has already been used implicitly in (3.9). In the same model, when the condition is made that the allele of interest is eventually lost,

$$a^{**}(x) = -x. \quad (4.58)$$

The arguments leading to these formulas can be made more rigorous by suitable handling of small-order terms. The conditional density $f^*(x; p, t)$ now satisfies the forward equation

$$\frac{\partial f^*(x; p, t)}{\partial t} = -\frac{\partial\{a^*(x)f^*(x; p, t)\}}{\partial x} + \frac{1}{2}\frac{\partial^2\{b^*(x)f^*(x; p, t)\}}{\partial x^2} \quad (4.59)$$

and the backward equation

$$\frac{\partial f^*(x; p, t)}{\partial t} = a^*(p)\frac{\partial f^*(x; p, t)}{\partial p} + \frac{1}{2}b^*(p)\frac{\partial^2 f^*(x; p, t)}{\partial p^2}. \quad (4.60)$$

Using (4.53), (4.55) and (4.56) it is easy to check that these are consistent with (4.6) and (4.9). The conditional mean absorption time may now be found by using $a^*(x)$ and $b^*(x)$ in (4.40) and (4.41), and the resulting value agrees with that found from (4.49) and (4.50). This final approach is more general in that it uses the defining equations (4.59) and (4.60) and thus can be used to find higher moments of the conditional absorption time. We take this point up later when considering specific applications.

Parallel calculations apply, with the obvious changes, to find the conditional density function $f^{**}(x; p, t)$ when the condition is made that the allele of interest is eventually lost from the population.

4.7 Diffusion Theory

As mentioned in Section 4.1, there exists a deep mathematical theory of diffusion processes. Expositions of this theory are given by Ito and McKean (1965), Freedman (1971) and Mandl (1968). In this section we consider those parts of the theory that are of use to us in genetic processes. Because the random variable of interest to us is the frequency of some allele, we consider only diffusion processes on the interval $[0, 1]$.

The drift and diffusion functions $a(x)$ and $b(x)$ were introduced in (4.4) and (4.3). They may be used to define the important functions $p(x)$ and

$m(x)$, defined respectively by

$$p(x) = \int_c^x \exp(-2 \int_c^y a(z)/b(z) dz) dy, \tag{4.61}$$

$$m(x) = 2 \int_c^x \{b(y)\}^{-1} \exp(2 \int_c^y a(z)/b(z) dz) dy, \tag{4.62}$$

for some arbitrary constant c . Up to a linear transform, $p(x)$ is identical to the fixation probability $P_1(x)$. A diffusion is said to be on its natural scale if $p(x) = x$, which, from (4.61), is equivalent to $a(x) = 0$. For any diffusion not on its natural scale it is possible to find a transformed random variable (indeed the transformation is $x \rightarrow p(x)$) that is, and this explains the intimate link between $p(x)$ and $P_1(x)$. For this reason, $p(x)$ is called the *scale function* of the diffusion process. We give a name to the function $m(x)$ in a moment.

The functions $p(x)$ and $m(x)$ are central to many properties of diffusion processes, and we now show how they can be used to elucidate boundary behavior. Let r be an arbitrary point in $(0, 1)$ and s be one or other boundary point (that is $s = 0$ or $s = 1$). From $p(x)$ and $m(x)$ we compute the functions

$$u(s) = \int_r^s m(x) dp(x), \tag{4.63}$$

$$v(s) = \int_r^s p(x) dm(x). \tag{4.64}$$

The nature of the boundary s is exhibited as follows:

$u(s)$	$v(s)$	boundary type	accessible?	absorbing?	
$< \infty$	$< \infty$	regular	yes	no	(4.65)
$< \infty$	$= \infty$	exit	yes	yes	
$= \infty$	$< \infty$	entrance	no	no	
$= \infty$	$= \infty$	natural	no	yes	

A boundary is accessible if there exists positive probability that it can be reached in finite time from a given interior point, and is absorbing if the process remains forever at the boundary if it should reach it. We later give genetic examples of all four of these various boundaries. The terminology of boundary type follows Feller (1954), and other terminologies are possible, for example that of Prohorov and Rozanov (1969).

We have shown above why the description “scale function” is appropriate to $p(x)$. The definition of $m(x)$ shows that for a process on its natural scale,

$$\frac{dm(x)}{dx} = 2[b(x)]^{-1}. \quad (4.66)$$

It is a standard result for a diffusion process with $a(x) = 0$, $b(x) = b$, that the mean time for the random variable to reach $c \pm \delta$ from the value c is $b^{-1}\delta^2$, and is thus inversely proportional to the diffusion coefficient b . While in our process $b(x)$ is not constant, it may be so regarded, to a sufficient approximation, in any small interval $c \pm \delta$. This shows that to this level of approximation, $dm(x)/dx$ is proportional to the mean time that the diffusion process takes to leave this interval. This leads to the term *speed measure* for $m(x)$, although since larger values of $dm(x)/dx$ correspond to larger mean times for leaving the interval $c \pm \delta$, a better name for $m(x)$ would perhaps be “inertia measure”.

In our informal derivations in previous sections we have assumed without proof that density functions for diffusion processes exist. McKean (1956) has shown that a diffusion does have a density function $f(x; p, t)$, where p is the initial value of the diffusion random variable. Specifically, if the value of the diffusion random variable at time t is denoted $X(t)$,

$$\text{Prob}(c \leq X(t) \leq d) = \int_c^d f(x; p, t) dx, \quad (4.67)$$

for all p, c and d other than boundary points. When there are no natural boundaries, $f(x; p, t)$ possesses (Elliott (1955)) the eigenfunction expansion

$$f(x; p, t) = w(x) \left(C + \sum_{n=1}^{\infty} e^{\lambda_n t} \phi_n(x) \phi_n(p) \right), \quad (4.68)$$

where $w(x) = dm(x)/dx$ and $0 > \lambda_1 > \lambda_2 > \dots$ are distinct eigenvalues. The constant C is zero unless there exists a stationary distribution, when it takes the value

$$C = \left(\int_0^1 w(x) dx \right)^{-1}. \quad (4.69)$$

In this case,

$$\lim_{t \rightarrow \infty} \text{Prob}(c \leq X(t) \leq d) = C \int_c^d w(x) dx, \quad (4.70)$$

thus defining the stationary distribution and confirming our more informally derived (4.45).

When there is no stationary distribution, (4.68) can be written as

$$f(x; p, t) = \sum_{n=1}^{\infty} e^{\lambda_n t} (w(x)\phi_n(x))\phi_n(p). \quad (4.71)$$

We give a specific example of this expression in (5.11).

The eigenfunctions $\phi_n(x)$ satisfy the normalizing property

$$\int_0^1 \{\phi_n(y)\}^2 w(y) dy = 1, \quad (4.72)$$

this equation holding whether or not there exists a stationary distribution. The function $w(x)\phi_1(x)$ is analogous to the left eigenvector corresponding to the leading nonunit eigenvalue of the transition matrix P (see Section 4.2), while $\phi_1(p)$ is analogous to the corresponding right eigenvector.

4.8 Multi-dimensional Processes

So far we have considered diffusion processes in one dimension only. In a number of cases in population genetics theory we are, however, required to consider a vector of random variables rather than a single variable, and this leads to the consideration of multi-dimensional diffusion processes. We now informally extend to the multivariate case some of the derivations given in Section 4.2.

Consider first a set of linearly independent, jointly Markovian variables x_1, \dots, x_k for which, after a suitable re-scaling of time and space axes,

$$\begin{aligned} E(\delta x_i) &= a_i(x_1, \dots, x_k)\delta t + o(\delta t), \\ \text{var}(\delta x_i) &= b_i(x_1, \dots, x_k)\delta t + o(\delta t), \\ \text{covar}(\delta x_i, \delta x_j) &= c_{ij}(x_1, \dots, x_k)\delta t + o(\delta t), \end{aligned} \quad (4.73)$$

with higher absolute moments of order $o(\delta t)$. Let $f(x_1, \dots, x_k; t)$ be the joint density function of these random variables at time t . Then proceeding as in Section 4.2, this density function satisfies the forward equation

$$\begin{aligned} \frac{\partial}{\partial t} f(x_1, \dots, x_k; t) &= - \sum_i \frac{\partial}{\partial x_i} \{a_i(x_1, \dots, x_k) f(x_1, \dots, x_k; t)\} \\ &+ \frac{1}{2} \sum \frac{\partial^2}{\partial x_i^2} \{b_i(x_1, \dots, x_k) f(x_1, \dots, x_k; t)\} \\ &+ \sum_{i < j} \sum \frac{\partial^2}{\partial x_i \partial x_j} \{c_{ij}(x_1, \dots, x_k) f(x_1, \dots, x_k; t)\}. \end{aligned} \quad (4.74)$$

There will also exist a backward equation of obvious form corresponding to (4.9). Suppose now a joint stationary density $f = f(x_1, \dots, x_k)$ exists.

Then from (4.74) this density function will satisfy the equation

$$-\sum_i \frac{\partial}{\partial x_i} \{a_i f\} + \frac{1}{2} \sum_i \frac{\partial^2}{\partial x_i^2} \{b_i f\} + \sum_{i < j} \frac{\partial^2}{\partial x_i \partial x_j} \{c_{ij} f\} = 0. \quad (4.75)$$

Unfortunately the concept of a probability flux in several dimensions is more complex than in one dimension and, perhaps as a consequence, no simple explicit formula is known for the stationary distribution generalizing (4.45).

A most important question in multi-dimensional diffusion processes concerns the possible existence of a “second-order diffusion” or “quasi-Markovian variable.” We illustrate this concept by an example. In the two-sex model of Section 3.7 the pair (x_t, y_t) , where $x_t(y_t)$ is the frequency of A_1 among males (females) in generation t , is jointly Markovian. One suspects that x_t and y_t will seldom differ significantly from each other and that some weighted average of the two would behave in a “quasi-Markovian” manner. Such a possibility was investigated by Moran (1958) and Watter-son (1962); we present here an outline of the definitive work of Norman (1975c) on this point, simplified to cover specifically genetical applications.

Consider a population of size N reproducing at time points $n = 0, 1, 2, 3, \dots$, and suppose there exists at time n a random variable X_n ($0 \leq X_n \leq 1$) having the properties

$$\begin{aligned} \mathbf{E}\{X_{n+1} - X_n\} &= \tau_N a(X_n) + e_{1,n}^N, \\ \mathbf{E}\{X_{n+1} - X_n\}^2 &= \tau_N b(X_n) + e_{2,n}^N, \\ \mathbf{E}\{|X_{n+1} - X_n|\}^3 &= e_{3,n}^N. \end{aligned} \quad (4.76)$$

Here all expectations are conditional on X_0, X_1, \dots, X_n , $\tau_N > 0$ and $\tau_N \rightarrow 0$ as $N \rightarrow \infty$, and the “error” terms $e_{i,n}^N$ are all $o(\tau_N)$ in the sense that for any finite t ,

$$\sum_{n < \lfloor t/\tau_N \rfloor} \mathbf{E}\{|e_{i,n}^N|\} \rightarrow 0 \quad \text{as } N \rightarrow \infty. \quad (4.77)$$

The conditions (4.76) are reminiscent of the conditions (4.3)–(4.5), although we emphasize that X_n is not necessarily a Markovian variable. In the two-sex model just mentioned, for example, the quantity we use later for X_n (see Section 5.2) is not Markovian. A function X_n satisfying (4.76) does not necessarily exist, and if it does it is not necessarily unique: There may be several “quasi-Markovian” variables satisfying conditions like (4.76). We expect that under certain reasonable conditions the behavior of X_n will mimic that of a diffusion variable, and make this expectation more precise by specializing to the genetic case a general theorem of Norman (1975c).

Theorem 4.1. Suppose in (4.76) that $a(x)$ and $b(x)$ are polynomials with $a(0) \geq 0$, $a(1) \leq 0$, $b(0) = b(1) = 0$ and $b(x) > 0$, $0 < x < 1$. Define $X(t)$ as

a diffusion variable having initial value $X(0) = X_0$ and drift and diffusion coefficients $a(x)$, $b(x)$ respectively. Then for any time points $0 < t_1 < t_2 < \dots < t_j$, the joint distribution of $X_{n_1}, X_{n_2}, \dots, X_{n_j}$ converges to that of $X(t_1), X(t_2), \dots, X(t_j)$ as $N \rightarrow \infty$, $n_i \rightarrow \infty$ and $n_i \tau_N \rightarrow t_i$.

We do not prove this remarkable theorem here and note only the simplicity of the conditions for its applicability. In particular, as we see in the next chapter, the two-sex model of Chapter 3 satisfies these conditions, and this will lead to a definition of a variance effective population size for this model.

4.9 Time Reversibility

In this section discuss, informally and briefly, the extent to which a diffusion process is time reversible in the sense outlined for Markov chains in Chapters 2 and 3. We recall the definition (2.163) of time reversibility for Markov chains, and observe that this implies that

$$\phi_i p_{ij}^{(t)} = \phi_j p_{ji}^{(t)} \quad (4.78)$$

for any positive integer t .

Consider now a diffusion process on $[0, 1]$ possessing a stationary distribution $f(x)$, given by (4.45). This diffusion process is time-reversible. We have noted that for Markov chains certain questions involving time reversibility can be considered even though no stationary distribution exists. Various devices enable us to do the same thing for diffusion processes. In this book we shall consider only a useful general procedure, due to Norman (1978), which we take up in more detail in the next chapter when considering genetical examples. It is possible that one boundary of the diffusion process is absorbing and thus that no stationary distribution exists. When this is so the reversibility argument cannot be applied directly. On the other hand, it is sometimes possible to alter the diffusion process by inserting a small parameter ϵ such that the original process corresponds to $\epsilon = 0$ and, when $\epsilon > 0$, a stationary distribution does exist. Thus for $\epsilon > 0$ the time reversibility argument holds and if we now let $\epsilon \rightarrow 0$ it is sometimes possible to derive meaningful results for the original process by continuity. A specific example of this is given in Section 5.9.

4.10 Expectations of Functions of Diffusion Variables

In (4.27) we found the expected values of the integral of a function $g(x)$ over the entire time taken before absorption has been reached. In some

cases it is of interest to find the expected value at a single time point t , that is

$$h(t) = E_t[g(x)] = \int_0^1 g(x)f(x;p,t) dx, \quad (4.79)$$

where $f(x;p,t)$ is the density function of the diffusion random variable at time t and p its initial frequency. This expectation can be used in a variety of ways and has been exploited with particular success by Ohta and Kimura (1969a, 1971a): See also Kimura and Ohta (1971, pp. 183–190). Suppose at time $t + \delta t$ that the random variable takes the value $x + \delta x$. Then

$$\begin{aligned} h(t + \delta t) &= E(g(x + \delta x)) \\ &= E_t E_x(g(x + \delta x)) \end{aligned} \quad (4.80)$$

where E_x refers to the expectation operator conditional on the observed value x at time t and E_t refers to expectation with respect to the distribution of x . Now

$$E(g(x + \delta x)) \approx g(x) + E(\delta x)g'(x) + \frac{1}{2}E(\delta x)^2g''(x).$$

Inserting these values in (4.80) we get

$$h(t + \delta t) \approx h(t) + E_t[(a(x)g'(x) + \frac{1}{2}b(x)g''(x))\delta t]$$

and hence

$$\frac{d}{dt}h(t) = E_t(a(x)g'(x) + \frac{1}{2}b(x)g''(x)). \quad (4.81)$$

If the diffusion process admits a stationary distribution, the limiting case $t \rightarrow \infty$ in (4.81) yields

$$E[a(x)g'(x) + \frac{1}{2}b(x)g''(x)] = 0. \quad (4.82)$$

A generalization of these equations is possible for multi-dimensional diffusions. If the diffusion process involves linearly independent variables x_1, x_2, \dots, x_k , and if $h(t)$ is the expected value of some function $g(x_1, \dots, x_k)$ at time t , then in an obvious notation

$$\begin{aligned} \frac{d}{dt}h(t) &= E_t\left(\sum a_i(x_1, \dots, x_k) \frac{\partial g}{\partial x_i} + \frac{1}{2} \sum b_i(x_1, \dots, x_k) \frac{\partial^2 g}{\partial x_i^2} + \right. \\ &\quad \left. + \sum \sum c_{ij}(x_1, \dots, x_k) \frac{\partial^2 g}{\partial x_i \partial x_j}\right) \end{aligned} \quad (4.83)$$

If a stationary distribution exists then at stationarity

$$\begin{aligned} E \left[\sum a_i(x_1, \dots, x_k) \frac{\partial g}{\partial x_i} + \frac{1}{2} \sum b_i(x_1, \dots, x_k) \frac{\partial^2 g}{\partial x_i^2} \right. \\ \left. + \sum \sum c_{ij}(x_1, \dots, x_k) \frac{\partial^2 g}{\partial x_i \partial x_j} \right] = 0. \end{aligned} \quad (4.84)$$

We give examples later (see in particular Section 6.6) of the use of these formulas.

5

Applications of Diffusion Theory

5.1 Introduction

In this chapter we apply some of the diffusion theory considered in the previous chapter to various Markov chain models arising in population genetics in order to arrive at various conclusions of evolutionary interest.

Our first aim is to see how the behavior of a given Markov chain can be approximated by a diffusion process on $[0, 1]$. To do this it is convenient to start with the general Wright–Fisher model specified by (3.16) and (3.29). In this model the variable considered is the number j of A_1 genes at some locus A in a diploid population of fixed size N , and thus has state space $\{0, 1, 2, \dots, 2N\}$. To work with a variable whose state space is closer to that of the diffusion process, we consider instead the fraction x of A_1 genes in the population, whose state space is $\{0, (2N)^{-1}, \dots, 1\}$. We assume the notation x for the frequency of A_1 throughout this chapter except in Section 5.10, where more complex expressions are required. We also write p for the initial frequency of A_1 .

So far as other notation is concerned it is convenient to adopt the notation given in (1.25), so that the genotype fitnesses are denoted by

$$w_{11} = 1 + s, \quad w_{12} = 1 + sh, \quad w_{22} = 1. \quad (5.1)$$

Further, when mutation exists, we assume mutation rates $u(A_1 \rightarrow A_2)$ and $v(A_2 \rightarrow A_1)$.

The diffusion model we concentrate on requires that s , u and v are all $O(N^{-1})$. We make this assumption throughout, and then put

$$\alpha = 2Ns, \quad \beta_1 = 2Nu, \quad \beta_2 = 2Nv \quad (5.2)$$

where α , β_1 , and β_2 are all $O(1)$. Then standard binomial formulas for the model (3.16) show that

$$\begin{aligned} E(\delta x | x) &= (\alpha x(1-x)\{x+h(1-2x)\} - \beta_1 x + \beta_2(1-x))(2N)^{-1} \\ &\quad + o(N^{-1}), \\ \text{var}(\delta x | x) &= x(1-x)(2N)^{-1} + o(N^{-1}), \\ E\{|\delta x|^3\} &= o(N^{-1}). \end{aligned} \quad (5.3)$$

These moments fit into the format (4.3) - (4.5) provided that we choose

$$\delta t = (2N)^{-1}, \quad (5.4)$$

$$b(x) = x(1-x), \quad (5.5)$$

$$a(x) = \alpha x(1-x)\{x+h(1-2x)\} - \beta_1 x + \beta_2(1-x). \quad (5.6)$$

The requirement (5.4) is met by taking unit time in the diffusion process to correspond to $2N$ generations in the Markov chain. It is important to keep this scaling in mind when considering the relation between “time” properties in the diffusion process and those in the Markov chain. We now consider some properties of the diffusion process on $[0, 1]$ with drift and diffusion coefficients given respectively by (5.6) and (5.5).

Before proceeding we observe that in practical applications the idealized model (3.16) will probably have to be replaced by something more complex, perhaps one or other of the models discussed in Chapter 3 in connection with effective population sizes. At the end of the next section we pursue this point for one particular such complex model. Although the theory is by no means clear, it seems likely that all the diffusion results given below will continue to hold, at least to a good approximation, when N is replaced by the variance effective population size $N_e^{(v)}$. Except for the case considered at the end of the next section, we make no further explicit mention of this point in this chapter.

The first step in discussing properties of the diffusion process with the drift and diffusion coefficients (5.5) and (5.6) is to compute the scale function and speed measure of the process, defined by (4.61) and (4.62). These become

$$p(x) = \int_c^x y^{-2\beta_2} (1-y)^{-2\beta_1} \exp\{\alpha(2h-1)y^2 - 2\alpha hy\} dy, \quad (5.7)$$

$$m(x) = 2 \int_c^x y^{2\beta_2-1} (1-y)^{2\beta_1-1} \exp\{-\alpha(2h-1)y^2 + 2\alpha hy\} dy, \quad (5.8)$$

for an arbitrary constant c . We first use these expressions to consider boundary behavior. Use of (5.7) and (5.8) in (4.63) and (4.64) shows that near $x = 1$, the functions $u(x)$ and $v(x)$ take the form (for $\beta_1 \neq \frac{1}{2}$)

$$u(x) = A + O(1-x)^{1-2\beta_1}, \quad v(x) = B + O(1-x)^{2\beta_1}.$$

Here A and B are constants whose precise values are unimportant. It follows that $v(x)$ is always finite at $x = 1$, but that $u(x)$ is finite at this point only if $\beta_1 < \frac{1}{2}$. From this we conclude that the boundary $x = 1$ is regular (accessible but nonabsorbing) if $\beta_1 < \frac{1}{2}$ and entrance (inaccessible and nonabsorbing) if $\beta_1 > \frac{1}{2}$. The same conclusion holds for the boundary $x = 0$, with β_2 replacing β_1 . The values of α and h are irrelevant to these boundary descriptions. The case $\beta_1 = \frac{1}{2}$ is easily handled separately.

The intuitive meaning of these conclusions is clear enough. If the mutation rate from A_1 to A_2 and the population size are jointly large enough there is zero probability that the frequency of A_1 can ever achieve the value unity. Of course this conclusion applies for the diffusion process and it not true for the Markov chain (3.16), a fact we shall take up again later when considering the accuracy of diffusion approximations, particularly at a boundary.

If $\beta_1 = 0$ the boundary $x = 1$ is found to be exit (accessible and absorbing), and this again accords with what we expect since, if the boundary is reached, the absence of mutation from A_1 to A_2 means that the frequency of A_1 remains forever at unity. The fact that the boundary is accessible is less obvious intuitively: In Section 5.8 natural boundaries will be encountered which, although absorbing, are not accessible, that is for which there is zero probability that they are reached by diffusion from within $(0, 1)$.

The functions $p(x)$ and $m(x)$ are also central to the calculation of fixation probabilities and stationary distributions respectively, when these are appropriate. We defer consideration of these until we take up specific cases later.

We conclude this section by emphasizing that our main interest is in Markov chain models such as (3.16), and we view diffusion processes mainly as approximations to these. Usually the approximations are excellent, but in some instances, particularly near the boundaries $x = 0, x = 1$ they are less so, and for these cases some care is needed in proceeding. We take up this matter in more detail in Section 5.7.

5.2 No Selection or Mutation

When there is no selection or mutation the model defined by (3.16) and (3.29) reduces to (1.48). Rather complete knowledge of the diffusion approximation to this model is available, and in this section we explore this

in some detail. Clearly we have

$$a(x) = 0, \quad b(x) = x(1 - x), \quad (5.9)$$

and the forward equation becomes

$$\frac{\partial f(x; t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} \{x(1 - x)f(x; t)\}. \quad (5.10)$$

The solution of this equation, and others more complex, was achieved in a series of papers by Kimura (1955a, b, c, 1956, 1957), which heralded a rebirth of the mathematical theory of population genetics. Most of the results in this section were given in these papers. The explicit solution of (5.10), subject to the requirement $x = p$ when $t = 0$, is

$$f(x; p, t) = \sum_{i=1}^{\infty} \frac{4(2i+1)p(1-p)}{i(i+1)} T_{i-1}^1(1-2p) T_{i-1}^1(1-2x) \\ \times \exp\left\{-\frac{1}{2}i(i+1)t\right\}. \quad (5.11)$$

Here $T_{i-1}^1(x)$ is a Gegenbauer polynomial defined in terms of the hypergeometric function by

$$T_{i-1}^1(x) = \frac{1}{2}i(i+1)F\left(i+2, 1-i, 2, \frac{1}{2}(1-x)\right),$$

so that in particular

$$T_0^1(x) = 1, \quad T_1^1(x) = 3x. \quad (5.12)$$

The speed measure $m(x)$ for the coefficients (5.9) is such that

$$w(x) = dm(x)/dx = 2x^{-1}(1-x)^{-1}. \quad (5.13)$$

We use this to confirm that (5.11) is of the form defined by (4.67) and (4.68) with

$$\phi_i(y) = 2(2i+1)^{1/2} \{i(i+1)\}^{-1/2} y(1-y) T_{i-1}^1(1-2y). \quad (5.14)$$

The probabilities $P_0(t)$ and $P_1(t)$ that the diffusion has reached 0 or 1 respectively by time t are

$$P_0(t) = 1 - p + \sum_{i=1}^{\infty} (2i+1)p(1-p)(-1)^i F(1-i, i+2, 2, 1-p) \\ \times \exp\left(-\frac{1}{2}i(i+1)t\right), \quad (5.15)$$

$$P_1(t) = p + \sum_{i=1}^{\infty} (2i+1)p(1-p)(-1)^i F(1-i, i+2, 2, p) \\ \times \exp\left(-\frac{1}{2}i(i+1)t\right). \quad (5.16)$$

The probability of ultimate fixation at $x = 1$ can be found by letting $t \rightarrow \infty$ in (5.16) or else by computing (4.17), with $\psi(x)$ defined by (4.16) and (5.9). Evidently $\psi(x) = 1$ and hence

$$\text{Prob(ultimate fixation at } x = 1) = p. \quad (5.17)$$

The mean fixation time can be found from (4.22) and (4.23). These equations give

$$\begin{aligned}\bar{t}(x;p) &= 2(1-p)/(1-x), & 0 \leq x \leq p, \\ \bar{t}(x;p) &= 2p/x, & p \leq x \leq 1,\end{aligned}\tag{5.18}$$

so that the mean absorption time is

$$\bar{t}(p) = -2\{p \log p + (1-p) \log(1-p)\}\tag{5.19}$$

time units, or $-4N\{p \log p + (1-p) \log(1-p)\}$ generations. This agrees with the value (3.5) found without recourse to diffusion processes, and yields (3.6) and (3.7) as cases of particular interest.

The variance of the absorption time can be found from (4.31) and is

$$4\left(p \int_p^1 \lambda(x) dx - (1-p) \int_0^p \lambda(x) dx - \bar{t}(p)^2\right),\tag{5.20}$$

where

$$\lambda(x) = -2 \int_0^x [(1-y)^{-1} \log y + y^{-1} \log(1-y)] dy.\tag{5.21}$$

The value (5.20) is in terms of (squared) time units and must be multiplied by $4N^2$ to be brought to a (squared) generation basis.

The complete distribution of the absorption time is implicit in (5.15) and (5.16), since

$$\text{Prob}\{\text{absorption time} \leq t\} = P_0(t) + P_1(t).\tag{5.22}$$

Because of the form of the solutions (5.15) and (5.16), this expression is of most use when t is large. We show later how this solution may be supplemented by an asymptotic expansion the accuracy of which is best for small values of t . This asymptotic expansion, together with (5.22), then yields a rather complete picture of the distribution of the absorption time.

What do these diffusion results mean for the Markov chain model (1.48)? The fixation probability (5.17) is exactly correct for this model, since we have seen that this value can be reached directly. The mean absorption time approximation has been confirmed. We have, however, arrived at the more detailed information, from (5.18) and (4.26), that if the initial number of A_1 genes in the Markov chain model is k , the mean number of generations for which this number assumes the value j , before reaching 0 or $2N$, is approximately

$$\begin{aligned}\bar{t}_{k,j} &= 2(2N-k)/(2N-j), & j \leq k, \\ \bar{t}_{k,j} &= 2k/j, & j \geq k.\end{aligned}\tag{5.23}$$

The particular case $k=1$, of particular interest to Fisher and Wright, gives $\bar{t}_{1,j} = 2j^{-1}$, in agreement with (1.56).

We turn now to the spectral expansion (5.11). Recalling the difference in time scale between the Markov chain (1.48) and the diffusion process (5.10), it is clear that the eigenvalue $\exp\{-\frac{1}{2}i(i+1)\}$ is the analogue of the Markov chain value

$$\begin{aligned} \left[\left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) \cdots \left(1 - \frac{i}{2N}\right) \right]^{2N} &\approx \exp\{-1 + 2 + \cdots + i\} \\ &= \exp\{-\frac{1}{2}\{i(i+1)\}\}. \end{aligned}$$

There is also a parallel between the eigenfunctions in (5.11) and the eigenvectors of (1.48). For large t we may write

$$\begin{aligned} f(x; p, t) &= 6p(1-p)\exp(-t) + 30p(1-p)(1-2p)(1-2x) \\ &\quad \times \exp(-3t) + \cdots. \end{aligned} \tag{5.24}$$

The function $p(1-p)$ in the leading term of the expression on the right-hand side is clearly the analogue of the right eigenvector (1.51). Since this leading term is independent of x , the analogue of the corresponding left eigenvector is $\ell_k = \text{constant}$. This shows that the asymptotic ($t \rightarrow \infty$) conditional ($x \neq 0, 1$) distribution of x is uniform over $(0, 1)$, in agreement with the approximation of the model Wright–Fisher model (1.48) noted after equation (1.54). The complete expansion (5.24) shows, as was first observed by Kimura, that the extensive attention paid to this distribution was misplaced. The leading term in (5.24) does not dominate the second term until $t \approx 2$, that is $4N$ generations, and the distribution of the fixation time, given by (5.22), shows that fixation of one or other allele is likely to have occurred by this time, especially in the interesting case $p = (2N)^{-1}$.

We have seen that the eigenfunction solution to equations such as (5.11) appear in the form

$$f(x; p, t) = \sum_n \{\phi_n(p)\} \{\phi_n(x)w(x)\} \exp(\lambda_n t). \tag{5.25}$$

Thus the eigenfunctions corresponding to initial and current points bear a simple relationship to each other. Now in the model (1.48) it is quite easy to find the right eigenvectors exactly, but very difficult to find the left eigenvectors, and (5.25) suggests an approximation relation between the two. Since for the process (5.10)

$$w(x) = 2x^{-1}(1-x)^{-1}, \tag{5.26}$$

we may make the approximation for the Wright–Fisher model (1.48)

$$\ell_{ij} = j^{-1}(2N-j)^{-1}r_{ij}, \tag{5.27}$$

where $\ell_{ij}(r_{ij})$ is the j th element in the i th left (right) eigenvector of the transition matrix. Since we know $r_{2j} = j(2N-j)$, this derives the uniform approximation to the leading left eigenvector, and allows a rapid approximation to the remaining eigenvectors.

Although the solution (5.11) is exact for all t , it is most useful for large values of t (say $t > 1$). For small values of t , for example for $t < 0.1$, the infinite series converges slowly, and many terms must be calculated to obtain satisfactory approximations. Fortunately, an asymptotic expansion solution of equation (5.10) dovetails nicely with the solution (6.22) near $t = 1$, and the two solutions together then allow rather complete knowledge of the solution of (5.10). We do not give the derivation of this asymptotic expansion, which is due to Voronka and Keller (1975). A wider range of applications is given by Tier and Keller, (1978). Here we observe only that for $t < 1$,

$$P_1(p; t) \sim \{p(1-p)/C\}^{1/4} \exp(-2Ct^{-1}), \quad (5.28)$$

where

$$C = \frac{1}{4} \{\arccos(2p-1)\}^2.$$

Clearly, by symmetry,

$$P_0(p; t) \sim \{p(1-p)/C'\}^{1/4} \exp(-2C't^{-1}) \quad (5.29)$$

where

$$C' = \frac{1}{4} \{\arccos(1-2p)\}^2.$$

These two values can be combined to give an asymptotic expansion for the probability of fixation by time t . When $p = \frac{1}{2}$, $t = 0.65$, (5.28) gives $P_1(\frac{1}{2}, t) \approx 0.119$, whereas the correct value, found after much computation from (5.16), is 0.117. For $t = 1$, (5.28) gives $P_1(\frac{1}{2}, t) \approx 0.232$ whereas the approximation

$$P_1(\frac{1}{2}; t) \approx \frac{1}{2} - \frac{3}{4}e^{-1},$$

found by taking the two leading terms only in (5.16), gives the value 0.224. (The correct value is 0.223.) Remarks parallel to these apply for the density function $f(x; p, t)$: The asymptotic expansion is very accurate up to $t = 1$, where it agrees with the expression found by taking the three leading terms in (5.11). The latter then provide excellent approximations for large t values.

We consider now processes conditional on the event that a specified boundary is eventually reached. We suppose for definiteness that $x = 1$ is the absorbing state ultimately reached. Equations (4.53) and (5.11) show that the density function of x at time t is

$$\begin{aligned} f^*(x; p, t) &= \sum_{i=1}^{\infty} \frac{4(2i+1)x(1-p)}{i(i+1)} T_{i-1}^1(1-2p) T_{i-1}^1(1-2x) \\ &\times \exp\left\{-\frac{1}{2}i(i+1)t\right\}. \end{aligned} \quad (5.30)$$

For large t and small p this gives

$$f^*(x; p, t) \sim 6x \exp(-t), \quad (5.31)$$

so that

$$\lim_{t \rightarrow \infty} f(x \mid x \neq 0, 1, \text{ eventual fixation at } x = 1) = 2x. \quad (5.32)$$

It is interesting to compare this with the expression in (1.86). There is no similarity between the two expressions, and we conclude that this is a case where the nature of the branching process on which (1.86) is based, and that of the model (1.48), are sufficiently different so that one gives little information about the other for large t values.

The functions $t^*(x)$ defined in (4.49) and (4.50) become

$$\begin{aligned} t^*(x) &= 2(1-p)x/\{p(1-x)\}, & 0 \leq x \leq p, \\ t^*(x) &= 2, & p \leq x \leq 1. \end{aligned} \quad (5.33)$$

The conditional mean absorption time, found by integration, is then

$$t^*(p) = -2p^{-1}(1-p) \log(1-p). \quad (5.34)$$

In the Markov chain (1.48) this suggests the approximation that if k is the initial value of the Markov variable,

$$\begin{aligned} t_{k,j}^* &\approx 2(2N-k)j/k(2N-j), & j \leq k, \\ t_{k,j}^* &\approx 2, & j \geq k, \end{aligned} \quad (5.35)$$

and a conditional mean fixation time of $-4Np^{-1}(1-p) \log(1-p)$ generations. One interesting case arises when $k = 1$, so that there is initially only one gene of the allele of interest. One case of this concerns a unique selectively neutral new mutant destined for fixation. Equation (5.35) shows that on average, this allele spends two generations at each possible frequency value, ($k = 1, 2, \dots, 2N - 1$, so that if t_1^* is the conditional mean fixation time,

$$\bar{t}_1^* = 4N - 2 \quad (5.36)$$

generations. It is instructive to see how easily information about the conditional process can be found from information concerning the unconditional process.

The conditional variance of the absorption time can be found by solving (4.30), subject to appropriate boundary conditions. Here we must use the conditional process drift coefficient

$$a^*(x) = 1 - x$$

rather than the unconditional value. It is found that

$$\begin{aligned} (\sigma^*)^2(p) &= \frac{4\pi^2}{3} + 8p^{-1}(1-p) \log(1-p) \{1 - (2p)^{-1}(1-p) \log(1-p)\} \\ &\quad - 8 \sum_{i=1}^{\infty} \frac{p^i}{i^2}. \end{aligned} \quad (5.37)$$

In the limiting case $p \rightarrow 0$ this gives

$$(\sigma^*)^2 \approx 8 \left(\frac{\pi^2}{6} - 1.5 \right) \approx 1.16, \quad (5.38)$$

or, for the process (1.48), $4.64N^2$ (squared generations). The complete distribution of the conditional absorption time can be found immediately from (5.16). We have

$$\begin{aligned} & \text{Prob}\{\text{absorption at } x = 1 \text{ before time } t \mid \text{eventual absorption at } x = 1\} \\ &= \frac{\text{Prob}\{\text{absorption at } x = 1 \text{ before time } t\}}{\text{Prob}\{\text{eventual absorption at } x = 1\}} \quad (5.39) \\ &= 1 + \sum_{i=1}^{\infty} (2i+1)(1-p)F(i+2, 1-i, 2, p)(-1)^i \exp\{-i(i+1)t\}. \end{aligned}$$

The expressions (5.34) and (5.38) can in principle be found from this distribution, but it is far simpler to arrive at them in the manner we have shown. Kimura (1970) established (5.39) and discussed the nature of the corresponding density function for various p values.

The asymptotic expansion of Voronka and Keller may be used immediately for conditional processes. Thus, for example, (5.28) gives

$$P^*(p; t) = p^{-1}\{p(1-p)C\}^{1/4} \exp(-2Ct^{-1}), \quad (5.40)$$

where $P^*(p; t)$ is the conditional probability of fixation at $x = 1$ by time t , given fixation eventually occurs. Similarly the conditional density function $f^*(x; p, t)$ can be accurately approximated by small t , and this leads, as with $f(x; p, t)$, to rather complete knowledge of its nature.

We consider briefly the case where we condition on eventual loss of the allele A_1 . Since for this case $a(x) = 0$, $b(x) = x(1-x)$ and $P_0(x) = 1-x$, the analogue of (4.55) gives $a^{**} = -x$. The analogue of (5.34) is, in terms of generations, $t^{**} = -4N(p \log p)/(1-p)$. This is identical to the value given in (3.22), and this is not surprising since the value of the drift coefficient a^{**} given above is identical to that given in (5.61) below for the one-way mutation case when $\theta = 2$. Since the diffusion coefficients are also the same in the two cases, the entire stochastic behavior of the conditional process without mutation and the unconditional process with mutation (for $\theta = 2$) are identical. This seems for the moment to be no more than a curiosity, but it will have a more interesting interpretation when considering age and retrospective properties in Chapter 9.

We conclude this section by discussing a model generalizing (1.48). Consider for example the two-sex model introduced in Section 3.7. Using the notation of that section, the quantities $k(t)$ (the number of A_1 genes among the males in generation t) and $\ell(t)$ (the number of A_1 genes among the females in generation t), are jointly Markovian. We make progress by using the Norman theory of quasi-Markovian variables introduced in Section 4.8.

The weighted average gene frequency $x(t)$ is

$$x(t) = k(t)(4N_1)^{-1} + \ell(t)(4N_2)^{-1} \quad (5.41)$$

and it is easy to check, for the model considered, that

$$\begin{aligned} E\{x(t+1) - x(t) \mid x(t)\} &= 0 \\ \text{var}\{x(t+1) - x(t) \mid x(t) = x\} &= x(1-x)(2N_e)^{-1} + e_{2,t}, \\ E\{|x(t+1) - x(t)|^3\} &= e_{3,t}, \end{aligned} \quad (5.42)$$

where $N_e = 4N_1N_2/(N_1 + N_2)$ and, for large N_1 and N_2 , the error terms $e_{i,t}$ can be shown to satisfy (4.77) with $\tau_N = (2N_e)^{-1}$. Thus $x(t)$ is a quasi-Markovian variable, and the conclusions given in Section 4.8 for such variables can be applied. In particular the probability of fixation of A_1 is

$$k(0)(4N_1)^{-1} + \ell(0)(4N_2)^{-1}.$$

The variance formula confirms the value (3.124) for the variance effective population size.

5.3 Selection

Suppose now that the three genotypes have fitnesses given by (5.1). If there is no mutation, the drift coefficient (5.6) becomes

$$a(x) = \alpha x(1-x)\{x + h(1-2x)\}. \quad (5.43)$$

From this the scale function and speed measure are calculated as

$$p(x) = \int_{c_1}^x \psi(y) dy, \quad (5.44)$$

$$m(x) = 2 \int_{c_2}^x y^{-1}(1-y)^{-1}\{\psi(y)\}^{-1} dy, \quad (5.45)$$

where

$$\psi(y) = \exp \alpha \{(2h-1)y^2 - 2hy\}. \quad (5.46)$$

Both boundaries $x = 0$, $x = 1$ are exit, and the probability that one or the other boundary is eventually reached is unity. The respective probabilities are given by (4.15) and (4.17), with $\psi(y)$ defined by (5.46).

These expressions simplify significantly only in the case of no dominance ($h = \frac{1}{2}$), for which

$$P_1(p) = \{1 - \exp(-\alpha p)\} / \{1 - \exp(-\alpha)\}. \quad (5.47)$$

This agrees with the approximation (3.31) found without using diffusion methods. Some numerical values calculated from (5.47) are given in Table 5.1.

	$p = 0.001$			$p = 0.5$		
	$N = 10^4$	$N = 10^5$	$N = 10^6$	$N = 10^4$	$N = 10^5$	$N = 10^6$
	0.01	0.181	0.865	1.000	1.000	1.000
	0.001	0.020	0.181	0.865	1.000	1.000
s	0.0001	0.002	0.020	0.181	0.731	1.000
	0.00001	0.001	0.002	0.020	0.525	0.731

Table 5.1. Values of $P_1(p)$, for various values of N , s , and p , calculated from (5.47)

The conclusions to be drawn from this table are straightforward. When N , s and p are jointly sufficiently large, fixation of the favored allele is essentially certain: This occurs approximately when $Nsp > 5$. As N , s or p decreases, the fixation probability decreases, and if $Ns < 0.1$ it does not differ (relatively) by more than 10% from the neutral value p . Perhaps the most striking conclusion is the very strong effect of selection in influencing fixation probabilities: As noted below (3.31), selective differences far too small to be found in the laboratory can nevertheless have a decisive effect on evolutionary behavior, at least in populations that are not too small. The same conclusion holds, at least qualitatively, when there is no dominance (that is $h \neq \frac{1}{2}$), although some minor modifications to the numerical values are necessary, especially when dominance is complete ($h = 0$ or $h = 1$). Even in the overdominance case ($sh > s > 0$) fixation of one or the other allele is certain although, as we see later, this will normally take an extremely long time, and in practical terms one must then question the appropriateness of the assumptions made, in particular that there is no mutation and that the population size, selective differences, and dominance relationship remain unchanged throughout the entire fixation process. So far as fluctuations in population size are concerned, it seems likely, for any fixed selection scheme, that (5.47) and (4.17) still apply if N is replaced by the variance effective population size, although the theory for this has not been verified.

The complete solution of the forward equation (4.6), with $a(x)$ and $b(x)$ defined by (5.43) and (5.5), is very complex. Nevertheless solutions were found by Kimura (1955a, b, c; 1957) initially for the no dominance case and subsequently for the general case. Unfortunately the very complexity of the solutions makes examination of their implications difficult, although this does not detract from the influence that the derivation of these solutions has had on population genetics theory. For more details concerning these solutions, see Crow and Kimura (1970, pp. 396–414).

5.4 Selection: Absorption Time Properties

Despite the very complex form of $f(x; t)$ referred to at the end of the previous section, a rather simple expression exists for the function $\bar{t}(x; p)$, defined in (4.25), and since this function summarizes perhaps the most important features of the transient behavior of the process, we now compute it for the selective model we are considering. All that is required to do this is to substitute (5.43) and (5.5) into the general formulas (4.22) and (4.23). For $h = \frac{1}{2}$ we get

$$\begin{aligned}\bar{t}(x; p) &= 2P_0(p)\{\alpha x(1-x)\}^{-1}\{\exp(\alpha x) - 1\}, & 0 \leq x \leq p \\ \bar{t}(x; p) &= 2P_1(p)\{\alpha x(1-x)\}^{-1}\{1 - \exp(-\alpha(1-x))\}, & p \leq x \leq 1\end{aligned}\tag{5.48}$$

where $P_1(p)$ is found from (5.47) and $P_0(p) = 1 - P_1(p)$. For the Markov chain defined by (3.16) and (3.29), this implies that the mean number of generations for which there are $j = 2Nx$ A_1 alleles, given an initial number $k = 2Np$, is approximately

$$\begin{aligned}\bar{t}_{k,j} &= 2\{\exp(-2p) - \exp(-\alpha)\}\{\exp(\alpha x) - 1\} \\ &\quad \times [\alpha x(1-x)\{1 - \exp(-\alpha)\}]^{-1} \quad (j \leq k), \\ \bar{t}_{k,j} &= 2\{1 - \exp(-\alpha p)\}\{1 - \exp(-\alpha(1-x))\} \\ &\quad \times [\alpha x(1-x)\{1 - \exp(-\alpha)\}]^{-1} \quad (j \geq k).\end{aligned}\tag{5.49}$$

For $k = 1$ this gives (1.62) if we make the approximation $1 - \exp(-\alpha/2N) = \alpha/2N$.

The mean time for fixation is found jointly from (4.21) and (5.48), but unfortunately no explicit evaluation of the integrals is possible, and numerical computation is necessary. There is, however, one case where useful progress can be made. If α and p are jointly sufficiently large so that fixation of the favored allele can be taken as being almost certain, we get

$$\begin{aligned}\bar{t}(x; p) &\approx 0, & 0 \leq x \leq p \\ \bar{t}(x; p) &\approx 2\{\alpha x(1-x)\}^{-1}, & p \leq x \leq 1 - 4\alpha^{-1}, \\ \bar{t}(x; p) &\approx 2\{\alpha x(1-x)\}^{-1}\{1 - \exp(-\alpha(1-x))\}, & 1 - 4\alpha^{-1} \leq p \leq 1.\end{aligned}\tag{5.50}$$

The first equation shows that in the case considered, the frequency of the favored allele spends negligible time less than its initial value. The second equation is perhaps the most interesting. Converting to generations, it implies that in the Markov chain the mean time spent in the frequency range (x_1, x_2) where $p \leq x_1 < x_2 \leq 1 - 4\alpha^{-1}$, is approximately

$$\int_{x_1}^{x_2} \left\{ \frac{1}{2} s x (1-x) \right\}^{-1} dx$$

generations. This is identical to the value (1.29) found for the corresponding deterministic process, and we can conclude that the behavior of the process $(p, 1 - 4\alpha^{-1})$ is “quasi-deterministic”. When the frequency exceeds $1 - 4\alpha^{-1}$ the deterministic value no longer gives an adequate guide to the stochastic behavior. In particular, the mean number of generations (in the Markov chain) for which $x = 1 - i(2N)^{-1}$, for small integers i , is essentially equal to the “neutral” value 2. This is severely overestimated by the deterministic formula, and clearly at this stage of the process selective forces have become of secondary importance, and random sampling almost wholly determines the gene frequency behavior.

For general values of h in $(0, 1)$, the expressions (4.22) and (4.23) do not simplify readily. However, the general behavior just noted for the no dominance case continues to apply. Quasi-deterministic behavior obtains for sufficiently large p and α , at least until the frequency x of A_1 approaches unity, when selective forces once more can be ignored. The value of x where this occurs will depend to some extent on the level of dominance but will not differ materially from the value $1 - 4\alpha^{-1}$ found in the no dominance case.

The value $k = 1$ is of particular interest. Here we may approximate the probability $P_1\{(2N)^{-1}\}$ by $(2N)^{-1} \left(\int_0^1 \psi(y) dy \right)^{-1}$ and this leads, in the Markov chain, to the approximation (1.60).

The variance of the absorption time is given in principle by (4.31), but evaluation of this will certainly require numerical methods.

The overdominance case presents features of special interest. Here the deterministic theory gives a stable polymorphism while the stochastic theory predicts eventual loss of one or other allele. On the other hand, it is plausible, in the stochastic case, that extremely long periods of time are spent near the quasi-equilibrium point, at least when selection is strong, so that in some sense the deterministic theory provides a useful guide to the stochastic behavior. (Of course, if mutation is allowed, an entirely different stochastic behavior arises and one which should be well described by the deterministic theory.) We now discuss the stochastic process in more detail, and show in particular that the plausibility argument given above does not necessarily apply.

We start from the observation just made, that in the stochastic process fixation of one or other allele is eventually certain if there is no mutation, and ask how much time fixation takes to occur. This is best answered by considering the mean fixation time or, more crudely, the leading eigenvalue in the spectral expansion of the density function $f(x; t)$. The latter approach was taken by Miller (1962) and Robertson (1962) and produced a surprising answer. Define the quasi-equilibrium point by x^* (see (1.31): We now assume the notation (1.25c)). If x^* is close to 0 or 1 it is possible that this leading eigenvalue λ_1 , in an expansion of the form (4.68), is larger in

absolute value than for the selectively neutral case. This suggests a more rapid fixation process under overdominance than under neutrality. This perhaps surprising conclusion is explained by noting that if x^* is close to 0 (or 1), selection tends to drive the frequency of A_1 close to 0 (or 1) comparatively rapidly, and then random sampling effects which, as we have noted, play a predominant role near the boundaries, lead to loss or fixation of A_1 . The magnitude of this effect can be measured by taking the ratio of the absolute values of the leading eigenvalue to its neutral counterpart. When x^* is close to 0 or 1, this ratio increases with s for values of s of order a few percent, although for very large s values the ratio ultimately decreases as s increases. The discussion just given shows why this behavior should occur. For intermediate values of x^* (approximately $0.2 < x^* < 0.8$) the ratio always decreases as s increases, so that here heterosis always slows down the fixation process. It is clear that for these values of x^* , selection does not provide a thrust towards the boundaries sufficient to speed up fixation. Tables and graphs illustrating this behavior are given by Robertson (1962); Robertson centered attention on retardation behavior and thus considers the reciprocal of the ratio defined above.

A more complete analysis is provided by considering the mean fixation time, although a further degree of complexity arises here since this depends, unlike the eigenvalues, on the initial frequency p of A_1 . It is perhaps natural to pay special attention to the case $p = x^*$, although a general analysis is quite straightforward. The mean fixation time is given by (4.22), (4.23), (4.15) and (4.17), where $b(x) = x(1 - x)$ and

$$\psi(x) = \exp\{\alpha(2h - 1)(x - x^*)^2\}. \quad (5.51)$$

The retardation factor as measured by mean fixation times corresponding to Robertson's eigenvalue ratio is the ratio of the mean fixation time as just calculated to its neutral theory counterpart (5.19). We do not give here details of the numerical values found: For these see Ewens and Thomson (1970). The conclusions are in general agreement with those of Robertson, at least for $p = x^*$. For general values of p the behavior can be quite complex, the retardation factor sometimes increasing, then decreasing, then increasing again as s increases.

We turn now to mean absorption times conditional on eventual fixation (or loss) of a specified allele. The formulas appropriate to calculate this are (4.49) and (4.50) or (4.51) and (4.52). Perhaps the case of greatest interest is when $0 < h < 1$ and the condition is made that the favored allele fixes. When $h = \frac{1}{2}$, equations (4.49) and (4.50) give

$$t^*(x; p) = 2e^{-\alpha x} \{1 - e^{-\alpha(1-p)}\} \{e^{\alpha x} - 1\}^2 \\ \times [\alpha x(1-x) \{1 - e^{-\alpha}\} \{e^{\alpha p} - 1\}]^{-1}, \quad 0 \leq x \leq p, \quad (5.52)$$

$$t^*(x; p) = 2\{e^{\alpha x} - 1\} \{e^{\alpha(1-x)} - 1\} [\alpha x(1-x) \{e^{\alpha} - 1\}]^{-1}, \quad p \leq x \leq 1. \quad (5.53)$$

Similarly, if the condition is made that eventually A_1 is lost,

$$t^{**}(x; p) = 2\{e^{\alpha x} - 1\}\{e^{\alpha(1-x)} - 1\}[\alpha x(1-x)\{e^{\alpha} - 1\}]^{-1},$$

$$0 \leq x \leq p, \quad (5.54)$$

$$t^{**}(x; p) = 2e^{-\alpha(1-x)}\{1 - e^{-\alpha p}\}\{e^{\alpha(1-x)} - 1\}^2$$

$$\times [\alpha x(1-x)\{1 - e^{-\alpha}\}\{e^{\alpha(1-p)} - 1\}]^{-1}, \quad p \leq x \leq 1. \quad (5.55)$$

There are several interesting points about these equations. First, the value of $t^*(x; p)$ for $x \geq p$ is identical to that of $t^{**}(x; p)$ for $x \leq p$. We explain why this should be so when considering time-reversal properties in Section 5.9. The second point concerns the nature of the formula for $t^*(x; p)$ for very small p , or correspondingly $t^{**}(x; p)$ for very large p , and is relevant when considering a selectively favored new mutant destined for fixation. We observe that $t^*(x; p)$ is symmetric about $x = 0.5$; the mean time spent in any interval $(x, x + \delta x)$ is the same as the mean time spent in $(1 - x - \delta x, 1 - x)$. Even more surprisingly, $t^*(x; p)$ remains unchanged if α is replaced by $-\alpha$, so that a selectively disadvantageous mutant, if destined for fixation, spends as much time, on the average, in any frequency range as a corresponding selectively advantageous mutant destined for fixation. This remarkable fact, noted first in effect by Maruyama (1974), will again be reconsidered later in the light of time-reversal properties. It is indeed easy to see that the entire behavior of the conditional process is independent of the sign of s , since the diffusion coefficient $b^*(x)$, calculated from (4.56) and (5.5), is independent of s while the drift coefficient $a^*(x)$, calculated from (4.55), (5.43), and (5.47), is

$$a^*(x) = \frac{1}{2}\alpha x(1-x)/\tanh(\frac{1}{2}\alpha x).$$

Clearly $a^*(x)$ is independent of the sign of α : This more detailed conclusion was first noted by Watterson (1977b). However, despite the symmetry of $t^*(x)$ around $x = \frac{1}{2}$, it is not true that $a^*(x) = a^*(1-x)$.

For arbitrary levels of dominance, (4.50) shows that with $p = (2N)^{-1}$,

$$t^*(x; (2N)^{-1}) = 2(b(x)\psi(x) \int_0^1 \psi(y) dy)^{-1} \int_0^x \psi(y) dy \int_x^1 \psi(y) dy, \quad (5.56)$$

where $\psi(y)$ is defined by (5.46). If this expression is written more fully as $t^*(x; \alpha, h, (2N)^{-1})$, it follows that

$$t^*(x; \alpha, h, (2N)^{-1}) = t^*(1-x; \alpha, 1-h, (2N)^{-1}). \quad (5.57)$$

This implies that conditional mean fixation time properties for a favored allele are the same as those for the corresponding disadvantageous allele, provided the dominance relation is reversed. This generalizes the conclusion just reached for the case of no dominance.

Using the notation (5.1), the fitnesses will display overdominance if $sh > s > 0$ or underdominance if $sh < s < 0$. In either of these cases the

quasi-equilibrium point x^* , defined above, may be written more fully as $x^* = x^*(h) = h/(2h - 1)$. Then (5.57) may be written as

$$t^*(x, x^*, (2N)^{-1}) = t^*(1 - x, 1 - x^*, (2N)^{-1}), \quad (5.58)$$

an equation first noted in this form by Nei and Roychoudhury (1973).

For general levels of dominance it is no longer true (as it was with no dominance) that $t^*(x, \alpha, h, (2N)^{-1}) = t^*(x, -\alpha, 1 - h, (2N)^{-1})$, nor is it true that $a^*(x, \alpha, h) = a^*(x, -\alpha, 1 - h)$. There is, however, one relation, first noted by Maruyama and Kimura (1974), that does remain true. Keep the fitness scheme (5.1) fixed and consider two cases, one where the initial frequency of A_1 is $(2N)^{-1}$ and the condition is made that A_1 eventually fixes, and the other where the initial frequency of A_1 is $1 - (2N)^{-1}$ and the condition is made that A_1 is eventually lost. By considering A_2 rather than A_1 it is clear that the equation

$$t^*(1 - x, -\alpha, 1 - h, (2N)^{-1}) = t^{**}(x, \alpha, h, 1 - (2N)^{-1}) \quad (5.59)$$

must be true, and this may be used with (5.58) to show that

$$t^*(x, \alpha, h, (2N)^{-1}) = t^{**}(x, \alpha, h, 1 - (2N)^{-1}). \quad (5.60)$$

(We noted above the special case of this equation when $h = \frac{1}{2}$.) Thus the mean time spent in any frequency range is the same for both processes. On the other hand, it is not true that $a^*(x) = a^{**}(x)$ so that despite (5.60), the two processes have quite different properties. Again, these perhaps paradoxical conclusions will be reconsidered, and to a large extent resolved, when we consider time-reversal properties of diffusion processes.

5.5 One-Way Mutation

Until now in this chapter we have ignored the possibility of recurrent mutation from one allelic type to another when considering allele-frequency behavior. In some circumstances this might cause little inaccuracy but in general, especially from a macro-evolutionary rather than a micro-evolutionary point of view, it is essential that mutation be taken into account. In this section we make a start on this by supposing a model such as (3.16) where A_1 mutates to A_2 (at rate u), with no reverse mutation. The drift and diffusion coefficients for the diffusion process approximating this Markov chain are, when there is no selection,

$$a(x) = -\frac{1}{2}\theta x, \quad b(x) = x(1 - x), \quad (5.61)$$

where $\theta = 4Nu$. Clearly A_1 eventually becomes lost from the population and interest centers entirely on properties of the time until this loss occurs. These properties are defined in large measure by the function $t(x; p)$, and insertion of the coefficients (5.61) into (4.38) and (4.39) gives this function immediately. The values so calculated are given in (3.19), where allowance

must be made for the fact that the time-scale assumed there assumes unit time for one generation. Perhaps the case of most interest is when $p = (2N)^{-1}$, so that to a close approximation, the mean time that A_1 exists in the population is

$$\bar{t}\{(2N)^{-1}\} \approx \int_{(2N)^{-1}}^1 2y^{-1}(1-y)^{\theta-1} dy \quad (5.62)$$

generations. This is of order $2 \log(2N)$ generations for moderate values of θ : A new mutation A_1 will not, on average, remain in the population for very long, or to attain a high frequency, if there is no recurrent mutation $A_2 \rightarrow A_1$.

The process we are considering, since it admits the possibility of only two alleles, is perhaps of limited interest. However, several of its properties throw considerable light on important features of the infinitely many alleles model (3.72). Some of these were already given in (3.92) and (3.93). It is clear that in the infinitely many alleles model, we may normally expect several low-frequency alleles in the population. For example if $\theta = 1$, $2N = 10^6$, there will typically be about fifteen alleles present in the population at any time, and of these, typically about ten will have a frequency less than 0.01. If θ is small enough, the most likely situation is where there is one allele at high frequency together with several alleles at a very low frequency. This is confirmed by observing, from (3.92), that

$$\begin{aligned} & \text{Prob}(\text{there exists an allele with frequency greater than } 0.99) \\ &= \text{mean number of alleles with frequency greater than } 0.9 \\ &= \theta \int_{0.99}^1 x^{-1}(1-x)^{\theta-1} dx \quad (5.63) \\ &> \approx \theta \int_{0.99}^1 (1-x)^{\theta-1} dx \\ &= (0.01)^\theta. \end{aligned}$$

For $\theta = 0.1$ this probability is about 0.63. For larger values of θ ($\theta > 1$ approximately) it becomes rather unlikely that such a high-frequency allele will exist, and the most likely configuration is one where a number of alleles exist at low but unequal frequencies. In all cases the least likely situation is one where two, three or four alleles exist with approximately equal frequencies. These arguments suggest an approach to testing whether a neutral model such as (3.72) and not, for example, one involving selection is adequate to explain observed allelic frequencies. We take up this question at greater length in Chapter 9 when discussing the concept of polymor-

phism and in Chapter 11 when considering tests for neutrality using allele frequency data.

There are two further points that are of interest in considering the model (3.72) and its evolutionary behavior. The first concerns the nature of the boundary $x = 1$ for the two allele model originally considered. Use of (5.61) in (4.65) shows that this boundary is entrance if $\theta \geq 1$. This implies that in this case, it is impossible to reach this boundary by diffusion from the interior of $(0, 1)$. It is therefore impossible to consider behavior conditional on the requirement that this boundary is reached, and further it is unnecessary to impose the condition that the boundary is not reached and then consider conditional behavior: This latter condition is already implicit and formulas such as (5.62) apply immediately. When $\theta < 1$ the boundary $x = 1$ is regular and hence attainable and now new behavior arises under the condition that this boundary is not reached. Again assuming $p = (2N)^{-1}$ we find that (5.62) must be replaced, conditional on $x = 1$ not being reached, by

$$\bar{t}\{(2N)^{-1}\} = \int_{(2N)^{-1}}^1 2y^{-1}(1-y)^{1-\theta} dy \quad (5.64)$$

generations. The integrand in (5.64) has the usual interpretation that its integral over any frequency range provides the mean time that the allele frequency spends in this range before the allele is eventually lost.

The second point concerns the frequency of the most frequent allele. The argument that led to (5.63) shows that for $0.5 \leq x \leq 1$, the probability density function of the frequency of the most frequent allele in the infinitely many alleles model is, at equilibrium,

$$f(x) = \theta x^{-1}(1-x)^{\theta-1}. \quad (5.65)$$

For values of x less than 0.5 a deeper argument is clearly required: Nevertheless, the probability density function of the most frequent allele can be found for these values also (Watterson and Guess (1977), Watterson (1976b)). Further details are given in Section 5.10. The use of (5.65) to approximate the value of the monomorphism probability P_{mono} is of particular interest to us. This approximation gives

$$P_{\text{mono}} = \text{Prob (only one allele present in the population in any generation at equilibrium)}. \quad (5.66)$$

If we make the approximation

$$P_{\text{mono}} \approx \int_{1-(2N)^{-1}}^1 f(x) dx, \quad (5.67)$$

then immediately from (5.65)

$$P_{\text{mono}} \approx (2N)^{-\theta}. \quad (5.68)$$

This approximation has been made on several occasions in the literature, and in Section 5.7 we examine its accuracy.

In the Moran infinitely many alleles model an exact expression (see (3.99)) is available for P_{mono} . In that expression the definition of θ follows the standard Moran model definition given in (3.98). It is interesting to use this expression, with however θ now defined in the standard Wright–Fisher model form as $\theta = 4Nu$, to obtain a heuristic Wright–Fisher model approximation

$$P_{\text{mono}} \approx (2N - 1)! / \{(1 + \theta)(2 + \theta) \cdots (2N - 1 + \theta)\}. \quad (5.69)$$

Although there is no justification for this heuristic approximation, we shall see later that, surprisingly, (5.69) gives a better approximation to P_{mono} than does the more frequently used (5.68).

5.6 Two-Way Mutation

Suppose now in the model (3.16) that mutation both from A_1 to A_2 (at rate u) and from A_2 to A_1 (at rate v) occurs, with no selection. As we have already seen, there will now exist a stationary distribution for the frequency x of A_1 for which we already have an exact expression (3.25) for the mean and an approximation expression (3.26) for the variance.

Our aim now is to approximate the entire distribution by diffusion methods. The drift and diffusion coefficients are found from (5.6) (putting $\alpha = 0$) and (5.5) and then (4.45) leads to the stationary distribution

$$f(x) = \frac{\Gamma\{2\beta_1 + 2\beta_2\}}{\Gamma\{2\beta_1\}\Gamma\{2\beta_2\}} x^{2\beta_2-1}(1-x)^{2\beta_1-1}. \quad (5.70)$$

The mean and variance of this distribution are $\beta_2/(\beta_1 + \beta_2)$ and $\beta_1\beta_2/\{(\beta_1 + \beta_2)^2(2\beta_1 + 2\beta_2 + 1)\}$ respectively, and these agree with the exact and approximate values given in (3.25) and (3.26), once allowance is made for a change of scale.

This stationary distribution allows a third derivation of (3.27) and (3.28). If $u = v$ and $4Nu = \theta$, then $2\beta_1 = 2\beta_2 = \theta$ and thus

$$f(x) = \frac{\Gamma(2\theta)}{\Gamma(\theta)\Gamma(\theta)} x^{\theta-1}(1-x)^{\theta-1}.$$

From this, the probability that two genes drawn at random from the population are of the same allelic type is

$$\int_0^1 f(x)\{x^2 + (1-x)^2\} = \frac{1 + \theta}{1 + 2\theta}, \quad (5.71)$$

in agreement with (3.27) and (3.28).

The general form of the stationary distribution (5.70) is clear. For small β_1 and β_2 , that is small mutation rates and/or population sizes, most of the probability mass is in the extremes of the distribution, so that the most likely situation is one where one or other allele is at a low frequency or is even temporarily absent from the population. When β_1 and β_2 are large the variance becomes small and the behavior is “quasi-deterministic”: Only small deviations are likely from the deterministic theory equilibrium point. This can be illustrated by supposing $u = v$ so that the mean of the stationary distribution is 0.5 irrespective of the population size. Thus supposing $u = v = 2.5 \times 10^{-6}$, the stationary probability that the frequency of A_1 is between 0.4 and 0.6 rises from 0.2 for $N = 10^5$ to 0.8 for $N = 10^6$ and is essentially unity for $N = 10^7$ or more.

The probability distribution (5.70) allows no atoms of probability at the boundaries $x = 0$, $x = 1$. This is in accordance with what is expected from the coefficients (5.5) and (5.6), which suggest instantaneous reflection from these boundaries. Nevertheless for the discrete process (3.16) there must exist nonzero stationary probabilities for the states $\{0\}$ and $\{2N\}$, and if u and v are small these probabilities will be quite large. It therefore becomes a matter of some interest to find how these boundary probabilities can be approximated from (5.70). This matter is taken up in the next section.

When selection is also allowed, together with two-way mutation, there will still exist a stationary distribution, although its form is naturally more complicated than that in (5.70). Use of the complete expressions (5.5) and (5.6) gives, from (4.45), the formula

$$f(x) = \text{const } x^{2\beta_2-1}(1-x)^{2\beta_1-1} \exp\{2\alpha hx - \alpha(2h-1)x^2\} \quad (5.72)$$

for this distribution, where the constant is a function of β_1 , β_2 , α , and h and may be found in principle by normalization. Since $\beta_1 = 2Nu$, $\beta_2 = 2Nv$, the expression in (5.72) is identical to that in (1.68). The form of this distribution is of most interest when overdominance occurs or when one allele is at a selective advantage to the other. The former case was discussed in some detail below (1.68), so we consider here only the latter. Assume for definiteness that $s < sh < 0$ so that A_1 is at a selective disadvantage and consequently usually at a low frequency. Assuming then that x is small, we may ignore the term $(1-x)^{2\beta_1-1}$ as well as the term in x^2 in the exponent in (5.72) to get

$$f(x) \approx \text{const } x^{2\beta_2-1} \exp(2\alpha hx).$$

Since x is a frequency, its value cannot exceed 1. However for sufficiently large values of $|\alpha h|$ ($|\alpha h| > 3$ should normally suffice), the function $f(x)$ is negligibly small when $x > 1$, and the normalizing constant may be evaluated to a sufficient approximation by supposing $0 \leq x < \infty$. This leads

to

$$f(x) \approx |2\alpha h|^{2\beta_2} \{\Gamma(2\beta_2)\}^{-1} x^{2\beta_2-1} \exp\{-|2\alpha h|x\}. \quad (5.73)$$

From this the mean and variance of the stationary distribution of the frequency of A_1 are found to be, approximately,

$$\beta_2/|\alpha h|, \quad \frac{1}{2}\beta_2/(\alpha h)^2 \quad (5.74)$$

respectively. Allowing for changes in notation, the mean value agrees with the deterministic equilibrium point (1.36), while the variance provides new information and gives some idea of the extent of stochastic variation that can be expected around the mean. Parallel values may be calculated when $s > sh > 0$.

5.7 Diffusion Approximations and Boundary Conditions

The aim of this section is to consider the extent to which various formulae derived from, or suggested by, diffusion methods provide accurate approximations to the true but unknown corresponding values for the Markov chain specified by (3.16) and (3.29). It must of course be kept in mind that this brings us no closer to an evaluation of how close our results are to “reality”: The diffusion process may well provide a better reflection of real population processes than does the Markov chain model.

If one adopts the view that the primary process of interest is a Markov chain such as (3.16), approximating diffusion formulae can usually be obtained in two different ways. The first is to approximate the Markov chain process by a diffusion process by finding the appropriate drift and diffusion coefficients, using the theory of Section 5.1, to calculate the required quantity for this diffusion process, and then to use the value so found as an approximation for the Markov chain. The second way is more direct, and was used several times in Chapter 3: There is no concept of an approximating diffusion process, and the quantity of interest is approximated by considering only the leading terms in a Taylor series expansion. The two approaches give the same formulas (compare, for example, (3.5) and (5.19)), and an extension of the second approach, using higher-order Taylor series approximations, leads to an assessment of the accuracy of diffusion formulas, using standard techniques.

Consider for example the diffusion approximation (5.47) for the probability of fixation of a favored allele A_1 in the absence of dominance and mutation. This equation was also found, as (3.31), without using diffusion theory. Now (3.31) was reached by ignoring certain small-order terms in the deviation of (3.30), and a formula somewhat more accurate than (3.30)

is found from

$$0 = \sum_{i=1}^4 \mathbf{E}(\delta x)^i \pi^{(i)}(x)/i!, \quad (5.75)$$

as this equation incorporates terms of order N^{-2} as well as terms of order N^{-1} . An even better approximation would arise by replacing the derivatives in (5.75) by finite differences. If we now put

$$\pi(x) = \{1 - \exp(-\alpha x)\}\{1 - \exp(-\alpha)\}^{-1} + N^{-1}g(x) \quad (5.76)$$

in (5.75), a second order differential equation for $g(x)$ will be obtained. Since $g(0) = g(1) = 0$, this equation can be solved for $g(x)$ and thus a correction term to $\pi(x)$ of order N^{-1} obtained. More details are given in Ewens (1964).

Similar corrections may be made to the mean absorption times, although here difficulties arise for very large or very small values of p , since the higher derivatives of $\bar{t}(p)$, defined by (3.5) or (5.19), become increasingly large for these p values. Nevertheless, even for $p = (2N)^{-1}$ the diffusion approximation (3.6) is remarkably accurate: A more precise value, found in Fisher (1958, p. 98) and confirmed by Watterson (1975), is

$$\bar{t}\{(2N)^{-1}\} = 1.355076 + 2 \log 2N \quad (5.77)$$

generations. An almost identical correction occurs when there is one-way mutation (cf. (5.62)).

It is also possible to consider corrections to complete distributions arrived at by diffusion theory, in particular to the stationary distribution (4.45). Here it must first be decided in what way the diffusion formula is used as an approximation to a discrete distribution. If the stationary distribution $\{\alpha_i\}$ defined below (3.24) is approximated by a continuous distribution $f(x)$, both approximations

$$\alpha_i \approx \text{const} \int_{(i-1/2)/2N}^{(i+1/2)/2N} f(x) dx, \quad \alpha_i \approx \text{const} f(i/2N) \quad (5.78)$$

could be used. Although in general the two approximations will give similar values for moderate values of i , problems can arise with both definitions at $i = 0$ and $i = 2N$. Thus for the stationary distribution (5.72), the latter definition leads to a zero or infinite value for α_j , both clearly unsatisfactory. However, the first approximation in (5.78) requires adjustments to the terminals of integration for $i = 0$ and $i = 2N$. Further, the integration involved often cannot be completed exactly and numerical methods, which reduce to evaluation of $f(x)$ at discrete point values, are then required. Altogether the best way to view the diffusion approximation is probably to use the second approximation in (5.78) for all values of i other than 0 and

$2N$ and to estimate α_0 and α_{2N} through the stationary equations

$$\alpha_0 = \sum \alpha_j p_{j,0}, \quad \alpha_{2N} = \sum \alpha_j p_{j,2N}. \quad (5.79)$$

The constant is now chosen so that $\sum \alpha_i = 1$. This approach normally leads to quite accurate diffusion approximations to the stationary distribution of Wright–Fisher models.

In considering approximations to α_0 and α_{2N} , Wright (1931, p. 123), (1969, pp. 356–357) replaced (5.79) by the approximation

$$2Nv\alpha_0 \approx \frac{1}{2}\alpha_1, \quad 2Nu\alpha_{2N} \approx \frac{1}{2}\alpha_{2N-1}. \quad (5.80)$$

These approximations were suggested by parallel approximations for the asymptotic conditional distribution ℓ_j (see (1.54)), considered in great detail by Wright and Fisher. In subsequent work Wright did not regard the approximations in (5.80) as necessarily being accurate, and checked carefully in each case whether they were reasonable. Unfortunately these approximations have often been used uncritically by other authors, and this has led to rather inaccurate expressions for α_0 and α_{2N} . Similar uncritical use of the approximation

$$\alpha_0 \approx \int_0^{(2N)^{-1}} f(x) dx, \quad \alpha_{2N} \approx \int_{1-(2N)^{-1}}^1 f(x) dx \quad (5.81)$$

has also led to estimates of large (relative) error.

This latter point may be illustrated by discussing approximations to the quantity P_{mono} , defined in Section 5.5, for the infinitely many alleles model (3.72). The approximation (5.68) for this quantity can be reached by using (5.81) as a starting point: This was essentially the approach of Kimura (1971), who computed the corresponding value in the K -allele model (3.68) and then let $K \rightarrow \infty$. This approach, however, uses diffusion approximations for precisely those values when they are most suspect, and a more detailed computation of P_{mono} is needed. This was provided by Watterson (1975), who arrived at the approximation

$$P_{\text{mono}} \approx \exp(-0.1003\theta)\Gamma(1+\theta)(2N)^{-\theta}. \quad (5.82)$$

We have already noted the approximation (5.69) derived formally by putting $i = 2N$ in (3.78). Table 5.2 displays exact values of P_{mono} for the case $2N = 1000$ found numerically, as well as the approximate values found from (5.82), (5.69), and (5.68). Clearly (5.82) gives an excellent approximation for all values of θ considered while (5.68) gives a good approximation only when $\theta < 1$.

In considering diffusion approximations it has been assumed above that the population size, mutation rate, and selective differences are all such that the parameters α , β_1 , and β_2 in (5.2) are $O(1)$. Not only is this condition rather imprecise: It may also not apply in several cases of interest. An

θ	0.1	0.5	1	5	10
exact	0.472	0.267×10^{-1}	0.902×10^{-3}	0.669×10^{-13}	0.979×10^{-24}
approx. (5.82)	0.472	0.267×10^{-1}	0.905×10^{-3}	0.727×10^{-13}	1.331×10^{-24}
approx. (5.69)	0.477	0.280×10^{-1}	1.000×10^{-3}	1.188×10^{-13}	3.470×10^{-24}
approx. (5.68)	0.501	0.316×10^{-1}	1.000×10^{-3}	0.010×10^{-13}	1.000×10^{-30}

Table 5.2. Values of P_{mono} for $2N = 1000$ and various θ

attempt to overcome this problem has been made by Ethier and Norman (1977), who provide bounds on diffusion approximations irrespective of the order of magnitude assumptions discussed above. More specifically, for the model defined by (3.16) and (3.24), Ethier and Norman provide an explicit upper bound for the difference between the expectation of any infinitely differentiable function of the gene frequency x and the value as calculated from the approximating diffusion process, for any values of N , u and v . This bound is uniform over time and thus applies also to the stationary distribution. For further details see Ethier and Norman (1977), in particular their equation (7).

An interesting case arises for the heterozygosity measure $2x(1 - x)$. The stationary expectation of this quantity for the diffusion process may be found immediately from (5.70). If we assume for convenience that $2\beta_1 = 2\beta_2 = \theta$, the value found for this expectation is $\theta/(2\theta + 1)$. However, the stationary expectation for the Markov chain defined by (3.16) and (3.24) can be found exactly and is, explicitly,

$$\frac{\theta(1 - u)\{1 - (2N)^{-1}\}}{2\theta - 2u\theta + (1 - 2u)^2}.$$

The difference between this and the diffusion approximation is

$$\frac{u\{2 - 2u + \theta\}}{(2\theta + 1)\{2\theta - 2u\theta + (1 - 2u)^2\}}.$$

For the Ethier and Norman theory the upper bound provided for the error in the diffusion approximation is found by applying their equation (7) for the function $2x(1 - x)$. The bound is

$$\max(u, v) + (4N)^{-1} + 27 \max(u^2, v^2) + (7/16)N^{-2}.$$

In our case this may be written

$$u\{1 + \theta^{-1} + 27u + 7u\theta^{-2}\},$$

and it is not hard to verify that this function does indeed bound the exact error in the diffusion approximation given above.

A further remark should be made about order of magnitude assumptions. The diffusions we have considered assume that $E(\delta x)$ for the Markov chain (3.16) is of the same order of magnitude as $\text{var}(\delta x)$, namely N^{-1} . This assumption is not always justified, for example if in the model defined by (3.16) and (3.24) N and u are jointly large enough so that Nu is not small and cannot be taken as of order unity. One case of special interest is that when $E(\delta x)$ is of order ϵ and $N\epsilon$ is large. Processes for which this is the case have been discussed by Karlin and McGregor (1974) and Norman (1974, 1975a). For these processes the gene frequency clusters around its deterministic value given by the infinite population theory outlined in Chapter 1. Deviations of gene frequency from this value are, asymptotically, normally distributed with standard deviation of order $(N\epsilon)^{-1/2}$. For certain parameter values this diffusion approximation and the one we have discussed earlier overlap. This situation is analogous to the overlapping domains of applicability of the Poisson and normal approximations to the binomial distribution in statistical theory, and in such cases the two approximations approximate not only the discrete process but also each other.

We observe in conclusion that diffusion theory can give a quite false impression not only quantitatively but also qualitatively about the boundary behavior in some Markov chains, and illustrate this by considering the stationary distributions (5.72). The criteria in (4.65) shows that the boundary $x = 0$ is unattainable if $\beta_2 \geq \frac{1}{2}$, that is if the mutation rate $A_2 \rightarrow A_1$ is sufficiently large. Further, this conclusion remains unchanged whatever the selective parameters. Suppose now $\alpha \ll \alpha h \ll 0$. The discussion centered around (5.73) shows that the frequency in A_1 will usually be very small, and in a Markov chain model such as that defined by (3.16) and (3.29) there will be a substantial probability that at any time A_1 is absent from the population, in contrast to the diffusion theory prediction. Thus suppose $2N = 10^6$, $v = \frac{1}{2} \times 10^{-6}$, $s = -0.2$, $sh = -0.1$ so that $\beta_2 = 0.5$, $\alpha = -2 \times 10^5$. Equation (5.74) shows that the mean and standard deviation of the number of A_1 genes in the population at any time are 5 and 5, respectively, and this certainly suggests a non-negligible probability that there are no A_1 genes present. The distribution (5.73) implies that for small positive integers i , the stationary probability of i A_1 genes is approximately $0.2 \exp(-0.2i)$, and the approximation (5.79) suggests that the stationary probability of the value 0 is about 0.13. This is not negligibly small, and we conclude that the diffusion theory boundary behavior gives a rather misleading picture of the behavior of the Markov chain for very small numbers, including zero, of A_1 genes.

5.8 Random Environments

So far in this book it has been supposed that stochastic changes in gene frequencies have been brought about solely by random sampling effects in finite populations. There are however further sources of stochastic variation, and perhaps the most important of these is that brought about by random temporal changes in the selection parameters, due perhaps to fluctuations in the environment. Models for this form of randomness were introduced by Kimura (1954), who analyzed (somewhat incorrectly) a model when the selection parameters w_{ij} in the (infinite) population model (1.24) are of the form

$$w_{11} = 1, \quad w_{12} = 1 - s, \quad w_{22} = (1 - s)^2. \quad (5.83)$$

Here s is a random variable with mean $\mu\delta$, variance $\sigma^2\delta$, and with higher moments $O(\delta^2)$ or less. It is assumed that δ is a small parameter, so that small-order terms in δ are ignored.

Let s_t be the value assumed by the random variable s in generation t . Then with a slight change of notation, (1.24) can be written, for this model, as

$$x_{t+1} = x_t / \{1 - s_t(1 - x_t)\}. \quad (5.84)$$

If $y_t = x_t / (1 - x_t)$, this becomes

$$y_{t+1} = y_t(1 - s_t)^{-1}, \quad (5.85)$$

and putting $z_t = \log y_t$, (5.85) leads to

$$z_t = z_0 - \sum_{i=0}^{t-1} \log(1 - s_i). \quad (5.86)$$

Suppose now that the s_i are independently and identically distributed random variables. Then apart from the constant z_0 , z_t is the sum of independently and identically distributed random variables and thus the central limit theorem may be applied. Since

$$\begin{aligned} E\{\ln(1 - s)\} &\approx E\left\{-\left(s + \frac{1}{2}s^2\right)\right\} = -\delta\left(\mu + \frac{1}{2}\sigma^2\right), \\ \text{var}\{\ln(1 - s)\} &\approx \text{var}(-s) = \delta\sigma^2, \end{aligned}$$

z_t will have an approximate normal distribution with mean

$$z_0 + t\delta\left(\mu + \frac{1}{2}\sigma^2\right) = \mu_z \quad (5.87)$$

and variance

$$t\delta\sigma^2 = \sigma_z^2. \quad (5.88)$$

A standard statistical transformation now gives the corresponding density function for x_t as

$$f(x; t) = \frac{1}{\sqrt{2\pi\sigma_z x(1-x)}} \exp \left[-\frac{1}{2} \left\{ \log \frac{x}{1-x} - \mu_z \right\}^2 / \sigma_z^2 \right]. \quad (5.89)$$

This conclusion is reached by more or less exact methods, the only approximation involved being the normal distribution assumption is justified by the central limit theorem. We now show how it can be obtained by diffusion methods. From (5.84),

$$\begin{aligned} E(\delta x) &= E\{sx(1-x) + s^2x(1-x)^2 + 0(s^3)\} \\ &= \delta x(1-x)\{\mu + \sigma^2(1-x)\} + O(\delta^2), \end{aligned} \quad (5.90)$$

$$E(\delta x)^2 = \delta\sigma^2x^2(1-x)^2 + O(\delta^2), \quad (5.91)$$

with higher moments $O(\delta^2)$ or less. Equations (5.90) and (5.91) provide the drift and diffusion coefficients for an approximating diffusion process. Several authors have incorrectly omitted the term in σ^2 in $E(\delta x)$ and thus have obtained incorrect solutions to the diffusion equation. The forward Kolmogorov equation for this process then becomes

$$\begin{aligned} \frac{\partial}{\partial t} f(x; t) &= -\frac{\partial}{\partial x} \{\delta x(1-x)\{\mu + \sigma^2(1-x)\}f(x; t)\} \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{\delta\sigma^2x^2(1-x)^2f(x; t)\}. \end{aligned} \quad (5.92)$$

It may be checked by substitution that the solution of (5.92) is (5.89), and thus the diffusion approximation leads to exactly the same solution as the central limit theorem approximation.

We make several remarks concerning the solution (5.89). First, if $\mu + \frac{1}{2}\sigma^2 > 0$, the density function (5.89) increasingly concentrates near $x = 1$ as $t \rightarrow \infty$, while for $\mu + \frac{1}{2}\sigma^2 < 0$ it concentrates increasingly near $x = 0$. This behavior was termed “quasi-fixation” by Kimura (1954), and was defined more rigorously by Karlin and Liberman (1974) through the concept of “stochastic local stability”. An equilibrium x^* is said to be stochastically locally stable (Karlin and Liberman (1974)) if for any $\gamma > 0$ there exists a neighborhood $x^* - \xi, x^* + \xi$ such that for any initial frequency x_1 in this neighborhood,

$$\text{Prob} \left\{ \lim_{n \rightarrow \infty} x_n = x^* \right\} > 1 - \gamma. \quad (5.93)$$

Thus in the model (5.83) the boundary $x = 1$ is stochastically locally stable if $\mu + \frac{1}{2}\sigma^2 > 0$ and the boundary $x = 0$ is stochastically locally stable if $\mu + \frac{1}{2}\sigma^2 < 0$.

This shows that even when $\mu = 0$, so that “on the average” A_1 and A_2 have equal fitnesses, the density function of the frequency x of A_1 still concentrates increasingly near $x = 1$. This reveals an important new observation: The variance in fitness of any genotype over time is just as important

as the (arithmetic) mean fitness in determining evolutionary behavior. Further, if two alleles have equal arithmetic mean fitnesses, the allele with the smaller variance in fitness is in effect selectively favored. The true fitness is in fact measured best by the geometric mean fitness. In the above case, A_1 and A_2 are selectively equivalent only if $\mu + \frac{1}{2}\sigma^2 = 0$, and this corresponds to the fact that when terms of order δ^2 are ignored, $-(\mu + \frac{1}{2}\sigma^2)$ is the geometric mean selective disadvantage of A_2 .

Finally, the solution (5.89), in contrast to other solutions such as (5.11) of diffusion equations, does not have the form of an eigenfunction expansion. This is confirmed by the theory of Section 4.7. In the above process, application of (4.65) shows that the boundaries $x = 0$, $x = 1$ are natural, and for such boundaries it is not necessary that the solution be in eigenfunction form.

A second selection scheme perhaps reveals more of the flavor of random environment models. Suppose the fitnesses in each generation are of the form

$$w_{11} = 1 + s, \quad w_{12} = 1, \quad w_{22} = 1 + \alpha s \quad (5.94)$$

where $0 < \alpha \leq 1$. This model has been studied in detail by Karlin and Liberman (1974) in discrete time and Levikson and Karlin (1975) in the diffusion case. The conclusions are analogous in the two situations, and we present only several discrete-time results. We again assume s is a random variable having mean $\delta\mu$, variance $\delta\sigma^2$, and higher moments $O(\delta^2)$ or less.

Suppose first that $\alpha = 1$. In this case the geometric mean fitness of the heterozygote is 1, whereas that of each homozygote is, to the order of approximation used, $1 + \delta\mu - \frac{1}{2}\delta\sigma^2$. If $\mu = 0$ the homozygotes have the same arithmetic mean fitness as the heterozygote but a lower geometric mean fitness and thus, from the previous discussion, can be regarded as being at a selective disadvantage to the heterozygote. This is confirmed by Karlin and Liberman (1974), who show that with probability one, each trajectory of gene frequency converges to 0.5. This behavior occurs even for positive μ so long as $\mu - \frac{1}{2}\sigma^2 < 0$. Thus 0.5 is stochastically locally stable for this case.

When $\mu - \frac{3}{4}\sigma^2 < 0 < \mu - \frac{1}{2}\sigma^2$, it is possible for the frequency of A_1 to converge to 0, to 0.5, or to 1. Which type of convergence occurs depends on the initial frequency of A_1 . If $\mu - \frac{3}{4}\sigma^2 < 0$ the only two limiting possibilities are convergence of the frequency of A_1 to 0 or 1. In all cases the actual convergence behavior is not deterministic in the sense that it will depend on the values taken by s in the early generations.

When $\alpha < 1$ the picture is far more complex, and it is then possible in suitable circumstances that a stationary distribution for the frequency of A_1 can arise. The condition for this is that

$$3\alpha/(1 + \alpha) < 2\mu/\sigma^2 < 1 \quad (5.95)$$

and that the initial frequency x_0 of A_1 be between $\alpha/(1 + \alpha)$ and 1. (The condition (5.95) requires $\alpha < \frac{1}{2}$.) The condition (5.95) follows from equation (4.8) in Karlin and Liberman (1974) but is rather less general than their condition, applying only when the moments of s have the properties we have assumed. The drift and diffusion coefficients for the model (5.94) are

$$\begin{aligned} a(x) &= x(1-x)\{(1+\alpha)x - \alpha\}\{\mu - \sigma^2(x^2 + \alpha(1-x)^2)\}, \\ b(x) &= \sigma^2 x^2(1-x)^2\{(1+\alpha)x - \alpha\}^2. \end{aligned} \quad (5.96)$$

If these values are formally inserted into the stationary distribution formula (4.45), the result is

$$f(x) = \text{const } x^{-2\mu/\alpha\sigma^2} (1-x)^{-2\mu/\sigma^2} \{x - \alpha/(1+\alpha)\}^{2(1+\alpha)\mu/\alpha\sigma^2 - 4} \quad (5.97)$$

which, for the parameters specified by (5.95), is integrable over $(\alpha/(1 + \alpha), 1)$. Is it however the required stationary distribution? Tanaka (1957) has shown that for diffusion with inaccessible boundaries formal calculation of $f(x)$ in this way does indeed provide the correct stationary distribution, provided that the resulting function is integrable and that $\psi(x)$, defined by (4.16), is non-integrable at both boundaries. We have already checked the former condition and since in this case

$$\psi(x) = \text{const } x^{-2+2\mu/\alpha\sigma^2} (1-x)^{-2+2\mu/\sigma^2} \{x - \alpha/(1+\alpha)\}^{2-2(1+\alpha)\mu/\alpha\sigma^2}$$

we see that the second requirement is also satisfied. Further, (4.65) shows that both $x = 1$ and $x = \alpha/(1 + \alpha)$ are inaccessible and thus (5.97) does provide the required stationary distribution. The distribution (5.97) also applies for the diffusion case (Levikson and Karlin (1975)).

The choice $\alpha = -1$ implies that fitnesses in (5.94) are additive. For fixed additive fitnesses the situation is straightforward: A_1 will become fixed when $s > 0$ and A_2 will become fixed when $s < 0$. When s is a random variable, however, neither the boundary $x = 0$ nor the boundary $x = 1$ is necessarily stochastically locally stable in this additive case. The respective conditions that the boundary $x = 0$ and the boundary $x = 1$ be stochastically locally stable are, for small s ,

$$-\mu - \frac{1}{2}\sigma^2 > 0, \mu - \frac{1}{2}\sigma^2 > 0, \quad (5.98)$$

respectively. These conditions show that if the variance of s is sufficiently large compared to the mean of s , neither the boundary $x = 0$ nor the boundary $x = 1$ is stochastically locally stable.

A model more general than (5.83) and (5.94) allows the fitness of the various genotypes to be

$$\begin{array}{ccc} A_1 A_1 & A_1 A_2 & A_2 A_2 \\ 1 + \eta_n & 1 & 1 + \epsilon_n \end{array} \quad (5.99)$$

Since only ratios of fitnesses are relevant we do not lose generality by fixing the fitness of $A_1 A_2$ at unity, and this is done in (5.99). The fitnesses η_n and

ϵ_n are the realized values, in generation n , of the random variables η and ϵ . These variables are possibly correlated for any given n (for example, the model (5.94) is a special case of (5.99), and for the case $\alpha = 1$ in that model, $\eta_n \equiv \epsilon_n$). On the other hand it is assumed that (η_n, ϵ_n) are independent of (η_m, ϵ_m) when $n \neq m$. We assume throughout that

$$E(\eta) = \mu_1\delta, \quad \text{var}(\eta) = \sigma_1^2\delta, \quad E(\epsilon) = \mu_2\delta, \quad \text{var}(\epsilon) = \sigma_2^2\delta \quad (5.100)$$

for some small parameter δ , while higher moments of η and ϵ are $O(\delta^2)$ or less.

Let x_n be the frequency of A_1 in generation n . Our objective is to determine how properties of the random sequence x_1, x_2, x_3, \dots depend on the joint distribution of η and ϵ . The starting point for doing this is the recurrence relationship

$$x_{n+1} = x_n(1 + \eta_n x_n) / \{1 + \eta_n x_n^2 + \epsilon_n(1 - x_n)^2\}. \quad (5.101)$$

The concept of stochastic local stability was defined in connection with (5.93). For the model (5.99), the boundary $x = 0$ is stochastically locally stable if $E \log(1 + \epsilon) > 0$ and the boundary $x = 1$ is stochastically locally stable if $E \log(1 + \eta) > 0$. In other words, since

$$E \log(1 + \eta) = (\mu_1 - \frac{1}{2}\sigma_1^2)\delta + O(\delta^2), \quad (5.102)$$

the condition that $x = 1$ be stochastically locally stable is that $\mu_1 - \frac{1}{2}\sigma_1^2 > 0$, showing the importance of the variance in determining stochastic local stability behavior at $x = 1$. A parallel remark applies for $x = 0$.

These conclusions confirm those reached above for the model (5.83). In this model the boundary $x = 1$ was claimed to be stochastically locally stable when $E \log(1 - s) < 0$ and the boundary $x = 0$ was claimed to be stochastically locally stable if $E \log(1 - s) > 0$. The selective parameters in the model (5.83) are equivalent to $(1 - s)^{-1}$, 1 and $1 - s$, and these are particular cases of the parameters in (5.99). The stochastic local stability behavior of this model then agrees with that of the more general model (5.99).

We now turn to finite populations. Here stochastic fluctuations in gene frequency occur for two reasons, random sampling effects and stochastic variation in fitness. We consider as an example the behavior of diploid models with fitness scheme (5.99). If $E(\eta) = E(\epsilon)$, $\text{var}(\eta) = \text{var}(\epsilon) = \sigma^2$, $\text{corr}(\eta, \epsilon) = r$ and A_1 mutates to A_2 at rate u with reverse mutation also at rate u , there will exist a symmetric stationary distribution about $x = 0.5$ of the frequency of A_1 for any finite population size N . This distribution can be found explicitly (Avery (1977)) and increasingly concentrates near $x = \frac{1}{2}$ as σ^2 increases. In other words, increasing the variance of the homozygote fitnesses increases the degree of heterozygosity in the population. The degree of heterozygosity also increases with r . These interesting and important conclusions differ qualitatively from those found by incorrect analyses of Kimura (1954) and Ohta (1972).

If there is no mutation, interest centers on fixation probabilities and mean fixation times. Avery (1977) found that the probability of fixation of a new mutant increases substantially as σ^2 increases, whereas it is almost independent of the value of r . The conditional mean times to fixation and to loss also increase substantially with σ^2 and with r : Again, the earlier conclusions of Kimura and Ohta are qualitatively incorrect. The general observation in all cases is that increasing the variance of homozygote fitness tends to increase genetic variation in populations, often by substantial amounts. This observation is clearly of some relevance in possibly explaining at least part of the large genetic variation observed in natural populations.

We turn now to spatial variation. Although normally this is considered when temporal variation also occurs, we first consider a model, due to Levene (1953), for which fixed fitness regimes occur in each of M habitats, the fitnesses in habitat i being $1 + \eta^{(i)}$, 1 , $1 + \epsilon^{(i)}$. In this "fixed Levene model" the entire population mates at random in a common area and then disperses to the various habitats, a fraction c_i entering habitat i . The recurrence relation for the frequency x of A_1 is

$$x_{n+1} = \sum_i c_i \{x_n + \eta^{(i)} x_n^2\} / \{1 + \eta^{(i)} x_n^2 + \epsilon^{(i)} (1 - x_n)^2\}. \quad (5.103)$$

The equilibrium solutions of this system are of particular interest. Karlin (1977b) proved that in general at most three internal equilibria of (5.103) can exist, and that in certain cases at most one can exist. In particular, when

$$\eta^{(i)} + \epsilon^{(i)} + \eta^{(i)} \epsilon^{(i)} \leq 0 \quad (5.104)$$

for $i = 1, 2, 3, \dots, M$, at most one internal equilibrium exists, and if such an equilibrium exists does exist it is globally stable. If no such equilibrium exists there is a unique globally stable fixation point.

One case where (5.104) plainly applies is where $\eta^{(i)} < 0$, $\epsilon^{(i)} < 0$ for all i . Here there always exists a unique globally stable internal equilibrium, analogous to (1.31). A second case where (5.104) holds is the linear scheme $\eta^{(i)} = -\epsilon^{(i)}$. Whether or not an internal equilibrium exists depends in a complicated way on the $\eta^{(i)}$ and c_i values. A parallel remark holds when fitnesses are of the multiplicative form for which $1 + \epsilon_i = (1 + \eta_i)^{-1}$. The location of the equilibrium point must be found numerically from the recurrence relation (5.103).

As might be expected, if $1 + \eta^{(i)} > 1 > 1 + \epsilon^{(i)}$ for all i , A_1 becomes fixed in the population, with a corresponding conclusion for A_2 when $1 + \eta^{(i)} < 1 < 1 + \epsilon^{(i)}$.

We consider finally models involving both spatial and temporal variation. Gillespie (1974a,b, 1976a,b, 1977a,b, 1978) and Gillespie and Langley (1974, 1976) have analyzed an increasingly complex series of such models, culminating in a "stochastically additive scale, concave fitness function" (SAS-CFF) model, that lead to broad predictions concerning natural pop-

ulations that fit well with observed genetic patterns. While various other spatial-temporal models also have to be considered, we concentrate here on the Gillespie–Langley model and its practical implications.

Consider first a single population with fitnesses of the additive form

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ 1 + s & 1 + \frac{1}{2}(s + t) & 1 + t \end{array} \tag{5.105}$$

Here the values of s and t vary from generation to generation according to a stochastic process for which

$$E(s) = \mu_s, \text{ var}(s) = \sigma_s^2, E(t) = \mu_t, \text{ var}(t) = \sigma_t^2, \text{ corr}(s, t) = \rho. \tag{5.106}$$

Values of s and t at different time points are assumed to be independent. By suitably amending (5.98) or by direct argument, Gillespie (1974a) proved that a polymorphic stationarity distribution for the frequency of A_1 exists if $|\alpha| < 1$, where

$$\alpha = 4(\mu_s - \mu_t + \frac{1}{2}(\sigma_t^2 - \sigma_s^2)) / \{\sigma_s^2 + \sigma_t^2 - 2\rho\sigma_s\sigma_t\}. \tag{5.107}$$

He further showed that when the stationary distribution of the frequency x of A_1 exists, it is of the form

$$f(x) = \text{const } x^\alpha(1 - x)^{-\alpha}. \tag{5.108}$$

We next consider a random Levene model which is identical to the fixed Levene model considered above, except that the fitnesses in any subpopulation vary randomly through time. Suppose that at time n the fitnesses in the i th subpopulation are $1 + s_n^{(i)}$, $1 + \frac{1}{2}(s_n^{(i)} + t_n^{(i)})$, $1 + t_n^{(i)}$, respectively. These fitnesses are assumed to be independent over time but not necessarily from one subpopulation to another at any given time. Then (Gillespie (1974a)) the appropriate generalization of the above result is that a stationary polymorphic distribution for the frequency x of A_1 exists if and only if

$$4|\mu_s - \mu_t + \frac{1}{2}(\sigma_s^2 - \sigma_t^2)| / \{(\sigma_s^2 + \sigma_t^2 - 2\rho\sigma_s\sigma_t)(1 + \pi(1 - k))\} < 1, \tag{5.109}$$

where

$$\pi = 2 \sum_{i>j} c_i c_j, \quad k = (\sigma_{ss} + \sigma_{tt} - 2\sigma_{sst}) / (\sigma_s^2 + \sigma_t^2 - 2\rho\sigma_s\sigma_t),$$

$\sigma_{ss}(\sigma_{tt}) =$ covariance of the $s(t)$ values in two sub-populations at the same time,

$\sigma_{sst} =$ covariance of the s value from one subpopulation with the t value from another subpopulation at the same time.

The stationary distribution is again of the form given in (5.108). Increased patchiness as measured by increased π increases the probability that polymorphism will arise, as does a decrease in the spatial correlations. For

completely correlated environments $k = 1$, in which case the condition (5.109) reduces to $|\alpha| < 1$. The means and variances of s and t are both important determinants of whether or not a stationary polymorphic distribution exists.

All the above assumes additive fitnesses. Gillespie (1976a,b, 1978) and Gillespie and Langley (1974) argue more generally that it is more likely that the enzymatic activities of the various genotypes are additive and that the fitnesses of any genotype are concave functions of these activities. Under this argument the activities of the three genotypes are of the form z_1 , $\frac{1}{2}(z_1 + z_2)$, z_2 and the fitnesses are $\phi(z_1)$, $\phi(\frac{1}{2}(z_1 + z_2))$, $\phi(z_2)$, where $\phi(z)$ is a concave function satisfying

$$\phi'(z) > 0, \quad \phi''(z) < 0, \quad \lim_{z \rightarrow \infty} \phi(z) = K < \infty. \quad (5.110)$$

Perhaps the simplest function having these properties is

$$\phi(z) = (1 + c)z / (c + z). \quad (5.111)$$

It is clear that the probability of polymorphism in this SAS-CFF model is increased compared to that in the cases when the fitnesses themselves are additive, since $\phi(\frac{1}{2}(z_1 + z_2)) > \frac{1}{2}\phi(z_1) + \frac{1}{2}\phi(z_2)$. If z_1 and z_2 have means μ_1 and μ_2 , common variance σ^2 , correlation ρ the polymorphism requirement generalizing (5.109) is

$$2|\mu_1 - \mu_2| / [\sigma^2(1 - \rho)\{\phi'(1)(1 + \pi(1 - k)) - \phi''(1)/\phi'(1)\}] < 1, \quad (5.112)$$

where k and π are as defined above, the covariances σ_{ss} , σ_{tt} and σ_{sstt} referring to the distribution of the z 's. The stationary distribution, when it exists, is a beta distribution generalizing (5.108).

Gillespie (1977a, pp. 305–311) demonstrated that the predictions of this model, especially for $\phi(z)$ of the form (5.111), fit in remarkably well with a large series of observations of natural polymorphisms. We do not pursue these comparisons here, nor the generalizations of the model to the multilocus case where the expected nature of observed linkage disequilibrium are also discussed, since they go beyond the purely mathematical analyses that are our primary concern.

We conclude this section by observing that several diffusion processes discussed in this section possess natural boundaries. Thus since regular, exit and entrance boundaries have already been encountered in previous sections, we have seen that even rather simple genetic models can lead to all four of the boundary classifications given in (4.65).

5.9 Time-Reversal and Age Properties

It has been remarked on several occasions earlier that information about the past behavior of diffusion processes allowing a stationary distribution can

be obtained by determining properties of the future behavior. We should therefore be able to use some of the conclusions reached above to discuss past behavior of various processes, and in particular to find properties of the “age” of an allele.

The time-reversal property states that for any diffusion on $[0, 1]$ admitting a stationary distribution, the probability of any sample path leading from x (at time 0) to y (at time t) is equal to that of the “mirror-image” path leading from y (at time $-t$) to x (at time 0). Unfortunately this observation is not immediately useful for several questions of interest in population genetics, since these questions refer to processes for which either the boundary $\{0\}$, or the boundary $\{1\}$, or both, are accessible absorbing states of the diffusion process, and thus for which no stationary distribution exists. This problem can be overcome in the following way.

Suppose that $\{0\}$ is an absorbing state but that $\{1\}$ is not: This will occur in practice, for example, if there is mutation from A_1 to A_2 but no reverse mutation. Now introduce mutation from A_2 to A_1 at rate ϵ : A stationary distribution now exists and reversibility arguments apply. In particular, given a current value x for the frequency of A_1 , the distribution of the time (in the future) until $\{0\}$ is next reached is identical to that of the time (in the past) that it was last left. Now let $\epsilon \rightarrow 0$: The distribution of the time (in the future) until the frequency reaches 0 converges to that applying when $\epsilon = 0$. The limiting distribution is then identical to the age distribution of an allele which arose as a unique new mutation and whose current frequency is x . This argument can be made more precise (Watterson (1977b), Levikson (1977a, b)) by introducing a “return” process whereby the frequency of A_1 is returned from 0 to δ ($\delta > 0$) whenever 0 is reached: In practice we put $\delta = (2N)^{-1}$ to correspond to the frequency of a new mutant. We now give some examples of the conclusions reached by this argument.

Consider first the case of no selection or mutation. Assume the allele A_1 arose by a unique mutation in an otherwise pure A_2A_2 population and is now observed with frequency x . The distribution of its age is thus the distribution of its time until loss, conditional on the event that eventual loss does occur. This distribution can be found by centering attention on A_2 (with current frequency $1 - x$) rather than A_1 , and is then given by (5.39) with $p = 1 - x$. The mean age can be found either through this distribution or alternatively by replacing p by $1 - x$ in (5.34). This leads to a neutral theory mean age of $-4Nx(1 - x)^{-1} \log x$ generations. The variance of the age is found by putting $p = 1 - x$ in (5.37).

A parallel formula can be found when we assume fitness values $1 + s$ for A_1A_1 , $1 + \frac{1}{2}s$ for A_1A_2 and 1 for A_2A_2 . Use of (5.54) (with $p = x$) shows that the mean age of A_1 , given that it is currently observed with frequency

x , is

$$\begin{aligned} & \int_0^x 4N[\alpha\{e^\alpha - 1\}]^{-1}\{e^{\alpha y} - 1\}\{e^{\alpha(1-y)} - 1\}\{y(1-y)\}^{-1} dy \\ & + \int_x^1 4N\{1 - e^{-\alpha x}\}[\alpha\{1 - e^{-\alpha}\}\{e^{\alpha(1-x)} - 1\}]^{-1}e^{-\alpha(1-y)}\{e^{\alpha(1-y)} - 1\}^2 \\ & \quad \times \{y(1-y)\}^{-1} dy \end{aligned} \quad (5.113)$$

generations. This converges to the neutral theory expression as $\alpha \rightarrow 0$, as we expect, and the form of the integrand allows calculation of the mean time, in the past, that the frequency of A_1 assumed a value in any arbitrary interval (y_1, y_2) .

Suppose now A_1 mutates to A_2 at rate u with no reverse mutation. If one initial A_1 gene occurred by a unique mutation and the frequency of A_1 is currently observed at x , the mean age of A_1 is, from (3.19),

$$\begin{aligned} & 4N(1-\theta)^{-1} \int_0^x y^{-1}\{(1-y)^{\theta-1} - 1\} dy \\ & + 4N(1-\theta)^{-1}\{1 - \{(1-x)^{1-\theta}\}\} \int_x^1 (1-y)^{\theta-1} dy \end{aligned} \quad (5.114)$$

generations. A case of particular interest is that for which $x = 1$, corresponding to temporary fixation of A_1 : This evaluation is allowed only when $\theta < 1$.

It is also possible to consider the mean age of A_1 conditional on the requirement that the frequency of A_1 was never unity in the past. This is identical to the mean time for loss of A_1 given that its future frequency never achieves the value unity. This is given by (5.114) for $\theta > 1$, since then the condition that the frequency of A_1 never reaches unity is automatically satisfied. For $\theta < 1$ the probability that the frequency of A_1 never reaches unity given a current value x is found from (4.15) to be $(1-x)^{1-\theta}$. Use of (4.51) and (4.52) then shows that the conditional mean age of A_1 is

$$\begin{aligned} & \int_0^x 4Ny^{-1}(1-\theta)^{-1}\{1 - (1-y)^{1-\theta}\} dy \\ & + \int_x^1 4Ny^{-1}(1-\theta)^{-1}(1-y)^{1-\theta}\{(1-x)^{\theta-1} - 1\} dy. \end{aligned} \quad (5.115)$$

This reduces to the expression (5.64) for $x = (2N)^{-1}$. Equation (5.115) was first given by Kimura and Ohta (1973), using an approach different from that just given.

The conclusions reached above can be reached in a different way (Sawyer (1977)). If the Markovian random variable is approximated by a diffusion variable, the appropriate drift and diffusion coefficients are

$$a(x) = 1 - x - \frac{1}{2}\theta x, \quad b(x) = x(1 - x). \quad (5.116)$$

In Sawyer's approach the random variable is moved at random times immediately to the boundary $x = 0$. These moves occur at times that are independent of the current value of the variable, and the times between consecutive moves are independent random variables with density function

$$f(t) = \frac{1}{2}\theta \exp(-\frac{1}{2}\theta t), \quad 0 \leq t < \infty. \quad (5.117)$$

This approach to finding age properties arises because, if a given gene in the ancestral line considered is a new mutant, the frequency of its allelic type in the population is $(2N)^{-1}$ when it first occurs, irrespective of the frequency x of the allelic type of its parent gene in the previous generation. Here the boundary $x = 1$ is not absorbing while the boundary $x = 0$ is not accessible, that is, it cannot be reached by drift from the interior of $(0, 1)$, although we have seen this boundary can be reached because of the discrete moves in the process. The frequency x of the allelic type in the ancestral line has a stationary distribution which is found by Sawyer using renewal theory arguments. Sawyer found that

$$\lim_{t \rightarrow \infty} \text{Prob}\{x_t \leq p\} = 1 - (1 - p)^{\theta/2}, \quad (5.118)$$

and the reversibility argument shows immediately that the distribution of frequencies seen in the past is identical to that to be seen in the future, given a current frequency x . From this observation we can derive the expression (5.114) as well as find further age properties of the process.

We conclude this section by discussing the various symmetry properties in Section 5.4 in the light of time-reversal arguments. We take as our starting point the observation that any two of the equations (5.57), (5.59) and (5.60) imply the remaining equation. Now (5.59) is true by direct argument, so the observation of prime interest is that (5.57) implies (5.60) and vice versa. Our starting point in Section 5.4 was that (5.57) was true by computation of both sides in the formula, and this then implies the truth of (5.60). Our starting point here is that time-reversal arguments imply the truth of (5.60). This is true because the reflection of a path leading from ϵ (ϵ small) to $1 - \epsilon$ is a path leading from $1 - \epsilon$ to ϵ . The identity of the probability of the two paths leads directly to (5.60), and we conclude that time-reversibility implies (5.60) and hence (5.57) and (5.58). The further facts that $t^*(x; p)$ defined in (5.53) is symmetric about 0.5 and independent of the sign of α do not appear to follow directly from time-reversal arguments, although use of (5.57), which we have seen can be arrived at through time-reversal arguments, shows that when $h = \frac{1}{2}$ either property implies the other.

5.10 Multi-Allele Diffusion Processes

In this section we consider diffusion approximations to finite Markov chain K -allele models of the form (3.68).

The simplest version of the model (3.68) arises when the function ψ_i takes the value $X_i/2N$. It is clear that in this Markov chain model the probability of fixation of any allele is initial frequency, and we also have the eigenvalue formula (3.69) concerning the rate of decrease of the probability that j or more alleles exist at time t . To obtain further results we turn to the diffusion approximation to (3.68).

We write $x_i = X_i/2N$ ($i = 1, 2, \dots, K - 1$) and let δx_i be the change in x_i from one generation to the next. Then elementary theory shows that given x_1, \dots, x_{K-1} ,

$$E(\delta x_i) = 0, \quad \text{var}(\delta x_i) = (2N)^{-1}x_i(1-x_i), \quad \text{covar}(\delta x_i, \delta x_j) = -(2N)^{-1}x_i x_j.$$

These parameters, in conjunction with (4.73), lead to the following partial differential equation for the joint density function $f = f(x_1, \dots, x_{K-1}; t)$ of x_1, \dots, x_{K-1} at time t , where unit time corresponds to $2N$ generations:

$$\begin{aligned} \frac{\partial f}{\partial t} = & \frac{1}{2} \sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2} \{x_i(1-x_i)\} \\ & - \sum_{i < j} \sum \frac{\partial^2 f}{\partial x_i \partial x_j} \{x_i x_j\}. \end{aligned}$$

This is a generalization of (5.10) and admits an eigenfunction solution generalizing (5.11), which has been found by Littler and Fackerall (1975) and Griffiths (1979c). The corresponding backward equation is

$$\frac{\partial f}{\partial t} = \frac{1}{2} \sum_i p_i(1-p_i) \frac{\partial^2 f}{\partial p_i^2} - \sum_{i < j} \sum p_i p_j \frac{\partial^2 f}{\partial p_i \partial p_j},$$

where p_i is the initial value of x_i . This equation may be used to find various fixation probabilities. The probability π ($= \pi(p_1, p_2, \dots, p_{K-1})$) of any fixation event satisfies

$$\frac{1}{2} \sum p_i(1-p_i) \frac{d^2 \pi}{dp_i^2} - \sum_{i < j} \sum p_i p_j \frac{d^2 \pi}{dp_i dp_j} = 0, \quad (5.119)$$

subject to the appropriate boundary conditions. For example, the probability that A_i eventually fixes satisfies (5.119) together with the boundary conditions

$$\begin{aligned} \pi(p_1, \dots, p_{K-1}) &= 1 \quad \text{if } p_i = 1, \\ \pi(p_1, \dots, p_{K-1}) &= 0 \quad \text{if } p_j + p_m + \dots + p_u = 1 \quad (j, m, \dots, u \neq i). \end{aligned}$$

The solution of these equations is $\pi = p_i$, which we know also to be exactly correct for the model (3.68) with $\psi_i = X_i/2N$. Suppose now that we wish

to find the probability π that ultimately A_i and A_j are the last two alleles to exist. Here the boundary conditions are

$$\begin{aligned} \pi(p_1, \dots, p_{K-1}) &= 1 && \text{if } p_i + p_j = 1, \\ \pi(p_1, \dots, p_{K-1}) &= 0 && \text{if } p_m + p_s + \dots + p_u = 1 \quad (m, s, \dots, u \neq i, j), \end{aligned}$$

and the solution of (5.119) satisfying these conditions is

$$\pi = p_i p_j \{ (1 - p_i)^{-1} + (1 - p_j)^{-1} \}.$$

In the case $K = 3$ this shows, for example, that the probability that A_1 is the first allele to be lost is

$$p_2 p_3 \{ (1 - p_2)^{-1} + (1 - p_3)^{-1} \}.$$

Similar probabilities may be found for other fixation events.

We turn now to questions concerning the time until various fixation events occur. The development is easiest when $K = 3$, so we discuss the analysis in detail in this case only, and quote results for larger values of K . Complete details are available in Littler (1975).

Define T_i as the time required until exactly i ($i = 1, 2$) alleles exist in the population. We first find an expression for $E(T_1)$. Conditional on the event that A_1 is the last remaining allele, the mean of T_1 is

$$E(T_1 | A_1) = 2p_1^{-1}(1 - p_1) \log(1 - p_1),$$

from (5.34). Since the probability is p_1 that indeed A_1 is the last remaining allele we have

$$E(T_1) = -2[(1 - p_1) \log(1 - p_1) + (1 - p_2) \log(1 - p_2) + (1 - p_3) \log(1 - p_3)].$$

Clearly this value can be extended immediately to the case of K alleles to get

$$E(T_1) = -2 \sum (1 - p_i) \log(1 - p_i). \tag{5.120}$$

It is equally straightforward to use the analysis leading to (5.37) to find an expression for the variance of T_1 .

In the three-allele case we find $E(T_2)$ as follows. The event $T_2 \leq t$ implies that at time t , at least one p_i value is zero. Standard probabilistic formulas for unions of events give

$$E(T_2) = -2 \left[\sum p_i \log p_i + \sum (1 - p_i) \log(1 - p_i) \right]. \tag{5.121}$$

In the particular case $p_i = 1/3$ these formulas give

$$E(T_1) \sim 3.2N \text{ generations}, \quad E(T_2) \sim 1.6N \text{ generations}.$$

Littler (1975) found that in the K -allele case,

$$E(T_i) = -2 \left(\sum_{s=1}^i (-1)^{i-s} \binom{K-1-s}{i-s} \left(\sum (1 - p_{i_1} - \dots - p_{i_s}) \right) \times \log(1 - p_{i_1} - \dots - p_{i_s}) \right) \quad (5.122)$$

where the inner sum is taken over all possible values $1 \leq i_1 < i_2 < \dots < i_s \leq K$. This reduces to (5.120) when $i = 1$ and generalizes (5.121) to arbitrary K when $i = 2$. It is of some interest to note that if $p_i = K^{-1}$,

$$\lim_{K \rightarrow \infty} E(T_j) = 2/j, \quad j = 1, 2, \dots \quad (5.123)$$

This conclusion may be compared to the ‘‘eigenvalue’’ expression (3.69), and this comparison shows that the eigenvalues give a poor indication of the way in which $E(T_j)$ changes as a function of j .

A further asymptotic result of interest is given by Littler (1975). If $M^{(t)}$ is the number of alleles present at time t in a K -allele model with $p_i = K^{-1}$, then

$$E(M^{(t)}) \sim 1 + 3e^{-t} + 5e^{-3t} + 7e^{-6t} + \dots + (2j + 1)e^{-\frac{1}{2}j(j+1)t} + \dots \quad (5.124)$$

as $K \rightarrow \infty$.

Suppose now A_i mutates to A_j at positive rate u_{ij} ($i \neq j$). It follows that

$$\begin{aligned} E(\delta x_i) &= -x_i \sum_j u_{ij} + \sum_j x_j u_{ji} \\ &= (2N)^{-1} m_i(x_1, \dots, x_{K-1}) \end{aligned}$$

say, where

$$m_i(x_1, \dots, x_{K-1}) = -x_i \sum \beta_{ij} + \sum x_j \beta_{ji}$$

and $\beta_{ij} = 2Nu_{ij}$. If each $u_{ij} > 0$, ($i \neq j$), there will exist a stationary distribution for the joint frequency of A_1, \dots, A_{K-1} . This distribution has not been found in general, although it must clearly satisfy the stationarity equation

$$\begin{aligned} 0 &= - \sum \frac{\partial}{\partial x_i} \{m_i(x_1, \dots, x_{K-1})f(x_1, \dots, x_{K-1})\} \\ &\quad + \frac{1}{2} \sum \frac{\partial^2}{\partial x_i^2} \{x_i(1 - x_i)f(x_1, \dots, x_{K-1})\} \\ &\quad - \sum_{i < j} \sum \frac{\partial^2}{\partial x_i \partial x_j} \{x_i x_j f(x_1, \dots, x_{K-1})\}. \end{aligned} \quad (5.125)$$

Fortunately, in one case of special interest, (5.125) can be solved explicitly. This is the equal-mutation model for which u_{ij} is independent of i and j .

Suppose

$$u_{ij} = u/(K - 1),$$

so that u is the total mutation rate for any gene. Then the appropriate solution of (5.125) is

$$f(x_1, \dots, x_{K-1}) = \frac{\Gamma\{K\epsilon\}}{\{\Gamma(\epsilon)\}^K} \{x_1 x_2 \cdots x_K\}^{\epsilon-1}, \quad (5.126)$$

where we write for convenience $x_K = 1 - x_1 - x_2 - \cdots - x_{K-1}$ and $\epsilon = 4Nu/(K - 1)$. This (Dirichlet) distribution arises in several areas of statistics, and its properties are well known. It may be used to re-derive various formulae already found by other methods and, as we soon note, to find new ones. Thus the probability that two genes drawn at random are of the same allelic type is

$$\sum_{i=1}^K \int \cdots \int x_i^2 f(x_1, \dots, x_{K-1}) dx_{K-1},$$

and this leads, after some calculation, to the expression in (3.70).

Suppose now the gene frequencies are arranged in decreasing order

$$x_{(1)} \geq x_{(2)} \geq x_{(3)} \geq \cdots \geq x_{(K)} \geq 0. \quad (5.127)$$

These frequencies are called the order statistics of x_1, x_2, \dots, x_k . Their joint distribution is, directly from (5.126),

$$f(x_{(1)}, \dots, x_{(K-1)}) = \frac{K! \Gamma\{K\epsilon\}}{\{\Gamma(\epsilon)\}^K} \{x_{(1)} x_{(2)} \cdots x_{(K)}\}^{\epsilon-1}. \quad (5.128)$$

From this distribution the joint distribution of the first j order statistics $x_{(1)}, x_{(2)}, \dots, x_{(j)}$ may be found, although we do not give the formula here.

The limiting case $K \rightarrow \infty$ is of special interest. While the joint distribution (5.126) has no nontrivial limit, Kingman (1975, 1977b) found that the distribution of the first j order statistics *does* converge, for any j , to a nontrivial limit, called by Kingman the Poisson–Dirichlet distribution. (A more appropriate name is the *Kingman* distribution.)

The form of this distribution coincides with the joint distribution of the first j order statistics in the infinitely many alleles model. This remarkable result is most important as it allows us, so long as we concentrate on order statistics and functions derived from them, to move freely between the K -allele model in Section 3.5 and the infinitely many alleles model of Section 3.6. This allows us to approach certain problems in the infinitely many alleles model either directly or through the K -allele model, whichever we prefer. To illustrate this, the probability that two genes taken at random in the infinitely many alleles model are of the same allelic type has been computed directly in (3.74), and also can be found by letting $K \rightarrow \infty$ in (3.70). The reason why the two approaches yield the same value is that the

probability in question can be expressed in terms of the order statistics as

$$E(F) = E(x_{(1)}^2 + x_{(2)}^2 + x_{(3)}^2 + \dots), \quad (5.129)$$

where F is the (random) population homozygosity. The eigenvalue set (3.90) for the infinitely many alleles model can also be found through a limiting process from the K -allele case.

The expression for the Poisson–Dirichlet distribution is rather complex. The distribution of $x_{(1)}$, at least over the range (0.5, 1), has already been noted in (5.65). For $(2N)^{-1} \leq x_{(1)} \leq 0.5$, Watterson and Guess (1977) show that the density function of $x_{(1)}$ is of the form

$$f(x_{(1)}) = \Gamma(\theta + 1)e^{\gamma\theta} x_{(1)}^{\theta-2} g((1 - x_{(1)})/x_{(1)}), \quad (5.130)$$

where $\theta = 4Nu$ and $g(\cdot)$ is a complicated function which is best defined through the Laplace transform equation

$$\int_0^\infty e^{-tz} g(z) dz = \exp\left(\theta \int_0^1 y^{-1}(e^{-ty} - 1) dy\right). \quad (5.131)$$

More generally the joint density function of $x_{(1)}, x_{(2)}, \dots, x_{(r)}$ is

$$f(x_{(1)}, \dots, x_{(r)}) = \theta^r \Gamma(\theta) e^{\gamma\theta} g(y) \{x_{(1)}x_{(2)} \cdots x_{(r)}\}^{-1} x_{(r)}^{\theta-1}, \quad (5.132)$$

where $y = (1 - x_{(1)} - x_{(2)} - \dots - x_{(r)})/x_{(r)}$. These conclusions are noted by Watterson (1976b). The expression (5.132) simplifies, when $x_{(1)} + \dots + x_{(r-1)} + 2x_{(r)} \geq 1$, to

$$f(x_{(1)}, \dots, x_{(r)}) = \theta^r \{x_{(1)}x_{(2)} \cdots x_{(r)}\}^{-1} (1 - x_{(1)} - \dots - x_{(r)})^{\theta-1}. \quad (5.133)$$

One interesting conclusion to be reached from (5.130) concerns the age of the most frequent allele currently observed in the population. In particular, the following question can be asked: What is the probability that the most frequent allele in the population is also the oldest? Time-reversal arguments show that this is identical to the probability that the most frequent allele will last longer into the future. Given the current allele frequency configuration this is, by symmetry, the current frequency $x_{(1)}$ of the most frequent allele. The unconditional probability in question is then just the mean value of $x_{(1)}$, namely

$$\int_{(2N)^{-1}}^1 x_{(1)} f(x_{(1)}) dx_{(1)} \quad (5.134)$$

where $f(x_{(1)})$ is given by (5.130). The simple form for this density in (0.5, 1) shows that a lower bound to this probability is

$$\theta \int_{0.5}^1 x_{(1)} \{x_{(1)}^{-1} (1 - x_{(1)})^{\theta-1}\} dx_{(1)} = \left(\frac{1}{2}\right)^\theta. \quad (5.135)$$

Watterson and Guess (1977) compute a more accurate value using (5.130) and (5.134). In a similar way the probability that the i th most frequent allele is the oldest can be computed from the expected value of the i th order statistic.

A further use for (5.132) arises in deriving the distribution of the population homozygosity $F = x_1^2 + x_2^2 + \dots$. As we have seen, F can equally well be expressed as the sum of squares of ordered frequencies, so its distribution for the K -allele model converges to that for the infinite allele model as $K \rightarrow \infty$. The complete distribution of F in the K -allele model is not simple. Nevertheless it is comparatively easy to find the mean and variance of F from (5.125): We obtain

$$E(F) = (1 + \epsilon)/(1 + K\epsilon) \quad (5.136)$$

$$\text{var}(F) = 2\theta(1 + \epsilon)/[(1 + K\epsilon)^2(2 + K\epsilon)(3 + K\epsilon)], \quad (5.137)$$

where $\epsilon = \theta/(K - 1)$. The value for the mean is in effect computed by (5.127) and is identical to that reached in (3.70) by different methods. By letting $K \rightarrow \infty$ and using the convergence theory, the infinitely many alleles model mean and variance of F are found as

$$E(F) = (1 + \theta)^{-1}, \quad \text{var}(F) = 2\theta/[(1 + \theta)^2(2 + \theta)(3 + \theta)]. \quad (5.138)$$

The value for the mean is a standard result, cf. (3.74).

Suppose now in the K -allele model that selective differential exist. Let the fitnesses of $A_i A_j$ be $1 + s_{ij}$ and suppose s_{ij} is of order N^{-1} . Writing $\alpha_{ij} = 2Ns_{ij}$ and assuming mutation as above, the joint density function of x_1, \dots, x_{K-1} can be found by solving (5.125) where now $m_i(x_1, \dots, x_{K-1})$ contains a further turn due to selective differences. The explicit solution (Wright, 1949, p. 383) is

$$f(x_1, \dots, x_{K-1}) = \text{const} \exp\left(\sum_{i=1}^K \sum_{j=1}^K x_i x_j \alpha_{ij}\right) \{x_1 \cdots x_K\}^{\epsilon-1}. \quad (5.139)$$

In particular, if all heterozygotes have equal fitness $1 + s$ ($s > 0$) and all homozygotes have fitness 1, and if $\alpha = 2Ns$, (5.139) becomes

$$f(x_1, \dots, x_{K-1}) = \text{const} \exp\{-\alpha F\} \{x_1 \cdots x_K\}^{\epsilon-1}. \quad (5.140)$$

We have noted that the limiting ($K \rightarrow \infty$) behavior can be analyzed only by transferring attention to the order statistic $x_{(1)}, x_{(2)}, \dots, x_{(j)}$. Unfortunately the joint distribution of the order statistics is here too complicated to yield useful information. Considerable progress can be made, however, when the α_{ij} are small, and thus for the rest of this chapter we assume this is the case. We thus put, in (5.139),

$$\exp\left\{\sum \sum x_i x_j \alpha_{ij}\right\} = 1 + \sum \sum x_i x_j \alpha_{ij} + O(\alpha_{ij}^2).$$

The joint distribution of the order statistics may be found by summing in (5.139) over all possible permutations of $1, 2, \dots, K$. This yields

$$f(x_{(1)}, x_{(2)}, \dots, x_{(K)}) = \frac{K! \Gamma\{K\epsilon\}}{\{\Gamma(\epsilon)\}^K} \left(1 + A\left(F - \frac{\epsilon + 1}{\theta + 1}\right) + O(\alpha_{ij}^2)\right) \times \{x_{(1)}x_{(2)} \cdots x_{(K)}\}^{\epsilon-1}, \quad (5.141)$$

where

$$A = K^{-1} \sum \alpha_{ii} - \{K(K-1)\}^{-1} \sum_{i \neq j} \sum \alpha_{ij}, \quad F = \sum x_{(i)}^2.$$

We are interested in two particular fitness schemes. The first, or “heterotic”, scheme has just been noted: All heterozygotes have fitness $1 + s$ and all homozygotes have fitness 1. Here $\alpha_{ii} = 0$, $\alpha_{ij} = \alpha = 2Ns$ so that

$$f(x_{(1)}, \dots, x_{(K)}) = \frac{K! \Gamma\{K\epsilon\}}{\{\Gamma(\epsilon)\}^K} \left(1 - \alpha \left\{F - \frac{\epsilon + 1}{\theta + 1}\right\} + O(\alpha_{ij}^2)\right) \times \{x_{(1)} \cdots x_{(K)}\}^{\epsilon-1}. \quad (5.142)$$

In the second selective model (the “deleterious alleles” model) a fraction γ of the K alleles are deleterious: individuals carrying i deleterious genes ($i = 0, 1, 2$) have fitness $1 - is$. Here $A = 0$ and the terms in α_{ij}^2 in (5.141) must be computed. This leads to

$$f(x_{(1)}, \dots, x_{(K-1)}) = \frac{K! \Gamma\{K\epsilon\}}{\{\Gamma(\epsilon)\}^K} \left(1 + 2\alpha^2\gamma(1-\gamma)\left(F - \frac{\epsilon + 1}{\theta + 1}\right) + O(\alpha_{ij}^2)\right) \times \{x_{(1)} \cdots x_{(K)}\}^{\epsilon-1}. \quad (5.143)$$

From (5.142) and (5.143) the joint distribution of $x_{(1)}, \dots, x_{(j)}$ may, in principle be found. From Kingman’s theory this joint distribution will converge, as $K \rightarrow \infty$, to a nontrivial limit, namely that of the first j order statistics in the infinite allele heterosis and deleterious alleles models. (In the latter model γ is defined as the probability that any new mutant allele is deleterious.) We do not pursue these distributions here other than to note that since to the order of approximation considered (5.143) can be obtained from (5.142) by replacing α by $-2\alpha^2\gamma(1-\gamma)$, the same must be true in the limiting order statistics distribution and any function derived from it. The information of most use to us may be found from one such function, namely the frequency spectrum of the two models considered. The frequency spectrum $\phi(x)$ was introduced in Section 3.6: It has the interpretation that $\phi(x)\delta x$ is the mean number of alleles in the population with frequency in $(x, x + \delta x)$. From this frequency spectrum we can find

three quantities of some interest in theoretical population genetics:

$$\int_0^1 \phi(x) dx = \text{mean number of alleles in the population}, \quad (5.144)$$

$$\int_0^1 x^2 \phi(x) dx = \text{mean population homozygosity}, \quad (5.145)$$

$$\int_0^1 \{1 - (1 - x)^n\} \phi(x) dx = \text{mean number of alleles in a sample} \\ \text{of size } n \text{ from the population.} \quad (5.146)$$

The mean population homozygosity, $E(F)$, is the probability that two genes taken at random from the population are of the same allelic type. For the neutral infinitely many alleles model,

$$\begin{aligned} \phi(x) &\approx 2N\theta, & 0 \leq x \leq (2N)^{-1}, \\ \phi(x) &\approx \theta x^{-1}(1-x)^{\theta-1}, & (2N)^{-1} \leq x \leq 1. \end{aligned} \quad (5.147)$$

Use of (5.147) in (5.144), (5.145) and (5.146) re-derives the quantities (3.92), (3.74) and (3.85) found in Chapter 3 by other methods. Our aim is to consider the corresponding three values in the “heterotic” and “deleterious alleles” selection models.

Watterson (1977a) demonstrated that in the heterotic model the frequency spectrum $\phi(x)$ becomes, for $(2N)^{-1} \leq x \leq 1$,

$$\phi(x) = \theta x^{-1}(1-x)^{\theta-1} (1 + \alpha x \{2 - (2 + \theta)x\} (1 + \theta)^{-1} + O(\alpha^2)). \quad (5.148)$$

It follows immediately from (5.144)–(5.148) that the mean number of alleles in the population exceeds its neutral theory value by an amount

$$\alpha\theta(1 + \theta)^{-2} + O(\alpha^2), \quad (5.149)$$

that

$$E(F) = (1 + \theta)^{-1} - 2\alpha\theta\{(1 + \theta)^2(2 + \theta)(3 + \theta)\}^{-1}, \quad (5.150)$$

and that the mean number of alleles observed in a sample of n genes exceeds its neutral theory value by

$$\alpha\theta(1 + \theta)^{-2} - 2\theta(\theta + 2n)\{(1 + \theta)(n + \theta)(n + 1 + \theta)\}^{-1}. \quad (5.151)$$

Clearly, for large n , this is approximately equal to the expression in (5.149), as we expect.

We have noted that for the deleterious alleles model the parallel quantities may be found by replacing α by $-2\alpha^2\gamma(1 - \gamma)$. Thus for this model the mean number of alleles in the population falls short of the neutral value by

$$2\theta\alpha^2\gamma(1 - \gamma)(1 + \theta)^{-2}, \quad (5.152)$$

the mean homozygosity is given by

$$E(F) = (1 + \theta)^{-1} + 4\alpha^2\gamma(1 - \gamma)\theta\{(1 + \theta)^2(2 + \theta)(3 + \theta)\}^{-1} \quad (5.153)$$

while the mean number of alleles in a sample of n falls short of its neutral theory value by

$$\begin{aligned} & 2\alpha^2\gamma(1 - \gamma)\theta(1 + \theta)^{-2} - 2\alpha^2\gamma(1 - \gamma)\theta(\theta + 2n) \\ & \times \{(1 + \theta)(n + \theta)(n + 1 + \theta)\}^{-1}. \end{aligned} \quad (5.154)$$

6

Two Loci

6.1 Introduction

Most of the theory of this book so far has assumed that the fitness of any individual depends on his genetic make-up at a single locus. Although for certain specific purposes this assumption may give reasonable approximations, it is in general a gross simplification, in particular when epistatic, that is interactive, effects arise between loci. In this chapter we suppose that the fitness of any individual depends on his genetic constitution at two (or sometimes three) loci. Although this assumption is hardly less realistic than the previous one, it does allow substantial advance to be made, as has been noted in Section 2.10, on assessing the evolutionary effect of recombination between loci. It also allows us to assess the extent to which two-locus behavior is predictable from combining two single-locus analyses. We shall also see later in this chapter that it allows an investigation of the effects of modifier genes.

The model we use is that described in Section 2.10. We assume viability selection only, with fitness scheme given by (2.90) or, equivalently, by (2.91), random mating, no fitness differentials between sexes and a discrete time parameter. Thus the recurrence relations (2.94) describe the evolution of the frequencies of the gametes A_1B_1 , A_1B_2 , A_2B_1 and A_2B_2 , that is of gametes of types 1, 2, 3, 4 respectively.

We have already observed, in Section 2.10, two major consequences of these recurrence relations. The first is the the mean fitness increase theorem, claiming that under random mating mean fitness increases from one

generation to the next, or at worst remains stable, is no longer true as a mathematical theorem. The second is that the equilibrium points of the recurrence system can depend on the recombination fraction between the loci. We start by examining these conclusions in greater detail.

6.2 Evolutionary Properties of Mean Fitness

Our aim in this section is to discuss the implications of the recurrence relations (2.94) for the evolution of mean fitness, defined by (2.93). The first substantial analysis of multilocus mean fitness behavior was given by Kimura (1958), although here we adopt a rather different approach than his. A second important early discussion of the question, not perhaps sufficiently highly appreciated, is that of Kojima and Kelleher (1961); The fact that mean fitness can decrease in two-locus systems was first explicitly mentioned in their paper.

If there is linkage disequilibrium at any stable equilibrium point, as is true, for example, at the equilibria (2.97) or (2.98), it is always possible to find a neighborhood of the equilibrium point such that, starting from any point in this neighborhood, the mean fitness decreases. Thus the MFIT cannot be true as a mathematical theorem, or in some cases be even approximately correct. Karlin (1975) indeed asserts that it “usually fails” in the sense that for almost all sets of possible genotype fitness values, it is possible to find gamete frequencies for which the mean fitness decreases, at least for a few generations.

These considerations, however, are probably of lesser importance from a practical standpoint. We are mainly interested in the behavior of mean fitness during those generations when substantial changes in gene frequency occur, and here it is possible to rescue in large part the spirit of the MFIT, at least in a wide variety of cases of practical interest. That such an attempt is not over-optimistic is supported by the observations of Karlin and Carmelli (1975), who observed that when the entries in the fitness matrix (2.90) are chosen randomly, the mean fitness increases for most generations for most fitness configurations. It is important, therefore, to emphasize that our aim is not to prove a mathematical theorem but to determine circumstances in which increases in mean fitness tend to occur.

The first attempt at finding some principle along these lines was through the introduction of the principle of quasi-linkage equilibrium (QLE) (Kimura (1965)). The essence of this principle is as follows. If we define, in the notation of Section 2.10, a quantity Z by

$$Z = c_1 c_4 / c_2 c_3, \tag{6.1}$$

then to a first order of approximation the change in the value of $\log Z$ between consecutive generations is

$$\Delta \log Z \approx c_1^{-1} \Delta c_1 + c_4^{-1} \Delta c_4 - c_2^{-1} \Delta c_2 - c_3^{-1} \Delta c_3. \quad (6.2)$$

The values of Δc_i can be found from (2.94) as

$$\Delta c_i = \bar{w}^{-1} \{c_i(w_i - \bar{w}) + \eta_i R w_{14} (c_2 c_3 - c_1 c_4)\}. \quad (6.3)$$

Substitution of these values into (6.2) eventually leads to the approximation

$$\bar{w} \Delta \log Z \approx \bar{\epsilon} - R w_{14} (Z - 1) \{c_2 + c_3 + Z^{-1} (c_1 + c_4)\}, \quad (6.4)$$

where

$$\bar{\epsilon} = w_1 - w_2 - w_3 + w_4.$$

Suppose now that $Z > 1$. If $\bar{\epsilon}$ can be treated, approximately, as a constant, there will be a tendency for Z to decrease, at least for values of Z sufficiently large compared to $\bar{\epsilon}$. Similarly when $Z < 1$ there will be a tendency for Z to increase, at least for very small Z . We may thus hope that Z approaches a constant value at which

$$\Delta \log Z = 0. \quad (6.5)$$

The change in mean fitness can be approximated, if small-order terms are ignored, by

$$\Delta \bar{w} \approx 2 \sum w_i \Delta c_i,$$

and substitution from (6.3) then gives

$$\Delta \bar{w} \approx 2 \bar{w}^{-1} \left\{ \sum c_i w_i (w_i - \bar{w}) + R w_{14} (c_2 c_3 - c_1 c_4) \bar{\epsilon} \right\}. \quad (6.6)$$

If now (6.5) holds we must have, from (6.4),

$$R w_{14} (c_2 c_3 - c_1 c_4) = -\bar{\epsilon} \left(\sum c_i^{-1} \right)^{-1},$$

and substituting this into (6.6) we find

$$\Delta \bar{w} \approx \bar{w}^{-1} \left(2 \sum c_i (w_i - \bar{w})^2 - 2 \bar{\epsilon}^2 \left(\sum c_i^{-1} \right)^{-1} \right). \quad (6.7)$$

The two terms in the parentheses on the right-hand side have arisen in our previous discussion. The first is the total gametic variance defined in (2.106), and the second is the epistatic gametic variance defined in (2.107). The difference between the two must be the additive gametic variance, or equivalently, in view of the discussion in Section 2.10, the additive genetic variance. We then conclude that whenever (6.5) holds,

$$\Delta \bar{w} \approx \bar{w}^{-1} \sigma_A^2. \quad (6.8)$$

Provided the above reasoning is satisfactory, we may therefore expect the system to evolve rapidly to a state where the approximation (6.8) is true,

and if this is so we have succeeded in our aim of rescuing the MFIT as a reasonable general principle.

The above reasoning, however, requires much closer examination. Clearly any well-behaved function Y of gamete frequencies will converge to its equilibrium as the system itself approaches its equilibrium state, and at the moment we have no reason to prefer calculations using Z to those using the function Y . But when Y and Z are different functions, the value of $\Delta\bar{w}$ found by assuming $\Delta Y = 0$ will be different from that assuming $\Delta Z = 0$. Indeed numerical calculations by Kimura (1965), and Ewens (1976) show that mean fitness usually changes at a much slower rate than does Z . It is therefore unreasonable to consider changes in mean fitness assuming $\Delta Z = 0$. Instead it would be more reasonable to consider changes in Z assuming $\Delta\bar{w} = 0$.

Despite these comments, it is possible to arrive at the approximation (6.8) by a deeper argument, at least in cases of biological interest. This was done by Nagylaki (1976): See also Hoppensteadt (1976) and Conley (1972). We now outline the main points of Nagylaki's argument.

Suppose that the fitness differences in the system (2.90) are small, so that w_{ij} can be written as $1 + sa_{ij}$, where s is a small parameter and the a_{ij} are moderate or small. We consider the linkage disequilibrium measure $D (= c_1c_4 - c_2c_3)$. From the recurrence relations (2.94),

$$\Delta D = -RD + sf(c_i, a_{ij}), \quad (6.9)$$

where the exact form of the function f is not important. This leads to

$$D(t) = (1 - R)^t D(0) + s(1 - R)^t \sum_{u=1}^t (1 - R)^{-u} f(c_i(u-1), a_{ij}), \quad (6.10)$$

where $D(t)$ is the value of D in generation t . Clearly, for R moderate, there exists a time t_1 for which $D(t_1)$ is of order s so we can write, for $t \geq t_1$, $D(t) = sD^*(t)$ and hence, from (6.9),

$$\Delta D^*(t) = -RD^*(t) + f(c_i(t), a_{ij}). \quad (6.11)$$

We also know, from (2.94), that Δc_i is of order s at most, and hence

$$f(c_i(t), a_{ij}) - f(c_i(t-1), a_{ij}) \quad (6.12)$$

is of order s at most. Hence, from (6.11) and (6.12),

$$\begin{aligned} \Delta D^*(t) - R^{-1}\{f(c_i(t), a_{ij}) - f(c_i(t-1), a_{ij})\} \\ = -R\{D^*(t) - R^{-1}f(c_i(t), a_{ij})\} + O(s), \end{aligned}$$

or

$$\Delta G(t) = -RG(t) + O(s), \quad (6.13)$$

where

$$G(t) = D^*(t) - R^{-1}f(c_i(t), a_{ij}). \quad (6.14)$$

There will also exist a time t_2 such that

$$G(t_1)(1 - R)^{t_2 - t_1} < s,$$

and from (6.14) and (6.11) it then follows that for $t \geq t_2$,

$$\Delta D = O(s^2). \tag{6.15}$$

Since $D = c_2 c_3 (Z - 1)$, it follows that

$$\Delta Z = O(s^2) \quad \text{for } t \geq t_2. \tag{6.16}$$

We turn now to changes in mean fitness. A more exact computation than that leading to (6.8) yields

$$\Delta \bar{w} = (\sigma_A^2 - 2RD\bar{\epsilon} + 2\bar{\epsilon}^2 \left(\sum c_i^{-1} \right)^{-1}) + O(s^3) \tag{6.17}$$

for $t \geq t_1$. Further, a more exact computation than that leading to (6.4) gives

$$Z^{-1} \Delta Z = \bar{\epsilon} - RD \sum c_i^{-1} + O(s^2). \tag{6.18}$$

If we use this equation, (6.17) becomes

$$\Delta \bar{w} = \left(\sigma_A^2 + 2\bar{\epsilon} \left(\sum c_i^{-1} \right)^{-1} Z^{-1} \Delta Z \right) + O(s^3). \tag{6.19}$$

Now $\bar{\epsilon}$ is of order s , and then (6.16) shows that for $t \geq t_2$, ΔZ is of order s^2 . Thus the second term in the parentheses on the right-hand side of (6.19) is $O(s^3)$ for $t \geq t_2$. Since in general σ_A^2 is $O(s^2)$, it follows that for $t \geq t_2$,

$$\Delta \bar{w} \approx \sigma_A^2 \tag{6.20}$$

during these epochs of the process for which the various order of magnitude arguments hold. On the other hand, σ_A^2 is very close to zero when the system is near an equilibrium point, so that it is possible at that stage of the process that the second term in parentheses in (6.19) dominates the first term, leading to possible decreases in mean fitness. These decreases are probably small and of little evolutionary consequence.

The correct statement of the QLE principle is thus the following. For small selective differences (of order s) and loose linkage, a state soon arises where the change in mean fitness is given by the right-hand side in (6.19), where ΔZ is of order s^2 and $\bar{\epsilon}$ is of order s . Since σ_A^2 is usually of order s^2 to a leading order of approximation, (6.20), embodying the main concepts of the MFIT, is usually true. It is not correct to say, as some versions of the QLE principle claim, that (6.20) is true because changes in Z are smaller than changes in \bar{w} and can be ignored: It is possible (see Kimura (1965)) that $\Delta Z \rightarrow \infty$ and yet (6.20) still holds.

It is of some interest to illustrate this conclusion by some numerical examples. Table 6.1 shows the values of $\Delta \bar{w}$, ΔZ , σ_A^2 and $\bar{\epsilon}$ for the evolution of a system with fitness matrix (2.96), with $R = 0.5$ and with various

Generation	$\Delta\bar{w} (\times 10^4)$	$\Delta Z (\times 10^4)$	$\sigma_A^2 (\times 10^4)$	$P (\times 10^4)$	$\bar{\epsilon} (\times 10^2)$
Case 1: Initial gamete frequencies 0.11, 0.16, 0.39, 0.34					
1	-1.67	1,750.0	0.532	-2.12	0.793
2	-0.462	1,020.0	0.496	-0.541	0.728
5	0.376	140.0	0.450	0.350	0.676
10	0.386	3.39	0.402	0.400	0.690
20	0.310	-0.526	0.322	0.322	0.719
30	0.246	-0.401	0.254	0.255	0.744
40	0.193	-0.301	0.200	0.201	0.758
50	0.152	-0.224	0.157	0.157	0.775
100	0.045	-0.046	0.046	0.046	0.805
200	0.005	0.002	0.005	0.005	0.807
Case 2: Initial gamete frequencies 0.42, 0.09, 0.11, 0.38					
1	-8.26	128,000.0	0.096	-6.27	0.073
2	1.23	15,600.0	0.054	-0.334	-0.384
5	0.693	777.0	0.034	0.648	-0.773
10	0.047	19.9	0.026	0.047	-0.821
20	0.019	0.032	0.020	0.020	-0.822
30	0.014	0.043	0.014	0.015	-0.818
40	0.010	0.038	0.011	0.011	-0.821
50	0.008	0.034	0.008	0.008	-0.818
100	0.002	0.015	0.002	0.002	-0.810
Case 3: Initial gamete frequencies 0.00001, 0.48386, 0.51612, 0.00001					
1	-78.3	1,180.0	0.000	1,130,000.0	-2.288
2	-23.6	2,560.0	0.002	-47.2	-1.533
5	-1.36	970.0	0.004	-1.51	-0.895
10	-0.030	32.4	0.004	-0.030	-0.812
20	0.003	0.042	0.003	0.003	-0.808
30	0.002	0.013	0.002	0.002	-0.808
40	0.002	0.010	0.002	0.002	-0.807
50	0.001	0.009	0.002	0.002	-0.807

Table 6.1. Parameters associated with the evolution of the two-locus system with fitness matrix (2.96); $R = 0.5$

initial gamete frequencies. The table also gives the value of $P = \sigma_A^2 + 2\bar{\epsilon}(\sum c_i^{-1})^{-1}Z^{-1}\Delta Z$, the right-hand side in (6.19), which from the above discussion we expect to provide a close approximation to $\Delta\bar{w}$. For these fitness values we may take $s \sim 0.02$.

The values in the table illustrate the various points made above. First, the mean fitness may decrease in the early generations, but eventually it begins to increase. Second, the value ΔZ and σ_A^2 are, after the early generations, both of order s^2 while $\bar{\epsilon}$ is of order s . Finally, and most important,

values of $\Delta\bar{w}$ are closely approximated by the quantity P throughout the process and, after the initial generations, by σ_A^2 . The only exception to this rule arises in case 3 where the population starts in extreme linkage disequilibrium. Here the early generations of the process show large values for ΔZ and $\Delta\bar{w}$ and exceptionally large values of P . The additive genetic variance, however, is quite small. This is clearly an extreme case where the behavior of the system before time t_1 cannot be predicted from the above analysis.

In his original discussion of the QLE principle, Kimura (1965) provided an example in which the approximation (6.20) is quite accurate after a number of generations has passed, even though $\Delta Z \rightarrow \infty$ as $t \rightarrow \infty$. The state of QLE does not arise for some time, approximately 100 generations in this example, due to the very low value 0.0001 of R that is assumed. In this example the numerical values in the fitness matrix (2.91) are

$$\begin{array}{ccc} 1.00 & 1.00 & 0.95 \\ 1.00 & 1.00 & 0.95 \\ 0.95 & 0.95 & 1.10 \end{array}$$

and the initial gamete frequencies are all 0.25. In generation 100, $\Delta Z \approx 400$ and yet $\Delta\bar{w} = 11.59 \times 10^{-5}$, $\sigma_A^2 = 11.01 \times 10^{-5}$. Clearly we cannot say $\Delta\bar{w} \approx \sigma_A^2$ because $\Delta Z \approx 0$. Despite the large value of ΔZ , the quantity $P (= \sigma_A^2 + 2\bar{\epsilon}(\sum c_i^{-1})^{-1}Z^{-1}\Delta Z)$ takes the value 12.6×10^{-5} . This occurs because of the extremely small values of $\bar{\epsilon}$ and Z^{-1} . This shows that (6.19) holds and the QLE principle in the Nagylaki interpretation applies.

There are several points one can make in conclusion. First, we have considered only two alleles at the loci in question: The extension of the above arguments to many alleles has been made by Nagylaki (1977b). Second, it is of some interest to calculate the total change in mean fitness from the initial point to the equilibrium point. This may sometimes be negative because large decreases in mean fitness during the early generations outweigh the consistent but small increases during later generations. Nagylaki (1977) examined this question and suggests that this happens comparatively seldom: In most cases the total course of the evolutionary process is to increase the mean fitness.

Finally, one may ask whether there are any special classes of fitness matrices for which the mean fitness always increases. Our class of fitness matrices where this is the case has been provided by Ewens (1969a, b). Suppose that the fitness matrix (2.91) is in the "additive" form

$$\begin{array}{cccc} & B_1B_1 & B_1B_2 & B_2B_2 \\ A_1A_1 & \alpha_1 + \beta_1 & \alpha_1 + \beta_2 & \alpha_1 + \beta_3 \\ A_1A_2 & \alpha_2 + \beta_1 & \alpha_2 + \beta_2 & \alpha_2 + \beta_3 \\ A_2A_2 & \alpha_3 + \beta_1 & \alpha_3 + \beta_2 & \alpha_3 + \beta_3 \end{array} \quad (6.21)$$

The fitness of any individual is here the sum of two components, one characterizing the genotype at the A locus and the other characterizing the genotype at the B locus. For this fitness matrix the mean fitness \bar{w} becomes

$$\begin{aligned} \bar{w} = & \alpha_1(c_1 + c_2)^2 + 2\alpha_2(c_1 + c_2)(c_3 + c_4) + \alpha_3(c_3 + c_4)^2 \\ & + \beta_1(c_1 + c_3)^2 + 2\beta_2(c_1 + c_3)(c_2 + c_4) + \beta_3(c_2 + c_4)^2. \end{aligned} \quad (6.22)$$

Suppose the gamete frequencies c_1, c_2, c_3, c_4 take any arbitrary values in generation t . The mean fitness in generation $t+1$ is found by replacing c_i by c'_i in (6.22). Now (6.22) depends on the gamete frequencies only through the gene frequencies $c_1 + c_2, c_3 + c_4, c_1 + c_3$, and $c_2 + c_4$. However, from the basic recurrence relation (2.94), the gene frequencies $c'_1 + c'_2, c'_3 + c'_4, c'_1 + c'_3$, and $c'_2 + c'_4$ are independent of R once the c_i are given and thus, in particular, are the same as for the special case $R = 0$. But when $R = 0$ the system (2.94) is identical to a four-allele single-locus system, and then Kingman's theorem (see Section 2.4) shows that mean fitness is nondecreasing. It follows that in the two-locus system (6.21) the mean fitness is nondecreasing. While we have used single-locus theory to obtain this result, it is not true that the complete evolution of the system is identical to that of any four-allele single-locus system: We have used the parallel with the latter merely to assert that mean fitness is nondecreasing.

This argument can be extended immediately to cover an arbitrary number of alleles at each of the two loci and indeed an arbitrary number of loci with an arbitrary recombination pattern. Later we shall use this conclusion to derive properties of additive fitness models. The result has been extended in a different way by Lyubich (1992), who found that the "increase in mean fitness" result also holds for a set of fitnesses generalizing the additive form given above (his fitnesses (9.5.11)).

6.3 Equilibrium Points

In the previous section we have been concerned with a specific property of the dynamics of the recurrence system (2.94). In the present (brief) section we turn to static properties and introduce the machinery whereby we examine the equilibrium properties of this system.

If we write (2.94) in the form

$$c'_i = \phi_i(c_1, c_2, c_3, c_4, R) \quad (6.23)$$

the point $c_i = c_i^*$ is an equilibrium point if the c_i^* satisfy the equations

$$c_i^* = \phi_i(c_1^*, c_2^*, c_3^*, c_4^*, R), \quad (i = 1, 2, 3, 4). \quad (6.24)$$

It is clear that the system (2.94) may possess several equilibrium points and, further, that these points may depend on the value of R . We discuss these observations in more detail later. When the fitness matrix (2.90) possesses special properties the equilibrium equation (6.24) can often be solved

explicitly, but in general this cannot be done and numerical methods are required. In the next section we present examples of equilibria found by both methods, and we use both sets of results to discuss equilibrium properties in some detail. We have already noted several connections between mean fitness and the equilibrium points. Thus any unique stable equilibrium point when $R = 0$ corresponds to a maximum of mean fitness, while if the coefficient of linkage disequilibrium is nonzero at any stable equilibrium for $R > 0$, the equilibrium mean fitness must be less than the maximum possible fitness. Further, there are gamete frequency trajectories near such equilibria along which mean fitness is decreasing, at least over some generations.

Suppose an equilibrium point of the system (2.94) has been found. It is then necessary to examine its stability behavior, since unstable equilibrium points are of little interest. The local linear stability of the system is tested by standard methods which we here outline. Suppose in any generation $c_i = c_i^* + \delta_i$, (with $\sum \delta_i = 0$), where the δ_i are small deviations from the equilibrium value. If the corresponding derivations in the following generation are δ'_i , then from (2.94),

$$\begin{aligned} c'_i + \delta'_i &= \phi_i(c_1^* + \delta_1, c_2^* + \delta_2, c_3^* + \delta_3, c_4^* + \delta_4) \\ &= \phi_i(c_1^*, c_2^*, c_3^*, c_4^*) + \sum_j \delta_j \left[\frac{\partial \phi_i}{\partial c_j} \right]^* + O(\delta^2). \end{aligned} \quad (6.25)$$

Here $[f]^*$ means the function f evaluated at the equilibrium point. If we ignore small-order terms, we get

$$\delta' = A\delta, \quad (6.26)$$

where A is a 4×4 matrix whose (i, j) term is $[\partial \phi_i / \partial c_j]^*$. Since

$$\delta^{(n)} = A^n \delta^{(0)}, \quad (6.27)$$

the spectral expansion of the matrix A shows that the equilibrium point is locally linearly stable if and only if all eigenvalues of A are less than unity in absolute value. These eigenvalues can be evaluated numerically or, in special cases, found algebraically. The condition $\sum \delta_i = 0$ can be used to simplify the calculations under both approaches.

Our definition of stability is a local one only. Questions concerning global stability and domains of attraction are far harder to answer and, especially concerning the latter, little is known about them.

6.4 Special Models

Historically the first analyses, both static and dynamic, of the recurrence system (2.94) assumed special forms for the fitness matrix (2.90). While this fact often allowed explicit expressions for the equilibria and explicit

criteria for their stability which suggested rather general conclusions for other fitness matrices, there was no certainty that the conclusions reached were not an artifact of the simple forms assumed for the fitness values, and thus no certainty of the generality of the conclusions reached. In this section we attempt to overcome these difficulties by presenting explicit conclusions for certain special matrices, as well as presenting numerical conclusions found from a number of more or less arbitrary fitness matrices.

We have already noted the additive fitness model (6.21). For the multiplicative fitness model the matrix (2.91) appears in the form

$$\begin{array}{ccc} \alpha_1\beta_1 & \alpha_1\beta_2 & \alpha_1\beta_3 \\ \alpha_2\beta_1 & \alpha_2\beta_2 & \alpha_2\beta_3 \\ \alpha_3\beta_1 & \alpha_3\beta_2 & \alpha_3\beta_3 \end{array} \quad (6.28)$$

while a third special class of fitness matrices is given by the “symmetric viability” model, for which the fitnesses are

$$\begin{array}{ccc} 1 - \delta & 1 - \beta & 1 - \alpha \\ 1 - \gamma & 1 & 1 - \gamma \\ 1 - \alpha & 1 - \beta & 1 - \delta \end{array} \quad (6.29)$$

In the analysis of this model it is usually assumed that $\alpha, \beta, \gamma, \delta > 0$, and we also make this assumption throughout. These models are not mutually exclusive: For example, if $\beta + \gamma = \alpha = \delta$, the symmetric viability model is also an additive model. Several models which do not initially appear in one of these forms can in fact be so written by suitable re-parameterization. Thus the model

$$\begin{array}{ccc} 1 + s & 1 + t & 1 - s \\ 1 & 1 + t & 1 \\ 1 - s & 1 + t & 1 + s \end{array} \quad (6.30)$$

of Kimura (1956b) can be cast in the form (6.29) by putting $\alpha = (s + t)/(1 + t)$, $\beta = 0$, $\gamma = t/(1 + t)$, $\delta = (t - s)/(1 + t)$. (Kimura’s analysis of the model (6.30) marked the beginning of the mathematics of two-locus models as discussed in this book.) However, a second model studied by Kimura (1956b), for which the fitness matrix is

$$\begin{array}{ccc} 1 + s & 1 + s + t & 1 - s \\ 1 + s & 1 + s + t & 1 - s \\ 1 - s & 1 - s + t & 1 + s \end{array} \quad (6.31)$$

cannot in general be cast in any of the three forms above. Similarly the model (1.90) does not fall into any of these forms, whereas the models (1.91) and (1.92) do. In examining both static and dynamic properties of the additive, multiplicative and symmetric viability models and of various numerical models we assume throughout that $0 < R \leq 0.5$: the case $R = 0$

reduces to a one-locus four-allele model to which the theory of Section 2.4 can be applied, while cases with $R > 0.5$ are not of biological interest.

We first take the additive fitness scheme (6.21). Using the fact that mean fitness is nondecreasing in this model, Karlin and Feldman (1970b) demonstrated, in the case $\alpha_2 > \alpha_1, \alpha_3$ and $\beta_2 > \beta_1, \beta_3$, that there exists a unique internal equilibrium point at

$$\begin{aligned} c_1 &= (\alpha_2 - \alpha_3)(\beta_2 - \beta_3) / ((2\alpha_2 - \alpha_1 - \alpha_3)(2\beta_2 - \beta_1 - \beta_3)), \\ c_2 &= (\alpha_2 - \alpha_3)(\beta_2 - \beta_1) / (((2\alpha_2 - \alpha_1 - \alpha_3)(2\beta_2 - \beta_1 - \beta_3))), \quad (6.32) \\ c_3 &= (\alpha_2 - \alpha_1)(\beta_2 - \beta_3) / ((2\alpha_2 - \alpha_1 - \alpha_3)(2\beta_2 - \beta_1 - \beta_3)), \\ c_4 &= (\alpha_2 - \alpha_1)(\beta_2 - \beta_1) / ((2\alpha_2 - \alpha_1 - \alpha_3)(2\beta_2 - \beta_1 - \beta_3)). \end{aligned}$$

Further, this equilibrium is globally stable. The location of this equilibrium point is what would be expected by a composition of single-locus analyses, using (1.31). It is independent of R and the coefficient of linkage disequilibrium is zero at it. Since also mean fitness is nondecreasing in the additive fitness model considered, one might be tempted to conclude that for this model the gene frequencies loci evolve in an independent way, and that for practical purposes the two loci can be studied separately. This conclusion, however, is not quite correct. If the coefficient of linkage disequilibrium is zero in any generation it does not necessarily remain zero in future generations, and this can prevent gene frequencies converging monotonically to their equilibrium values given in (6.32).

If $\alpha_1 > \alpha_2 > \alpha_3$ the frequency of A_1 converges to unity. If at the same time $\beta_2 > \beta_1, \beta_3$ a polymorphism will be maintained at the B locus in accordance with (1.31). This case and the corresponding cases where fixation occurs at the B locus or at both loci essentially reduce, so far as equilibrium properties are concerned, to single-locus systems, so we consider them no further here. The criteria that these fixation events occur depend only on the selective parameters $\{\alpha_i\}$ and $\{\beta_i\}$, and are independent of the recombination fraction R .

An important question concerning the concept of marginal fitnesses arises with respect to the equilibrium point (6.32). Suppose that we observe the A locus only and compute the marginal fitnesses at this locus as defined by (2.96). These marginal fitnesses reduce to $\alpha_1 + \beta$, $\alpha_2 + \beta$, $\alpha_3 + \beta$ (for some β), as we might expect. A parallel observation applies for the B locus. The important observation for us is that, since in computing (6.32) it has been assumed that $\alpha_2 > \alpha_1, \alpha_2 > \alpha_3$ and $\beta_2 > \beta_1, \beta_2 > \beta_3$, the marginal fitnesses will also exhibit overdominance. The question of whether marginal overdominance always applies for stable internal two-locus equilibria is an interesting one, to which we shall return on several occasions.

We turn now to the multiplicative system (6.28). The properties of this model are more complex than those for the additive system. We again suppose that $\alpha_2 > \alpha_1, \alpha_2 > \alpha_3$ and $\beta_2 > \beta_1, \beta_2 > \beta_3$. Then for all values of R there exists an equilibrium point with gametic frequencies given in

(6.32). However, this equilibrium is stable only if R is large enough. Roux (1974), generalizing the analysis of Bodmer and Felsenstein (1967), found that stability applies if and only if

$$R > \frac{(\alpha_2 - \alpha_1)(\beta_2 - \beta_1)(\alpha_2 - \alpha_3)(\beta_2 - \beta_3)}{(2\alpha_2 - \alpha_1 - \alpha_3)(2\beta_2 - \beta_1 - \beta_3)\alpha_2\beta_2}. \quad (6.33)$$

This is also a sufficient condition for local stability. Moran (1968) found that a sufficient condition for global stability of (6.32) is

$$6\left(\frac{1}{2} - R\right) < \min \left\{ \frac{(\alpha_2 - \alpha_3)(\alpha_2 - \alpha_1)}{\alpha_2(2\alpha_2 - \alpha_1 - \alpha_3)}, \frac{(\beta_2 - \beta_3)(\beta_2 - \beta_1)}{\beta_2(2\beta_2 - \beta_1 - \beta_3)} \right\}. \quad (6.34)$$

It is not known how close this is to being a necessary condition. He also provided two further sufficient conditions, namely

$$\frac{1}{2} - R < \frac{1}{2} \min\left(\frac{\alpha_1}{\alpha_2}, \frac{\alpha_3}{\alpha_2}\right), \quad \frac{1}{2} - R < \frac{1}{2} \min\left(\frac{\beta_1}{\beta_2}, \frac{\beta_3}{\beta_2}\right).$$

The requirement (6.33) has been generalized in an elegant way to an arbitrary number of alleles at the two loci by Roux (1974). Suppose we associate a multiplicative fitness component a_{ij} with the genotype A_iA_j at locus A and a multiplicative component b_{ij} to B_iB_j at locus B . Suppose further that when treated as single-locus systems, each of these has an internal equilibrium point where

$$\text{freq}(A_i) = p_i, \quad \text{freq}(B_j) = q_j. \quad (6.35)$$

Then the two-locus system will have an equilibrium point at which $\text{freq}(A_iB_j) = p_iq_j$. We now define the A and B locus equilibrium mean fitnesses as

$$\bar{w}_A = \sum \sum p_i p_j a_{ij}, \quad \bar{w}_B = \sum \sum q_i q_j b_{ij}, \quad (6.36)$$

put

$$c_{ij} = a_{ij} \sqrt{p_i p_j} / \bar{w}_A, \quad d_{ij} = b_{ij} \sqrt{q_i q_j} / \bar{w}_B, \quad (6.37)$$

and let λ, ψ be the largest nonunit eigenvalues respectively of the matrices $\{c_{ij}\}, \{d_{ij}\}$. Then Roux demonstrated that the condition for the equilibrium at which $\text{freq}(A_iB_j) = p_iq_j$ to be stable is that

$$R > \lambda\psi / \{(1 - \lambda)(1 - \psi)\}. \quad (6.38)$$

In the case of two alleles at each locus, we may adopt our previous notation and put, for the A locus,

$$a_{11} = \alpha_1, \quad a_{12} = a_{21} = \alpha_2, \quad a_{22} = \alpha_3.$$

Then at an equilibrium point of this locus,

$$p_1 = (\alpha_2 - \alpha_3) / \{2\alpha_2 - \alpha_1 - \alpha_3\} = 1 - p_2 \quad (6.39)$$

and hence

$$\bar{w}_A = (\alpha_2^2 - \alpha_1\alpha_3)/(2\alpha_2 - \alpha_1 - \alpha_3). \quad (6.40)$$

It follows from (6.37)–(6.40) that

$$c_{11} = \frac{\alpha_1(\alpha_2 - \alpha_3)}{\alpha_2^2 - \alpha_1\alpha_3}, \quad c_{12} = c_{21} = \frac{\alpha_2\{(\alpha_2 - \alpha_3)(\alpha_2 - \alpha_1)\}^{1/2}}{\alpha_2^2 - \alpha_1\alpha_3},$$

$$c_{22} = \frac{\alpha_3(\alpha_2 - \alpha_1)}{\alpha_2^2 - \alpha_1\alpha_3}.$$

The eigenvalues of the matrix $\{c_{ij}\}$ are easily found to be 1 and

$$\lambda = -(\alpha_2 - \alpha_3)(\alpha_2 - \alpha_1)/(\alpha_2^2 - \alpha_1\alpha_3).$$

In a similar way we find for the B locus

$$\psi = -(\beta_2 - \beta_3)(\beta_2 - \beta_1)/(\beta_2^2 - \beta_1\beta_3).$$

Insertion of these values into (6.38) gives precisely the condition (6.33).

Suppose now that the condition (6.33) is not met. In the symmetric case where $\alpha_1 = \alpha_3$, $\beta_1 = \beta_3$ there will now exist two stable equilibria, both possibly exhibiting large numerical values of the coefficient of linkage disequilibrium. Thus, for example, if $\alpha_1 = \alpha_3 = \beta_1 = \beta_3 = .99$, $\alpha_2 = \beta_2 = 1$, and $R = 0.000009$, the two stable equilibria are

c_1	c_2	c_3	c_4	
0.45	0.05	0.05	0.45	(6.41)
0.05	0.45	0.45	0.05	

Such equilibria arise in multiplicative models only for very tightly linked loci. However, we shall observe in the next chapter that for interactive systems involving many loci, rather less stringent conditions of the value of the recombination fraction still lead to equilibria of the form (6.41), that is with $D \neq 0$.

As with the additive case, whenever the conditions $\alpha_2 > \alpha_1$, $\alpha_2 > \alpha_3$ and $\beta_2 > \beta_1$, $\beta_2 > \beta_3$ do not both hold, edge or corner equilibria may arise. These are of little interest to us and we consider them no further.

We turn now to the mean fitness. First, since equilibria of the form (6.41) have nonzero linkage disequilibrium values, mean fitness cannot be maximized at them. The equilibrium (6.32) is perhaps of greatest interest. Surprisingly, the mean fitness is not maximized at this point: Rather, the fitness surface has a saddle point at (6.32). These conclusions imply that mean fitness can be decreasing in the neighborhood of equilibrium points.

One special property of multiplicative fitness schemes concerns the coefficient of linkage disequilibrium D . If in any generation the equation $D = 0$ holds, then the gametic frequency recurrence relations show that D remains at the value 0 for all future generations. In this case the mean fitness

must be nondecreasing, since the recurrence relations then reduce to single-locus equations. More generally, for $R \leq 0.5$, the sign of D will not change throughout the entire evolution of the recurrence process (Karlin, (1975)).

We consider next the value of the equilibrium mean fitness as a function of R . When the inequality (6.33) holds, the equilibrium mean fitness is independent of R , since the location of the equilibrium point is independent of R . When (6.33) does not hold the location of the equilibrium point does depend on R . Since the equilibrium mean fitness decreases with R for $R \approx 0$, we may suspect that it is nondecreasing with R for all R , and this has been confirmed for this model by Karlin (1975).

We observe finally that if $\alpha_2 > \alpha_1, \alpha_2 > \alpha_3$ and $\beta_2 > \beta_1, \beta_2 > \beta_3$ and an internal equilibrium point, either (6.32) or of the form typified by (6.41), exists, the marginal fitness values as computed from (2.97) exhibit overdominance. This is easily checked for the equilibrium point (6.32), and can also be verified for the linkage-disequilibrium equilibrium points.

The last special case we consider in some detail is the symmetric viability model (6.29). This model possesses several unusual features not shared by additive or multiplicative schemes. Perhaps the most important of these is the unexpected existence of so-called asymmetric equilibria. The fitness matrix (6.29) possesses certain symmetry properties, and because of this one might expect that at any equilibrium point, the identities

$$c_1^* = c_4^*, \quad c_2^* = c_3^* \quad (6.42)$$

will hold. Strangely, while equilibria satisfying (6.42) usually do exist, there is a further class of equilibria for which (6.42) does not hold. This fact was discovered by Karlin and Feldman (1970a) and revealed previously unthought-of complexities in the model (6.29). Suppose for example that the fitness matrix (6.29) takes the form

$$\begin{array}{ccc} 0.97 & 0.96 & 0.98 \\ 0.96 & 1.00 & 0.96 \\ 0.98 & 0.96 & 0.97 \end{array} \quad (6.43)$$

and that $R = 0.04$. Then the recurrence system (2.94) admits five equilibrium points and at only one of these, the first noted below, do the equations in (6.42) hold. The five equilibria are

$$\begin{array}{cccc} c_1 & c_2 & c_3 & c_4 \\ 0.238 & 0.262 & 0.262 & 0.238 \\ 0.154 & 0.656 & 0.036 & 0.154 \\ 0.154 & 0.036 & 0.656 & 0.154 \\ 0.889 & 0.050 & 0.050 & 0.011 \\ 0.011 & 0.050 & 0.050 & 0.889 \end{array} \quad (6.44)$$

For smaller values of R the number of symmetric equilibria for which the equations in (6.42) hold increases to three. Thus for $R = 0.0004$, apart

from four asymmetric equilibria, there are symmetric equilibria at

c_1	c_2	c_3	c_4	
0.0034	0.4966	0.4966	0.0034	(6.45)
0.2734	0.2266	0.2266	0.2734	
0.4950	0.0050	0.0050	0.4950	

This does not automatically occur for every choice of α , β , γ , δ . Thus if the inequalities $\beta + \gamma > \alpha$, $\beta + \gamma > \delta$ do not both hold, there will be only one equilibrium of the form (6.42), whatever the value of R .

It is of particular importance for this model to ask two questions concerning stable equilibrium points. As noted above, up to seven internal equilibria can exist for certain parameter choices in the model (6.29). We therefore ask, first, how many of these can be stable, and second, will there always be at least one stable internal equilibrium, at least for certain values of R ? Karlin (1975) claimed that irrespective of R there can never be more than two stable internal equilibria for the system (6.29). Thus if seven such equilibria exist, at least five must be unstable. Can we be guaranteed that at least one internal stable equilibrium exists? Karlin and Feldman (1970a,b) found that this is so, at least provided R is sufficiently small (and, as we are assuming, that α , β , γ , $\delta \geq 0$). For larger values of R quite complex behavior is possible, and this behavior can be described explicitly in the special case $\delta = \alpha$. We restrict our attention to symmetric equilibria of the form (6.42), since Karlin and Feldman (1970a) demonstrated that the asymmetric equilibria are unstable in this case. Here the equilibrium point $c_i^* = 0.25$ exists for all R but is stable only if

$$R > \frac{1}{4}(\beta + \gamma - \alpha) \quad (6.46)$$

and

$$\alpha > |\beta - \alpha|. \quad (6.47)$$

If the right-hand side in (6.46) is negative and (6.47) holds, this of course implies stability for all R . When the inequality (6.46) does not hold there will exist two equilibria of the form

$$\begin{aligned} c_1^* &= c_4^* = 0.25 \pm 0.25\{1 - 4R/(\beta + \gamma - \alpha)\}^{1/2}, \\ c_2^* &= c_3^* = 0.25 \mp 0.25\{1 - 4R/(\beta + \gamma - \alpha)\}^{1/2}. \end{aligned} \quad (6.48)$$

The condition on the recombination fraction R that these be stable is that $\Phi(R)$, defined by

$$\begin{aligned} \Phi(R) &= 4R^2(\beta + \gamma - \alpha) + 2R\{2\alpha^2 - \beta^2 - \gamma^2 - \alpha(\beta + \gamma)\} \\ &\quad + \alpha(\beta + \gamma - \alpha)^2, \end{aligned} \quad (6.49)$$

be positive. This requirement is always satisfied for R sufficiently close to zero, so that the equilibria (6.48) are always stable if linkage is sufficiently tight. When $R = \frac{1}{4}(\beta + \gamma - \alpha)$, the upper limit for R for which equilibria of

the form (6.48) exist, condition (6.49) reduces to (4.36). If we denote the solutions of the quadratic equation

$$\Phi(R) = 0 \tag{6.50}$$

by R_1 and R_2 , with ($R_1 < R_2$), it follows that the equilibrium point (6.48) will be stable in the following cases:

- (1) if $R_1 > 0$, $|\beta - \gamma| > \alpha$, then (6.48) is stable for $0 < R < R_1$;
- (2) if $R_1 > \frac{1}{4}(\beta + \gamma - \alpha)$, then (6.48) is stable for $0 < R < \frac{1}{4}(\beta + \gamma - \alpha)$;
- (3) if $0 < R_1 < R_2 < \frac{1}{4}(\beta + \gamma - \alpha)$, then (6.48) is stable for $0 < R < R_1$ and for $R_2 < R < \frac{1}{4}(\beta + \gamma - \alpha)$.

In the latter case there exists a gap of instability (when $R_1 < R < R_2$) for which there are no internal stable equilibria. Clearly the stability behavior, even of the symmetric equilibria, is not straightforward.

Before leaving symmetric viability models we observe that the model (1.92) can be cast in the form (6.29) by suitable normalization and parameterization. It is easy to see that if there exists an equilibrium of the form (6.42), then $c_1^* = c$ satisfies the condition in (1.94). Further properties of this model follow from the general properties of symmetric viability models.

Given a point of stable equilibrium, what can be said about the behavior of the equilibrium mean fitness as a function of R ? Karlin (1975) asserted that this mean fitness is nonincreasing in R , so that the behavior here agrees with that for additive and multiplicative models. Further, it can be shown (Karlin (1975)) that marginal overdominance holds. That is, if the marginal fitnesses are calculated as in (2.96) for any stable equilibrium point of a symmetric viability system, the marginal heterozygote fitness, at both loci, will exceed that of the corresponding homozygotes.

It is convenient now to list the general conclusions and impressions drawn from our analyses of additive, multiplicative, and symmetric models.

- (i) Only in a restricted class of models (the additive class and more generally the class found by Lyubich (1992)) does the mean fitness increase theorem hold. In other models violations of this theorem are possible.
- (ii) When the double heterozygote is not the only viable genotype, but is at least is the most fit genotype, so that $\alpha_2 > \alpha_1$, $\alpha_2 > \alpha_3$ and $\beta_2 > \beta_1$, $\beta_2 > \beta_3$ in the additive and multiplicative models, α , β , γ , $\delta > 0$ in the symmetric model, there always exists a stable internal equilibrium point for sufficiently small values of R . At least for the symmetric viability model, the existence of stable equilibria for larger values of R depends on quite complicated criteria involving R and the selective parameters.

- (iii) Stable equilibria with small R often exhibit linkage disequilibrium whereas stable equilibria with large R often do not.
- (iv) For any given value of R there can be at most two stable polymorphic equilibria.
- (v) The value of the equilibrium mean fitness decreases (or at worst remains stable) as R increases, in accordance with the argument of Fisher outlined in Section 1.5.
- (vi) Whenever a stable equilibrium exists, the marginal fitnesses, computed from (2.96), exhibit induced overdominance.

How general are these conclusions? It may be argued, since they have been derived from fitness matrices possessing special properties, that they are artifacts of the special features assumed and do not apply to a wider class of fitness matrices. To check this, we present conclusions derived algebraically from other fitness matrices and also numerically from arbitrary numerical fitness matrices. The latter have been considered in particular by Karlin (1975) and Karlin and Carmelli (1975), and we draw heavily on their conclusions in our analysis.

We consider first the mean fitness increase theorem (MFIT). We have noted that if the coefficient of linkage disequilibrium D is nonzero at a stable equilibrium point, the MFIT cannot be true as a mathematical theorem. Now (2.94) shows that the requirement $D = 0$ at equilibrium implies $w_i = \bar{w}$ ($i = 1, 2, 3, 4$) at that equilibrium. These equations imply certain constraints on the fitness parameters which will only hold in special cases, and we can conclude that in this sense, as a mathematical statement, the MFIT is false. However, it is perhaps more important to note, from our analysis of QLE, that the result implied by the theorem is “usually” true.

Are there any general fitness schemes other than the additive scheme for which the MFIT is necessarily true? Karlin (1975) found that for the symmetric viability model with fitness matrix

$$\begin{array}{ccc} 0 & 1 - \beta & 0 \\ 1 - \gamma & 1 & 1 - \gamma \\ 0 & 1 - \beta & 0 \end{array} \quad (6.51)$$

with $\beta + \gamma < 1$, the theorem is necessarily true. This model, implying lethality of all double homozygotes, may be regarded as an unusual one, and since no other general class of models implying the mathematical truth of the MFIT has been found, we may conclude that in practice the additive model is the only important example where this theorem holds.

We turn next to the question of the existence of stable equilibria when the double heterozygote is the most fit genotype. In all cases that we considered above a stable equilibrium was guaranteed for small R , and we now ask whether this is true for arbitrary fitness matrices. Karlin (1975)

demonstrated that this is not so: Even when the double heterozygote is the most fit genotype, it is possible, for some fitness schemes, that no internal stable equilibrium point exists for any positive recombination rate.

One fitness scheme possessing this property is, in the notation of (2.91),

$$\begin{matrix} 0.98 & 0.998 & 0.98 \\ 0.976 & 1 & 0.965 \\ 0.97 & 0.995 & 0.96 \end{matrix} \tag{6.52}$$

Are there special conditions on the fitness values that ensure a stable internal equilibrium, at least for small R ? When all four double homozygotes are more fit than any single heterozygote but less fit than the double heterozygote, there are two such equilibria for small R . Perhaps of greater interest is in the case where the double homozygotes are the least fit, with single heterozygotes of intermediate fitness and the double heterozygote most fit. In this case there exists at least one stable internal equilibrium for small values of R . All these results, together with details on the method of proof of these statements are given by Karlin (1975).

We turn next to the conclusion that equilibria for small R usually exhibit linkage disequilibrium, whereas for large R the linkage disequilibrium is small or even zero. Of course this conclusion is not uniformly true: For additive models, for example, the (unique) internal equilibrium (6.32) obtains for all R and exhibits zero linkage disequilibrium. Nevertheless, as a general statement, this conclusion is broadly correct. Thus for $R = 0.5$ the fitness matrix (2.96) possesses a stable equilibrium at the point (2.98), and at this point the linkage disequilibrium is quite small (-0.000935). When $R = 0.001$ there exist two stable equilibria, at the points

$$c_1 = 0.447, \quad c_2 = 0.030, \quad c_3 = 0.022, \quad c_4 = 0.500$$

and

$$c_1 = 0.015, \quad c_2 = 0.503, \quad c_3 = 0.469, \quad c_4 = 0.013.$$

At these two points the coefficient of linkage disequilibrium D is 0.223 and -0.236 respectively, and clearly these values of D are far larger than the value at the point (2.98).

Despite these remarks, the relationship between R and the equilibrium value of D is not quite clear-cut. Franklin and Feldman (1977) provide an example, quite unexpected in view of our previous conclusions, of a fitness matrix for which, with certain values of R , there exist two stable equilibria, at one of which $D = 0$ while at the other $D \neq 0$. An example of a fitness matrix for which this occurs, written in the form (2.91), is

$$\begin{matrix} 0.78 & 0.82 & 0.77 \\ 0.82 & 0.79 & 0.805 \\ 0.77 & 0.805 & 0.795844. \end{matrix} \tag{6.53}$$

Here the equilibrium point $c_1 = 0.29989$, $c_2 = 0.143184$, $c_3 = 0.143184$, $c_4 = 0.683744$, for which $D = 0$, is stable for all R greater than 0.05. At the same time there exists, for each R , a second stable equilibrium for which $D \neq 0$. The location of this equilibrium depends on R . For $R = 0.1$ it is at $c_1 = 0.446233$, $c_2 = 0.223923$, $c_3 = 0.223923$, $c_4 = 0.105920$, and at this point $D = -0.002877$. On the other hand, when $R = 0.5$ the equilibrium point is at $c_1 = 0.443863$, $c_2 = 0.222814$, $c_3 = 0.222814$, $c_4 = 0.110508$, and at this point $D = 0.000596$. Even more unexpectedly, similar behavior can arise under multiplicative fitnesses (Karlin and Feldman, unpublished).

Next we consider the question of the maximum number of stable internal equilibria. Karlin (1975) demonstrated that for small values of R , no more than two such equilibria are possible. For larger R values no mathematical results are available, although the many simulations of Karlin and Carmelli (1975) and others, for arbitrary fitness matrices, suggest strongly that no more than two stable equilibria are possible for any value of R .

The next point concerns the behavior of the equilibrium mean fitness \bar{w}^* as a function of R . In the special cases examined above, \bar{w}^* is nonincreasing with R , and this accords with the verbal discussion of Fisher (1958). Unfortunately, this property does not always hold. While \bar{w}^* is locally (and indeed globally) maximized at $R = 0$, it is possible that for certain ranges of R values, \bar{w}^* increases as R increases. An example of a fitness scheme, in the notation (2.91), for which this occurs is provided by Karlin and Carmelli (1975):

$$\begin{array}{ccc}
 0.462245 & 0.403142 & 0.188776 \\
 0.136754 & 0.481281 & 0.391682 \\
 0.182915 & 0.245957 & 0.182463
 \end{array} \tag{6.54}$$

The equilibrium mean fitness, as a function of R , is shown in Table 6.2. In those cases where \bar{w}^* increases with R over certain R values, the behavior of \bar{w}^* as a function of R is often of the form displayed in this table: An initial decrease in \bar{w}^* is followed by an increase over a small range of R values with an ultimate flattening out of the values of \bar{w}^* as R increases to 0.5. If two stable equilibria exist for certain R values there will exist two curves of \bar{w}^* against R , and it is possible that both such curves exhibit the form of behavior just described.

The form of fitness scheme where this behavior tends to arise is typified by the fitness values in (6.54). One double homozygote and the double heterozygote have larger fitnesses than the remaining genotypes, and the double heterozygote has the largest fitness. This ensures a stable equilibrium at $R = 0$ with two gametes only and, by continuity, for R small a stable equilibrium with two gametes predominating in frequency. For intermediate values of R all genotypes exist at positive frequency, and because of the lower fitness of most genotypes the equilibrium mean fitness here takes its minimum value. For large values of R there is again a fairly high

R	Equilibrium mean fitness
0	0.463385
0.005	0.462887
0.010	0.462480
0.015	0.462168
0.020	0.461958
0.027	0.461845
0.030	0.461866
0.035	0.462003
0.037	0.462095
≥ 0.042	0.462245

Table 6.2. Values of equilibrium mean fitness for the viability matrix (6.54), for various values of R

mean fitness. The conditions required for this behavior are perhaps rather special and, at least in the numerical example given, the effect of R on \bar{w}^* is quite small. Nevertheless it is of some interest to note that in principle at least, this curious behavior can occur.

We turn finally to the question of induced overdominance. In all cases considered above it can be shown that if a stable internal equilibrium exists, the marginal fitness at each locus must exhibit overdominance. Although no numerical counter-example for arbitrary fitness matrices has yet been found, no proof exists that this behavior applies generally. The converse of this proposition is false: There may exist an internal equilibrium exhibiting marginal overdominance which is *unstable*. Thus starting near this equilibrium the gamete frequencies will change and this can produce behavior, if one of the two loci only and its marginal fitnesses are observed, contrary to that of single-locus theory. An example of this is given by Ewens and Thomson (1978). For the symmetric fitness matrix

$$\begin{array}{ccc}
 0.79 & 0.60 & 0.79 \\
 0.90 & 1.00 & 0.90 \\
 0.79 & 0.60 & 0.79
 \end{array} \tag{6.55}$$

written in the form (2.91), the equilibria (6.48) become

$$\begin{aligned}
 c_1^* &= c_4^* = 0.25 \pm 0.25\{1 - R/0.0725\}^{1/2}, \\
 c_2^* &= c_3^* = 0.25 \mp 0.25\{1 - R/0.0725\}^{1/2}.
 \end{aligned}$$

These equilibria exist whenever $R < 0.0725$ and, from (6.49), they are stable whenever $R < 0.05756$. The marginal fitnesses at the B locus always exhibit overdominance, but those at the A locus do only if $R < 0.05971$. Thus whenever $0.05756 < R < 0.05971$, both marginal fitnesses exhibit overdominance at the equilibrium and yet the equilibrium is not stable.

It will be clear that difficulties arise in enunciating general principles for equilibria of two-locus systems. While several conclusions, as discussed above, are generally true, counterexamples are usually possible. These sometimes refer to cases of unlikely biological interest, and it remains a challenge to discover a principle that describes normal behavior in biologically realistic circumstances.

6.5 Modifier Theory

One of the most interesting applications of two- (or multi-) locus theory arises when one of the loci considered is a modifier locus, that is the genes present at that locus modify in some way the values of various genetic parameters (for example mutation rate, recombination fraction) at or between other loci. These other loci are called the primary loci, and our main interest is to consider the way in which evolutionary processes at these loci are affected by the existence of the modifier locus.

Two general classes of modifier locus theory may be defined. In the first class it is supposed that an individual's fitness depends in part on his genetic constitution at the modifier locus. The modification process for the evolution of dominance, mentioned already and discussed in more detail below, falls in this class, as the fitnesses in (1.92) show. In the second class of modification schemes the fitness of any individual is assumed to be independent of his genetic constitution at the modifier locus. Any evolution at the modifier locus is then a result of its interaction with the primary loci and may then be described as being due to secondary selection. This class of modification schemes was introduced into the literature by Nei (1967), who considered a modifier controlling the recombination fraction between two primary loci.

Our first example is of the former class and concerns Fisher's (1928a,b, 1930b, 1931, 1934) theory of the evolution of dominance through the action of a modifier locus. We have already proposed in (1.90) a fitness scheme for this situation. A proper description of the joint evolution at A and M loci requires consideration of the frequencies of the four gametes A_1M_1 , A_1M_2 , A_2M_1 and A_2M_2 , but, if s is small and the recombination fraction between primary and modifier locus is not small, no serious error is made by making the approximation that linkage equilibrium obtains throughout the joint evolutionary process at the A and the M loci. Under this approximation the frequency of any gamete may be written as the product of the frequencies of the two constituent alleles. A more refined analysis (Feldman and Karlin (1971)) confirms that the error involved in making this approximation is negligible.

Suppose that the frequency of A_1 is close to unity and that A_1 mutates to A_2 at rate u . Then under our simplifying assumptions, if x is the frequency

of A_1 and y that of M_1 ,

$$\Delta y \approx sx(1-x)y(1-y)\{4ky + 2h - 2hy - 2k\}, \quad (6.56)$$

$$\Delta x \approx sx(1-x)\{1-x + (2x-1)(1-y)(2ky + h - hy)\} - ux. \quad (6.57)$$

The final term in (6.57) does not of course come from (2.94) but arises from the recurrent mutation from A_1 to A_2 . Clearly both the frequency x and the frequency y change over successive generations, and our aim is to use (6.56) and (6.57) to find an expression for the change Δy that is independent of x . Before doing this we observe from (1.34), with a slight change of notation, that the equilibrium value of x when $y = 0$ is $1 - u/sh$ and when $y = 1$ is $1 - (u/s)^{1/2}$. Thus $x(1-x)$ always lies in the interval

$$\left\{ (u/sh)(1-u/sh), (u/s)^{1/2}\{1 - (u/s)^{1/2}\} \right\}, \quad (6.58)$$

and is consequently always bounded above by $(u/s)^{1/2}$. This implies that Δy is always bounded above by

$$(us)^{1/2}y(1-y)\{4ky + 2h - 2hy - 2k\}. \quad (6.59)$$

For $k = \frac{1}{2}h$ this yields the upper bound

$$\Delta y \leq (us)^{1/2}hy(1-y), \quad (6.60)$$

while for $k = 0$ it yields the upper bound

$$\Delta y \leq 2h(us)^{1/2}y(1-y)^2. \quad (6.61)$$

One way of obtaining a more accurate assessment of the value of Δy would be to use (6.56) and (6.57) to form the differential equation

$$\frac{dy}{dx} = \frac{s(1-x)y(1-y)\{4ky + 2h - 2hy - 2k\}}{s(1-x)\{1-x + (2x-1)(1-y)(2ky + h - hy)\} - u}. \quad (6.62)$$

If this equation could be solved for x (as a function of y), the resulting solution could be inserted in (6.56) to obtain a very accurate value for Δy . Unfortunately no solution of (6.62) has yet been found, and the best that appears possible in this direction is to solve (6.62) numerically: This would, however, be equivalent to a joint numerical solution of (6.56) and (6.57).

A slightly less accurate method is to argue as follows. For any fixed value of y there will exist an equilibrium value of x , somewhere in the interval (6.58). This value is found by solving the equation $\Delta x = 0$. Although we cannot expect x to have reached this equilibrium value for any current value of y , a reasonable approximation is obtained by assuming that it has. The solution of the equation $\Delta x = 0$, that is of the equation

$$(1-x)^2 + (1-x)(2x-1)(1-y)(2ky + h - hy) = u/s, \quad (6.63)$$

is not as straightforward as might initially appear. Since $x \approx 1$ one might be tempted to ignore the term in $(1-x)^2$ and, putting $2x-1 \approx 1$, to write

$$1-x \approx (u/s)\{(1-y)(2ky + h - hy)\}^{-1}.$$

Insertion of this in (6.56) gives, for the important particular case $k = h/2$,

$$\Delta y \approx uy, \quad (6.64)$$

which is the value given by Fisher (1928b). Suppose we write $\Delta y = y(1 - y)\phi(y)$, and (with Fisher) call $\phi(y)$ the “selective intensity in favor of the modifier”. Then (6.64) shows that

$$\phi(y) \approx u(1 - y)^{-1} \quad (6.65)$$

so that $\phi(y)$ becomes indefinitely large as $y \rightarrow 1$. This conclusion was stressed by Fisher (1928c) as an essential point of his argument, since it appears to imply a strong selective force on the modifying allele.

The above analysis, however, is incorrect. This can be seen immediately by observing that it leads to a violation of the upper bound (6.60). For $y \approx 1$ the term $(1 - x)^2$ becomes the dominant factor on the left-hand side of (6.63), and it therefore should not be ignored when $y \approx 1$. It is, however, possible to replace the term $2x - 1$ in (6.63) by 1 with little loss of accuracy, since the error involved in doing this is of order $(1 - x)^2(1 - y)$ and is thus always extremely small. Solving the resulting equation gives

$$1 - x = \frac{1}{2}(- (1 - y)(2ky + h - hy) + \{(1 - y)^2(2ky + h - hy)^2 + 4u/s\}^{1/2}),$$

and insertion of this value for $1 - x$ in (6.56) shows that to a very close approximation,

$$\begin{aligned} \Delta y \approx \frac{1}{2}sy(1 - y)(4ky + 2h - 2hy - 2k) \times & \{-(1 - y)(2ky + h - hy) \\ & + \{(1 - y)^2(2ky + h - hy)^2 + 4u/s\}^{1/2}\}. \end{aligned} \quad (6.66)$$

For the special case $k = h/2$ this becomes

$$\Delta y \approx \frac{1}{2}shy(1 - y)(-h(1 - y) + \{h^2(1 - y)^2 + 4u/s\}^{1/2}). \quad (6.67)$$

Since

$$-h(1 - y) + \{h^2(1 - y)^2 + 4u/s\}^{1/2} \leq (4u/s)^{1/2},$$

the approximation (6.67) gives

$$\Delta y \leq (us)^{1/2}hy(1 - y), \quad (6.68)$$

in agreement with (6.60). Defining $\phi(y)$ as above we get, to a close approximation,

$$\phi(y) \leq (us)^{1/2}h. \quad (6.69)$$

Parallel calculations for the case $k = 0$ show that, to a close approximation,

$$\Delta y \leq y(1 - y)\{1.24h^{1/2}u^{3/4}s^{1/4}\} \quad (6.70)$$

for this case. Similar expressions arise for other values of k . Numerical calculations, derived from the recurrence relations governing gamete frequencies,

show (Ewens (1967)) that the approximations (6.67)–(6.70) are very accurate, and confirm our belief that linkage equilibrium can be assumed to a close approximation throughout this process.

What value do the above calculations have for the evolution of dominance question? It is clear that $\phi(y)$ is always very small, and certainly does not become infinitely large, as Fisher's analysis claimed. It is essentially because of this observation that Wright (1929a,b) originally cast doubt on Fisher's theory. We have noted, in Chapter 1, Wright's emphasis on the pleiotropic effects of genes. If, in line with his view, the modifier gene is subject to a primary selection pressure quite independent of its modification action, the very small selective pressure due to dominance modification cannot control its evolution and is essentially irrelevant. Fisher (1929) resists this viewpoint, but, quite apart from the bias in his argument induced by the mathematical error noted above, this author finds his position unconvincing.

The above analysis has assumed that from the start, the favored primary allele A_1 is always at a high frequency and the small effect of dominance modification is in large part due to the very low frequency of heterozygotes A_1A_2 throughout the process. In some cases, for example with the classic evolution of the melanic form in the moth *Biston betularia* due to industrial pollution, the eventually favored form starts at a low frequency. Thus during the course of its frequency increase many heterozygotes will appear upon which the force of dominance modification can act. The extent to which the frequency of the modifier is changed in this way depends on the degree of linkage between primary and modifying loci. If the two loci are closely linked there is some possibility for a substantial increase in the frequency of the modifier, and this tendency is magnified the larger the primary locus selective differences are. At the same time, once the frequency of A_1 reaches a high value the induced selective force on the modifier becomes very weak and the argument of Wright outlined above will again prevail.

We turn now to other ways in which modifier loci can act, considering in particular modification of linkage and mutation rate. In this way we wish to give an explanation for the evolution of these characteristics which is independent of arguments involving inter-population competition at least implicit in the early discussions of them. The classic papers introducing modifiers which do not change fitnesses were those of Nei (1967, 1969), who considered modification of linkage. We follow here, however, the discussion of this topic given by Feldman (1972).

Consider two primary loci A and B and a modifier locus M , the effect of which is to influence the recombination fraction between A and B loci. Let these loci lie on a chromosome in the order MAB , with recombination values R between M and A loci, R_{ij} between A and B loci and $R + R_{ij} - 2RR_{ij}$ between M and B loci. Here R_{ij} depends on the genotype M_iM_j at the modifier locus.

Our mode of approach is the following. Suppose the various genotypes at A and B loci have fitnesses specified by (2.90). If all individuals at the M locus are M_1M_1 the recombination fraction between A and B loci is R_{11} , and we suppose that the population has reached a stable equilibrium point for this recombination fraction with gamete frequencies c_1^* , c_2^* , c_3^* and c_4^* .

Suppose next that the allele M_2 is now introduced at a low frequency at the M locus. The frequencies of the gamete $M_1A_1B_1$, $M_1A_1B_2$, $M_1A_2B_1$ and $M_1A_2B_2$ then become $c_1^* + \delta_1$, $c_2^* + \delta_2$, $c_3^* + \delta_3$, $c_4^* + \delta_4$, and we write the frequencies of $M_2A_1B_1$, $M_2A_1B_2$, $M_2A_2B_1$ and $M_2A_2B_2$ as δ_5 , δ_6 , δ_7 , δ_8 . We now set up recurrence relations extending (2.94) for the eight gamete frequencies which, if terms in δ_i^2 , $\delta_i\delta_j$ are ignored, become linear recurrence relations in δ_5 , δ_6 , δ_7 and δ_8 . If all the eigenvalues of the matrix governing this recurrence system are less than unity, then $\delta_i \rightarrow 0$ and the system returns to its original equilibrium with M_2 absent. If at least one eigenvalue is greater than unity in absolute value the frequency of M_2 will increase, and the recombination fraction between A and B loci, for those individuals carrying the M_2 gene, will change. Clearly our objective is to find the circumstances, in terms of these eigenvalues, which lead to this increase.

In general this proves to be rather difficult, although we present later a general argument that at least suggests what the nature of these eigenvalues is. In the additive, multiplicative and symmetric viability models, Feldman showed that provided $c_1^*c_4^* \neq c_2^*c_3^*$, the frequency of M_2 will increase if and only if $R_{12} < R_{11}$, that is if and only if the modifier heterozygote M_1M_2 leads to tighter linkage between A and B loci than does M_1M_1 .

There is no reason to suppose that this conclusion does not apply for general fitness matrices. That this is so is suggested by a conclusion of Feldman and Krakauer (1976). Let the fitness matrix (2.90) at A and B loci be arbitrary and suppose, in the above notation, that $R_{12} < R_{11}$, R_{22} . Now define

$$m_1 = (R_{22} - R_{12}) / (R_{11} + R_{22} - 2R_{12}) = 1 - m_2,$$

$$R^* = m_1^2 R_{11} + 2m_1 m_2 R_{12} + m_2^2 R_{22}.$$

Suppose now that $(c_1^*, c_2^*, c_3^*, c_4^*)$ is a solution of the equilibrium equations (2.95) if $R = R^*$. Then Feldman and Krakauer demonstrated that there is an equilibrium of the three-locus system at which

$$\begin{aligned} \text{freq}(M_i A_1 B_1) &= m_i c_1^*, & \text{freq}(M_i A_1 B_2) &= m_i c_2^*, \\ \text{freq}(M_i A_2 B_1) &= m_i c_3^*, & \text{freq}(M_i A_2 B_2) &= m_i c_4^*, \end{aligned}$$

for $i = 1, 2$. The stability properties of this equilibrium have not yet been determined but, if this equilibrium is stable at least for certain recombination values, it strongly suggests the evolution of tighter linkage between A and B loci by secondary selection. We note in passing the curious re-

semblance between the formula for m_1 and the equilibrium gene frequency (1.31) in a one-locus selective scheme.

We turn now to modification of mutation rate. Suppose the fitnesses at the primary locus A are given by (1.25b) with $s > sh > 0$; cases with complete dominance in fitness can be considered similarly. Suppose that the mutation rate $A_1 \rightarrow A_2$ is controlled by the genes present at a modifier locus and is u_{ij} for individuals of genotype M_iM_j . We may suppose that initially the frequency of M_1 is unity and that the frequency of A_1 is at the mutation-selection equilibrium value $1 - u_{11}/s(1-h)$. The population mean fitness is now $1 - 2u_{11}$. The allele M_2 is now introduced at a low frequency. By considering linearized recurrence relations for the four gametic types, it is found that the frequency of M_2 increases if and only if $u_{12} < u_{11}$. More generally the frequency of M_2 will steadily increase to unity if $u_{22} < u_{12} < u_{11}$, so that the mutation rate $A_1 \rightarrow A_2$ becomes u_{22} and the population mean fitness becomes $1 - 2u_{22}$. If $u_{12} < u_{11}$, u_{22} a polymorphism is established at the M locus. All these conclusions are true irrespective of the linkage arrangement between primary and modifier loci. Again, we attempt to restate these conclusions later as particular applications of a general modifier principle.

Of course the question of the establishment of an optimal mutation rate requires arguments more complex than these, and must take into account the need for long-term flexibility perhaps enhanced by a higher mutation rate. These matters were discussed in Chapter 1, and our interest in them here is that inter-population selection arguments are not required to arrive at an agency reducing the mutation rate.

It is also possible to discuss the dynamics of modifiers of sex-ratio, migration and selfing (Feldman and Krakauer (1976), Karlin and McGregor (1974)). Instead of discussing these cases specifically, we turn instead to the question of whether there exists a general principle for modifier loci embracing the conclusions just reached as particular cases. Such a general principle has been proposed by Karlin and McGregor (1974). Although this principle does not have the status of a mathematical theorem, it nevertheless applies widely for modifier loci. Consider a primary locus (or loci) with evolution determined at least in part by a parameter θ , for example a mutation rate. Suppose the value of θ is determined by a selectively neutral modifier locus and that for individuals of genotype M_iM_j ($i, j = 1, 2$) the parameter takes the value θ_{ij} . Assume that random mating obtains, and let $\bar{w}(\theta_{ij})$ be the mean fitness of the primary system at a stable equilibrium when $\theta = \theta_{ij}$. Then if

$$\bar{w}(\theta_{11}) < \bar{w}(\theta_{12}) < \bar{w}(\theta_{22}) \quad (6.71)$$

the allele M_2 will become fixed at the modifier locus. If the inequalities in (6.71) are reversed, M_1 will become fixed, while if $\bar{w}(\theta_{12}) > \bar{w}(\theta_{11})$, $\bar{w}(\theta_{22})$ a stable polymorphism will arise at the modifier locus. Thus this principle essentially asserts that the evolution at the modifier locus is such

as to maximize mean fitness, and this has been observed in the specific examples above. We have proved earlier that in the multiplicative and symmetric viability models the mean fitness is nonincreasing in R , and this principle then indicates that modifiers decreasing the recombination fraction become fixed. This is in agreement with the conclusion reached by Feldman (1972) described above, and generalizes that conclusion by not restricting attention to the frequency of the modifier M_2 when it is small. This principle suggests that there are circumstances where *increased* recombination between loci is favored.

6.6 Two-Locus Diffusion Processes

In this section we consider multidimensional diffusion analogues to the two-locus two-allele Markov chain models (3.130) and (3.131), using throughout this section the same notation as that used in Section 3.7. The evolution of the models described by (3.130) and (3.131) can be described by considering the linearly independent frequencies c_1 , c_2 and c_3 . However, there is an obvious asymmetry in doing this, and in any event we find it more convenient to work with the generation t frequencies $x(t) = c_1 + c_2$, $y(t) = c_1 + c_3$ and $D(t) = c_1c_4 - c_2c_3$. To use the diffusion theory of Section 5.1 we must assume that means, variances and covariances of the changes of these frequencies between consecutive generations are of order N^{-1} . This will require us to assume in particular that R is $O(N^{-1})$.

We suppose first that there is no mutation, and consider initially the RUZ model. Then from (3.139),

$$E\{D(t+1) - D(t) \mid c_i(t)\} = -(2N)^{-1}\{1 - 2NR\}D(t). \quad (6.72)$$

If R is of order N^{-1} this gives $(1 + 2NR)D$ as the drift coefficient corresponding to D , and it is found that the same value applies for the RUG model. Indeed it is found that all the coefficients for the diffusion process approximating the RUG model are the same as for the RUZ model, so that the diffusion approximations to the two processes are identical.

Using (6.72) and the corresponding values for p and q , as well as the variance and covariance terms for these quantities, it is found for both models that if we scale time so that unit time T corresponds to N generations of the Markov chain, the backward Kolmogorov equation for the joint density of the frequency of A_1 , the frequency of B_1 , and the linkage disequilibrium at time t , when there is no selection, and given the initial values p , q and

γ	0.5	5	10	25
λ_1	-0.813	-0.9926	-0.9978	-0.9996
$\exp(\lambda_1/50)$	0.984	0.980	0.980	0.980

Table 6.3. The largest solution (λ_1) of (6.74) for various values of γ

D for these variables, becomes (Ohta and Kimura, 1969a)

$$\begin{aligned}
 \frac{\partial f}{\partial t} = & \frac{1}{4}p(1-p)\frac{\partial^2 f}{\partial p^2} + \frac{1}{4}q(1-q)\frac{\partial^2 f}{\partial q^2} + \frac{1}{2}D\frac{\partial^2 f}{\partial p\partial q} + \frac{1}{2}D(1-2p)\frac{\partial^2 f}{\partial p\partial D} \\
 & + \frac{1}{2}D(1-2q)\frac{\partial^2 f}{\partial q\partial D} + \frac{1}{4}\{pq(1-p)(1-q) \\
 & + D(1-2p)(1-2q) - D^2\}\frac{\partial^2 f}{\partial D^2} \\
 & - \frac{1}{2}(1+2NR)\frac{\partial f}{\partial D}. \tag{6.73}
 \end{aligned}$$

Here f is the joint density function of $x(t)$, $y(t)$ and $D(t)$, given initial values p, q and D for these frequencies. Our aim is to use this equation, in conjunction with the theory of Section 4.10, to find diffusion analogues for various quantities established for the Markov chain models of Section 3.9. We focus here on the diffusion process eigenvalues and the diffusion process expectation of $\{D(t)\}^2$. These have been found by Ohta and Kimura (1969a), and we follow their analysis closely.

Our point of departure for both problems is to consider the expectations of the three quantities $x(t)(1-x(t)) \times y(t)(1-y(t))$, $D(t)(1-2x(t))(1-2y(t))$, and $D(t)^2$. Simultaneous equations for these expected values can be found by using (4.83). By inserting trial eigenvalue solutions with undetermined coefficients, Ohta and Kimura found that the expectations of these quantities converge to zero at rate $\exp(\lambda_1 t)$, where λ_1 is negative and is the largest root of the equation

$$4\lambda^3 + (20 + 12\gamma)\lambda^2 + (27 + 38\gamma + 8\gamma^2)\lambda + (9 + 26\gamma + 8\gamma^2) = 0. \tag{6.74}$$

In this equation $\gamma = NR$ and from our assumptions is $O(1)$. While an explicit solution of this cubic equation is possible, it is perhaps preferable to solve (6.74) numerically for selected values of γ , and some specific solutions are given in Table 6.3.

It is naturally of interest to compare these values with those in Table 3.1. Since unit time in the diffusion corresponds to N generations in the Markov chain, we should compare $\exp \lambda_1 t$ with μ^{Nt} , where μ is the Markov chain eigenvalue, or equivalently can be compared to the final line in Table 3.1. It will be noted that the agreement is excellent, thus showing that the leading eigenvalue for the Markov chain is closely approximated by that for the diffusion process.

The value of $E\{D(t)\}^2$ turns out to be a complicated expression involving all three eigenvalue solutions of (6.74). We do not give an explicit expression here: It is sufficient to note that the value obtained is in excellent agreement with the Markov chain solution given in Table 3.1. It appears, with both conclusions, that the requirement that R be $O(N^{-1})$ does not appear to be necessary for the agreement between Markov chain and diffusion solutions. This no doubt occurs because, in any expression where it occurs, R is invariably multiplied by the coefficient of linkage disequilibrium which, when R is not small, is usually small itself.

Suppose now that mutation exists, so that there is a stationary distribution of gene frequencies. We now change notation so that x , y and D denote these stationary distribution frequencies. The expectation $E(D^2)$ in this stationary distribution is of particular interest to us. We write the mutation rates as u_1 from A_1 to A_2 , v_1 from A_2 to A_1 , u_2 from B_1 to B_2 and v_2 from B_2 to B_1 , and assume that all mutation rates are $O(N^{-1})$. The drift and diffusion coefficients for the changes in $x(t)$, $y(t)$ and $D(t)$ (now including terms in the mutation rates) can be inserted in (4.84) to find the stationary expectation of any function $g(x, y, D)$ of these variables. The equation so obtained (Ohta and Kimura, 1969b) is

$$\begin{aligned} E\left(x(1-x)\frac{\partial^2 g}{\partial x^2} + y(1-y)\frac{\partial^2 g}{\partial y^2} + 2D\frac{\partial^2 g}{\partial x\partial y} + 2D(1-2x)\frac{\partial^2 g}{\partial x\partial D} \right. \\ \left. + 2D(1-2y)\frac{\partial^2 g}{\partial y\partial D} + \{xy(1-x)(1-y) \right. \\ \left. + D(1-2x)(1-2y) - D^2\}\frac{\partial^2 g}{\partial D^2} \right. \\ \left. + 4N\{v_1 - (u_1 + v_1)x\}\frac{\partial g}{\partial x} + 4N\{v_2 - (u_2 + v_2)y\}\frac{\partial g}{\partial y} \right. \\ \left. - D(2 + 4NK)\frac{\partial g}{\partial D}\right) = 0, \end{aligned} \quad (6.75)$$

where $K = R + u_1 + u_2 + v_1 + v_2$ and the drift and diffusion coefficients implied by the mutation rates are displayed in the equation. The expectation is with respect to the joint stationary distribution of x , y , and D . Our aim is to make suitable choices for g so that three simultaneous equations can be found from which $E(D^2)$ can be obtained. Three choices of g which do this are

$$g_1 = xy(1-x)(1-y) \quad (6.76)$$

$$g_2 = D(1-2x)(1-2y) \quad (6.77)$$

$$g_3 = D^2. \quad (6.78)$$

Inserting $g = g_3$ from (6.78) into (6.75), we obtain

$$E(g_1 + g_2 - g_3(3 + 4NK)) = 0. \quad (6.79)$$

The remaining two equations yield expectations for g_1 , g_2 and g_3 in terms of expectations for “lower-order” quantities such as $x(1-x)$ and xy . The expectations of these lower-order quantities may be found from (6.75) by choosing $g = x(1-x)$, $g = xy$, and so on. The joint solution of the resulting equations and (6.79) gives

$$E(D^2) = \frac{\{2.5 + N(K + U)\}NA}{(1 + 2NU)(3 + 4NK)(2.5 + NU + NK) - 3 - 4NU}, \quad (6.80)$$

where

$$U = u_1 + u_2 + v_1 + v_2 \quad \text{and} \\ A = \frac{8Nu_1u_2v_1v_2}{(u_1 + u_2)(v_1 + v_2)} \left(\frac{1}{4N(u_1 + v_1) + 1} + \frac{1}{4N(u_2 + v_2) + 1} \right). \quad (6.81)$$

We may usually assume mutation rates are sufficiently small so that NU is moderate. Unless the two loci are very closely linked, NR and hence NK will both be large, and in this case we find that

$$E(D^2) \sim A/\{4R(1 + 2NU)\}. \quad (6.82)$$

This expression is of the same order of magnitude as the mutation rate, and will thus usually be very small. This reinforces the conclusion we have reached above, that random processes in finite populations are of minor importance in causing nonzero values of the coefficient of linkage disequilibrium.

Some interest also attaches to the standardized linkage disequilibrium σ_D^2 , defined by

$$\sigma_D^2 = E(D^2)/\{Exy(1-x)(1-y)\}. \quad (6.83)$$

The expectations in the denominator can be computed using (6.75)–(6.78), and we find that if NR is large,

$$\sigma_D^2 \approx (4NR)^{-1}. \quad (6.84)$$

This is again small. If the two loci are tightly linked so that NR is not large, a more accurate expression (Ohta and Kimura (1969b)) is

$$\sigma_D^2 = [3 + 4NK - 2\{2.5 + NK + NU\}^{-1}]^{-1}. \quad (6.85)$$

6.7 Associative Overdominance and Hitchhiking

We consider in this section two concepts of potential practical interest which arise in finite populations with nonzero values of the coefficient of linkage disequilibrium.

The first concept is that of associative overdominance. This was introduced by Frydenberg (1963) to explain secular gene frequency changes in

certain experimental *Drosophila* populations. The essence of the notion is that whereas the genes at the locus of interest may be selectively equivalent, they appear to exhibit overdominance by being linked (with nonzero values of linkage disequilibrium) to a locus where true overdominance does occur. The most detailed theoretical treatment of this concept is by Kimura and Ohta (1971a, pp. 110–116).

Consider two loci A and B for which true overdominance occurs at the A locus while the B locus is selectively neutral, so that the fitnesses of the various genotypes are

$$\begin{array}{ccc} A_1A_1B - B- & A_1A_2B - B- & A_1A_2B - B- \\ 1 - s_1 & 1 & 1 - s_2. \end{array}$$

We denote the frequencies A_1 and B_1 by x and y respectively. Then the frequencies of the gametes A_1B_1 and A_2B_1 are $xy + D$ and $(1 - x)y - D$, and thus the marginal fitness of B_1B_1 individuals (see (2.96)) is

$$\begin{aligned} & y^{-2}[(xy + D)^2(1 - s_1) + 2(xy + D)\{(1 - x)y - D\} \\ & \quad + (1 - s_2)\{(1 - x)y - D\}^2] \\ & = 1 - s_1(x_y^{-1}D)^2 - s_2(1 - x - y^{-1}D)^2. \end{aligned} \quad (6.86)$$

Similarly the marginal fitness of B_1B_2 individuals is

$$\begin{aligned} & 1 - s_1(x - y^{-1}D)\{x - (1 - y)^{-1}D\} \\ & \quad - s_2\{1 - x + (1 - y)^{-1}D\}(1 - x - y^{-1}D), \end{aligned} \quad (6.87)$$

while that of B_2B_2 individuals is

$$1 - s_1\{x - (1 - y)^{-1}D\}^2 - s_2\{1 - x + (1 - y)^{-1}D\}^2. \quad (6.88)$$

The apparent selective advantage of B_1B_2 over B_1B_1 is

$$s_1D(x + y^{-1}D)\{y(1 - y)\}^{-1} - s_2D(1 - x - y^{-1}D)\{y(1 - y)\}^{-1} \quad (6.89)$$

and over B_2B_2 is

$$\begin{aligned} & -s_1D\{x - (1 - y)^{-1}D\}\{y(1 - y)\}^{-1} \\ & \quad + s_2\{1 - x + (1 - y)^{-1}D\}\{y(1 - y)\}^{-1}. \end{aligned} \quad (6.90)$$

There is one case of these formulas of special interest. If the selection at the A locus is so strong that we may assume $x = x^* = s_2/(s_1 + s_2)$, the apparent selective disadvantages become

$$(s_1 + s_2)D^2/\{y^2(1 - y)\} \quad \text{and} \quad (s_1 + s_2)D^2/\{y(1 - y)^2\} \quad (6.91)$$

respectively. These are non-negative, so that in this case, if there is nonzero linkage disequilibrium between selected and neutral loci, apparent (or associative) overdominance will exist at the neutral locus. Clearly the extent of this effect will depend on the value of D^2 or, in the more general case where we cannot assume $x = s_2/(s_1 + s_2)$, on D and D^2 . We now discuss how

large this effect might be when linkage disequilibrium is generated from stochastic processes in finite populations.

The formula we have derived for $E(D^2)$ in the previous section assumes no selective effects at either locus and must therefore be generalized to cover the present model. This has been done by Ohta and Kimura (1970) assuming that selection is so strong that $x = x^*$. It is found that

$$E(D^2) = \frac{x^*(1-x^*)E\{y(1-y)\}}{1 + 4N(R+u+v) + \frac{(1-2x^*)^2}{x^*(1-x^*)} \frac{N(R+2u+2v)}{1+N(R+2u+2v)}}. \quad (6.92)$$

Here u and v are the mutation rates at the B locus. If R is not extremely small, this expression is $O(N^{-1})$ and hence is very small. We may thus expect little effect of associative overdominance in this case. Similarly, when there is no mutation and fixation at both loci eventually occurs, the effect of linkage disequilibrium, while perhaps initially nontrivial, eventually becomes negligibly small, so that associative overdominance is, in this case, a transient phenomenon. Extensions of these conclusions to the case where several overdominance loci are linked to the neutral locus are given by Ohta and Kimura (1970).

We turn now to the concept of hitchhiking. Hitch-hiking occurs when the gene frequencies at a neutral locus are affected by those at a linked selected locus where a favorable allele is proceeding towards fixation. As the name implies, we are mainly interested in the extent to which the frequency of one allele at the neutral locus increases through linkage to the favored allele. Aspects of this possibility have been examined by Maynard Smith and Haigh (1974), Haigh and Maynard Smith (1976), Ohta and Kimura (1975, 1976) and Thomson (1977).

Haigh and Maynard Smith consider a somewhat different question than do Ohta and Kimura. They assume an initial polymorphism at the neutral locus and, supposing that a substitution then occurs at the selected locus, focus attention on the expected final value of heterozygosity at the neutral locus when the substitution ceases. Ohta and Kimura, on the other hand, imagine a new mutation to arise at a neutral locus while a selected locus is substituting, and consider the effect on the expected total heterozygosity at the neutral locus of the selected substitution. Which consideration is the more relevant biologically is not clear, and both sets of authors argue for their own viewpoint.

The purely mathematical discussion is less controversial, and we consider first the analysis of Ohta and Kimura. In order to have a standard of reference we consider the model (1.48), which concerns a selectively neutral locus without reference to any linked loci. If a single A_1 mutant arises in an otherwise pure A_2A_2 population, the number of A_1 genes will be j on an average for $2j^{-1}$ generations (see (1.56)). This means, on average, that

the total heterozygosity created by this mutation is

$$\sum_{j=1}^{2N-1} 2j(2N-j)(2N)^{-2}2^{j-1} = 2. \quad (6.93)$$

We shall take this value as a standard against which values computed under hitchhiking may be compared.

Suppose the A locus is selectively neutral while, at the B locus, the favored allele B_1 is steadily replacing B_2 . We may assume, to a reasonable approximation, that this replacement is deterministic, so that the frequency y of B_1 satisfies the differential equation

$$\frac{dy}{dt} = sy(1-y). \quad (6.94)$$

(For convenience, this assumes no dominance at the B locus.) We denote by x_1 the frequency of A_1 among B_1 chromosomes and by x_2 the frequency of A_1 among B_2 chromosomes. The total frequency of A_1 is thus $x = yx_1 + (1-y)x_2$, and our aim is to compute the expected value of the function H , defined by

$$H = \int_0^{\infty} 2x(1-x) dt. \quad (6.95)$$

Ohta and Kimura approach this problem by using (4.83). Differential equations for the expected value of x_1^2 , x_1x_2 , and x_2^2 are found from (4.83) by successively using these functions for $g(\cdot)$. These equations must be solved numerically and the solutions inserted in (6.95). These depend on the initial values of x_1 and x_2 , the selective coefficient s , the recombination fraction R between A and B loci and the value y_0 assumed for y when the initial mutation at the A loci takes place. Ohta and Kimura (1975) give separate values for $E(H)$ depending on whether the initial mutant lies on a B_1 or B_2 chromosome. For our purposes it is probably convenient to consider the weighted average

$$\bar{E}(H) = y_0E_1(H) + (1-y_0)E_2(H), \quad (6.96)$$

where $E_i(H)$ is the expected value of H assuming the initial mutant is on a B_i chromosome. In Table 6.4 we give values of $\bar{E}(H)$ for various values of s and y_0 for the values $R = 0.1$, $N = 100$, computed from those of Ohta and Kimura (1975). It will be seen that $\bar{E}(H)$ does not differ substantially from the value 2, computed without taking linked loci into account, and from this point of view we may conclude, with Ohta and Kimura in this case, that hitchhiking is of comparatively small importance in altering the value of total mean heterozygosity at the neutral locus. Although we have considered only one value for N and one value for R in Table 6.4, the general conclusion reached applies for a very wide range of R and N values, and

	y_0	0.1	0.2	0.5
s	0.05	1.97	1.97	1.97
	0.10	1.94	1.94	1.96

Table 6.4. Values of $\bar{E}(H)$ for various values of s and y_0 with $R = 0.1$, $N = 100$

is appreciably in error only when $NR < 5$ and $Ns < 100$. The minimum expected value of H is about 1.2, and occurs when $NR \sim 0$, $Ns \approx 5$.

Maynard Smith and Haigh considered it more relevant biologically to consider the effect on an existing neutral polymorphism of a selective substitution and do so by comparing the expected final heterozygosity in this case with that existing before the selective substitution starts. They show that if $R \ll s \ll 1$, the ratio of the final heterozygosity H_∞ (when the selective substitution has taken place) to the initial heterozygosity H_0 is of the form

$$\text{const} \times Rs^{-1}, \quad (6.97)$$

where the constant depends on the initial gamete configuration. Under the assumptions made the quantity (6.97) is quite small. However, Ohta and Kimura (1975) computed H_∞/H_0 for a far wider range of R and s values and conclude that unless Ns is small (< 100 , approximately) then $H_\infty/H_0 \sim 1$. In particular this is true if $R > s$, and thus in this case the effect of hitchhiking is negligible.

One of the theoretical results of Maynard Smith and Haigh (1974) can be used in a test for a hitchhiking event, and is discussed in more detail in Section 11.3.5. Consider a diploid population of N individuals and suppose that initially only A_2A_2 individuals exist in the population. A single favorable new A_1 mutant now arises at this locus. The fitnesses of A_1A_1 , A_1A_2 and A_2A_2 are $1 + 2s$, $1 + 2$ and 1 respectively, with $s > 0$, and as a result A_1 increases in frequency, under a deterministic analysis, from a frequency $(2N)^{-1}$ to a value close to 1. We call this event a selective sweep. Suppose that the recombination between this selected locus and a closely linked neutral locus is R and that at the neutral locus there are two alleles B_1 and B_2 , with frequencies x and $1 - x$ at the start of the selective sweep. Then under a deterministic theory, if the new mutant A_1 gene is on the same gamete as B_1 before the selective sweep, the frequency of B_1 will increase from x to the value $1 - c + cx$ at the conclusion of the selective sweep, where, to a reasonable approximation,

$$c = \frac{R \log 2N}{s}. \quad (6.98)$$

If the new mutant A_1 gene is on the same gamete as B_2 before the selective sweep, the frequency of B_1 will decrease to cx at the conclusion of the selective sweep.

We return to this calculation in Section 11.3.5, and remark here only that the possibility of hitchhiking during a selective sweep is one reason why we should view much of the single-locus theory of population genetics, where loci are treated individually with no regard to events at linked loci, with some caution.

6.8 The Evolutionary Advantage of Recombination

We have noted above that the mean fitness of a random-mating population is maximized when the recombination fraction between the two loci considered is zero. (This conclusion may be generalized to cover an arbitrary number of loci with an arbitrary number of alleles at each locus.) Why then have populations not evolved so that recombination does not exist? Even if we were to discount the use of mean fitness as a measure of success in intergroup competition and, further, assert that in any event recombination is determined more by the evolution of modifier genes than by such competition, we must still find an answer to this question, since in the analyses we have so far mentioned, such modifiers often act so as to reduce recombination.

Our aim then is to find what advantage the existence of recombination might be for a population. In this context we shall mean sexual recombination: We do not consider the possible advantages of recombination in asexual populations. Thus, in what follows we assume the existence of two sexes with identical fitness patterns. A generalization of the theory of Section 2.3 shows that the frequency of any gamete will be the same in males and females, so that in the quantitative discussion of recombination no explicit recognition of the existence of the two sexes is necessary.

The classical argument for the existence of sexual recombination is that of Fisher (1930a) and Muller (1932), namely that recombination favors the incorporation into the population of favorable new alleles arising at different loci, since recombination is more efficient in allowing such favored genes to occur in the same individual. A verbal discussion proceeds as follows. Suppose a favorable mutation A_1 arises at a locus A and begins to spread throughout a population. If a favorable mutation B_1 subsequently arises at a locus B , then without recombination A_1 and B_1 cannot both become simultaneously fixed unless the initial B_1 mutation happens to arise on an A_1 chromosome. This is unlikely to occur until the frequency of A_1 is substantial, and thus either the evolution at other loci is slowed down by the evolution of the A locus or the favorable mutation A_1 is lost through the increase in frequency of B_1 at the B locus, and hence of the linked allele A_2 . With recombination, both A_1 and B_1 genes can eventually arise on the same chromosome so that evolution, under this argument, proceeds more rapidly than with no recombination.

This argument clearly assumes that ultimately the advantage to the population with recombination will arise through intergroup competition. (Recall that Fisher's original argument for decreased recombination also makes this assumption.) It would be fitting, in line with our discussion of the evolution of modifier genes, to attempt to produce an argument that does not rely ultimately on such an assumption: We mention such arguments later. We emphasize again that our aim is to compare systems with no recombination ($R = 0$) to those with positive recombination ($R > 0$): This is a different question to comparing two populations with positive recombination rates R_1, R_2 respectively. It may well be that, since high recombination breaks up favorable gene complexes as well as creating them, the incorporation of favorable new mutants proceeds best, at least in some circumstances, in populations with low but positive recombination. To repeat, this is not the comparison that is being made.

The first attempt to quantify the Fisher-Muller theory was by Crow and Kimura (1965). Crow and Kimura assume a population in which favorable new mutations arise in a population of size $\frac{1}{2}N$ at total rate NU per generation. The new favorable mutations are all at different loci, and the mutation is nonrecurrent. We may suppose for convenience that each new mutant has selective advantages with no dominance. While (see Section 1.4) most favorable new mutations will be lost from the population, we may expect an equal fraction to be lost with and without recombination, so that this random loss may safely be ignored and all processes treated as deterministic. We assume finally finally that in a population without recombination, on average g generations pass between the occurrence of a favorable new mutation and the occurrence of a second favorable mutation in a descendant of the first. Then in such a population favorable new mutation are incorporated into the population at a rate of one per g generations.

In a population with recombination, all favorable new mutations during g generations can be incorporated, and hence since NU favorable mutations arise per generation, NUg arise during g generations. As far as the rate of incorporation of favorable new mutations is concerned, then, the advantage of recombination is $NUg : 1$, and in order to discuss this ratio more usefully, it is necessary to find a formula for g in terms of N, U and s .

We have assumed no dominance and a selective advantage s to single mutants. The frequency x of individuals carrying a favorable new mutant is then given by (1.27) if we put $h = \frac{1}{2}$ and replace s by $2s$. Under the initial condition $x = N^{-1}$, the solution of (1.27) is clearly

$$x = (1 + (N - 1)e^{-st})^{-1}. \quad (6.99)$$

Thus, in the first i generations after the occurrence of this mutation the total number of its descendants is

$$N \int_0^i x dt = N s^{-1} \log\{(N - 1 + e^{si})/N\}. \quad (6.100)$$

The total number of favorable new mutations in these descendants is found by multiplying this quantity by U , and thus g is found as the solution of the equation

$$1 = N U s^{-1} \log\{(N - 1 + e^{sg})/N\}.$$

This gives immediately

$$g = s^{-1} \log\{N(e^{s/NU} - 1) + 1\},$$

and thus the rate of incorporation of advantageous mutations in populations with recombination to those without becomes, under this analysis,

$$N U s^{-1} \log\{N(e^{s/NU} - 1) + 1\} : 1. \quad (6.101)$$

Several limiting cases of this formula are of interest. Suppose U is extremely small, so that favorable new mutations arise very rarely. We may then expect that each favorable mutation is incorporated in the population before the next arises, and in this case there is no advantage to recombination. This argument is confirmed by noting that the ratio (6.101) approaches unity as $U \rightarrow 0$. Similarly for very large s the incorporation of each new favorable mutant should be very rapid, and we again confirm that the ratio (6.101) approaches unity as $s \rightarrow \infty$. Clearly the situation when recombination is most favored is when U/s is large and N is large. Crow and Kimura (1965) give a table of values of (6.101) for various combinations of N , U and s values which document this.

This conclusion was challenged by Maynard Smith (1968), who produced a "counter-example" in which the existence of recombination made no difference to the rate at which two favorable alleles increased in frequency. Maynard Smith considered unfavorable alleles at two loci which are maintained at low frequency in a population through recurrent mutation and showed that the gamete frequencies would then be in linkage equilibrium. Suppose now that the environment alters and that both rare alleles are favored, with a multiplicative fitness scheme of the form (6.28), and steadily increase in frequency. We have stated earlier that with a multiplicative fitness scheme a population having zero linkage disequilibrium initially will persist in a state of zero linkage disequilibrium. In this case, the value of the recombination fraction R is irrelevant to the rate of increase of frequency of the two alleles if there is no linkage disequilibrium, since R appears only as a multiplier of the coefficient of linkage disequilibrium. Hence the rate of incorporation of the rare alleles is unaffected by the existence of recombination.

As pointed out by Maynard Smith (1968) and Crow and Kimura (1969), there is an essential difference between the assumptions made in the two analyses. Maynard Smith's analysis assumes that gametes carrying both initially unfavored alleles exist at positive frequency. Crow and Kimura's analysis assumes zero initial frequency for such gametes and, as is clear from (2.94) with $R = 0$ or by general reasoning, such gametes can never arise in a population without recombination if their initial frequency is zero. More generally, Maynard Smith's claim is that the essential difference between the two analyses arises because in his analysis for favorable new alleles arise in a recurrent process, whereas in that of Crow and Kimura they arise uniquely. Clearly if favorable new alleles do arise recurrently at a sufficiently high frequency, then even without recombination a favored new mutation at the B locus can arise on a chromosome carrying a favored mutation at the A locus in the course of fixation.

Crow and Kimura's (1969) claim is that their argument does not assume unique favorable mutants, but rather that these occur sufficiently rarely so that double mutants arise very seldom (or, more generally, if many loci are substituting simultaneously, that n -tuple mutants arise with completely negligible frequency). Thus the real essence of their argument remains unchanged. Maynard Smith (1971) carried out an analysis dropping the assumption of zero initial linkage disequilibrium but incorporating a recurrent mutation rate of favorable mutations, and concluded that for moderate populations ($N \approx 10^6$) there is little advantage to recombination, while for large populations ($N \approx 10^{10}$) populations with recombination can incorporate favorable new mutations about four or five times faster than populations without recombination.

It is clear that the final answers to these questions rely on biological arguments concerning the most likely circumstances from which a micro-evolutionary process begins and, more generally, on the main nature of evolution. If evolution is mainly of the "shifting balance" type of Wright, outlined in Chapter 1, where gene frequencies are high throughout, the argument of Crow and Kimura argument does not apply. If, however, evolution depends more on the incorporation of rare favorable mutations, their argument is much more important. The extent to which this is so will depend on the rate of occurrence of favorable new mutations, the population size, and the selective advantage of the new mutant.

A number of interesting quantitative conclusions have been found when mutation to favorable new alleles is recurrent. Thus when the double mutant is more fit than multiplicative fitnesses would imply, Eshel and Feldman (1970) demonstrate that with no recombination, the frequency of the double mutant gamete is always larger than it is with positive recombination, provided that the initial disequilibrium $c_1c_4 - c_2c_3$ is non-negative, where c_1 is the initial frequency of the double mutant gamete. They further show that when single mutants are deleterious but the double mutant advantageous, for suitable fitness values and sufficiently low mutation rate

the two mutants will increase in frequency only if the linkage between the two loci is sufficiently tight. This conclusion was essentially given also by Crow and Kimura (1965). Karlin (1973) considered stochastic versions of the process, paying particular attention to the mean time until the first double mutant gamete is formed and the mean time until fixation of the double mutant gamete.

All the above arguments concern long-term optimization and ultimately rely on intergroup competition for the establishment of the population with the optimal recombination value. Short-term arguments have been offered by Williams (1966, 1975) and Williams and Mitton (1973). These center around the claim that under intense selection, populations having high recombination have an immediate advantage over populations with no recombination because they produce more high-fitness genotypes, which are assumed to be the only genotype to survive. This argument has been contested by Maynard Smith (1971). Felsenstein and Yokoyama (1976) introduce a locus which modifies recombination between primary loci and discuss verbally and by simulation the fate of the allele causing no recombination as favorable mutations arise and become fixed at the primary loci. This argument is in the spirit of Section 6.5 and avoids group-competition arguments. Unfortunately the complexities of the argument make a mathematical analysis well-nigh impossible. The argument relies on the existence of randomly generated linkage disequilibrium in finite populations, and thus the analysis of Section 6.6, which suggests that unless population sizes are small such linkage disequilibrium is rarely large, becomes relevant.

6.9 Summary

In the preceding sections we have outlined several conclusions concerning the joint evolution at two linked loci. A number of topics have not been discussed. These include the effect of population subdivision (Feldman and Christiansen (1975), Nei and Li (1973)), the effect of allowing several alleles at one locus (Feldman et al. (1975)) and the effect of different recombination fractions in the two sexes (Strobeck (1974)). We cannot hope to cover here all possible extensions and generalizations. Far and away the most important question concerns the degree of linkage disequilibrium in natural populations. We have observed, and will note again in the next chapter, that if linkage disequilibrium can normally be taken as negligibly small, a great simplification can be made to the theory. Loci can essentially be examined one by one, with interactive effects between loci being of minor importance. Ohta and Kimura (1975), for example, claimed that linkage disequilibrium in nature is comparatively rare and that such simplifying assumptions, which allow us to carry the theory a considerable distance, can reasonably be made. Thus they assert that “for large and stable pop-

ulations the concept of quasi-linkage equilibrium together with the single locus theory is sufficient to treat most problems realistically". Lewontin (1974), on the other hand, emphasized the role of linkage and linkage disequilibrium in evolution, and Wright, as we have noted, also emphasized interactive effects of loci. Under the latter view evolution is far more complex than would appear under the "single-locus reductionist" approach, and its quantitative assessment is, as a result, extremely difficult. It cannot be claimed, even now, after thirty years, that a weight of evidence has yet accumulated on either side. In the next chapter we carry the theory to many loci and note the additional complications, compared to those in a two-locus analysis, that then arise.

7

Many Loci

7.1 Introduction

Our aim in this chapter is to outline certain properties of populations when it is assumed that the various characteristics, and in particular the fitness of any individual, depend on his genetic constitution at all loci in the genome. Although we often assume random mating and/or particular forms for various parameter values, since the analysis can be carried further when these assumptions are made, we also consider cases where no such assumptions are made. In particular we shall prove the fully general version of the Fundamental Theorem of Natural Selection, where no assumption is made about the mating scheme, random or otherwise, about the fitness values, the number of loci that fitness depends on or the number of possible alleles at each locus. Thus we consider here both general and specific cases, and base our overall conclusions on the results deriving from both.

This chapter has two intertwining themes. The first concerns the relationship between properties of the entire multilocus system and those of the various subsystems, in particular single-locus subsystems, that the entire system defines. The second concerns linkage disequilibrium and its effect on static and dynamic properties of the population. The former theme is of interest because, while many properties of individuals depend on genes at a large number of loci, experiments often involve one or a small number of loci, and it is clearly important to assess the extent to which valid inferences can be made from the loci investigated to the entire system involved. We shall find that to a great extent, the validity of these inferences

depends on the amount of linkage disequilibrium. For highly structured systems, that is those possessing high linkage disequilibrium values, these inferences can be of dubious value. For unstructured systems, with little or no disequilibrium, the inferences are more likely to be valid.

A large literature on multilocus theory is now available. We do not aim to cover this here, focussing on those aspects of the theory relating to the themes discussed above.

7.2 Notation

Since the notation for multilocus systems can become remarkably confusing, we collect here most of the notation that will be used in this chapter: This notation does not necessarily conform to that used in other chapters.

We suppose the entire genetic system, or genome, to consist of K loci, the generic symbol used for a locus being k (and on occasions where two loci are considered simultaneously, k and ℓ). Thus k takes the values $1, 2, \dots, K$: As far as possible we use upper-case symbols for fixed quantities such as the number of loci in a system, and the corresponding lower-case symbol for generic or typical values. We suppose there are I_k alleles possible at locus k , so that there will exist $I = \prod I_k$ different K -locus gametes, which are assumed labeled in some agreed fashion $1, 2, \dots, i, \dots, I$.

We are also interested in one-, two-, and in general H -locus subsystems of the entire K -locus system. The alleles at locus k are labeled $A_{k1}, A_{k2}, \dots, A_{kI_k}$. Two-locus gametes are described by the alleles at the two loci at which they are defined, for example $(A_{ku}, A_{\ell v})$. For general H -locus systems there will exist $Q = \prod I_k$ H -locus gametes, where the product is taken over all H loci in the subsystem. These are also assumed to be labeled in some agreed fashion $1, 2, \dots, q, \dots, Q$.

We now turn to the frequencies of the various alleles and gametes, and will consistently use the notation x for gene frequencies, y for frequencies of two-locus gametes, z for frequencies of H -locus gametes and c for frequencies of K -locus gametes. Finally we consider all G genotypes in the entire K -locus system, and assume that these are listed in some agreed order. The frequency of genotype s in this listing is denoted g_s . More explicitly we have

$$\begin{aligned}
 x_{ku} &= \text{frequency of the allele } A_{ku}, \\
 y_{ku,\ell v} &= \text{frequency of the two-locus gamete } (A_{ku}, A_{\ell v}), \\
 z_q &= \text{frequency of the } q\text{th } H\text{-locus gamete,} \\
 c_i &= \text{frequency of the } i\text{th } K\text{-locus gamete,} \\
 g_s &= \text{frequency of the } s\text{th genotype in the genome.} \quad (7.1)
 \end{aligned}$$

When referring to two or more H -locus gametes we use suffixes p, q and r ; when referring to two or more K -locus gametes we use suffixes h, i and

j . Clearly the frequencies x_{ku} , $y_{ku,\ell v}$ and z_q can be found by appropriate summation. Thus, for example,

$$z_q = \sum_{i \in S_q} c_i, \quad (7.2)$$

where S_q is the set of all K -locus gametes containing the same alleles at the loci of the H -locus system as the q th H -locus gamete.

7.3 The Random Mating Case

7.3.1 Linkage Disequilibrium, Means and Variances

Our initial analysis assumes random mating. This implies a focus on gametes and their frequencies, as well as on measures of linkage disequilibrium.

The concept of linkage disequilibrium was introduced in Chapter 2 for two-locus, two-allele systems. We use the symbol D for two-locus disequilibria: Thus

$$D_{ku,\ell v} = y_{ku,\ell v} - x_{ku}x_{\ell v}. \quad (7.3)$$

Higher-order linkage disequilibria have been defined by Geiringer (1944), Bennett (1954) and Slatkin (1972). Although linkage disequilibrium is a major concern of this chapter, we shall not introduce these measures here. We note, however, that if the frequency of every K -locus gamete is the product of the frequencies of its constituent alleles, all these measures are zero, and that large linkage disequilibrium values imply that gametic frequencies cannot be found, even approximately, by forming the products of the corresponding allele frequencies.

For the random-mating case it is convenient to think of the genotype as being made up of two gametes, one derived maternally and one paternally. Each pair of K -locus gametes then defines some genotype: For this random-mating analysis we denote the genotype defined by gametes i and j as genotype (i, j) . The value of some measured characteristic of an (i, j) individual is written m_{ij} : It is assumed that there is no environmentally caused variation in this measurement. A particular case is the fitness of an (i, j) individual, given the special notation w_{ij} . The marginal value m_i of gamete i for the character in question is defined by

$$m_i = \sum_j c_j m_{ij}, \quad (7.4)$$

and the mean value \bar{m} for the entire population is given by

$$\bar{m} = \sum c_i m_i = \sum \sum c_i c_j m_{ij}. \quad (7.5)$$

The total genetic variance σ_T^2 for the character is

$$\sigma_T^2 = \sum \sum c_i c_j (m_{ij} - \bar{m})^2, \quad (7.6)$$

and the additive component of this, defined in Section 7.3.3, is denoted σ_A^2 .

Similarly, each pair of H -locus gametes defines an H -locus genotype, the (marginal) value \bar{m}_{pq} for an individual (p, q) being defined by averaging over K -locus genotypes. Explicitly,

$$\bar{m}_{pq} = \frac{\sum_{j \in S_p} \sum_{j \in S_q} c_i c_j m_{ij}}{z_p z_q}. \quad (7.7)$$

The marginal value for gamete p is

$$\bar{m}_p = \sum_q z_q \bar{m}_{pq} = \frac{\sum_{i \in S_p} \sum_j c_i c_j m_{ij}}{z_p}, \quad (7.8)$$

and it follows that

$$\sum_p z_p \bar{m}_p = \sum_i \sum_j c_i c_j m_{ij} = \bar{m}. \quad (7.9)$$

The total variance in the character for the H -locus subsystem is

$$\sum \sum z_p z_q (\bar{m}_{pq} - \bar{m})^2, \quad (7.10)$$

and this can also be divided into additive and non-additive components. This is true in particular when $H = 1$, for which the z_p are the gene frequencies x_{k1}, x_{k2}, \dots at a single locus. Since the single-locus case is particularly important, we now exhibit the variance partition for it. The marginal value for the genotype $A_{ku}A_{kv}$ is defined (as in (7.7)) by

$$\bar{m}_{ku,kv} = \frac{\sum_{i \in S_{ku}} \sum_{j \in S_{kv}} c_i c_j m_{ij}}{x_{ku} x_{kv}}. \quad (7.11)$$

From this we may compute the marginal value of A_{ku} as

$$\bar{m}_{ku} = \sum \bar{m}_{ku,kv} x_{kv} \quad (7.12)$$

and the marginal average excess of A_{ku} by

$$a_{ku} = \bar{m}_{ku} - \bar{m}. \quad (7.13)$$

Then the locus k marginal variance for the character, namely

$$\sum_u \sum_v x_{ku} x_{kv} (\bar{m}_{ku,kv} - \bar{m})^2, \quad (7.14)$$

may be partitioned into an additive component $\sigma_A^2(k)$ and a dominance component $\sigma_D^2(k)$ defined respectively by

$$\sigma_A^2(k) = 2 \sum_u x_{ku} a_{ku}^2, \tag{7.15}$$

$$\sigma_D^2(k) = \sum_u \sum_v x_{ku} x_{kv} d_{k,uv}^2 \tag{7.16}$$

where

$$d_{k,uv} = \bar{m}_{ku,kv} - \bar{m}_{ku} - \bar{m}_{kv} + \bar{m}. \tag{7.17}$$

For $H > 1$, additive \times additive and other higher-order components of variance arise: These are of interest only in Section 7.5, and we defer further discussion of them until that section.

7.3.2 Recurrence Relations for Gametic Frequencies

The genetic evolution of a random-mating population is described most efficiently by recurrence relations for the various K -locus gamete frequencies. These relations generalize those in (2.94), and depend on genotypic fitnesses and the recombination pattern between loci. We define the latter through the function $f(i, j \rightarrow h)$, defined as the probability that a randomly chosen one of the two gametes formed by recombination in gametes i and j is gamete h . Then if c_i, c'_i are the frequencies of gamete i in consecutive generations,

$$\bar{w}c'_i = w_i c_i - \sum^{(1)} w_{ij} c_i c_j f(i, j \rightarrow h) + \sum^{(2)} w_{hj} c_h c_j f(h, j \rightarrow i). \tag{7.18}$$

Here $\sum^{(1)}$ denotes a summation over all gametes h and j with $i, j \neq h$ and $\sum^{(2)}$ denotes summation over all gametes h and j with $h, j \neq i$. Summation of (7.18) over all i in S_p yields

$$\bar{w}z'_p = \bar{w}_p z_p - \sum^{(1)} \bar{w}_{pq} z_p z_q f(p, q \rightarrow r) + \sum^{(2)} \bar{w}_{rq} z_r z_q f(r, q \rightarrow p) \tag{7.19}$$

where $\sum^{(1)}$ and $\sum^{(2)}$ have meanings parallel to those just given. The similarity in form between (7.18) and (7.19) shows that the recurrence relations for H -locus gametes written down by formal analogy with (7.18) do indeed provide the correct recurrence values, but with one restriction: The fitnesses \bar{w}_{pq} (unlike the full multilocus values w_{ij}) are not fixed but normally change from generation to generation. It follows that the recurrence relation (7.19) can be used to predict H -locus gametic frequencies only one generation in advance. For long-term predictions the full system (7.18) must be used. We have noted in Chapter 2 that for systems where fitness depends on the alleles at two loci, decreases in mean fitness can occur and the mean fitness increase theorem is not valid. If, however, the loci involved are in linkage equilibrium, a decrease in mean fitness cannot occur. This is shown

by setting $c_1c_4 - c_2c_3 = 0$ in (2.94). Doing this yields the “single-locus type” (2.7), and this implies non-decrease in mean fitness, at least for one generation. The same holds for many loci: If all linkage disequilibria of all orders are zero, (7.18) will also reduce to equations in single locus form, and the same conclusion holds. We emphasize that this is a condition on gamete frequencies, not on the fitnesses w_{ij} .

We turn now to equilibrium behavior. The equation $c'_i = c_i$ implies that $z'_q = z_q$, a conclusion that is put more usefully in contrapositive form: If $z'_q \neq z_q$, the full K -locus system cannot be in equilibrium. This is a conclusion concerning the full system derived from a subset of loci. In particular if $H = 1$ and only two alleles A_{k1} and A_{k2} can occur at locus k , the full system cannot be in equilibrium unless the marginal fitness of the heterozygote $A_{k1}A_{k2}$ is outside the range of those of the homozygotes $A_{k1}A_{k1}$ and $A_{k2}A_{k2}$ and unless further the frequency of A_{k1} is at the value predicted by an equation of the form of (1.31).

Despite the above results, the equations $z'_q = z_q$ ($q = 1, 2, \dots, Q$) do not necessarily imply that the full K -locus system is in equilibrium. Further, if indeed $c'_i = c_i$, so that the full K -locus system is at an equilibrium point, the stability of this equilibrium cannot necessarily be gauged by the values of \bar{w}_{pq} . It is possible for the values of the \bar{w}_{pq} to suggest stability of the K -locus equilibrium and yet for that equilibrium to be unstable. For examples see Ewens and Thomson (1977): Clearly subsystem behavior does not necessarily give correct information about the full system.

7.3.3 Components of Variance

In this section we compute the full K -locus “additive genetic” and the “additive gametic” components of the variance (7.6), continuing to assume random mating. We show that the two are identical and then consider their relations to the sum of the single locus additive genetic variances. The additive genetic variance was defined for one locus (and two alleles) in (1.9) and for two loci in (2.102), while the additive gametic variance was defined after (2.106). We consider first the additive gametic variance. For an arbitrary number of loci, the total gametic variance for any character is defined as

$$2 \sum c_i (m_i - \bar{m})^2, \quad (7.20)$$

and is a measure of the differences in marginal gametic values for this measurement. The additive gametic variance measures the extent to which this variance can be accounted for by additive effects of genes. We attach an additive parameter α_{ku} to the allele A_{ku} , where since many loci are now involved the α_{ku} are subject to the constraints (generalizing those in

(2.101))

$$\sum_u \alpha_{ku} x_{ku} = 0 \quad (7.21)$$

for each k . Subject to the constraints in (7.21) we now minimize

$$2 \sum_i c_i (m_i - \bar{m} - \sum \alpha_{ku})^2, \quad (7.22)$$

with respect to the α_{ku} , the inner summation being over all alleles in gamete i . The minimizing value $\hat{\alpha}_{ku}$ of the α_{ku} satisfy the equation

$$x_{ku} \hat{\alpha}_{ku} + \sum_{\substack{\ell, v \\ \ell \neq k}} y_{ku, \ell v} \hat{\alpha}_{\ell v} = \sum^{(ku)} c_i (m_i - \bar{m}), \quad (7.23)$$

where $\sum^{(ku)}$ implies summation over all gametes containing the allele A_{ku} . Except in degenerate cases, (7.23) is a system of linear equations for the $\hat{\alpha}_{ku}$, and thus has a unique solution. Standard least-squares theory now shows that the additive gametic variance is

$$2 \sum_{k,u} \left(\sum^{(ku)} c_i (m_i - \bar{m}_i) \hat{\alpha}_{ku} \right). \quad (7.24)$$

Similarly the additive genetic variance is found by attempting to account for genotypic values as far as possible by the additive effects of alleles. Subject to the constraints in (7.21), this is done by minimizing the expression

$$2 \sum_i \sum_j c_i c_j (m_{ij} - \bar{m} - \sum \alpha_{ku})^2, \quad (7.25)$$

with respect to the α_{ku} , the inner summation now being over all alleles in the genotype (i, j) . If $A_{\ell v}$ occurs twice in any genotype, the contribution $\alpha_{\ell v}$ is counted twice in the summation. The minimizing values $\hat{\alpha}_{ku}$ can again be computed, and it is found (Ewens and Thomson (1977)) that these also satisfy (7.23) and that the sum of squares removed reduces to the expression in (7.24). Thus the additive genetic and additive gametic variances are equal. To find either we thus compute whichever is easier in any given circumstance and note that the identity of the two reinforces the view, given by Crow and Kimura (1970), that the expression "genetic variance" should be used for both.

It is not, in general, easy to determine the difference between the true additive genetic variance, given by (7.24), and the sum $\sum_k \sigma_A^2(k)$ of the single-locus values, where $\sigma_A^2(k)$ is defined in (7.15). In general the two values are not equal. If, however, all possible two locus coefficients of linkage disequilibrium are zero, so that

$$y_{ku, \ell v} = x_{ku} x_{\ell v} \quad (7.26)$$

for all k, ℓ, u, v , then equality does hold. This can be seen immediately from (7.23). When (7.26) holds, the second term on the left-hand side of (7.23) is zero, because of the constraints in (7.21), so that $\hat{a}_{ku} = a_{ku}$ and (7.24) reduces immediately to $\sum \sigma_A^2(k)$. More generally, Avery and Hill (1978) show that whether (7.26) holds or not, if only two alleles are possible at each locus and the measurements m_{ij} can be expressed as the sum of single-locus genotypic contributions, the true additive genetic variance is

$$\sigma_A^2 = \sum_k V_A(k) + 4 \sum_{k < \ell} \sum (a_{k1} - a_{k2})(a_{\ell 1} - a_{\ell 2}) D_{k,\ell}, \quad (7.27)$$

where $V_A(k) = \sum x_{ku} \alpha_{ku}^2$, with α_{ku} being the true multilocus average effect of A_{ku} and $D_{k,\ell}$ is the coefficient of linkage disequilibrium between loci k and ℓ . We check that if $D_{k,\ell} = 0$ for all k and ℓ , the true additive genetic variance and the sum of the single-locus values are identical.

The terms in the second sum in (7.27) can be both positive and negative, and thus considerable cancellation is possible in the summation. However, this might not occur in highly structured genetic situations, when the second term can dominate the first term. It is therefore of some interest to assess the circumstances under which each of these cases occurs. Important progress was made on this point by Bulmer (1976). Bulmer simulated a genetic system in which the value of a given character is determined by the alleles at twelve loci as well as an independent environmental component, the contributions being additive across loci with $A_{k1}A_{k1}$, $A_{k1}A_{k2}$ and $A_{k2}A_{k2}$ contributing 0, 1 and 2 respectively to the measurement. The genetic contribution to the character thus ranges from 0 to 24, and there are no dominance contributions to the total genetic variance.

We are interested here in the effects of stabilizing, disruptive and directional selection schemes on the additive variance σ_A^2 and its two components as given in (7.27). If x_{k1} and x_{k2} are both moderate, $\sigma_A^2(k)$ was found by Bulmer to be approximately 0.4 or 0.5, so that $\sum \sigma_A^2(k)$ is approximately 5 or 6. Under both stabilizing and symmetrical disruptive selection $\sum \sigma_A^2(k)$ takes values of this order of magnitude, and thus provides a useful standard against which the value of the second term on the right-hand side of (7.27) can be compared. Under stabilizing selection this term is approximately -1.5 or -2.0 . Thus linkage disequilibrium lowers the true additive genetic variance somewhat from the value calculated without disequilibrium. However, under disruptive selection this term is extremely large, being ten or eleven times larger than $\sum \sigma_A^2(k)$ itself. It is clear why this is so. Under strong disruptive selection, two gametes, one consisting entirely of A_{k1} genes, ($k = 1, 2, \dots, 12$) the other, of A_{k2} genes ($k = 1, 2, \dots, 12$), occur in high frequency, and the genetic system acts very much like a single-locus system with two alleles having values 0, 12 and 24 for the three genotypes. For such a system the additive genetic variance is 72 if the two alleles are of equal frequency, and the contribution to variance of the linkage disequi-

librium terms makes up most of the difference between this value and that given by $\sum \sigma_A^2(k)$.

The relation between σ_A^2 and $\sum \sigma_A^2(k)$ under directional selection is not so easily arrived at intuitively. In Bulmer's simulations $\sum \sigma_A^2(k)$ is approximately 25% less than σ_A^2 during the rather small number of generations before fixation of the favored allele at each locus.

7.3.4 Particular Models

It is clear from the above that the degree of linkage disequilibrium in a genetic system influences considerably the extent to which multilocus properties of the system can be determined from a consideration of single-locus properties. It is thus important to assess the extent to which linkage disequilibria will arise in natural populations, particular at equilibrium, and to make a partial assessment of this we now consider equilibrium properties for certain special fitness models. We continue to assume random mating and the only character we consider is fitness. Since the models considered often have special properties, for example of symmetry, some caution is necessary in drawing inferences from them: We gain some generality by considering four different models, noting in particular when the same inference is suggested by all four.

Three of these models require the definition of a "fitness contribution" from each locus given, for the genotype $A_{ku}A_{kv}$ at locus k , by

$$w_k(u, v). \quad (7.28)$$

We assume the $w_k(u, v)$ are such that for a single-locus system with fitness parameters (7.28) there exists a unique internal stable equilibrium point with the frequency of A_{ku} being \hat{x}_{ku} .

Assume now the $w_k(u, v)$ are used in some way to define fitnesses for the K -locus genotypes. Any equilibrium point of such a system for which the frequency of the gamete $(A_{1u1}, A_{2u2}, A_{3u3}, \dots)$ is $\hat{x}_{1u1}\hat{x}_{2u2}\hat{x}_{3u3}\dots$ is called a "product" equilibrium: We shall be particularly interested in stability conditions for equilibria of this type. All coefficients of linkage disequilibrium, of all orders, are zero at such an equilibrium. Roux (1974) and Karlin (1977a) called these "Hardy-Weinberg" equilibria, but we prefer here the term "product" to avoid confusion with the slightly different single-locus meaning of the term "Hardy-Weinberg frequencies".

Suppose first that in the K -locus system, the fitness of any individual is in the additive form

$$\sum_k w_k(u, v) \quad (7.29)$$

where the sum is taken over his genotypes at all loci. An example of such a scheme (with different notation) is given in (6.21): The notational

equivalence is

$$\begin{aligned}\alpha_1 &= w_1(1, 1), & \alpha_2 &= w_1(1, 2), & \alpha_3 &= w_1(2, 2), \\ \beta_1 &= w_2(1, 1), & \beta_2 &= w_2(1, 2), & \beta_3 &= w_2(2, 2).\end{aligned}\tag{7.30}$$

A generalization of the discussion following (6.21) shows that mean fitness depends on gene frequencies only and is thus nondecreasing from generation to generation: This, and the class of models discussed by Lyubich (1992), together form perhaps the only broad class of models of practical interest for which the mean fitness increase theorem holds for an arbitrary value of K . Further, Karlin and Liberman (1979) found that the product equilibrium is the only equilibrium of the K -locus system for nonzero recombination rates, and it is globally stable at least for the case $K = 2$. At this equilibrium the additive genetic variance in any character can be found as the sum of single-locus values, but even though fitnesses are additive over loci, this is not generally true for the character "fitness" for nonequilibrium values.

For the second model it is supposed that fitness is in the multiplicative form

$$\prod_k w_k(u, v).\tag{7.31}$$

This scheme generalizes (6.28), to which it reduces with the identifications (7.30). Considerable speculation on the behavior of real genetic systems has followed from investigation of this model, and we therefore consider its properties in some detail.

The product equilibrium exists for this model but is not a maximum point of mean fitness. Thus mean fitness can decrease in the neighborhood of the equilibrium, and the mean fitness increase theorem fails. We show later that the mean fitness at equilibria other than the product equilibrium is rather higher than at the product equilibrium point.

A general analysis is difficult and unrevealing, so we consider a simplified case. Suppose that the recombination fraction between adjoining loci is R , that there is no interference, and that for all k , u and v

$$w_k(u, u) = 1 - a, \quad w_k(u, v) = 1, \quad (u \neq v)\tag{7.32}$$

where $a > 0$. For $K = 2$ this is a particular case of (6.28), and then (6.33) shows that the product equilibrium is stable only if

$$R > a^2/4.\tag{7.33}$$

When (7.33) is violated there exist complementary pairs of stable equilibria, with

$$\begin{aligned}\text{freq}(A_{11}A_{21}) &= \text{freq}(A_{12}A_{22}) = \frac{1}{4} \pm \frac{1}{4}\{1 - 4R/a^2\}^{1/2}, \\ \text{freq}(A_{11}A_{22}) &= \text{freq}(A_{12}A_{21}) = \frac{1}{4} \mp \frac{1}{4}\{1 - 4R/a^2\}^{1/2},\end{aligned}\tag{7.34}$$

$$D = \frac{1}{4}\{1 - 4R/a^2\}^{1/2}.\tag{7.35}$$

For $K = 3$, Feldman et al. (1974) showed that the product equilibrium is stable whenever (7.33) holds; when (7.33) does not hold, there exist four stable equilibria analogous to (7.34). Curiously, for a small range of R values ($0.01 < R < 0.0104272$ when $a = 0.2$), in excess of the bound $a^2/4$, these equilibria continue to be stable, but for sufficiently large R ($R > 0.0104272$ when $a = 0.2$) the product equilibrium is the only stable equilibrium.

For $K = 5$, Lewontin (1964a,b) showed by simulation for the case $a = 0.5$ that whereas the product equilibrium is stable for $R > 0.0625$, the bound given by (7.33), there exist several stable equilibria exhibiting linkage disequilibrium when $0 < R < 0.065$. There will thus be a small interval of R values for which “ $D = 0$ ” and “ $D \neq 0$ ” stable equilibria exist. For $K = 36$, Franklin and Lewontin (1970) show by simulation that, when $a = 0.1$, a large number of stable equilibria exists when $0 < R < 0.0025$; note that 0.0025 is the bound given by (7.33). All of these are in linkage disequilibrium. For $0.0025 < R < 0.01$ approximately, these stable equilibria persist with a stable product equilibrium also, while for $R > 0.01$ approximately, only the product equilibrium is stable. These conclusions jointly suggest that the range of R values for which there exist linkage disequilibrium stable equilibria increases steadily over the bound $a^2/4$ as the number of loci in the system increases. However, a very powerful theorem of Roux (1974) shows that, for *any* multiplicative fitness scheme, the conditions on recombination fraction values that ensure stability of the pairwise product equilibrium for all adjacent loci is sufficient to ensure stability of the product equilibrium in the complete K -locus multiplicative system. In particular, in the present example, (7.33) is sufficient for the stability of the K -locus product equilibrium for all K . This is an important conclusion and, in conjunction with the simulation conclusions of Franklin and Lewontin, suggests that for very large K a wide range of R values will exist for which stable linkage equilibrium and linkage disequilibrium equilibria occur. It is important to note that Roux’s theorem requires the condition (7.33) and its generalizations to hold for *all* pairs of adjacent loci. This was confirmed by Feldman et al. (1974) for the case $K = 3$, $a = 0.2$. For this value of a the inequality (7.33) becomes $R > 0.01$, and the product equilibrium is not stable if the recombination fraction between loci 1 and loci 2 is 0.0099 and that between loci 2 and 3 is 0.0103.

Lewontin (1964a) and subsequently Franklin and Lewontin (1970) noticed two further important properties of stable points of 5- and 36-locus systems respectively. The first is that loci far apart on the chromosome can be held in linkage disequilibrium when the recombination fraction between them is considerably in excess of the limit set by (7.33). This occurs because of loci in the system segregating between these end loci. Thus, for example, when $K = 5$ the recombination fraction between loci 1 and 5 is $R = 0.065$ ($4R = 0.26$). Second, the value of D for adjacent interior loci is greater than the value predicted by (7.35). This effect is most marked for large values of R . Thus in the 5 locus case the equilibrium value of D

between loci 2 and 3 is about 1.06 times as large as the values predicted by (7.35) for $R = 0.01$ and about 2.91 times as large for $R = 0.002$ the value of D from (7.35) is 0.112 whereas Franklin and Lewontin found an average value of $|D|$ for adjacent loci of 0.22. Since necessarily $D < 0.25$, this is an extremely large value. For $R > 0.0025$, two-locus theory does not predict stable linkage disequilibrium equilibria, and yet for $R = 0.004$ the average value of $|D|$ for adjacent loci was found to be as high as 0.185. The average of D^2 for all pairs of loci correspondingly decreases from about 0.05 at $R = 0.0027$ to 0.025 at $R = 0.007$ and essentially zero for $R \geq 0.01$.

These latter observations (for $K = 36$) arise because, at the equilibria of the system investigated, the equilibrium gametic frequencies for small R arise in a highly structured form, with two gametes each having frequency of about 0.4 and all 10^9 remaining gametes having total frequency of about 0.2. The two high frequency gametes are complementary in that for the great majority of loci, they carry the alternative forms of the alleles at each locus. (If stochastic loss of alleles at some loci had not occurred, these gametes would be perfectly complementary.) This suggests complex and interesting behavior for equilibria of multiplicative systems for large K , an argument which is taken up in the final section of this chapter.

A further property of highly structured systems is that the mean fitness is considerably higher than at a product equilibrium. In the present model the mean fitness at a product equilibrium is $(1 - \frac{1}{2}a)^K$, which is 0.158 for $a = 0.1$, $K = 36$. For equilibria with two almost complementary gametes present in high frequencies, Franklin and Lewontin (1970) found mean fitnesses in excess of 0.4. If only two complementary gametes occur, each with frequency 0.5, each individual is equally likely to be a complete homozygote (fitness = $(0.9)^{36} = 0.0225$) or a complete heterozygote (fitness 1), leading to a mean fitness of 0.511. Of course with recombination between loci, this cannot an equilibrium value of mean fitness, but the equilibrium value will not be much less than 0.511, and this agrees with the observation of Franklin and Lewontin.

The third model we consider is the "generalized nonepistatic" model of Karlin and Liberman (1978). Here the fitness of any individual is a linear combination of various multiplicative, additive and neutral components, so that this model can be thought of as being intermediate between the two just considered. We consider in detail only the two-locus case in which the fitness of the genotype $A_{1u}A_{1v}A_{2s}A_{2t}$ is

$$b_1 w_1(u, v) w_2(s, t) + b_2 w_1(u, v) + b_3 w_1(s, t) + b_4. \quad (7.36)$$

The sufficient condition that the product equilibrium be stable is that the recombination fraction between the loci exceed

$$\max_{a,b} \frac{b_1 \bar{w}_A \bar{w}_B \lambda_a \mu_b}{b_1 \bar{w}_A \bar{w}_B (1 - \lambda_2)(1 - \mu_b) + b_2 \bar{w}_A (1 - \lambda_a) + b_3 \bar{w}_B (1 - \mu_b) + b_4}. \quad (7.37)$$

Here the notation is that of (6.36) and (6.37), with

$$a_{uv} = w_1(u, v), \quad b_{uv} = w_2(u, v), \quad A_u = A_{1u}, \quad B_s = A_{2s},$$

and with $\lambda_1, \lambda_2, \dots, \mu_1, \mu_2, \dots$ defined as the nonunit eigenvalues of the matrices $\{c_{uv}\}, \{d_{uv}\}$ respectively. When $b_2 = b_3 = b_4 = 0$ (so that the model is multiplicative) this requirement reduces to (6.38), while if $b_1 = 0$ (so that the model is additive) it reduces to the known requirement $R > 0$. The condition on the value of R for stability of the product equilibrium is clearly less stringent than the corresponding condition for the multiplicative case with $b_2 = b_3 = b_4 = 0$.

The requirement (7.37) can be extended to an arbitrary number of loci, and an explicit condition (Karlin and Liberman (1979)) can be found for stability of the product equilibrium for general nonepistatic schemes which generalizes the condition (20) of Roux (1974) for purely multiplicative schemes. Although these conditions are not simple, two important conclusions emerge. First, the higher the mix of additive components in the fitness, as compared to multiplicative components, the less restrictive are the requirements on recombination for stability of the product equilibrium. This generalizes the discussion below (7.37). Second, stability of the product equilibrium obtains if there is "enough" recombination, whatever the mix may be, and in particular if all loci are unlinked the product equilibrium is stable for any generalized nonepistatic scheme.

We turn finally to a fitness scheme not defined in terms of the $w_k(u, v)$. If, in the two-locus model (6.29), the parameters β and γ are equal, the fitness of any individual depends solely on the number of loci at which he is heterozygous. We consider now the generalization of such a scheme to an arbitrary number K of loci, assuming two alleles possible at each locus and that the fitness of an individual heterozygous at k loci is γ_k ($k = 0, 1, 2, \dots, K$). All results given below for this model were obtained by Karlin (1977a).

By symmetry, the frequency of each gamete at the product equilibrium is 2^{-K} . If this equilibrium is stable for zero recombination, it is the unique and globally stable equilibrium for all recombination patterns. The conditions for stability with no recombination is

$$\sum_{k=0}^K \gamma_k \sum_{m=0}^k (-1)^m \binom{n}{m} \binom{K-n}{k-m} < 0, \quad n = 1, 2, \dots, K, \quad (7.38)$$

and for free recombination between all loci is

$$\sum_{k=0}^K \gamma_k \left\{ 2 \binom{K-1}{k} - \binom{K}{k} \right\} < 0. \quad (7.39)$$

In line with the discussion above, if (7.38) hold then automatically (7.39) holds. By contrast, the requirement (7.39) can hold and the requirement (7.38) not hold, as the multiplicative fitness scheme shows.

Suppose next that $\gamma_0 < \gamma_1 < \dots < \gamma_k < \dots < \gamma_K$ so that fitness increases with increasing heterozygosity. Then (7.39) holds, so that the product equilibrium is stable for free recombination between loci and, more generally, usually for rather loosely linked loci. However, the condition (7.38) need not necessarily hold, as the multiplicative case again shows. But when the γ_k increase in a *concave* fashion, so that $\gamma_k > \frac{1}{2}(\gamma_{k-1} + \gamma_{k+1})$, the condition (7.38) will hold: This was also noted by Slatkin (1972). In this case the additional fitness component for each additional heterozygous locus decreases with the number of current heterozygous loci, and the product equilibrium is the unique and globally stable equilibrium for all recombination patterns. When the γ_i form a *convex* system, so that $\gamma_k < \frac{1}{2}(\gamma_{k-1} + \gamma_{k+1})$, the product equilibrium is stable only for sufficiently large recombination. Thus Lewontin (1964a) showed that if $\gamma_0 : \gamma_1 : \gamma_2 : \gamma_3 : \gamma_4 : \gamma_5$ are in the ratios 2 : 3 : 6 : 11 : 18 : 27 and the recombination fraction between adjacent loci is R , the product equilibrium is stable only if $R > 0.038$.

Suppose now that the product equilibrium is not stable, so that the γ_i do not form a concave sequence and R is sufficiently small. For $R = 0$ and multiplicative fitness there are many stable equilibria, each one consisting of a pair of complementary gametic types. For sub-multiplicative fitnesses (for example, $\gamma_k = (k+1)^2/(K+1)^a$, $1 < a < K-1$), there are stable equilibria with a number of gametic types present, each in moderate frequency. By continuity, for $R \approx 0$ the stable equilibria in the multiplicative model are such that two complementary gametes occur in high frequency, with all other gametes at extremely low frequency, as the simulations of Franklin and Lewontin, (1970) suggest. For sub-multiplicative fitnesses a number of moderate frequency gametes exist at stable equilibrium points.

7.4 Non-Random Mating

7.4.1 Introduction

The need to analyze genomic data, in particular that from the human population, leads to the need for theory which relates to the evolutionary properties of the entire genome in populations that do not mate at random. In this section aspects of the multilocus non-random-mating theory will be developed, building on the theory given earlier in this chapter for the random-mating case and also on the theory given in Section 2.8 for the one-locus non-random-mating case. As above, we consider some quantitative character that is entirely determined for any individual by the genetic make-up of that individual. It is convenient to carry out the discussion by assuming that the character in question is fitness. This will have the benefit of providing theory for evolutionary processes and for the full multilocus generalization of the Fundamental Theorem of Natural Selection, described

below in Section 7.4.5. Nevertheless, much of the discussion applies to any arbitrary character. It will be found that surprisingly many random-mating results continue to hold in the non-random-mating case, although several do not. It is also surprising that several one-locus results carry over almost immediately to the multilocus case, although again several do not.

7.4.2 Notation and Theory

We recall the notation of Section 7.2, in particular the notation g_s for the frequency of sth multilocus genotype, which we call genotype G_s . Since random mating is no longer assumed, we may not assume that g_s can be found from the frequencies of the the two gametes, one maternally derived and one paternally, constituting the genotype G_s .

As in the random-mating case, the multilocus genotype frequencies define, by summation, the allelic frequencies at any chosen locus. For example, suppose we single out locus k and some allele A_{ku} at this locus. Then the frequency x_{ku} of this allele can be found from the various genotype frequencies $\{g_s\}$ through the formula

$$2x_{ku} = \sum^s c(ku, s)g_s, \quad (7.40)$$

where $c(ku, s)$ is the number of times (0, 1 or 2) that the allele A_{ku} occurs in genotype G_s .

The fitness of the genotype G_s is denoted w_s , so that the population mean fitness \bar{w} is $\sum g_s w_s$. The frequency of the genotype G_s will, in general, change between the time of conception and the time of reproduction of the parental generation, because of selective differentials: The intra-generational change in this frequency is $\Delta g_s = g_s w_s / \bar{w} - g_s$. The intra-generational change Δx_{ku} in the the frequency of A_{ku} is then found, from (7.40), to be given by

$$2\Delta x_{ku} = \sum^s c(ku, s)\Delta g_s. \quad (7.41)$$

This change Δx_{ku} is also the inter-generational change in the frequency of A_{ku} , where frequencies in both generations are taken at the time of conception.

The multilocus genotype fitnesses and frequencies also define "entire genome" average effects $\{\alpha_{ku}\}$ of the various alleles at the various loci. These are found by minimizing an expression generalizing that in (2.57), namely

$$\sum g_s \left(w_s - \bar{w} - \sum^{\{s\}} \alpha_{ku} \right)^2. \quad (7.42)$$

The inner sum in this expression is the sum of all the average effects of the various alleles, at all loci, in the genotype G_s , with the average effect of

any allele counted in twice if this allele occurs twice in this genotype. In this minimization procedure the constraints in (7.21) must be applied.

This minimization leads to a set of simultaneous equations defining the $\{\alpha_{ku}\}$ values, typified by

$$x_{ku}\alpha_{ku} + \sum_v X_{ku,kv}\alpha_{kv} + \sum_{r \neq k} \sum_t Q_{ku,rt}\alpha_{rt} = \bar{w}(\Delta x_{ku}). \quad (7.43)$$

In these equations, $X_{ku,kv}$ is the frequency of the ordered k -locus genotype containing alleles A_{ku} and A_{kv} . There are various (equivalent) definitions of $Q_{ku,rt}$. Perhaps the simplest (Lessard, (1997)) is that $Q_{ku,rt}$ is twice the probability that, in an individual chosen at random, a gene chosen at random from locus k is A_{ku} and a gene chosen independently at locus r is A_{rt} . This collection of equations generalizes those given in (2.65), and can also be written in matrix and vector form as

$$(D + P + Q)\alpha = \bar{w}\Delta. \quad (7.44)$$

Clearly this equation system provides a natural generalization of the equation system (2.65), but it should be noted that the meanings of the symbols in the two equations differ. In (7.44), for example, the diagonal matrix D has all frequencies of all alleles at all loci in the genome, whereas in (2.65) the diagonal matrix D contains only the frequencies of the alleles at the locus being considered. Similar comments apply to P , α and Δ .

It is important to note that the equation system (7.44) has a unique solution (Lessard, (1997)), as do equations (2.65).

The multilocus additive genetic variance, namely the sum of squares removed by fitting the α_{ku} values in (7.42), can be written in terms of the average effects in the form

$$\sigma_A^2 = 2\bar{w} \sum_k \sum_u \alpha_{ku}(\Delta x_{ku}), \quad (7.45)$$

the summation on the right-hand side being over all alleles at all loci in the genome. This is the natural multilocus generalization of the one-locus formula (2.63).

7.4.3 Marginal Fitnesses and Average Effects

The various multilocus genotype frequencies and fitnesses can be used to find single locus marginal fitnesses, single locus average effects, and single locus additive genetic variances, all of which are defined below. Having found these, we shall compare them to the corresponding true multilocus values. The main reason for doing this is that some theoretical calculation, or the data from some experiment might, focus on a particular locus, and it is necessary to know what the relation is between one-locus marginal values for some quantity estimated from this experiment and the true multilocus values.

The frequency of the single-locus genotype $A_{ku,kv}$ at locus k , namely $X_{ku,kv}$, is found by summing the frequencies of all multilocus genotypes containing $A_{ku,kv}$. The marginal fitness of the genotype $A_{ku}A_{kv}$, denoted here $\hat{w}_{\{ku,kv\}}$, is defined as a weighted average by

$$\hat{w}_{\{ku,kv\}} = \frac{\sum^{\{ku,kv\}} g_s w_s}{X_{ku,kv}}, \quad (7.46)$$

the sum again being taken over all multilocus genotype containing the genotype $A_{ku}A_{kv}$.

The mean fitness $\hat{w}(k)$ as calculated from the marginal fitnesses at locus k is found by replacing w_{ij} by $\hat{w}_{ku,kv}$ in equation (2.56). This leads to the value $\hat{w}(k) = \sum_u \sum_v X_{ku,kv} \hat{w}_{ku,kv}$. Equation (7.46) shows that this marginal value is identical to the true multilocus mean fitness $\bar{w} = \sum \sum g_s w_s$. This conclusion allows us to use \bar{w} rather than $\hat{w}(k)$ in the calculations below. The change $\hat{\Delta}x_{ku}$ in the frequency of the allele A_{ku} as calculated from the marginal fitnesses at locus k is found from (2.60) and (2.62) as

$$\hat{\Delta}x_{ku} = \sum_{kv} X_{ku,kv} \frac{\hat{w}_{ku,kv}}{\bar{w}} - x_{ku}. \quad (7.47)$$

Equations (7.46) and (7.47) jointly show that this is identical to the true change in the frequency of this allele, as given in (7.41). This allows us to use the true change Δx_{ku} rather than the value $\hat{\Delta}x_{ku}$ in the calculations below. It also implies that the average excess of any allele at any locus can also be calculated correctly from single-locus marginal fitnesses.

The marginal k -locus additive effects estimates $\{\hat{\alpha}_{ku}\}$ are defined by minimization of the expression (2.57), with w_{ij} replaced by $\hat{w}_{ku,kv}$. These are thus found as the solutions of the simultaneous equations

$$x_{ku} \hat{\alpha}_{ku} + \sum_{kv} X_{ku,kv} \hat{\alpha}_{kv} = \bar{w} \Delta p_{ku}. \quad (7.48)$$

In these equations we have used, on the right-hand side, the fact that the k -locus marginal mean fitness, and each change in gene frequencies as calculated from k -locus marginal fitnesses, are respectively equal to the true multilocus values.

Equations of this form may then be written down for all loci and the resulting equations formed into one large system of simultaneous equations. With an appropriate ordering of loci and alleles within loci, this system of equations may be written as

$$(D + P)\hat{\alpha} = \bar{w}\Delta, \quad (7.49)$$

where D and P now have the same interpretation as in equation (7.44). This system of equations differs from the true multilocus system of equations (7.44), so we conclude that one-locus marginal frequencies and fitnesses do *not* lead to correct calculations of the average effects of the various alleles at

the various loci in the entire genomic system. The marginal k -locus additive genetic variance estimate is defined by (2.63), with α_i replaced by $\hat{\alpha}_{ku}$. We denote this quantity by $\sigma_A^2(k)$. The sum of these marginal estimates, taken over all loci in the genome, is thus

$$\sum_k \sigma_A^2(k) = 2\bar{w} \sum_k \sum_u \hat{\alpha}_{ku}(\Delta x_{ku}). \quad (7.50)$$

7.4.4 Implications

In this section we consider some implications of the above results.

First, suppose that the true multilocus additive genetic variance σ_A^2 , defined in equation (7.45), is zero. This implies that the sum of squares removed by fitting the α_{ku} values in (7.42) is zero, which implies that the α_{ku} values themselves are zero. This in turn implies, from equations (7.43), that each Δx_{ku} is zero, so that for one generation at least, gene frequencies do not change. However this does *not* necessarily imply that the full multilocus system is in equilibrium: Genotype frequencies can change without any resulting gene frequency change.

The converse is also true. If each Δx_{ku} is zero, then the uniqueness of the solution of equations (7.43) implies that each α_{ku} is zero, and this in turn implies that the true multilocus additive genetic variance is zero.

Next, we have already observed that the true multilocus average effects are not in general correctly calculated from marginal k -locus fitness values. Equality between the true and the values calculated from marginal fitnesses will arise if

$$Q_{ku,rt} = 2x_{ku} x_{rt} \quad (7.51)$$

for all pairs (ku, rt) . In practice, this can be taken as a necessary condition also, and is a condition for total linkage equilibrium of all alleles at all pairs of loci. This condition is unlikely to hold even in the random-mating case, and is even less likely under non-random mating. Thus, in practice, multilocus average effects will not be estimated correctly from one-locus marginal fitness values.

This conclusion, together with the close connection between additive genetic variances and average effects, as shown for example in equation (7.45), implies that in general, the true total additive genetic variance is not correctly found by summing one-locus marginal estimating values. This is despite the fact that the contrary assumption is frequently made in the classical literature.

Despite this conclusion, some multilocus conclusions concerning additive genetic variances can be found from single-locus marginal results. If the k locus marginal additive genetic variance is zero, then each $\hat{\alpha}_{ku}$ is zero. The converse is also true. Further, if each $\hat{\alpha}_{ku}$ is zero, then each Δp_{ku} is zero.

The converse of this statement is also true. Thus each of these three results implies the other two.

If every single-locus marginal additive genetic variance is zero then every Δx_{ku} is zero, for all k and u . This implies that every α_{ku} is zero, by uniqueness, and thus that the true total additive genetic variance is zero. This sequence of implications also works in reverse order.

No mention has been made in this section of the additive gametic variance. This is because, when mating is non-random, gametic frequencies are not of value in determining genotype frequencies, so that the additive gametic variance, while it can be defined, is not useful.

7.4.5 *The Fundamental Theorem of Natural Selection*

Both the Price (1972) and the Lessard (1997) one-locus interpretations of the FTNS given in Section 2.9 can be generalized to the multilocus case. In the notation of Sections 7.2 and 7.4.2, the generalization of the Price one-locus interpretation (2.72) of the FTNS is that

$$\Delta_p(\bar{w}) = \sum^s \Delta g_s (w_s)_\alpha = \sigma_A^2 / \bar{w}, \quad (7.52)$$

where

$$(w_s)_\alpha = \bar{w} + \sum_k \sum_u c(ku, s) \alpha_{ku}, \quad (7.53)$$

the sum being taken over all alleles in the genome, and with $c(ku, s)$ being as defined immediately below (7.40).

The Lessard (1997) multilocus generalization of the Lessard single-locus statement (2.74) of the FTNS is that

$$\Delta_p(\bar{w}) = \sum (\Delta g_s)_\alpha w_s = \sigma_A^2 / \bar{w}. \quad (7.54)$$

In this expression the partial change $(\Delta g_s)_\alpha$ in the frequency of the genotype G_s is defined by

$$(\Delta g_s)_\alpha = g_s \frac{\sum_k \sum_u c(ku, s) \alpha_{ku}}{\bar{w}}. \quad (7.55)$$

The interpretations of $(w_s)_\alpha$ and $(\Delta g_s)_\alpha$ in the Price and the Lessard interpretations parallel the one-locus interpretations given in Section 2.9.

The proofs of these two claims are almost immediate. The Price concept of the partial change in mean fitness is the middle term in (7.52), and since $\sum \Delta g_s \bar{w} = 0$, (7.52) and (7.53) jointly show that this is

$$\sum_k \sum_u \left(\sum^s \Delta g_s c(ku, s) \right) \alpha_{ku}. \quad (7.56)$$

Equation (7.41) shows that this expression reduces to

$$2 \sum_k \sum_u (\Delta x_{ku}) \alpha_{ku},$$

and then (7.45) shows that this is equal to σ_A^2/\bar{w} . These simple steps complete the proof of the multilocus version of the FTNS in the Price interpretation. The proof of the Lessard version of the theorem is similar.

There are several remarks to make about the full multilocus FTNS. Perhaps the more important of these is that the two interpretations of the theorem given above FTNS are identical in that they both make the same algebraic statement, as was the case for the one-locus version of the theorem. To this extent there is a parallel between the one-locus and the multilocus versions of the theorem. There are, however, two aspects of the multilocus case which differ from those of the corresponding one-locus case. First, Fisher's "proof" of the theorem in the multilocus case rested on a simple summation of one-locus additive genetic variances over all loci in the genome. Doing this is tantamount to claiming that multilocus average effects are the same as one-locus marginal estimates and that the multilocus additive genetic variance is the sum of one-locus values. The discussion in Section 7.4.4 shows that in practice, both claims are true only in very unlikely event that all equations of the form (7.51) hold. It is then perhaps surprising that the theorem, when analyzed in algebraic detail, does indeed generalize to the multilocus case. The effects of the two incorrect assumptions cancel each other out.

This comment is connected to the second aspect of the relation between one-locus and multilocus results. It was shown in Section 2.9 with reference to the single-locus case that if fitness depends on the genotype at some given locus only and the equations in (2.76) hold for the genotypes at that locus, then the *total* change in mean fitness is equal to the *partial* change in mean fitness, defined either through (2.73) or (2.74). This conclusion no longer holds in the multilocus case.

To see this, suppose that equations of the form (2.76) holds for all triples of one-locus genotypes at all loci in the genome. Then equations of the form (2.77), (2.78) and (2.79) also hold for all possible genotype triples at all loci. Multiplying throughout by \bar{w} in equation (2.79) and changing to multilocus notation, we get

$$\bar{w}(x_{ku}\beta_{ku} + \sum_v X_{ku,kv}\beta_{kv}) = \bar{w}(\Delta x_{ku}). \quad (7.57)$$

The right-hand sides in equations (7.43) and (7.57) are equal so that, equating the two left-hand sides,

$$x_{ku}\alpha_{ku} + \sum_v X_{ku,kv}\alpha_{kv} + \sum_m \sum_t Q_{ku,mt}\alpha_{mt} = \bar{w}(x_{ku}\beta_{ku} + \sum_v X_{ku,kv}\beta_{kv}). \quad (7.58)$$

In contrast to the corresponding conclusion in the one-locus case, this equation does not necessarily imply that $\beta_{ku} = \alpha_{ku}/\bar{w}$, and thus does not necessarily imply that equations of the form (2.77) hold. This in turn implies that even if (2.76) holds for all triples of genotypes at all loci in the genome, the total change in mean fitness is not necessarily equal to the partial change. In practice, since the term $\sum_m \sum_t Q_{ku,mt} \alpha_{mt}$ in (7.58) is very unlikely to be zero, total and partial changes will be very unlikely to be equal.

In conclusion we emphasize the breadth of application of the multilocus FTNS. It is a whole-genome result and makes no assumption about the mating scheme, so that random mating is *not* assumed. Further, while the derivations given above all assume viability fitness, Lessard and Castilloux (1995) have shown that with the same assumptions as those made for viability selection, the theorem also holds in the fertility fitness case. On the other hand, all the various simplifying assumptions made in the single-locus version of the theorem continue to be made. For example, the complications caused by the sex chromosomes is ignored, as are those caused by the very existence of two sexes. This restricts for the moment the real-world applicability of the theorem, although a generalization of it to cover the case of two sexes is surely possible.

7.4.6 *Optimality Principles*

It is natural to follow a long-established practice in physics, associated there for example with least action principles, and to ask: What is optimized under the gene frequency changes brought about by natural selection?. One approach to this question was opened up (in the one-locus, random-mating case) by Kimura (1958) and has been taken up by, among others, Crow and Kimura (1970) Svirezhev (1972), Shahshahani (1979), Akin (1982) and Ewens (1992). We show below that the entire-genome, arbitrary-mating version of the FTNS leads to an optimality principle of natural selection generalizing Kimura's principle, which is restricted to the one-locus, random-mating multiple-allele case discussed in Section 2.4. A second, and quite different, optimality principle was introduced by Svirezhev (1972), and we discuss this principle below.

The discrete-time version of Kimura's (1958) principle derives from various results in Section 2.4, and depends in particular the gene frequency changes implicit in (2.7) and (2.8), the mean fitness defined in (2.10) and the additive genetic variance defined in (2.17). We also note that this expression for the additive genetic variance, together with the natural selection change in gene frequency

$$\Delta x_i = x_i(w_i - \bar{w})/\bar{w} \quad (7.59)$$

implied by (2.8), shows that the partial increase in mean fitness, described in (2.72), can be written as

$$2 \sum_i \Delta x_i (w_i - \bar{w}). \quad (7.60)$$

Suppose now that gene frequencies are changed by arbitrary amounts d_1, d_2, \dots, d_k , where necessarily $\sum_i d_i = 0$. We could then say, from (7.60), that the partial increase in mean fitness is

$$2 \sum_i d_i (w_i - \bar{w}). \quad (7.61)$$

Kimura (1958) then showed that under the restriction

$$\sum_i d_i^2 / x_i = \sigma_A^2 / 2\bar{w}, \quad (7.62)$$

the set $\{d_i\}$ of gene frequency changes that maximize the partial increase in mean fitness (7.61) is the natural selection set defined in (7.59).

While this is an interesting conclusion, it is incomplete, since no independent extrinsic reason is given for imposing the constraint (7.62). From a purely mathematical point of view there will always be some quadratic constraint such that a linear function such as (7.61) will be maximized at some specified set of $\{d_i\}$ values. Unless such a reason can be found for imposing the constraint, the maximizing conclusion loses its force. This was pointed out, for example, by Edwards (1974).

Moreover, the constraint is an important one. Crow and Kimura (1970) claimed that the result holds without the constraint, but this is not correct. This was pointed out by Svirezhev (1972). Svirezhev carried out his analysis in terms of the non-Euclidean distance measure $(ds)^2 = \sum_i d_i^2 / x_i$ appearing on the left-hand side of (7.62). This was done for mathematical convenience and no biological justification is needed for this procedure. We will see later that non-Euclidean distance measures are natural in the optimizing context.

Edwards (1974) noted that the Kimura optimizing principle can be cast in an equivalent dual form. This begins with the observation that when the arbitrary set of changes $\{d_i\}$ in gene frequency is equal to the natural selection set of changes $\{\Delta x_i\}$, the expression in (7.61) is equal to the partial increase σ_A^2 / \bar{w} in mean fitness. Then the dual form of the principle is that under the constraint

$$2 \sum_i d_i (w_i - \bar{w}) = \sigma_A^2 / \bar{w}, \quad (7.63)$$

together with the natural constraint $\sum_i d_i = 0$, the set $\{d_i\}$ of arbitrary changes in gene frequency is that which minimizes the quantity $\sum_i d_i^2 / x_i$ appearing on the left-hand side in (7.62) is the natural selection set $\{\Delta x_i\}$.

Since, from (2.16), the expression $w_i - \bar{w}$ appearing in (7.63) is identical to α_i , it is equivalent to write the constraint (7.63) as

$$2\mathbf{d}'\boldsymbol{\alpha} = 2\sum_i d_i\alpha_i = \sigma_A^2/\bar{w}. \quad (7.64)$$

Our aim in this section is to extend the Kimura maximizing principle both to the whole genome level and, simultaneously, to the non-random-mating case, and also to find a natural biological reason why a constraint of the form (7.62), or alternatively of the form (7.64), arises. We consider only the one-locus case in detail and use matrix methods in the analysis, since with a change in the definition of the matrices involved, the multi-locus extension follows almost immediately from the single-locus analysis. Specifically, we use the matrix notation defined immediately above (2.65), re-writing this equation as

$$\boldsymbol{\Delta} = (\bar{w})^{-1}(D + P)\boldsymbol{\alpha}. \quad (7.65)$$

The broad aim of the optimizing principle that we seek is to maximize, or to minimize, some function subject to some constraint. In order to suggest a starting point for doing this, we recall the derivation of the average effects of alleles through the minimization of the quadratic form (2.57) subject to the constraint (2.58) (in the single-locus case), or minimization of the quadratic form (7.42) subject to the constraint (7.21) (in the multilocus case). We therefore re-consider this procedure, and observe that the constraint (2.58), used in the derivation of (2.65) and thus of (7.65), is unnecessary, since this constraint is automatically satisfied at the unconditional minimum of the quadratic form (2.57). We are therefore free to impose an alternative constraint also satisfied at the unconditional minimum of this quadratic form. For this purpose we impose the constraint implied by (2.63), which we rewrite as

$$\boldsymbol{\alpha}'\boldsymbol{\Delta} = \sum_u \alpha_u \Delta x_u = \sigma_A^2/2\bar{w}. \quad (7.66)$$

It can then be shown that the calculation of the average effects $\boldsymbol{\alpha}$ is equivalent to the minimization of the quadratic form

$$\boldsymbol{\alpha}'(D + P)\boldsymbol{\alpha}, \quad (7.67)$$

subject to the constraint (7.66).

We now turn to finding an optimality principle generalizing that of Kimura. Let $\mathbf{d}' = (d_1, d_2, \dots, d_k)$ be an arbitrary vector of gene frequency changes. Generalizing the aim of the dual form of the Kimura principle, we seek to find some quadratic form $\mathbf{d}'T^{-1}\mathbf{d}$ in these gene frequency changes, generalizing the expression on the left-hand side of (7.62), which, subject to some natural constraint, is minimized at the natural selection values $\mathbf{d} = \boldsymbol{\Delta}$. The constraint we use is that given by the Edwards constraint in (7.64).

This minimum is obtained, using Lagrange multipliers, by finding the absolute minimization of the function

$$\mathbf{d}'T^{-1}\mathbf{d} - \lambda(2\mathbf{d}'\boldsymbol{\alpha} - \sigma_A^2/\bar{w}) \quad (7.68)$$

with respect to the elements in \mathbf{d} . Standard methods show that the minimum of this Lagrangian expression arises when

$$\mathbf{d} = \lambda T\boldsymbol{\alpha}, \quad (7.69)$$

where in this equation we may take λ as a disposable constant.

If we wish this conditional minimum to arise when $\mathbf{d} = \boldsymbol{\Delta}$, it is necessary to choose T so that

$$\boldsymbol{\Delta} = \lambda T\boldsymbol{\alpha}. \quad (7.70)$$

A comparison of this equation with (7.65) shows that we must take the matrix T to be some multiple of the matrix $D + P$. Whatever this choice of multiple might be, minimization of $\mathbf{d}'T^{-1}\mathbf{d}$ is equivalent to minimization of

$$\mathbf{d}'(D + P)^{-1}\mathbf{d}, \quad (7.71)$$

so we are now free to think of our aim as the minimization of this expression.

Suppose we choose $T = (\bar{w})^{-2}(D + P)$ and put $\mathbf{z} = \bar{w}(D + P)^{-1}\mathbf{d}$. Then

$$\mathbf{d}'T^{-1}\mathbf{d} = \mathbf{z}'(D + P)\mathbf{z}, \quad (7.72)$$

and the constraint $\mathbf{d}'\boldsymbol{\alpha} = \sigma_A^2/2\bar{w}$ given in (7.64) becomes

$$\mathbf{z}'(D + P)(\bar{w})^{-1}\boldsymbol{\alpha} = \sigma_A^2/2\bar{w}. \quad (7.73)$$

Since, from (2.65), $(D + P)\boldsymbol{\alpha} = \bar{w}\boldsymbol{\Delta}$, (7.73) may be written as

$$\mathbf{z}'\boldsymbol{\Delta} = \sigma_A^2/2\bar{w}. \quad (7.74)$$

In other words, we can think of our aim as the minimization of $\mathbf{z}'(D + P)\mathbf{z}$, subject to the constraint in (7.74).

But this is identical to the procedure defining the average effects of the alleles given above, that is the minimization of the expression in (7.67) subject to the constraints in (7.66). To the extent that we think of the calculation of the average effects as being a natural one, the minimization procedure that we have arrived at is also a natural one.

Further, it is easy to see that under random mating, the expression $\mathbf{d}'(D + P)^{-1}\mathbf{d}$ reduces to the Kimura expression $\sum_i d_i^2/x_i$, so that our procedure is a direct generalization of his to the non-random-mating case. Thus we have not only generalized his procedure but have also justified the conditioning restriction in that procedure.

In the random-mating case Svirezhev (1972) used the expression $\sum_i d_i^2/x_i$ as a natural non-Euclidian distance metric between parental and daughter generation. The above analysis shows that $\mathbf{d}'(D + P)^{-1}\mathbf{d}$ is the natural generalization of this to the non-random-mating case.

The multilocus generalization of this follows almost immediately. The parallel between the single-locus matrix equation (2.65) and the multilocus matrix equation (7.44) and the natural flexibility of matrix operations shows that the changes in gene frequency brought about by natural selection are those which minimize the non-Euclidean genetic distance $\mathbf{d}'(D + P + Q)^{-1}\mathbf{d}$ between parental and daughter generation multilocus gene frequency values, subject to the constraint that the partial increase in mean fitness between the two generations is σ_A^2/\bar{w} . As in the single-locus case, there is a natural reason why the quantity $\mathbf{d}'(D + P + Q)^{-1}\mathbf{d}$ can be regarded as a distance measure.

Another form of optimality was established by Svirezhev (1972). This refers only to the continuous-time, single-locus, random-mating case, and does not appear to generalize to discrete time, many loci or the non-random-mating case. It is based on the fact that the equations of motion in physics can be obtained by an optimizing principle. Specifically, the path taken by an object is that which minimizes the total difference between kinetic and potential energy over the path. Adopting the notation of Section 2.7, we let \dot{x}_i be the rate of change of the frequency of the allele A_i . Svirezhev (1972) then forms the function f defined by

$$f = \sum_i \frac{(\dot{x}_i)^2}{x_i} + \sum_i x_i(w_i - \bar{w})^2, \quad (7.75)$$

where w_i and \bar{w} are defined respectively in (2.9) and (2.10). This function is thought of as the difference between a kinetic and a potential energy, the latter being half the additive genetic variance and being zero at the equilibrium point(s) of the evolutionary system. The aim is to minimize the function

$$\int_{t_1}^{t_2} f dt,$$

subject to the natural conditions $\sum_i x_i = 1$, for given starting and finishing times t_1 and t_2 and given initial and final gene frequencies. Svirezhev finds that the minimum is found when $\dot{x}_i = x_i(w_i - \bar{w})$, which is exactly the equation of motion (2.41) that natural selection brings about.

It has been claimed that natural selection acts in such a way that the total integral of the additive genetic variance along the evolutionary path of gene frequencies. The claim is made since along this path the equation $\dot{x}_i = x_i(w_i - \bar{w})$ holds, and that substituting this value of \dot{x}_i into the equation (7.75) for f , it is found that $f = 2 \sum x_i(w_i - \bar{w})^2$, the additive genetic variance. However this argument is not correct, since the equation

$$\sum_i \frac{(\dot{x}_i)^2}{x_i} = \sum_i x_i(w_i - \bar{w})^2$$

does not hold for arbitrary evolutionary trajectories.

7.5 The Correlation Between Relatives

In this section we investigate properties of the correlation between relatives for characters that are assumed to depend on a large number of loci. A full discussion of this topic would take into account various forms of non-random mating and also environmental effects. A full consideration of these effects would take us far further into the biometrical aspects of population genetics than we wish to go. Further, correlations in the non-random mating case become extremely complicated, so we assume random mating throughout this section. Our aim is to restrict attention to the effects of linkage and linkage disequilibrium on standard formulas for these correlations under the simplifying assumptions we make. A brief discussion of the entire question of the correlation between relatives will be given in Chapter 8.

We first consider characters determined by a single locus and consider a more efficient method of arriving at (1.17)–(1.20). In this and in later generalizations to characters determined by many loci, it is convenient to calculate covariances rather than correlations: The latter can always be found from the former by dividing by the total variance σ^2 in the character considered.

Suppose that the alleles at the locus in question are A_1, A_2, \dots, A_k , with respective frequencies x_1, x_2, \dots, x_k , and that the value for the character for an $A_u A_v$ individual is m_{uv} . We define \bar{m} and the average excess a_u of A_u as in 2.60, and write

$$\bar{m}_{u,v} = \bar{m} + a_u + a_v + d_{u,v}, \quad (7.76)$$

so that

$$\sum_u a_u x_u = 0, \quad \sum_u d_{u,v} x_u = 0 \quad \text{for all } v. \quad (7.77)$$

The additive and dominance genetic variances σ_A^2 and σ_D^2 are given by (1.16) and (1.15). Consider now two individuals X and Y , with measurements $m(X)$ and $m(Y)$. The covariance between these measurements is

$$E\{(m(X) - \bar{m})(m(Y) - \bar{m})\}. \quad (7.78)$$

If the two individuals are unrelated, this covariance is zero. On the other hand, related individuals can possess genes in common that are identical in descent from one or more common ancestors. Thus, for example, the contribution a_u in (7.76) may be identical in both individuals, since both possess an A_u gene passed on from a common ancestor. Suppose the genes possessed by individual X are x_f, x_m , where the suffixes f and m denote the genes passed on from father and mother, respectively, and define y_f, y_m similarly for individual Y . We use the symbol “ \equiv ” to denote “identical

by descent", and define

$$\begin{aligned} P_{ff} &= \text{Prob}(x_f \equiv y_f), & P_{fm} &= \text{Prob}(x_f \equiv y_m), \\ P_{mf} &= \text{Prob}(x_m \equiv y_f), & P_{mm} &= \text{Prob}(x_m \equiv y_m). \end{aligned} \tag{7.79}$$

Then by inserting (7.76) in (7.78) it is found that

$$\begin{aligned} \text{covar}(X, Y) &= \frac{1}{2}(P_{ff} + P_{fm} + P_{mf} + P_{mm})\sigma_A^2 \\ &\quad + (P_{ff}P_{mm} + P_{fm}P_{mf})\sigma_D^2. \end{aligned} \tag{7.80}$$

This formula, due essentially to Malécot (1948), provides a simple and powerful method for deriving covariances for any two related individuals, and we now use it to re-derive (1.17) - (1.20).

Consider first the father-son correlation, with X being the father and Y the son. Since the mother and father are assumed to be unrelated, $P_{mm} = P_{fm} = 0$. Also $P_{ff} = P_{mf} = \frac{1}{2}$, and insertion of these values into (7.80) yields (1.17). If X and Y are full sibs, $P_{ff} = P_{mm} = \frac{1}{2}$, $P_{fm} = P_{mf} = 0$, and insertion of these values in (7.80) gives (1.18). Equations (1.19) and (1.20) can be found equally easily.

We next extend (7.80) to the case where the character in question depends on all K loci in the genome. Here additive \times additive, additive \times dominance, and further variance terms enter into the covariances, and it is necessary to develop notation for these. We write $\sum \sigma_A^2$ for the sum, over all of loci, of the (marginal) additive genetic variances at those loci, with a similar notation for dominance variances. We also write $\sum \sigma_{AA}^2$ for the sum, over all possible pairs of loci, of the additive \times additive variances for each pair, $\sum \sigma_{AD}^2$ for the sum of all additive \times dominance variances, and more generally

$$\sum \sigma_{A^r D^s}^2, \tag{7.81}$$

for $1 \leq r + s \leq K$, for the sum of all possible r -wise additive and s -wise dominance variances.

Consider first the simplest case when all loci involved are unlinked and all pairwise coefficients of linkage disequilibrium are zero. Kempthorne (1954, 1955) obtained the appropriate generalization of (7.80) as

$$\begin{aligned} \text{covar}(X, Y) &= \sum' \left(\frac{1}{2}(P_{ff} + P_{fm} + P_{mf} + P_{mm}) \right)^r \\ &\quad \times (P_{ff}P_{mm} + P_{mf}P_{fm})^s \sigma_{A^r D^s}^2 \end{aligned} \tag{7.82}$$

where the summation \sum' is over all r and s with $1 \leq r + s \leq K$. Writing

$$\alpha = \frac{1}{2}(P_{ff} + P_{fm} + P_{mf} + P_{mm}), \quad \beta = P_{ff}P_{mm} + P_{mf}P_{fm}, \tag{7.83}$$

this becomes, for a character depending on two loci,

$$\text{covar}(X, Y) = \alpha \sum \sigma_A^2 + \alpha^2 \sum \sigma_{AA}^2 + \beta \sum \sigma_D^2 + \beta^2 \sum \sigma_{DD}^2 + \alpha\beta \sum \sigma_{AD}^2, \tag{7.84}$$

where both summations are over the two loci involved. This confirms in an elegant way the correlations found in Section 2.10. It follows also from (7.82) that for a character depending on an arbitrary number of loci,

$$\begin{aligned} &\text{covar}(\text{father-son}) \\ &= \left\{ \frac{1}{2} \sum \sigma_A^2 + \frac{1}{4} \sum \sigma_{AA}^2 + \frac{1}{8} \sum \sigma_{A^3}^2 \right. \\ &\quad \left. + \cdots + \frac{1}{2^K} \sum \sigma_{A^K}^2 \right\}, \end{aligned} \tag{7.85}$$

$$\begin{aligned} &\text{covar}(\text{grandfather-grandson}) \\ &= \left\{ \frac{1}{4} \sum \sigma_A^2 + \frac{1}{16} \sum \sigma_{AA}^2 + \cdots + \frac{1}{4^K} \sum \sigma_{A^K}^2 \right\}. \end{aligned} \tag{7.86}$$

Clearly ancestral line covariances do not contain dominance terms. It is not clear how important the various terms in (7.85) and (7.86) are. While for large r each r -order additive interaction term is probably very small, there are $\binom{K}{r}$ terms in $\sigma_{A^r}^2$ and, even allowing for the factor 2^{-r} , the total contribution of r -order interaction terms need not be small.

Our next objective is to generalize (7.82) to allow for linkage between loci, continuing however to assume complete linkage equilibrium. This generalization has been considered for two loci by Cockerham (1956) and for K loci by Schnell (1963) and van Aarde (1975). The typical term in (7.82) is the sum of expressions of the form

$$\alpha^r \beta^s \sigma^2(k_1, k_2, \dots, k_r; \ell_1, \ell_2, \dots, \ell_s) \tag{7.87}$$

where $k_1 < k_2 < \dots < k_r, \ell_1 < \ell_2 < \dots < \ell_s$ with $k_p \neq \ell_q$ for all p, q and $\sigma^2(k_1, k_2, \dots, k_r; \ell_1, \ell_2, \dots, \ell_s)$ denotes the contribution to the variance of the interaction of the additive effects at loci k_1, k_2, \dots, k_r and the dominance effects at loci $\ell_1, \ell_2, \dots, \ell_s$. The coefficient $\alpha^r \beta^s$ is the probability of $r + s$ independent events, the independence arising from the assumptions that the loci are unlinked and that linkage equilibrium obtains, composed of events typified by

$E_{ff}^k = \{\text{genes derived by } X \text{ and } Y \text{ at locus } k \text{ from their respective fathers are identical by descent}\}.$

For unlinked loci events of the form E_{ff}^k and E_{ff}^ℓ are independent for $k \neq \ell$ if we continue to assume linkage equilibrium, but this is no longer true for linked loci. Thus, for example, the probability of the compound event

$$E_{fm}^{k_1} E_{mf}^{k_2} \cdots E_{ff}^{k_r} \tag{7.88}$$

can no longer be obtained by simple multiplication, but will involve the recombination fractions between loci k_1, k_2, \dots, k_r . The appropriate generalization of (7.82) for linked loci is found by replacing the typical term by the sum, for all possible r -wise additive and s -wise dominance

contributions, of

$$\begin{aligned} & \left(\frac{1}{2}\right)^r \text{Prob}\left(\prod_{p=1}^r \{E_{ff}^{k_p} + E_{fm}^{k_p} + E_{mf}^{k_p} + E_{mm}^{k_p}\}\right) \\ & \quad \times \prod_{q=1}^s \{E_{ff}^{\ell_q} E_{mm}^{\ell_q} + E_{fm}^{\ell_q} E_{mf}^{\ell_q}\} \\ & \quad \times \sigma^2(k_1, k_2, \dots, k_r; \ell_1, \ell_2, \dots, \ell_s). \end{aligned} \quad (7.89)$$

In the calculation of this probability, all product terms are expanded out and the probabilities of sums of compound events of the form (7.88) calculated. This formula, although explicit, yields rather complicated values when more than two loci are involved. In particular, the expression (7.89) does *not* in general provide a contribution to the covariance of the simple form $c \sum \sigma_{A^r D^s}^2$, for some constant c . We thus consider the application of (7.89) for two loci only, assuming the recombination fraction between these loci is R . The summations in the various formulas below are over these two loci.

Consider first the case of full sibs. Here the events E_{fm} and E_{mf} are impossible (for both loci) and the events E_{ff} , E_{mm} have the same probabilities for both loci. These observations lead to a covariance of

$$\begin{aligned} & \frac{1}{2} \text{Prob}(E_{ff}^k + E_{mm}^k) \sum \sigma_A^2 + \text{Prob}[E_{ff}^k E_{mm}^k] \sum \sigma_D^2 \\ & \quad + \frac{1}{4} \text{Prob}[(E_{ff}^k + E_{mm}^k)(E_{ff}^l + E_{mm}^l)] \sigma_{AA}^2 \\ & \quad + \frac{1}{2} \text{Prob}[(E_{ff}^k + E_{mm}^k) E_{ff}^l E_{mm}^l] \sigma_{AD}^2 \\ & \quad + \text{Prob}[E_{ff}^k E_{mm}^k E_{ff}^l E_{mm}^l] \sigma_{DD}^2. \end{aligned} \quad (7.90)$$

Events involving the subscripts ff and mm are independent. We know that $\text{Prob}(E_{ff}^k) = \text{Prob}(E_{mm}^k) = \frac{1}{2}$ so to compute (7.90) it is necessary to compute $\text{Prob}(E_{ff}^k E_{ff}^l)$. By considering the various cross-over possibilities it is found that this probability is

$$\frac{1}{2} - R + R^2, \quad (7.91)$$

and substitution of this value in (7.90), and the same value for $\text{Prob}(E_{mm}^k E_{mm}^l)$, leads immediately to the expression in (2.125).

It is of some interest to compute half-sib and ancestral line covariances. For half-sibs, assuming a common father, (7.90) still holds, since E_{fm} and E_{mf} are still both impossible events, but now E_{mm} is also an impossible event and (7.90) reduces easily to

$$\text{covar}(\text{half-sibs}) = \frac{1}{4} \sum \sigma_A^2 + \frac{1}{4} \left(\frac{1}{2} - R + R^2\right) \sigma_{AA}^2. \quad (7.92)$$

For ancestral lines there are no dominance components to the covariance but, interestingly, except for the father-son value, these covariances do depend on R . Thus, for example, (7.85) still applies for all R but, for two

loci, (7.86) must be replaced by

$$\text{covar}(\text{grandfather-grandson}) = \frac{1}{4} \sum \sigma_A^2 + \frac{1}{8}(1-R)\sigma_{AA}^2. \quad (7.93)$$

While the right-hand sides in (7.92) and (7.93) are identical for unlinked loci and also for completely linked loci ($R = 0$), they are not equal for $0 < R < \frac{1}{2}$. For these value of R , linkage causes a greater change in the grandfather-grandson covariance (compared to the unlinked loci case) that it does to the half-sib covariance.

Do simple limiting covariances hold for tightly linked loci? van Aarde (1975) demonstrated that, as $R \rightarrow 0$,

$$\begin{aligned} \text{covar}(X, Y) \rightarrow \alpha \left(\sum \sigma_A^2 + \frac{1}{2} \sigma_{AA}^2 \right) \\ + \beta \left(\sum \sigma_D^2 + \frac{1}{2} \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 \right), \end{aligned} \quad (7.94)$$

where α and β are defined in (7.83). This formula can also be deduced immediately from Schnell, 1963. Fisher (1918) asserted that for linked loci the pattern of covariances for traits depending on two loci would not differ significantly from those applying for traits depending on one locus. For tightly linked loci this view is supported by the form of the expression in (7.94), which is of the same as that in (7.80) with the single-locus additive and dominance variances being replaced by the "generalized" values

$$\sum \sigma_A^2 + \frac{1}{2} \sigma_{AA}^2, \quad \sum \sigma_D^2 + \frac{1}{2} \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 \quad (7.95)$$

respectively. A parallel remark no doubt holds for a number of closely linked loci. For values of R not close to 0 or $\frac{1}{2}$, however, slight deviations from the pattern suggested by (7.80) do occur, as may be seen by comparing (7.92) and (7.93). While these covariances are identical at $R = 0$, $R = \frac{1}{2}$, they differ at $R = \frac{1}{4}$ by $\sigma_{AA}^2/64$, a value that is however probably negligible for most traits.

Formulae for covariances when the loci involved are not in linkage equilibrium are extremely complicated. Even for two loci the expressions given by Gallais (1974) and Weir and Cockerham (1977) contain upward of a hundred terms, and there can be no hope of estimating the various components from even the most extensive data. Thus, while a useful general formula for correlations appears to be almost impossible to find, some progress can, however, be made by considering special cases. The models considered in Section 7.3.4 suggest that for large recombination fractions equilibrium values of the coefficient of linkage disequilibrium are likely to be small: This is confirmed by simulation for "random" fitness patterns. Thus most interest attaches to the case of small recombination fractions where, as we have noted, both linkage disequilibrium and linkage equilibrium equilibria are both sometimes stable.

For two loci and small R we may compare correlations between relatives by comparing the limiting value (7.94) for the linkage equilibrium case with

the corresponding limiting value found by direct computation when only two gametes exist, so that there is maximum linkage disequilibrium. It is found that there is no necessary relation between the two correlations and that they can differ considerably even for simple models. Thus if the matrix of measurements is of the simple additive form

$$\begin{array}{rcccl}
 & & A_{21}A_{21} & A_{21}A_{22} & A_{22}A_{22} \\
 A_{11}A_{11} & 1.1 & 1.1 & 0.8 & \\
 A_{11}A_{12} & 1.2 & 1.2 & 0.9 & (7.96) \\
 A_{12}A_{12} & 1.4 & 1.4 & 1.1 &
 \end{array}$$

the correlation between two relatives X and Y is

$$\text{corr}(X, Y) = 0.874\alpha + 0.126\beta \quad (7.97)$$

in the linkage equilibrium case where

$$y_{11,21} = 0.09, \quad y_{11,22} = 0.21, \quad y_{12,21} = 0.21, \quad y_{12,22} = 0.49.$$

Here α and β are defined in (7.83). By contrast, the correlation is

$$\text{corr}(X, Y) = 0.276\alpha + 0.724\beta \quad (7.98)$$

in the case of extreme linkage disequilibrium, where

$$y_{11,21} = 0.3, \quad y_{11,22} = y_{12,21} = 0, \quad y_{12,22} = 0.7.$$

Although this example represents two extreme cases of linkage disequilibrium, it is perhaps disquieting that the correlations (7.97) and (7.98) should be so very different.

This observation leads us to a more informal discussion of the correlation between relatives for a metrical trait determined by many loci and also by the environment. So far we have considered only the purely genetic contribution to these correlations, and have also assumed random mating with respect to the trait of interest. For practical applications it is important to discuss the consequences of assortative mating and environmental effects on these correlations, since assortative mating certainly occurs with respect to various metrical characters and a shared or a similar environment between close relatives implies an environmental component to the correlation between relatives. Clearly, any modeling of the environmental contribution is hazardous.

Almost all authors, when considering assortative mating and environmental effects, have assumed, usually implicitly, that the loci controlling the trait are unlinked, that complete linkage equilibrium holds, and that no epistatic effects arise. If these simplifying assumptions hold, then (7.82) becomes

$$\text{covar}(X, Y) = \alpha \sum \sigma_A^2 + \beta \sum \sigma_D^2, \quad (7.99)$$

where α and β are determined in (7.83) and both sums are taken over all loci involved in determining the trait in question. We rewrite this as

$$\text{corr}(X, Y) = \alpha H + \beta D, \quad (7.100)$$

where $H = \sum \sigma_A^2 / \sigma^2$, $D = \sum \sigma_D^2 / \sigma^2$, a notation more in line with that of biometrical genetics and one we follow for the remainder of this section. We emphasize the strong assumptions that have been made in replacing complex equations such as (7.89), which is itself far less complex than the corresponding equation when linkage equilibrium is not assumed, by the simple (7.100). In this connection we observe that in the biometrical literature, for example in Eaves et al. (1977), it is sometimes implied that absence of assortative mating ensures linkage equilibrium, but the earlier theory of this chapter shows that this claimed implication is incorrect.

The question of assortative mating without environmental effects was considered by Fisher (1918) and many subsequent authors. We assume a correlation r_{pp} between mating individuals for the trait in question: The case $r_{pp} = 0$ leads to equations like (7.100). When there is no dominance, Nagylaki (1978) arrived by very simple arguments at formulas for the correlation between relatives. Wright (1921), Crow and Felsenstein (1968) and Wilson (1973) arrived at extensions of these formulas for the more difficult case where dominance exists. Some examples of the correlations of an individual with the relative specified are the following:

$$\begin{aligned} \text{corr}(\text{with parent}) &= r_{p0} = \frac{1}{2}(1 + r_{pp})H, \\ \text{corr}(\text{with full sib}) &= r_{ss} = \frac{1}{2}(1 + r_{pp}H)H + \frac{1}{4}D, \\ \text{corr}(\text{with } n\text{th generation grandparent}) &= r_{p0} \left\{ \frac{1}{2}(1 + r_{pp}H) \right\}^{n-1}, \quad (7.101) \\ \text{corr}(\text{with uncle}) &= r_{un} = \left\{ \frac{1}{2}(1 + r_{pp}H) \right\}^2 H \\ &\quad + \frac{1}{8}r_{pp}DH, \\ \text{corr}(\text{with first cousin}) &= r_{fc} = \left\{ \frac{1}{2}(1 + r_{pp}H) \right\}^3 H \\ &\quad + D \left\{ \frac{1}{4}r_{pp}H \right\}^2. \end{aligned}$$

More complex formulas for models where selection acts against individuals with extreme values of the character have been given by Wilson (1973, 1976).

We turn now to environmental effects. These were considered by Fisher (1918), but only for the simplified model where the phenotypic value P of the trait can be written as $P = G + E$, G and E being genetic and environmental contributions, and for which

$$\sigma_p^2 = \text{var}(P) = \text{var}(G) + \text{var}(E). \quad (7.102)$$

The equation $P = G + E$ implies no interaction between genotype and environment, while (7.102) further implies no covariance. While these assumptions may be reasonable as approximations in various controlled

breeding experiments, they are severe assumptions for a trait such as IQ in humans. The entire work of Burt (1971) on this trait, for example, makes precisely these assumptions. Burt's approach is to define $E = \text{var}(E)/\sigma_p^2$ and to add E to each correlation displayed in (7.101), where now H and D are redefined using σ_p^2 instead of σ^2 as divisor. Burt also uses a variance term for assortative mating but this appears to be unjustified and is not followed by other authors. If any two correlations can be estimated from data, these formulas then give H , D and E , since $H + D + E = 1$. Burt uses the three correlations r_{p0} , r_{pp} and the correlation $(1 + E)^{-1}$ of identical twins, since under his model a further correlation is required to estimate the "assortive mating" component of variance.

A more realistic model clearly allows for both genotype-environment interaction and covariance, and also distinguishes the almost certainly different environmental correlations for individuals of different degrees of relatedness. Various models, of increasing complexity, have been put forward, largely in the psychological literature to this end. Jinks and Eaves (1974) assumed an added environmental correlation only for those individuals in the same family, that is, for parents and offspring living together; this applies for the first two correlations in (7.101). A revised model adds a correlation only for sibs raised together. Eaves et al. (1977) presented a model with two environmental correlations, E_1 for "within families" and E_2 for "between families": Here E_1 is added to the first two correlations in (7.101) and E_2 to the last three.

With empirical values for correlations between a variety of relatives, involving perhaps 10-15 relationships, the parameters of any reasonably simple model can be estimated by least squares and the goodness of fit of the model tested by chi-square. The argument in this procedure is that if a fit is not rejected by chi-square the model can be accepted and a more complex model is not needed. Perhaps the major difficulty with this procedure is the low power of the goodness-of-fit procedure. Thus Last (1976) found that for reasonable parameter values sample sizes in excess of 5,000 are required to be fairly certain of detecting genotype-environment interactions of some magnitude. The question of the adequacy of various models and of the above fitting procedure has led to much acrimonious discussion in the literature: See Eaves et al. (1977), Mather and Jinks (1977) and Goldberger (1978a,b) for summarizing and strongly contrasting views. Our interest here is more in the genetic aspects, and we conclude by emphasizing that essentially all models used in the biometrical literature, on all sides, have used simple formulas, such as (7.99), for the purely genetic component of correlation and that in practice far more complex formulas are surely required. While simplifying assumptions must be made in these analyses at some level, it is not yet certain that the level chosen, that is that leading to (7.99), is a satisfactory one. The entire question of using correlations to estimate heritability must be viewed with great caution.

7.6 Summary

The two main conclusions reached in this chapter are, first, that opposing views may reasonably be held about the likely extent of linkage disequilibrium in natural populations and, second, that the degree of linkage disequilibrium that does occur will alter, perhaps significantly, the numerical values of several important population genetic parameters. As a result, the extent of linkage disequilibrium that can be expected in natural populations will affect one's view of the likely dynamic and static behavior of these populations.

The view that extensive linkage disequilibrium might occur in natural populations has been promoted in particular by Franklin and Lewontin (1970) and Lewontin (1974). This view is supported in some special cases by other authors, for example in the case of disruptive selection by Bulmer (1976). While Franklin and Lewontin show that the correlation properties in multilocus systems can build up linkage disequilibrium values in excess of those predicted by two-locus theory, their analysis applies almost exclusively for multiplicative selective models. Our analysis shows that for many other models, rather less linkage disequilibrium can be expected. Even under the multiplicative scheme, stable equilibria with no linkage disequilibrium can exist simultaneously with stable equilibria with high linkage disequilibrium, and it is not certain in general what the domains of attraction of the two types are. For stabilizing selection schemes the numerical calculations of Bulmer (1976) confirm the prediction of Wright (1965b) that the true additive genetic variance is somewhat less than the value computed assuming no disequilibrium, but the decrease is only of order 25% in Bulmer's simulations.

The theory described above does not take into account the possibility that linkage disequilibrium can be due to factors other than selection, in particular by geographical structure, as discussed below in Section 8.4. The existence of so-called "haplotype blocks", especially in humans, indicates that substantial linkage disequilibrium exists in humans, and this is of particular interest to human geneticists wishing to map disease genes. These blocks might be due to selection, to geographical factors, or simply to the existence of recombinational hot-spots and nonrecombinational "cold regions". The haplotype block concept, and mathematical aspects of the problem of mapping disease genes, will be considered in detail in Volume II.

We have already recorded in the summary of Chapter 6 the view of Ohta and Kimura (1975) that on the whole, linkage disequilibrium is generally small in natural populations, and to a first order can be ignored: Broadly speaking, this view is also given in Crow and Kimura (1970). Even to this day, no uniform view yet exists on this point.

If extensive linkage disequilibrium does occur, the genetic properties of a population can be quite complex. The mean fitness increase theorem

need not hold, although one suspects that a generalization of the concept of quasi-linkage equilibrium ensures that mean fitness “mostly” increases. The mean fitness of a population with strong linkage disequilibrium can be quite high, with only a comparatively small number of possible genotypes represented with non-negligible frequencies. However, such a structured system has less flexibility and thus capacity to cope with altered environmental conditions than does an unstructured population. The true additive genetic variance may be rather less or considerably more than that calculated without linkage disequilibrium, and in general the properties of the genetic system cannot easily be found by combining single-locus analyses. The correlation between relatives for measurable characteristics is also affected by the degree of linkage disequilibrium between the loci determining these characteristics.

Many of the general multilocus results for the case of a random-mating population continue to hold in the non-random-mating case, but some do not. The Fundamental Theorem of Natural Selection holds whether mating is random or non-random.

We conclude with two remarks concerning Chapters 9 and 11. In Chapter 9 models are considered that recognize the gene as a sequence of nucleotides. The theory of the present chapter is relevant to this analysis, with the nucleotide replacing the gene as the fundamental unit. Of course entirely new values for mutation rates, recombination fractions and selective differentials are appropriate at the molecular level, and this will alter the way in which we view certain formulas involving these parameters. Second, in Chapter 11 various tests of the “neutral theory” are described. These tests consider the hypothesis that the alleles at some locus are selectively equivalent, and use only the frequencies of these alleles in the testing procedure. The theory thus ignores the possible effects on these frequencies of closely linked selective loci. If extensive linkage disequilibrium does occur gene frequencies reflect more the selective forces acting on segments of the chromosome rather than single loci, and thus much of the testing theory described in Chapter 11 becomes invalid under these circumstances.

8

Further Considerations

8.1 Introduction

In line with the aims of this book, the material given in preceding chapters has focused on purely mathematical considerations. However many topics in population genetics generally require an extensive verbal discussion with perhaps a minimum of mathematical treatment. Some of these topics are discussed in this chapter. Some mathematical material not naturally addressed in previous chapters is also taken up here. The content of this chapter may be viewed as an introduction to a more complete discussion, in Volume II, of the topics considered here.

8.2 What is Fitness?

In most of this book “fitness” has been taken to mean viability fitness, that is as a measure of the capacity of an individual of a given genotype to survive from the time of conception to the age of reproduction. In Section 2.6 fertility selection was discussed briefly, but nowhere have we considered the fitness component relating to mating success. Fitness properly embraces all three aspects and, to be considered in greater detail, should involve the population age distribution as well as ecological and other factors, very few of which are mentioned in the literature in any detail. Indeed the concept of fitness is possibly too complex to allow of a useful mathematical

development. Since it enters fundamentally into many population genetics considerations, it is remarkable how little attention has been paid to it.

The most complete discussion of the concept of fitness is that of Kempthorne and Pollak (1970), who emphasize several points worth repeating. First, while it is uniformly agreed that fitness is a property of the entire genome of an individual, it is also apparently agreed, with Wright (1931), that to a first approximation, for a short time, a constant net selective value of any allele may usefully be defined. The concept of the marginal fitness value for any one-locus genotype, introduced in (7.46), and the fact that these values do not usually lead to the correct values of the additive affects of the alleles at the locus at which these marginal fitnesses are defined, are relevant factors in this viewpoint. However, evolutionary arguments require consideration of very long time periods, and it is not certain that the approximation of a constant net selective value of any allele is adequate for long-term considerations. The analysis of the interactive effects of genes at several loci given in Chapters 6 and 7 is relevant to this point. Second, the lack of connection between the Fisherian definition of fitness through Malthusian parameters and the bulk of mathematical evolutionary theory is unfortunate. Finally the very concept of fitness, in particular when fecundity parameters are not in the multiplicative form (2.28), appears to be elusive.

8.3 Sex Ratio

The problem of the evolution of the sex ratio was discussed briefly in Section 1.5, and the “non-genetic” argument of Fisher concerning this evolution was noted. A number of genetically-based arguments for sex ratio adjustment have been put forward since Fisher’s time: See in particular Shaw and Mohler (1953), Shaw (1958) and Eshel (1975). For theoretical analysis of Fisher’s theory see Kolman (1960), Bodmer and Edwards (1960), Edwards (1961) and Verner (1965).

We now briefly describe an approach to this problem by Uyenoyama and Bengtsson (1979) which is of particular interest, since the analysis is genetically based and is close in spirit to Fisher’s original argument. Consider, in a diploid population, an autosomal locus admitting alleles A_1 and A_2 and hence genotypes A_1A_1 , A_1A_2 and A_2A_2 , which we call genotypes 1, 2, 3 respectively. The frequency of males (females) of genotype i is m_i (f_i) and $\sum m_i = \sum f_i = 1$. We make two assumptions concerning the female genotypes. The first is that different genotypes have different brood sizes: Suppose females of genotype i have brood size proportional to σ_i . Second, we suppose that the sex ratio among offspring depends on the maternal genotype, and that females of genotype i produce a fraction α_i of male offspring and $1 - \alpha_i$ of female offspring.

The values α_i and σ_i specify the recurrence relations of the f_i and m_i . We are particularly interested in the equilibria of these recurrence relations and will consider these only in the particular case where the condition

$$\sigma_1\sigma_2(\alpha_1 - \alpha_2) + \sigma_2\sigma_3(\alpha_2 - \alpha_3) + \sigma_3\sigma_1(\alpha_3 - \alpha_1) = 0 \quad (8.1)$$

holds. (We show later that an argument using the parental expenditure concept leads to (8.1).) The recurrence relations have three equilibria, one of which is symmetric ($m_i = f_i$) and the others asymmetric ($m_i \neq f_i$). We focus attention here on the asymmetric equilibria.

The frequencies M of males and F of females in the population are

$$M = \sum f_i\sigma_i\alpha_i / \sum f_i\sigma_i, \quad F = \sum f_i\sigma_i(1 - \alpha_i) / \sum f_i\sigma_i. \quad (8.2)$$

Shaw and Mohler (1953) and others define the mean fitness W for this model as

$$W = \sum f_i\sigma_i\{\alpha_i M + (1 - \alpha_i)F\}. \quad (8.3)$$

Subject to $\sum f_i = 1$, W may be maximized with respect to the f_i , and, assuming that (8.1) holds, the maximizing values are found to occur precisely at the asymmetric equilibria of the system. In other words, if the system evolves towards such an equilibrium, it is evolving in such a way as to optimize the sex ratio as measured by the mean fitness W . It is found that the optimizing value of the sex ratio is

$$F/M = \{\sigma_1(1 - \alpha_1) - \sigma_2(1 - \alpha_2)\} / \{\sigma_2\alpha_2 - \sigma_1\alpha_1\}. \quad (8.4)$$

It remains for us to justify that (8.1) will be true under the parental expenditure concept. Suppose the ratio of the parental expenditure required to raise a female offspring to maturity compared to that to raise a male offspring is $\phi : 1$. The mean expenditure per offspring of a female of genotype i is then proportional to $\alpha_i + \phi(1 - \alpha_i)$. If females of all genotypes make the same total expenditure for their entire brood, then the brood size σ_i must satisfy the requirement

$$\sigma_i = K\{\alpha_i + \phi(1 - \alpha_i)\}^{-1} \quad (8.5)$$

for some constant K . Values of σ_i satisfying (8.5) automatically satisfy (8.1), and furthermore lead to an F/M value of ϕ^{-1} , as might be expected. Thus the optimal sex ratio is determined entirely by the relative expenditure in rearing male and female offspring.

8.4 Geographical Structure

The importance of geographical structure to the evolutionary theories of Wright (1931, 1969a, b, 1977, 1978) was noted in Chapter 1. The effects of

structure have been considered at some length in the literature and here we refer, very briefly, to several features of the analyses made.

The two main types of model of geographical structure considered in the literature are an “island” model of distinct subpopulations, or demes, and a continuous cline model clines in one or sometimes two dimensions. A standard fact about an island model is that if random mating obtains within any island, the fraction of individuals who are heterozygotes at a single locus admitting two alleles is necessarily less than or equal to the corresponding fraction for a large random-mating population with the same mean allelic frequencies. If the frequency of A_1 in island i is x_i , and island i comprises a fraction f_i of the entire population, this (Wahlund) inequality can best be demonstrated through the equation

$$2\bar{x}(1 - \bar{x}) = 2 \sum f_i x_i (1 - x_i) + 2 \sum f_i (x_i - \bar{x})^2, \quad (8.6)$$

where $\bar{x} = \sum f_i x_i$. We use this fact below when considering mean fixation times in stochastic models involving geographical structure.

The Wahlund effect concerns the frequencies of the genotypes at a single gene locus. The effects of geographical subdivision are perhaps more important in the analysis of multilocus systems, and we now consider one such analysis, of particular interest to human geneticists, where the joint frequencies of the genotypes at two loci are of interest. To be concrete we assume alleles A_1 and A_2 , with frequencies x and $1 - x$, at the first locus and B_1 and B_2 , with frequencies y and $1 - y$, at the second locus. We also assume that the two loci are unlinked.

Suppose that a proportion f_1 of a population lives on one island and a proportion $f_2 = 1 - f_1$ lives on another. The frequencies of A_1 and A_2 in the first island are assumed to be x_1 and $1 - x_1$ and in the second island x_2 and $1 - x_2$. Similarly the The frequencies of B_1 and B_2 in the first island are assumed to be y_1 and $1 - y_1$ and in the second island y_2 and $1 - y_2$.

We assume random mating within each island, that a sufficiently long time has passed so that a stationary situation has been reached, and also that there is no selection at the two loci of interest. Then linkage equilibrium exists within each island, so that for example within island 1 the frequency of the gamete $A_1 B_1$ (more frequently called “haplotype” in human genetics) is $x_1 y_1$. Thus the overall frequency of the gamete $A_1 B_1$ is $c_1 = f_1 x_1 y_1 + f_2 x_2 y_2$, with similar calculations leading to overall frequencies c_2, c_3 and c_4 of $A_1 B_2, A_2 B_1$ and $A_2 B_2$.

Even though linkage equilibrium holds within any island, it does not necessarily hold over the entire population, so that it is not necessarily the case that $c_1 c_4 - c_2 c_3 = 0$. For example, suppose that $f_1 = f_2 = 1/2$, so that the two islands have equal population sizes, that the frequencies of the gametes $A_1 B_1, A_1 B_2, A_2 B_1$ and $A_2 B_2$ in island 1 are 0.56, 0.24, 0.14 and 0.06 and in island 2 are 0.02, 0.08, 0.18 and 0.72. Then linkage equilibrium holds within each island, but the overall gametic frequencies $c_1 = 0.29, c_2 =$

0.16, $c_3 = 0.16$ and $c_4 = 0.39$ are far from satisfying the linkage equilibrium requirement $c_1c_4 - c_2c_3 = 0$.

This calculation shows that geographical structure can lead to linkage disequilibrium. The extent to which, in humans, linkage disequilibrium between two loci might be due to geographical structure is controversial. This matter is discussed again in Section 10.8, where tests of linkage based on association will be discussed.

We turn now to stochastic behavior. One interesting class of problems concerns quantities that are not affected by geographical structure. For a finite population, Maruyama (1970, 1971, 1974) found two such quantities, at least for selectively neutral loci and certain genetic models. The first of these is the probability of fixation of a given allele, and the second is the mean total number of heterozygotes to appear as a result of a single new mutation. It then follows automatically from (8.6) that the mean time to fixation in the subdivided case is larger than that in the undivided case, and this was confirmed (Maruyama (1971)) by simulation. In the former case, on average, a smaller number of heterozygotes appears per generation, but for a greater number of generations, than in the latter case.

The eigenvalues in a genetic model involving geographical subdivision were given in (3.126), and the consequent effective population size was noted in (3.127). Except for very small migration rates, this effective size does not differ much from the actual population size. We might then be tempted to conclude that in this model the effect of subdivision is not important, and that for many purposes the population can be taken as one large random-mating population. Whether this view is correct or not is relevant to the evolutionary theory of Wright, depending as it does to some extent on population subdivision. However, in the model leading to (3.126), the migration is isotropic, and a much less extreme conclusion holds for structured populations where migration is most likely to occur to and from neighboring sub-populations.

Further eigenvalue questions have been discussed by Maruyama (1970, 1971, 1972), Nagylaki (1974b, 1976, 1977c), and Kimura and Maruyama (1971). Kimura and Maruyama also note one further important result: Even in the selectively neutral case clines of gene frequency can occur, the propensity for this depending on the subpopulation sizes and migration rates. Slatkin and Maruyama (1975) discuss the effect of stochastic gene frequency fluctuations on the slope of gene frequencies in a cline and show that the slope is decreased through such fluctuations.

A second form of stochastic fluctuation occurs in infinite populations where the migration rate and the gene frequencies of immigrants into any subpopulation are random variables. We do not discuss this case in here: Details of a model analyzing it were given by Nagylaki (1979).

A considerable literature also exists on the deterministic theory of geographically structured populations, originating with the remarkable pioneering paper of Fisher (1937). Of particular interest, in view of the theory

of Chapter 6, is the fact that association between frequencies of alleles at different loci can be generated by geographical subdivision, even in the absence of epistatic interactions between these loci. We consider here, however, an example of two questions specific to geographic subdivision, namely whether selection can maintain a cline of gene frequencies from west to east if A_1 is favored in the west and A_2 in the east, and second whether the frequency of A_1 can be sustained at positive values when A_1 is favored only in a finite interval of the entire east-west line. These questions have been considered in particular by Nagylaki (1975), and we follow his analysis of them closely.

Consider the line $[L, R]$, where possibly $L = -\infty$ or $R = +\infty$, and suppose at the point x on this line that the fitnesses of A_1A_1 , A_1A_2 and A_2A_2 are $1 + sg(x)$, $1 + hsg(x)$ and $1 - sg(x)$. (In this notation $h = 0$ corresponds to no dominance.) The small parameter s is a measure of the strength of selection. Each individual is assumed to migrate, from the time of birth to the time of reproduction, by a random amount y , where y has a normal distribution with mean 0, variance σ^2 . The migration distances of different individuals are assumed to be independent. The frequency $p = p(x; t)$ of A_1 at the point x at time t then satisfies the partial differential equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial x^2} + sg(x)p(1-p)\{1 + h - 2hp\}, \quad (8.7)$$

together with the boundary conditions $\partial p/\partial x = 0$ at $x = L$, $x = R$. This is a generalization of the formula of Fisher (1937). If a stationary cline exists there must be a solution of (8.7) with $\partial p/\partial t = 0$. In the important case $h = 0$, (8.7) shows that the equilibrium cline equation is

$$\frac{d^2 p}{dx^2} + \frac{2s}{\sigma^2} g(x)p(1-p) = 0. \quad (8.8)$$

The appropriate solution to this equation is found by using the boundary conditions $dp/dx = 0$ at $x = L$, $x = R$. This equation always has the trivial solutions $p(x) \equiv 0$, $p(x) \equiv 1$, and our aim is to find conditions for nontrivial solutions. The form of any solution will clearly depend on the numerical value of the parameter $2s/\sigma^2$ as well as the nature of $g(x)$. In the particular case where $L = 0$, $R = +\infty$, $s > 0$, and

$$g(x) = \begin{cases} 1, & 0 \leq x \leq a \\ -\alpha^2, & a < x < \infty \end{cases} \quad (8.9)$$

so that A_1 is favored when $x \leq a$ and A_2 when $x > a$, Nagylaki found that the necessary and sufficient condition for a unique nontrivial solution of (8.8) is

$$a\sqrt{2s} > \sigma \arctan \alpha. \quad (8.10)$$

Clearly the inequality (8.10) will apply for sufficiently large values of a and s and will not apply for sufficiently large σ , and, to a much smaller extent, α , since the right-hand side in (8.10) increases only from 0.78σ to 1.57σ as α increases from 1 to ∞ , but increases linearly with the migration distribution standard deviation σ . Clearly, while an inequality of the general form (8.10) is to be expected intuitively, the particular form of (8.10) is perhaps surprising, and indicates explicitly the relevance of the various parameters involved to the maintenance of A_1 .

This analysis can be taken over immediately to the case of a region, or “pocket”, in which A_1 is favored. By reflecting the interval $(0, a)$ about $x = 0$, we find that if

$$g(x) = \begin{cases} 1, & -a \leq x \leq a, \\ -\alpha^2, & |x| > a, \end{cases} \quad (8.11)$$

so that A_1 is favored in the “pocket” $(-a, a)$ but not elsewhere, the condition that A_1 can be maintained in the population is again (8.10).

Further analyses can be made for other functional forms for $g(x)$, but we do not pursue the details. Analyses of this sort are relevant to Wright’s theory of evolution and also to questions concerning theories of allopatric and sympatric models of speciation, as discussed for example by White (1978).

The above gives only a brief introduction to the complex theory surrounding geographically structured populations. Many further theoretical results are given by Nagylaki (1992), while Epperson (2003) provides a general discussion of the geographically structured case. These will be considered in more detail in Volume II.

8.5 Age Structure

So far all our analyses have ignored age structure. Perhaps curiously, the effect of age structure has been considered more frequently in the mathematical ecology literature than in the mathematical population genetics literature. To take account of age structure one must specify age-specific reproductive and survival schedules for all genotypes of both sexes, and must also make assumptions concerning the sex ratio and the mating process. While Norton (1928) and Haldane (1927) considered age-structured populations many years ago, it is only recently that further attention has been paid to them in any detail (see, for example, Demetrius (1971, 1974, 1975, 1976, 1977) and Charlesworth (1970, 1971, 1972, 1973, 1974)). An excellent summary of the topic is given by Charlesworth (1976).

Perhaps the most important aim, for age-structured populations, is to establish natural definitions of fitness that allow much of the classical theory to be applied. In this direction, Charlesworth (1976) found that, under

certain conditions, a natural definition of the fitness of any genotype exists, and that with this definition, natural selection leads to equilibria of the form (1.31) in the overdominance case and to dynamical equations of the form (1.26) if one allele is becoming fixed in the population. Similarly, Demetrius (1974) has given an analogue of the mean fitness increase theorem for age-structured populations. It thus appears likely that age structure does not introduce radically new behavior in populations compared to that expected from classical analyses. For this reason, perhaps inappropriately, we do not consider it in any further detail in this book.

8.6 Ecological Considerations

There is now a vast literature on the mathematical theory of ecological processes, including static and dynamical theories of the growth of a number of interacting populations. May (1975), (1976) and Pielou (1974) provide summaries of this literature. In particular the discrete-time Lotka–Volterra equation

$$N_i(t+1) = N_i(t)\left(1 + \alpha_i - \sum \beta_{ij}N_j(t)\right), \quad (8.12)$$

($i = 1, \dots, k$), modeling the dynamics of a community of k populations of respective sizes N_1, \dots, N_k , has been extensively analyzed. While a considerable verbal discussion exists in the literature on the relation between population genetics and ecology, rather less mathematical theory exists on this relationship. Thus the model (8.12) as it stands is free of genetic considerations.

However, if the parameters β_{ij} depend on the genetic constitutions of populations i and j , a description of the evolution in the model (8.12), concurrent with a description of the genetic evolution in the various populations, is possible in principle although no doubt normally difficult in practice. In this section we outline the analysis of Roughgarden (1976, 1977) of such a joint model: For analyses of related models see Fenchel and Christiansen (1977), Jayakar (1970) and Yu (1972).

Consider first the case of a single species in isolation and write (8.12) in the form

$$N(t+1) = N(t)\bar{w}, \quad (8.13)$$

where \bar{w} , the absolute mean fitness of the species, is a measure of the rate of increase in numbers of this species. Suppose that \bar{w} is determined by the alleles A_1 and A_2 at a given locus. If A_1 has frequency x , then from the elementary theory developed in Chapter 1,

$$\bar{w} = w_{11}x^2 + 2w_{12}x(1-x) + w_{22}(1-x)^2, \quad (8.14)$$

where w_{ij} is the absolute fitness of A_iA_j . The w_{ij} themselves are assumed to depend on the current size N of the population. Thus, quite apart from

the evolution of the population size N , there will be genetic evolution at the A locus determined by the standard equation (1.26). This equation, coupled with (8.13), then defines, for given functions $w_{ij}(N)$, the entire genetical-ecological evolution of the population. Anderson (1971) and more generally Roughgarden (1976) prove that this evolution is such that the equilibrium frequency x^* of A_1 is that producing the largest equilibrium population size and that also maximizes mean fitness. This is the first principle of genetical-ecological systems as enunciated by Roughgarden (1976).

Consider next a set of k co-evolving species. Here two different forms of behavior occur. The first of these arises when the fitnesses for any species are not directly functions of the allele frequencies in other species, although they may so indirectly by depending on the sizes of the other species, which in turn are determined by these frequencies. In the second case these fitnesses do depend directly on the allele frequencies in other species. The analysis of the second case is quite complex, although Roughgarden has been able to give explicit principles governing its evolutionary behavior. Here we concentrate on the first case, for which

$$\Delta N_i(t) = N_i(t)\{\bar{w}_i(N_1, \dots, N_k) - 1\}, \quad i = 1, \dots, k. \quad (8.15)$$

Here, as indicated, \bar{w}_i depends on $N_1 \dots N_k$, as well as on the frequency x_i of A_1 in species i , but not directly on allele frequencies for species other than species i . Define the gradient matrix $A = \{a_{ij}\}$ by

$$\begin{aligned} a_{ij} &= \partial(\Delta N_i)/\partial N_j \quad \text{at equilibrium} \\ a_{ij} &= N_i \partial \bar{w}_i / \partial N_j \quad \text{at equilibrium.} \end{aligned} \quad (8.16)$$

It is necessary to introduce the feedback F of the system, defined by

$$F = (-1)^{k+1} |A|, \quad (8.17)$$

where $|A|$ is the determinant of A . A sub-community of order $k-1$ may be defined by deleting species i from the system, and in this case the feedback in the sub-community is

$$F_i = (-1)^k |A_i|, \quad (8.18)$$

where A_i is obtained from A by striking out the i th row and i th column of A . For the equilibrium to be stable it is necessary that $F < 0$ and $\sum_i F_i < 0$.

Apart from the difference equation (8.15), there is also a second equation describing the genetic evolution in each population, namely

$$\Delta x_i = x_i(1-x_i)\{w_{11,i}x_i + w_{12,i}(1-2x_i) - w_{22,i}(1-x_i)\}/\bar{w}_i, \quad (8.19)$$

where

$$\bar{w}_i = x_i^2 w_{11,i} + 2x_i(1-x_i)w_{12,i} + (1-x_i)^2 w_{22,i}, \quad (8.20)$$

and the $w_{jl,i}$ depend, in a way we do not make explicit, on N_1, \dots, N_k . The joint ecological-genetical evolution of the system is now determined by

(8.15) and (8.18) and has the following important equilibrium properties as presented by Roughgarden (1976, 1977).

Principle 1. Suppose that, under the assumptions made above, there exists for any fixed x_1, x_2, \dots, x_k a unique locally stable equilibrium for the purely ecological model. Then an equilibrium point in the co-evolutionary model is locally stable if and only if \bar{w}_i is maximized locally with respect to x_i at that point.

This principle concerns gene frequencies. The second principle, stated below, concerns population numbers.

Principle 2. Under the above assumptions, the equilibrium size of species i is either maximized or minimized, at a stable equilibrium, at the equilibrium value of x_i . Maximization occurs if $F_i < 0$ and minimization if $F_i > 0$. If $F_i = 0$ then the equilibrium size of species i is not affected by genetic evolution in that species. Further, $F_i < 0$ for at least one species in that system. This last result follows immediately from the condition $\sum F_i < 0$ at a stable equilibrium.

We do not prove these remarkable principles and note only the dual optimality of both genetic and ecological parameters at stable equilibria. Roughgarden gives particular examples of the application of these principles, together with further generalizations, but we do not consider these less mathematical analyses here.

8.7 Sociobiology

Sociobiology is the study of the biological basis of social behavior. For this study to be meaningful it must be assumed that any behavior of interest has, at least in part, a genetical basis. Some behaviors, if they are genetically based, pose particular problems for evolutionary theory, the most outstanding example being that of altruism. Genes for altruism, if they exist, are at an immediate selective disadvantage in the population and should then be presumably lost from the population. Indeed Wilson (1975, p. 3) has claimed that the central theoretical problem of sociobiology is to explain how altruism can evolve by natural selection, assuming that there is a genetic basis for this character. For variations on this theme see Wilson (1977).

The sociobiological explanation for the existence of altruism, insofar as it is determined genetically, is through kin or group selection. While the behavior is disadvantageous to the individual exhibiting it, the altruistic act is sufficiently favorable to some small related or unrelated group so that the trait evolves by intergroup selection. In the *Origin of Species* Darwin also invoked intergroup selection arguments for similar traits. In this very

brief section we outline aspects of the theory on which this conclusion is based. For an excellent survey of group selection models, see Wade (1978).

The major quantitative construct necessary for kin selection arguments is some measure of the degree of relatedness between two individuals. Various measures are available for this, the most commonly used perhaps being the “coefficient of kinship”. This is defined as the probability that, for a given locus, a gene drawn at random from one individual is identical by descent to a gene drawn at random from the second individual. For any given degree of relatedness, this coefficient may be calculated by a standard path analysis method due to Wright (1921). The quantification for the kinship-based theory goes back to Haldane (1932a) and Fisher (1958, p. 178), and has been developed in detail by Hamilton (1964) through the concept of inclusive fitness. Under this theory it is claimed that the altruistic act is favored, for relatives of a given degree, if the number of relatives of this degree who survive as a result of it exceeds the reciprocal of the coefficient of kinship between them and the altruistic individual.

The relation between the inclusive fitness concept, altruism, multilocus evolutionary theory and evolution considered as an optimizing procedure is a controversial one. It has been discussed among others by Feldman and Cavalli-Sforza (1978, 1981), Grafen (1984), Hamilton (1996), Hammerstein (1996), Marrow et al. (1996) and Schwartz (2002). This topic will be discussed at length in Volume II.

It is interesting to observe that models for the evolution of altruism can be constructed, using group selection methods, where no concept of kin selection is invoked. Perhaps the most interesting of these models is that of Matessi and Jayakar (1976), which we now briefly describe. Consider an infinitely large population subdivided into finite groups of fixed size N . These groups are founded anew each generation in the following way. First, the entire population breeds at random in a common mating area and then splits up into groups of size N , the membership of each group being entirely random. Suppose within a given group there are i “altruists” and $N - i$ “non-altruists” and that the fitness of each altruist is then $\phi_A(i)$ and that of each non-altruist is $\phi_N(i)$. The altruism assumption is that

$$\phi_A(i) \leq \phi_N(i), \quad i = 1, 2, \dots, N - 1. \quad (8.21)$$

The mean fitness of a group containing i altruists is now

$$\phi(i) = N^{-1} \{ \phi_A(i) + (N - i)\phi_N(i) \}. \quad (8.22)$$

If it happens that $\phi(i + 1) > \phi(i)$, the existence of the altruists in the group favors the group as a whole, and if this advantage is sufficiently large compared to the disadvantage of altruists within each group, altruism will in some circumstances be favored. This argument is, at the moment, non-genetic but can be placed on a genetic basis by assuming certain genotypes for altruism.

The simplest possible forms for the functions $\phi_A(i)$, $\phi_N(i)$ are

$$\phi_A(i) = \beta_0 + \alpha_0 i, \quad \phi_N(i) = \beta_1 + \alpha_1 i. \quad (8.23)$$

The conditions (8.21) and $\phi(i+1) > \phi(i)$ jointly require

$$0 \leq \beta - (2N - 1)\alpha + N, \quad (10.54a)$$

$$-N\alpha \leq \beta \leq -\alpha + N, \quad (10.54b)$$

where $\alpha = (\alpha_1 - \alpha_0)/\alpha_1$, $\beta = (\beta_1 - \beta_0)/\alpha_1$.

For any given value of N , the conditions (10.54) define a set of α and β values within which, under this model, altruism can be expected to evolve. This set is a convex region in the (α, β) plane, and the smallest rectangle enclosing this set is

$$-N/(N - 1) \leq \alpha \leq N/(2N - 1), \quad 0 \leq \beta \leq N^2/(N - 1). \quad (8.55)$$

The area of the convex set relative to that of the rectangle defined by (8.55) defines a very crude measure of the “likelihood of altruism”. Despite the obvious limitations of this definition, it is interesting to observe that this measure decreases as a function of N from about 0.72 at $N = 2$ to 0.09 at $N = 20$. Clearly, even though kin selection is not involved, this form of altruism can arise in small, albeit temporarily formed, groups.

9

Molecular Population Genetics: Introduction

9.1 Introduction

In the preceding chapters of this book the basic genetic unit was taken as the gene, and the basic numerical quantity was the gene frequency. In particular, the fundamental unit step in evolution was taken as the replacement of one gene (more strictly, allele) by another in a population, and static genetic polymorphisms were usually described in terms of forces acting on gene frequencies. While in Chapters 6 and 7 the point was made that these polymorphisms are better viewed through forces acting on sets of genes at many loci, it remains true that no genetic unit finer than the gene has yet been considered in this book.

In this and the remaining chapters we consider the molecular population genetics theory arising from the recognition of the gene as a sequence of nucleotides. The task of placing population genetics theory on a molecular basis was begun by Kimura (1971); see also Nei (1975). To some extent the purely mathematical theory of the previous chapters carries through to the molecular level, with the nucleotide frequency replacing the gene frequency as the primary variable, but clearly, new models and viewpoints, as well as new “typical” values of various fundamental genetic parameters, are necessary at the molecular level.

Nucleotide sequences up to essentially the entire genome level are now available for many species. This chapter considers the theory relating to such sequences, since the theory relating to amino acid sequences is compli-

cated because of problems concerning the genetic code and its redundancy properties.

We now consider four points where the mathematical population genetics theory based on nucleotide frequencies differs from the classical theory based on gene frequencies. First, the molecular theory is dynamic, in contrast to the often static classical theory. Mutations are usually seen as leading to new allelic types rather than back to currently or previously existing types, since it is plausible that most nucleotide mutations will lead to sequences not currently existing in the population. Both the infinitely many alleles and the infinitely many sites models discussed in this chapter were originally proposed with this view in mind (Kimura and Crow (1964), Kimura (1969)). The dynamic nature of molecular population genetic models has been stressed in particular by Kimura (1971).

Secondly, because of extremely small intracistronic recombination rates, perhaps of order 10^{-5} or less, the assumption that the different sites within one gene evolve independently is particularly questionable. The mathematical theory of Chapters 6 and 7, there referring to genes and gametes rather than nucleotides and nucleotide sequences, shows that the evolution of tightly linked systems usually cannot be predicted from independent consideration of the separate loci (or, here, sites). Thus at the nucleotide sequence level various formulas in Chapters 3, 6 and 7 will be viewed differently than at the gametic level. For example, if (3.138) is used to compute fixation probabilities of two-locus gametes, the assumption $NR \gg 1$ may normally be made unless the loci are very close. In this case, the fixation probability (3.138) for gamete i becomes, essentially,

$$c_i(0) + \eta_i D(0). \quad (9.1)$$

This is just the product of the probabilities of fixation of the two alleles that make up gamete i , so that the fixation processes at the two loci are effectively independent. At the molecular level, on the other hand, it is possible that NR is small, since we might well be considering two nucleotides within the same gene, or cistron, and in this case the fixation probability for “gamete i ” is close to $c_i(0)$. Thus each two-site “gamete” evolves largely as a unit, and the fixation processes at the two sites are closely associated. Clearly the likely numerical values of the parameter R in the two cases affect the way in which (3.138) can be used to assess properties of concurrent nucleotide and gene fixation processes.

Thirdly, while the classical theory concerns the evolution of genes given labels “ A_1 ”, “ A_2 ”, etc., at the molecular level the actual genetic material is known, so that the symbols a , g , c , and t refer to specific rather than type entities. The fact that the theory thus concerns ultimate and real entities is of great importance, and further reference will be made to it in a moment. It also allows evolutionary inferences not closely associated with classical population genetics theory. For example, the considerable redundancy of the third nucleotide of a triplet in determining amino acids has been used by

Kimura (1977), Cornish-Bowden and Marson (1977), Barker et al., (1978), and Berger, (1978) for this purpose. We do not pursue these developments here.

Finally, and perhaps most important, molecular considerations often lead to *retrospective* rather than *prospective* evolutionary questions. The great work of Fisher, Haldane, and Wright was largely prospective: Given reasonable numerical values for various genetic parameters, they showed that evolution as a genetic process could and would occur. A hundred years ago such an undertaking was required. It is, however, no longer necessary to do this, and it now appears more useful to attempt to describe the course that evolution has taken by a retrospective analysis, and thus to gain empirical insight into evolutionary questions. This change of viewpoint has also led to the introduction of statistical methods for analyzing current genetical data, considered briefly in Chapters 10, 11, and 12. These matters will be discussed in greater detail in Volume II, taking up far more realistic cases than are considered in this book. The current emphasis on statistical inference procedures is perhaps the most important new direction in the theory in recent times. Knowledge of the actual genetic material is essential for these inferences, and the entire retrospective analysis must therefore be carried out in the framework of molecular population genetics.

9.2 Technical Comments

As stated above, two frequently used population genetic models have been inspired by the knowledge of the molecular structure of the gene, namely, the infinitely many alleles model and the infinitely many sites model. Population properties of these models are discussed in Sections 9.3 and 9.4 respectively. Our main interest is in properties of samples under both models, and these are discussed in Sections 9.5, 9.6, and 9.7. Various “time” and “age” properties are discussed in Section 9.9. Unless stated otherwise, selective neutrality, stationarity, and a constant population size are assumed throughout. The last assumption is clearly inappropriate for the human population, and models allowing for this expansion are a major topic of current research. So far as the assumption of neutrality is concerned, tests of this assumption are discussed in Chapter 11.

The sample size is denoted throughout by n (genes) and the population size by N . Since a diploid population is assumed, the number of genes in the population is $2N$. Because we often compare sample and population properties, we depart in this chapter from previous notation and write suffixes “ n ” and “ $2N$ ” when appropriate (for example, K_n and K_{2N}) to distinguish between sample and population quantities, respectively. (In later chapters, where sample properties only are discussed, we do not use a suffix.)

So far as the infinitely many sites model is concerned, because our main interest is in a collection of sites within a single gene, the assumption of no recombination between sites is the most appropriate one. This assumption is then made throughout this chapter when the infinitely many sites model is considered, except where otherwise stated. The theory for this case is largely due to the outstanding pioneer paper by Watterson (1975), to which we refer often. When there is no recombination between sites, the infinitely many sites model may also be viewed as an infinitely many alleles model, a connection that we explore below.

Much research currently centers around single nucleotide polymorphisms (SNPs). The infinitely many sites model is appropriate for the analysis of these. The classic definition of a polymorphism, given by Harris (1980, p. 331) in the context of protein polymorphism, is that a locus is polymorphic if the population frequency of the most frequent allele in the population of interest is no more than 0.99. However, this definition is, of course, arbitrary, and is not always implicit in published SNP polymorphism calculations, especially since observed SNP data refer to a sample, whereas the definition of polymorphism given above refers to a population. In this chapter we give calculations connecting sample data and the probability of population polymorphism.

Within each section, properties of the Wright–Fisher model, the nonoverlapping generations Cannings model, and the Moran model are discussed. Despite its importance, little work is available in the literature on the Cannings model, and it is assumed throughout that formulas for the nonoverlapping generation form of this model are close in form to those for the Wright–Fisher model, with an appropriate change in the definition of the parameter θ , which occurs in many formulas. For Wright–Fisher models the interpretation of this parameter is $\theta = 4Nu$, where u is the per-gene mutation rate, assumed to be small and in the diffusion approximation of order N^{-1} . When Wright–Fisher formulas are used for nonoverlapping generation Cannings models, the interpretation is $\theta = 4Nu/\sigma^2$, where σ^2 is defined in Section 3.3. These notational conventions are assumed throughout this chapter, without any further comment. For the Moran model the definition of θ is $2Nu/(1-u)$. The definition of u is straightforward in the infinitely many alleles model but less straightforward in the infinitely many sites model: The definition of u for this model is that u is the probability that there is at least one mutant nucleotide in any newborn in the DNA sequence under consideration.

Formulas for Wright–Fisher models, and thus for Cannings models, are all diffusion approximations, while those for Moran models are often exact. This has one important consequence. The diffusion process corresponding to the Wright–Fisher model, and the discrete Moran model process itself, are both time-reversible. This implies that many results found by going forward in time have an interesting interpretation going backward in time,

and conversely; one in effect gets two results for the price of one for these two processes.

Diffusion formulas concerning time or ages are usually given in diffusion time units. For the Wright–Fisher model, for example, the time unit is $2N$ generations. On some occasions it is convenient to convert diffusion times to generations, and on other occasions they are best left in diffusion time units.

Finally, we repeat a comment made above, that the theory considered here is introductory and does not consider complications due to variable population size, geographical subdivision, and so on. These complications must be taken into account in any significant data analysis, and will be addressed in detail in Volume II.

9.3 Infinitely Many Alleles Models: Population Properties

9.3.1 *The Wright–Fisher Model*

The neutral Wright–Fisher infinitely many alleles model was introduced and in part discussed in Sections 3.6 and 5.10. For example, in Section 5.7 a very close approximation for the monozygosity probability P_{mono} was found, and in Section 5.10 various other infinitely many alleles results, for example (5.123) and (5.124), were also described. In this section we discuss this model further. As noted in Section 9.2, the notation $\theta = 4Nu$ will be used, with the definition $\theta = 4Nu/\sigma^2$ applying when Wright–Fisher results are used for the Cannings model. Essentially all results given are diffusion approximations.

We first consider the number K_{2N} of alleles present in the population at any one time. If $K_{2N} = 1$, the population is monomorphic (see (5.66)), and it was shown in Section 5.7 that the expression in (5.82), which is the diffusion approximation (3.96) with the sample size n formally replaced by the population size $2N$, provides an excellent approximation for the probability of monomorphism. We have no right to expect this to occur, since the analysis of Section 3.6 assumes a sample size far less than the population size. We take up this point further in Chapter 10.

Of perhaps greater interest than the probability of monomorphism is the probability of population polymorphism as defined (by Harris) in Section 9.2. The calculations in (5.63) show that this probability is

$$\text{Probability of population polymorphism} = 1 - (0.01)^\theta. \quad (9.2)$$

If the Harris value 0.99 is replaced by the general value $1 - \delta$, for some small δ , then (9.2) is replaced by

$$\text{Probability of population polymorphism} = 1 - \delta^\theta. \quad (9.3)$$

We turn next to other properties of K_{2N} . No exact formula is known for the mean of K_{2N} , although the diffusion approximation to it is given in (3.92). More detailed information is provided by (3.93). Apart from this, little is known about the complete distribution of K_{2N} or the frequencies of the alleles present. Fortunately, our interest is mainly in properties of samples of genes from the population, rather than in the population itself, and here substantial information on sample numbers and frequencies is available, as discussed in Section 9.5.

Certain properties of the neutral infinitely many alleles model can be found immediately from the “two-allele” theory of Chapter 5. This is possible because, in the infinitely many alleles case, all alleles other than A_1 can be grouped simply as the “allele” “not- A_1 ”, and often this is sufficient to answer certain “infinitely many alleles” questions. Thus two-allele theory leads to the infinitely many alleles model probability (5.63), and several similar results are available.

For example suppose, in the Wright–Fisher two-allele model (3.16) with mutation from A_1 to A_2 at rate u , but with no reverse mutation, that the allele A_1 has a current frequency of unity. The mean time until A_1 is lost from the population can then be found immediately from (3.18) and (3.19), with $p = 1$. In the infinitely many alleles case, we can use these results by identifying A_1 with all the alleles initially in the population and A_2 with all new mutant alleles. From this we can find the mean time, in the infinitely many alleles model, until all the original alleles are lost from the population. The resulting mean time is, approximately,

$$4N \sum_{j=1}^{\infty} \{j(j + \theta - 1)\}^{-1} \text{ generations,} \quad (9.4)$$

as was given in (3.23). A slightly more accurate approximation is

$$4N \sum_{j=1}^{2N} \{j(j + \theta - 1)\}^{-1} \text{ generations.} \quad (9.5)$$

The case $\theta = 2$ is of some interest. For this value of θ the expression in (9.5) reduces to

$$4N - 2 \quad (9.6)$$

generations. This is identical to the conditional mean fixation time given in (5.36), which in turn is identical to the conditional mean loss time, given initially $2N - 1$ genes of the allele A_1 . The reason why the unconditional mean time in the mutation process and the conditional mean time in the nonmutation process are essentially identical for the case $\theta = 2$ can be seen from the fact that in the two corresponding diffusion processes, the drift and diffusion coefficients $a(x)$ and $b(x)$, given respectively in (5.61) and (4.58), are identical. Identical arguments show that the same mean time applies when there is a single initial A_1 gene when the condition is made,

in the no mutation case, that A_1 eventually fixes in the population. This mean time then has the interpretation as the mean time back to the most recent common ancestor gene of all genes in the current population, as we observe in the discussion surrounding (10.6).

We return to the expression (9.5) in Chapter 10, where it will be shown that the individual terms in (9.5) have an important interpretation regarding the past history of the population.

Returning to the case of a single allele A_1 with initial frequency 1, a calculation generalizing that leading to (9.5) can be made for selective models. Ohta (1974, 1976) has claimed that most gene fixation processes in evolution concern very slightly deleterious alleles. Consider then an infinitely many alleles model in which a given allele A_1 has initial frequency 1. We suppose that A_1A_1 individuals have fitness 1, that all A_1A_j heterozygotes have fitness $1 - s$, and that all other genotypes have fitness $1 - 2s$. The mean time until one or other deleterious allele fixes must exceed the mean time until loss of A_1 , and the latter mean time may be found immediately from two-allele theory using a generalization of (3.19) (see Ewens (1969c, equation (5.39)), Li and Nei (1977, equation 1). If $\alpha = |2Ns|$, this mean time is, in generations,

$$T(1) = 2N \int_0^1 t(x) dx, \quad (9.7)$$

where

$$t(x) = x^{-1}(1-x)^{\theta-1} \exp(2\alpha x) \int_0^x (1-y)^{-\theta} \exp(-2\alpha y) dy. \quad (9.8)$$

This mean time is calculated by Li and Nei (1977) for various (θ, α) combinations. As expected, it is extremely large even for moderate values of α , increasing (for $\theta = 1$) from $40N$ generations for $\alpha = 2.5$ to $5 \times 10^6 N$ generations for $\alpha = 10$. We conclude that the evolutionary role of these recurrent deleterious mutants is negligible if α is 5 or more.

Much interest now centers on retrospective properties of this and other models, as well as on “age” properties of the alleles in a population. These are discussed in Section 9.9.

9.3.2 The Moran Model

The Moran infinitely many alleles model was introduced in Section 3.6.4. In this section we consider further results for this model, focusing on the “age” and “time” results of interest in this chapter. As noted above, the definition of the parameter θ for the Moran model is $\theta = 2Nu/(1-u)$, and this definition applies throughout this section for this model.

If the population is monomorphic we say that the single allele present in the population is “quasi-fixed”. We do not use the expression “fixed”, since in an infinitely many alleles model this allele will eventually be lost from the population. Kelly (1976) has shown, for the Moran model, that the probability that a new mutant allele becomes quasi-fixed in the population is C^{-1} , where

$$C = \sum_{j=0}^{2N-1} \binom{2N-1+\theta}{j} \left(\binom{2N-1}{j} \right)^{-1}. \tag{9.9}$$

We now consider the mean number of birth and death events until all alleles present in the population at any time are lost. The value given in (9.5) for this mean is a diffusion approximation, applying for the Wright–Fisher model. In the case of the Moran model an exact calculation can be made by using the results of Section 3.4, regarding all the alleles in the population as A_1 and with initially $2N$ genes of this allelic type in the population. Watterson (1976a) found that the required mean number of birth and death events is

$$2N(2N + \theta)(\theta - 1)^{-1} \sum_{j=1}^{2N} j^{-1} \left(1 - \binom{2N}{j} \binom{2N + \theta - 1}{j}^{-1} \right). \tag{9.10}$$

(A formula different from (9.10), found by applying l’Hôpital’s rule, applies for the case $\theta = 1$.) In the case $\theta = 2$, the expression (9.10) gives, exactly, $8N^2(N + 1)/(2N + 1)$, or about $4N^2$, birth and death events. This can be thought of as corresponding to $4N$ “generations”, which appears to agree closely with the Wright–Fisher approximation in (9.6). This agreement is, however, misleading, since the definitions of θ differ in the two models.

We make several further comments about (9.10). First, as with the corresponding result for the Wright–Fisher model, we may think of (9.10) as providing, in this case exactly, the mean age of the oldest allele in the population. Second, the typical (j th) term in (9.10) is the mean number of birth and death events for which there are exactly j genes present of the various original alleles in the population before the eventual loss of all these alleles. Thus the expression (9.10) gives more information than might otherwise be thought.

Third, although the identity is not immediately obvious, the expression in (9.10) is identical to the expression

$$2N(2N + \theta) \sum_{j=1}^{2N} \frac{1}{j(j + \theta - 1)}. \tag{9.11}$$

We shall see in Chapter 10 that the individual terms in the sum also have an important interpretation, in this case concerning the past history of the population rather than its future evolution.

The expression in (9.11) may be written equivalently as

$$\sum_{j=1}^{2N} \frac{1}{v_j + w_j}, \quad (9.12)$$

where

$$v_j = \frac{ju}{2N}, \quad w_j = \frac{j(j-1)(1-u)}{(2N)^2}. \quad (9.13)$$

In Chapter 10 we shall explain why the mean age of the oldest allele can be expressed in the form defined by (9.12) and (9.13).

Finally, the Moran population model configuration process is reversible (Kelly, 1976). In other words, if we write down the possible states $E_1, E_2, \dots, E_{p(2N)}$ of the configuration process and the transition probabilities between them, (2.164) holds true when we interpret ϕ_i as the stationary probability of the i th configuration. Thus the prospective and retrospective behaviors of the process are identical, a fact we shall take advantage of later in discussing the past history of the population. The time reversibility property is what allows an interpretation for the terms in the sum in (9.11) in relation to the past history of the population.

The exact frequency spectrum (3.102) provides, almost immediately, two results of interest in this chapter. The first uses the concept of size-biased sampling, discussed in more detail Section 9.9. In the Moran model the probability that an individual drawn at random is of an allelic type having exactly j copies in the population is found by multiplying the j th term in (3.102) by $j/(2N)$. This gives a value of

$$\theta(2N)^{-1} \left(\binom{2N}{j} \binom{2N + \theta - 1}{j}^{-1} \right) \quad (9.14)$$

for this probability. This calculation will be of use later when we consider “age” properties of the alleles in the population.

Second, (3.102) allows an exact calculation of the probability of population polymorphism, as defined in Section 9.2. Any allele having a frequency exceeding 0.99 must be the most frequent allele in the population, and at most one allele can have such a frequency. Thus the probability that the most frequent allele in the population has frequency exceeding 0.99 is the mean number of alleles with frequency exceeding 0.99. Taking $0.99(2N)$ as an integer M , (3.102) shows that the probability of polymorphism is

$$1 - \theta \sum_{j=M+1}^{2N} j^{-1} \left(\binom{2N}{j} \binom{2N + \theta - 1}{j}^{-1} \right). \quad (9.15)$$

This is close to $1 - (0.01)^\theta$, the approximate value found above for the Wright–Fisher model using a diffusion approximation. As with other such

calculations, this apparent similarity is misleading because of the different definitions of θ in the two models.

Many further exact and elegant results can be found for the Moran model, but since our main interest is in samples of genes from a population rather than the entire population itself, we do not consider these further.

9.4 Infinitely Many Sites Models: Population Properties

9.4.1 Introduction

We turn now to the infinitely many sites model. This was in effect introduced by Kimura (1969) but only named as such by him later (Kimura, (1971)). In this model a gene is thought of as a long sequence of nucleotides and an allele refers to some specific such sequence, so that different allelic types are just different sequences. In using the words “gene” and “allele” when referring to this model we imply these definitions.

In this model a mutation is simply the change of one nucleotide type to another, and it is assumed in the model that any mutation arises at a site currently monomorphic in the population. (The concept of “infinitely many sites” is intended to formalize this assumption.) To compare infinitely many sites formulas with infinitely many alleles formulas we define a newborn to be a mutant if there is at least one mutant site in the newborn, and write the probability of this event as u .

Various formulas depend only on the mean number of mutant sites in the newborn, but other formulas depend on the complete distribution of the number of mutant sites. In his pioneering paper, Watterson (1975) assumed that the number of mutant sites in any newborn has a Poisson distribution with parameter ν . Following the definition of a mutant for the infinitely many sites model given in Section 9.2, the probability that a newborn is a mutant for this Poisson case is $u = 1 - e^{-\nu}$. Here we introduce the general model for which

$$\text{Prob}(j \text{ mutant sites in any newborn}) = q_j. \quad (9.16)$$

For this model the probability u that a newborn gene is a mutant is $1 - q_0$.

We follow the Watterson notation and denote the mean number $\sum j q_j$ of mutant sites in any newborn by ν . Since “sites” mutation rates are very small, u and ν differ only by small-order terms, and in the diffusion approximation u and ν may be used interchangeably. We repeat two comments made above, first that Wright–Fisher model results are diffusion approximations and Moran model results are often exact, and second, that unless otherwise stated, it is assumed that there is no recombination between sites.

9.4.2 The Wright–Fisher Model

In any infinitely many sites model the mean number of new sites at which segregation starts in the population in each generation is $2N\nu$. For the Wright–Fisher and Cannings models, where we use diffusion approximations, we replace this by $2Nu$. At stationarity the new alleles created by mutation will be balanced by an equal number of alleles that, because of random sampling and mutation, become lost from the population.

For this model the analogy of the formal mathematics of the segregation process at each site to that of the segregation process at each locus in classical genetics is particularly interesting. For example, since for the moment we assume selective neutrality, we can use the neutral Wright–Fisher model (1.48) to describe the segregation process at each site, since under our assumptions at most two nucleotide types are possible in the population at any site at any time. From (5.18), the mean number of generations for which the mutant nucleotide assumes the value j is $2/j$. Since on average $2Nu$ sites begin segregating in each generation, the mean number of sites at stationarity at which, at any time, there are j representatives of the mutant nucleotide is

$$\frac{4Nu}{j} = \frac{\theta}{j}, \quad j = 1, 2, \dots, 2N - 1. \quad (9.17)$$

Passing to a continuous approximation, we see that the mean number of generations for which the population frequency of the mutant nucleotide assumes a value in $(x, x + \delta x)$ is $2x^{-1}\delta x$, $((2N)^{-1} \leq x \leq 1)$. Thus at any time the mean number of sites $\phi(x)\delta x$ at which the mutant nucleotide assumes a value in $(x, x + \delta x)$ is

$$\phi(x)\delta x = \frac{4Nu}{x}\delta x = \frac{\theta}{x}\delta x. \quad (9.18)$$

We may call (9.17) in the discrete case, or (9.18) in the continuous case, the population frequency spectrum of the process. The stationary mean of the number S_{2N} of sites segregating in the population at any time may be found by integrating the function $1 - x^{2N} - (1-x)^{2N}$ with respect to the expression θx^{-1} given in (9.18) over the interval $((2N)^{-1}, 1)$. This calculation yields the value $\theta \log(2N)$; a more accurate expression (see (5.77)) is

$$E(S_{2N}) = \theta \log(2N) + 0.6775\theta. \quad (9.19)$$

This mean value applies for any recombination structure between sites. However, further properties of the distribution of S_{2N} depend on this recombination structure. If we were to assume unlinked sites and that the segregation processes at the various sites are independent, the number of segregating sites in the population would be, to a close approximation, a Poisson random variable with mean, and also variance, as given in (9.19). This would imply that the probability P_{mono} of no segregation at any site

would be

$$(2N)^{-\theta} e^{-0.6775\theta}. \quad (9.20)$$

Clearly, however, the independence assumption is quite unjustified for the case we consider, that is, where there is no recombination between sites. For the no recombination case, the infinitely many sites model reduces to the infinitely many alleles model. For this model the expression (5.82) gives a close approximation to P_{mono} . The ratio of the probabilities in (5.82) and (9.20) is

$$e^{\gamma\theta} \Gamma(1 + \theta), \quad (9.21)$$

where γ is Euler's constant $0.577216\dots$. This ratio differs from 1 by terms of order θ^2 when θ is small.

Watterson (1975) derived further properties of S_{2N} under the assumption of no recombination between sites. Most of these relate to samples and are thus discussed in Section 9.6. For the entire population he found that the variance of S_{2N} is approximately

$$\text{var}(S_{2N}) = E(S_{2N}) + \frac{\pi^2\theta^2}{6}, \quad (9.22)$$

and that for large N , the complete distribution of S_{2N} is approximately Poisson with mean (9.19). The variance (9.22) differs from that arising if independence between sites is assumed, namely $E(S_{2N})$, by terms of order θ^2 . This remark and the observation following (9.21) suggest that when θ is small, properties of the “no recombination” and the “free recombination” models are quite close. For general values of θ and with a small recombination fraction between adjacent sites, we expect the variance of S_{2N} to lie between $E(S_{2N})$ and the value given in (9.22). If this is so, the value of the “free recombination” variance calculation is that it provides a lower bound to the variance of S_{2N} when some recombination between sites is allowed.

Provided that the same fitness structure holds at all sites and that the stochastic processes at the various sites are assumed independent, the calculations leading to (9.19) can be generalized to take selection into account. Perhaps the most interesting selective scheme is that in which the mutant nucleotide is at a slight selective disadvantage s (< 0) with respect to the prevailing type, with no dominance. Here (5.48) shows that the right-hand side in (9.18) should be replaced by

$$\theta\{x(1-x)\}^{-1}\{e^{\alpha(1-x)} - 1\}\{e^\alpha - 1\}^{-1}\delta x, \quad (9.23)$$

where $\alpha = |4Ns|$. In principle this allows a calculation of the mean number of segregating sites in the population through an evaluation of the integral

$$\int_{(2N)^{-1}}^1 \theta\{x(1-x)\}^{-1}\{e^{\alpha(1-x)} - 1\}\{e^\alpha - 1\}^{-1} dx,$$

but unfortunately, no simple explicit form for this integral exists.

9.4.3 The Moran Model

As for the infinitely many alleles case, the Moran model admits several exact formulae in the infinitely many sites model. The first such formula that we discuss is that for the mean of the number S_{2N} of segregating sites. An ergodic argument similar to that leading to (3.102) shows that this is ν times the mean time that segregation continues at any site. This mean time is given exactly by (3.51) with $i = 1$, and this leads to the exact expression

$$\text{mean of } S_{2N} = 2N\nu \left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{2N-1} \right). \quad (9.24)$$

Equation (9.24) contains more information than initially appears, since the mean number of segregating sites for which the mutant nucleotide is represented exactly j times in the population is $2N\nu j^{-1}$. This is the exact Moran model analogue of the Wright–Fisher approximation in (9.17).

We now consider the exact “monomorphism” probability that $S_{2N} = 0$. This depends on the nature of the input mutation process assumed. Waterson (1975) found that in the Poisson mutation case, the monomorphism probability is

$$\frac{(2N-1)!}{(1+\theta)(2+\theta)\cdots(2N-1+\theta)}, \quad (9.25)$$

where θ is defined by

$$\theta = 2N(e^\nu - 1). \quad (9.26)$$

This definition is in line with the general definition of $\theta = 2Nu/(1-u)$ for the Moran model since, in the Poisson case, $u = 1 - e^{-\nu}$, and using this value for u in the general definition of θ we recover the expression in (9.26).

For the general mutation model (9.16) it can be shown that (9.25) holds if θ is defined by

$$\theta = 2N \left(\frac{1 - q_0}{q_0} \right) = \frac{2Nu}{1-u}. \quad (9.27)$$

In the infinitely many sites model without recombination, the event that $S_{2N} = 0$ is identical to the event that there is only one allele in the population, and it is therefore not surprising that the expression in (9.25) is identical to the infinitely many alleles expression (3.99) under the definition we assume for a mutant gene in the infinitely many sites model.

The fact that the coefficient $2N\nu$ on the right-hand side in (9.24) is in general different from θ leads to the question of when $\theta = 2N\nu$. Equation (9.27) shows that these two quantities are equal when

$$\nu = \frac{1 - q_0}{q_0}.$$

This will occur if the number of sites at which a mutation occurs in any newborn has the geometric distribution

$$\text{Prob}(j \text{ mutant sites in newborn}) = (1 - u)u^j, \quad j = 0, 1, 2, \dots \quad (9.28)$$

Although this is less natural than the Poisson distribution for the input mutation process, we shall see later that the model (9.28) has interesting exact properties for the Moran infinitely many sites model.

9.5 Sample Properties of Infinitely Many Alleles Models

9.5.1 Introduction

Any data concerning the genetic composition of a population derives from a sample from that population. Since these samples are central to the concept of the coalescent (Chapter 10), are used to test for selective neutrality (Chapter 11), to answer questions of interest to human geneticists (Section 10.8), and to investigate the phylogenetic relation between a number of species (Chapter 12), it is appropriate to consider sampling properties in detail. In this section we do this for various infinitely many alleles models. Parallel results for infinitely many sites models are given in Section 9.6. The infinitely many sites model is a natural one to consider in detail, since it is used to describe the stochastic behavior of DNA sequences. On the other hand, the infinitely many alleles model is sometimes used to model the evolutionary behavior of haplotypes, a topic of much current interest. (However, see the discussion in Section 9.7 concerning pitfalls in this procedure.)

Throughout this chapter, the number of genes in the sample is denoted throughout by n , and it is assumed that n is far smaller than the number of genes ($2N$) in the (diploid) population of size N .

9.5.2 The Wright–Fisher Model

We consider first the Wright–Fisher infinitely many alleles model. The diffusion approximation properties of a sample of n genes under this model are best summarized through the partition formula (3.83). This leads to the distribution of the number K_n of different allelic types observed in the sample as given in (3.84) and thus to the mean of K_n as given by (3.85). While (3.83) and thus (3.85) were found in Section 3.5 by using recurrence relations, the mean (3.85) can be found directly (see (3.94)) from the

frequency spectrum $\phi(x)$ in (3.95), using the calculation

$$E(K_n) = \int_0^1 \{1 - (1-x)^n\} \phi(x) dx. \quad (9.29)$$

We return to this comment below when considering time-dependent properties of this model.

There is currently much interest in estimating the parameter θ . Equations (3.83) and (3.84) show jointly that the conditional distribution of the vector $\mathbf{A} = (A_1, A_2, \dots, A_n)$ defined before (3.83), given the value of K_n , is

$$\text{Prob}\{\mathbf{A} = \mathbf{a} | K_n = k\} = \frac{n!}{|S_n^k| 1^{a_1} 2^{a_2} \dots n^{a_n} a_1! a_2! \dots a_n!}, \quad (9.30)$$

where $\mathbf{a} = (a_1, a_2, \dots, a_n)$.

Equation (9.30) implies that K_n is a *sufficient statistic* for θ . Standard statistical theory then shows that once the observed value k_n of K_n is given, no further information about θ is provided by the various a_j values, so that all inferences about θ should be carried out using the observed value k_n of K_n only. This includes estimation of θ or of any function of θ .

Since K_n is a sufficient statistic for θ we can use the probability distribution in (3.84) directly to find the maximum likelihood estimator $\hat{\theta}_K$ of θ . It is found that this estimator is the implicit solution of the equation

$$K_n = \frac{\hat{\theta}_K}{\hat{\theta}_K} + \frac{\hat{\theta}_K}{\hat{\theta}_K + 1} + \frac{\hat{\theta}_K}{\hat{\theta}_K + 2} + \dots + \frac{\hat{\theta}_K}{\hat{\theta}_K + n - 1}. \quad (9.31)$$

Given the observed value k_n of K_n , the corresponding maximum likelihood estimate $\hat{\theta}_k$ of θ is found by solving the equation

$$k_n = \frac{\hat{\theta}_k}{\hat{\theta}_k} + \frac{\hat{\theta}_k}{\hat{\theta}_k + 1} + \frac{\hat{\theta}_k}{\hat{\theta}_k + 2} + \dots + \frac{\hat{\theta}_k}{\hat{\theta}_k + n - 1}. \quad (9.32)$$

Numerical calculation of the estimate $\hat{\theta}_k$ using (9.32) is usually necessary.

The estimator implied by (9.31) is biased, and it is easy to show that there can be no unbiased estimator of θ . On the other hand, there exists an unbiased estimator of the population homozygosity probability $1/(1+\theta)$. If this estimator is denoted by $g(K_n)$, (3.84) shows that

$$\sum_{k=1}^n \frac{|S_n^k| \theta^k g(k)}{S_n(\theta)} = \frac{1}{1+\theta},$$

where $|S_n^k|$ is the absolute value of a Stirling number, defined below (3.84). From this, we see that

$$\sum_{k=1}^n |S_n^k| \theta^k g(k) = \theta(\theta+2)(\theta+3) \dots (\theta+n-1).$$

Since this is an identity for all θ , the expression for $g(k)$ for any observed value k_n of K_n can be found by comparing the coefficients of θ^k on both sides of this equation. In particular, when $k_n = 2$,

$$g(2) = \frac{\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}}{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}}. \tag{9.33}$$

Unbiased estimation of $1/(1 + \theta)$ for values of k_n larger than 2 is complicated, and it is then probably more convenient to use instead the estimator $(1 + \hat{\theta}_K)^{-1}$, where $\hat{\theta}_K$ is found from (9.31), even though this estimator is slightly biased.

It is sometimes preferred to estimate $(1 + \theta)^{-1}$ by f , defined in the notation of (3.88) by

$$f = \sum_i \frac{n_i^2}{n^2}. \tag{9.34}$$

This is a poor estimate in that it uses precisely that part of the data that is *least* informative about $(1 + \theta)^{-1}$. The estimate of θ derived from f , namely

$$\hat{\theta}_f = f^{-1} - 1, \tag{9.35}$$

has been shown (Ewens and Gillespie (1974)) to be strongly biased and to have mean square error approximately six or eight times larger than that of $\hat{\theta}$.

More generally, the only functions of θ allowing unbiased estimation are linear combinations of functions of the form

$$\{(a + \theta)(b + \theta) \cdots (c + \theta)\}^{-1}, \tag{9.36}$$

where a, b, \dots, c are integers with $1 \leq a < b < \cdots < c \leq n - 1$ (Ewens (1972)). While this fact derives mathematically from the form of the probability distribution (3.84), an argument in support of it, from an empirical sampling point of view, is as follows.

Suppose, for example, that $k_n = 2$ and write the unordered numbers of genes of the two alleles observed as N_1 and $n - N_1$. The probability distribution of the pair $(N_1, n - N_1)$ is identical to that of N_1 , and (9.30) shows that this is

$$\text{Prob}(N_1 = n_1) = \frac{n!}{|S_n^2| n_1 (n - n_1)}. \tag{9.37}$$

Given the observed values n_1 and $n - n_1$, the probability that two genes taken at random are of the same allelic type is

$$\frac{\binom{n_1}{2} + \binom{n - n_1}{2}}{\binom{n}{2}}.$$

Multiplying this expression by the right-hand side in (9.37) and summing over all possible values of n_1 gives the estimator (9.33). A similar argument

can be used to justify the fact that any function of the form in (9.36) admits unbiased estimation.

We now consider an approximation for the mean square error (MSE) of the estimator $\hat{\theta}_K$ as defined by (9.31). Writing the right-hand side of (9.31) as $\psi(\hat{\theta}_K)$, we have $K_n = \psi(\hat{\theta}_K)$ and also, from (3.85), $E(K_n) = \psi(\theta)$. Thus by subtraction,

$$K_n - E(K_n) = \psi(\hat{\theta}_K) - \psi(\theta).$$

A first-order Taylor series approximation for the right-hand side is $(\hat{\theta}_K - \theta)\psi'(\theta)$, so that

$$K_n - E(K_n) \approx (\hat{\theta}_K - \theta)\psi'(\theta).$$

Squaring and taking expectations, we get

$$\text{MSE}(\hat{\theta}_K) \approx \frac{\text{var}(K_n)}{\psi'(\theta)^2}. \quad (9.38)$$

The variance of K_n is given in (3.86), and it is immediate that

$$\psi'(\theta) = \sum_{j=1}^{n-1} \frac{j}{(\theta + j)^2}. \quad (9.39)$$

This leads to

$$\text{MSE}(\hat{\theta}_K) \approx \frac{\theta}{\sum_{j=1}^{n-1} \frac{j}{(j+\theta)^2}}. \quad (9.40)$$

The approximation (9.40) appears to be quite accurate, and we use it in Section 9.6 in comparing estimation of θ in the infinitely many alleles and the infinitely many sites models.

Griffiths (1979a,b) has found many time-dependent properties of the number and frequencies of alleles observed in a sample of n genes. These, of course, depend on the initial population frequencies chosen as well as on the mutation rate. At one extreme one can assume that initially only one allelic type exists in the population, and at the other extreme that $2N$ allelic types exist in the population. Many of these properties are found using the time-dependent frequency spectrum $\phi_t(x)$, which has the form

$$\phi_t(x) = \theta x^{-1}(1-x)^{\theta-1} \left(1 + \sum_{i=2}^{\infty} \lambda_i(t) \psi_i(x, \theta) g_i(p_1, p_2, \dots) \right). \quad (9.41)$$

In this expression the $\lambda_i(t)$ are eigenvalues whose values are given below, $\psi_i(x, \theta)$ is a function only of x, i , and θ , and $g_i(p_1, p_2, \dots)$ is a complicated function of the initial allelic frequencies p_1, p_2, \dots . The rate of convergence of this frequency spectrum to the stationary spectrum $\theta x^{-1}(1-x)^{\theta-1}$ depends on the eigenvalues $\lambda_j(t)$, which are given by

$$\lambda_i(t) = \exp\left(-\frac{1}{2}j(j-1+\theta)t\right), \quad j = 2, 3, 4, \dots, \quad (9.42)$$

		<i>t</i>						
		0.2		0.5		1.0		∞
		(i)	(ii)	(i)	(ii)	(i)	(ii)	(i) and (ii)
θ	0.1	1.31	10.12	1.40	4.62	1.47	2.77	1.57
	1.0	4.03	12.39	4.89	7.64	5.49	6.34	5.88
	1.5	5.51	13.62	6.74	9.25	7.54	8.18	7.90

Table 9.1. Mean number of alleles observed in a sample of 200 genes for various θ , t values. Unit time = $2N$ generations. Case (i): one initial allele. Case (ii): many initial alleles of equal frequency. From Griffiths (1979b).

and in particular on the largest eigenvalue $\exp\{-(1 + \theta)t\}$. These eigenvalues are the limiting values of the discrete configuration values given in (3.90) in the limit $N \rightarrow \infty$, $u \rightarrow 0$, with $4Nu = \theta$ held fixed.

The mean number of alleles in a sample of n genes can be found, following the same argument as that leading to (9.29), by evaluation of

$$\int_0^1 \{1 - (1 - x)^n\} \phi_t(x) dx. \quad (9.43)$$

An explicit expression for this mean is given by Griffiths (1979b, (2.10)), who also provides numerical calculations for various r , θ , t , and p_j values. We reproduce some representative calculations in Table 9.1 for two cases, first where there exists initially a single allele in the population and second where there exist initially many alleles of equal frequency. We observe that in the former case, the approach to the equilibrium point appears rather more rapid than in the latter.

Griffiths also found properties of two samples, one in each of two subpopulations, which split apart some time in the past. In particular, he found a formula for the mean number of alleles common to the two samples at time t after the split and the joint probability distributions of the sample frequencies of these alleles.

In Chapter 11 we shall consider various tests of the hypothesis of selective neutrality. These tests often reduce to a comparison of properties of the number of alleles, or of segregating sites, in a sample to some measure of population homozygosity (or, equivalently, heterozygosity). Unfortunately, the properties of the two measures under selection are often similar to their properties in a selectively neutral case in which the population has recently expanded in size after going through a bottleneck, or at the end of a selectively induced replacement process at a locus closely linked to the neutral locus (a “selective sweep”, discussed in Section 6.7). Thus these tests of selection can be rendered invalid at times closely following such historical events.

Table 9.1 can be used to find various properties of the number of alleles in a sample following a bottleneck or a selective sweep, since we might assume,

to a close approximation, that only one allele survives a tight bottleneck or a selective sweep. It shows, for example, that when $\theta = 1$, the mean number of alleles in a sample of 200 genes is 4.89 when N generations have passed after the bottleneck or selective sweep, which is about 83% of its stationary mean value 5.88.

The properties of the sample homozygosity should be close to those of the population homozygosity. We take 0 to be the time of the bottleneck and the population homozygosity at this time to be 1. With the mean homozygosity at time t diffusion time units denoted by $F^{(t)}$, (3.89) shows that

$$F^{(t)} = \frac{1}{1 + \theta} + \frac{\theta}{1 + \theta} \exp^{-(1+\theta)t}. \quad (9.44)$$

Thus $F^{(t)}$ depends only on the leading eigenvalue in the set (9.42) whereas the mean number of alleles depends on all the eigenvalues. When $\theta = 1$ the value of $F^{(t)}$ arising N generations after the bottleneck is 0.684, so that the mean heterozygosity at this time is 0.361. This is about 63% of its stationary value. The comparison of this with the corresponding value for the mean number of alleles in the sample is then relevant to the effect of a bottleneck on a test for selective neutrality conducted N generations after the bottleneck or the selective sweep.

9.5.3 The Moran Model

Many exact properties of a sample of genes can be found immediately from the sample configuration given in (3.83), since under the Moran infinitely many alleles model, (3.83) holds exactly if θ is defined by (3.98). This is in contrast to the situation for the Wright–Fisher model, where (3.83) is only an approximation. Thus with the Moran model definition of θ , (3.84), (3.85), (3.86), (3.87), and (3.88) are all exact, as is also the conditional distribution formula (9.30) that derives from them. It is interesting to ask why these formulas hold exactly in the Moran model, not only in a sample but also in the population, and also why sample formulas and population formulae are identical, with the replacement of n for $2N$. In Chapter 10 we shall see why these two properties of the Moran model hold.

Since for some simulation purposes it is necessary to derive a sample of genes that have the allelic partition formula (3.83), it is interesting to ask how such a sample may be generated efficiently. Perhaps the most interesting method to use is Hoppe's urn (Hoppe, (1984), (1987), Watterson (1984)). We imagine an urn containing one black ball of mass θ and a collection of balls of various colors, each of mass 1. Initially, the urn contains only the black ball. A ball is drawn at random from the urn with probability proportional to its mass. If it is the black ball, the black ball is replaced in the urn together with a new ball of a color not currently existing in the urn. If it is a colored ball, the ball drawn is replaced together with a new

ball of the same color as that drawn. The initial ball drawn must, of course, be the black ball. Thus if there are $j - 1$ colored balls in the urn at any one time, the probability that the next ball to enter the urn is of a new color, that is, the probability that the next ball drawn is the black ball is

$$\frac{\theta}{j - 1 + \theta}, \quad (9.45)$$

independently of the color composition of the $j - 1$ colored balls. The process stops when there are n colored balls in the urn, and the “color” partition formula for these n balls is given exactly by (3.83). This “urn” procedure allows rapid simulation of random variables having the distribution (3.83).

We can think of the Hoppe urn procedure as sampling “through space”, but we shall find in Chapter 10 that the procedure has an important interpretation as sampling “through time”.

A concept closely linked to Hoppe’s urn is that of a partition structure (Kingman, (1978)). There should be no particular significance attached to the sample size n , and we can regard a sample of size n genes as one arising from a sample of size $n + 1$, one of which was accidentally lost. We reasonably require a consistency of formulas for the two sample sizes. To formalize this we denote the left-hand side in the partition formula (3.83) by $P_n(a_1, a_2, \dots)$. The method of arriving at a sample of n genes as just described then implies that this must be equal to

$$\frac{a_1 + 1}{n + 1} P_{n+1}(a_1 + 1, a_2, \dots) + \sum_{j=2}^{n+1} \frac{j(a_j + 1)}{n + 1} P_{n+1}(a_1, \dots, a_{j-1} - 1, a_j + 1, \dots). \quad (9.46)$$

The right-hand side in (3.83) does satisfy this requirement, but Kingman raised the following much more general question: How may one characterize probability structures satisfying (9.46)? He called structures having this property “partition structures”, and showed that for all such structures of interest in genetics, $P_n(a_1, a_2, \dots)$ could be represented in the form

$$P_n(a_1, a_2, \dots) = \int P_n(a_1, a_2, \dots | \mathbf{x}) \mu(d\mathbf{x}), \quad (9.47)$$

where $P_n(a_1, a_2, \dots | \mathbf{x})$ is a complicated sum of multinomial probabilities whose exact form we do not write down. Kingman called μ the “representing measure” of $P_n(a_1, a_2, \dots)$ and found that for the partition formula (3.83), this representing measure is the Poisson–Dirichlet distribution, introduced in Section 5.10.

The consistency requirement (9.46) is a natural one for a sample of genes. We shall, however, find a perhaps more important interpretation for this requirement when considering, in Chapter 10, the past history of the population from which the sample was taken.

Kingman also took up the question of “noninterference”, defined by the requirement that if a gene is taken at random from the sample, and all r

genes of its allelic type are then removed from the sample, the partition probability structure of the remaining $n - r$ genes should be the same as that of an original sample of $n - r$ genes. Noninterference implies that $P_n(a_1, \dots, a_r, \dots)$ must satisfy the requirement

$$\frac{ra_r}{n} P_n(a_1, \dots, a_r, \dots) = c(n, r) P_{n-r}(a_1, \dots, a_{r-1}, \dots), \quad (9.48)$$

where $c(n, r)$ does not depend on a_1, a_2, \dots . Kingman then showed that of all partition structures of interest in genetics, the only one also satisfying the requirement (9.48) is (3.83).

These various results, including those relating to the Hoppe urn process, which might initially seem to be of purely mathematical interest, will appear to have a natural and important practical interpretation when we consider the coalescent process in Chapter 10.

9.6 Sample Properties of Infinitely Many Sites Models

9.6.1 Introduction

We now turn to sampling properties relating to the infinitely many sites model. The properties of samples in the infinitely many sites model are relevant to the theory of single-nucleotide polymorphisms, for which much data are currently gathered and from which many inferences about the population under consideration are made. As stated in Section 9.2, we assume throughout, unless otherwise stated, that there is no recombination between sites, and that selective neutrality and stationarity both obtain. In Section 9.6.2, where the Wright–Fisher and the Cannings models are considered, all results are diffusion approximations and the diffusion approximations definition of $\theta = 4Nu$ (for the Wright–Fisher model) and $\theta = 4Nu/\sigma^2$ (for the Cannings model) are used. The definition of θ for the Moran model, for which exact results are obtained, is more complex and is considered in Section 9.6.3.

9.6.2 The Wright–Fisher Model

If two nucleotides at a given site segregate in a population with current frequency x , $1 - x$, the probability that a given individual is heterozygous at this site is $2x(1 - x)$. The mean number of heterozygous sites per individual is found by averaging this over the function θx^{-1} found in (9.18), yielding

$$\theta \int_0^1 x^{-1} \{2x(1 - x)\} dx = \theta. \quad (9.49)$$

The same calculation applies, of course, for any two genes taken at random in the population.

It is interesting to observe that (9.19) shows that for the representative values $N = 500,000$, $u = 10^{-6}$, for which $\theta = 2$, there will be on average about 26 sites segregating in the population, while (9.49) shows that of these, on average 2 sites segregate in any given diploid individual.

Suppose now that a sample of 100 genes is taken. (We can think of the calculation in (9.49) as referring to a sample of two genes.) The argument leading to (9.19) shows that in the sample, the mean number of segregating sites is $\theta\{0.6775 + \log 100\}$. This allows an estimation of θ from an observed number of segregating sites in the sample, and then from (9.19) we are able to estimate the number of sites segregating in the population, assuming that the population size is known. We now explore this observation further.

Suppose that in a sample of size 100 we observe 10 segregating sites. The results in the previous section show that we could estimate θ from the equation

$$\hat{\theta}\{0.6775 + \log 100\} = 10,$$

giving $\hat{\theta} = 1.89$. This estimate in conjunction with (9.19) leads to the estimate of 27.44 for the number of sites segregating in a population of 500,000. It also leads to an estimate of the probability that no segregation occurs in the population at a randomly chosen site. Perhaps more important, it also leads to an estimate of the probability of population polymorphism at a given site, where we adopt the Harris definition of polymorphism given in Section 9.2 to the site rather than the gene level.

As an example of the calculations that are possible, we suppose that the gene consists of 2000 nucleotides. From the point of view of the individual sites relevant to single nucleotide polymorphisms, the value $\hat{\theta} = 1.89$ should be replaced by the "site" value $1.89/2000 = 0.000945$. In this case, an estimate of the probability that a given site is monomorphic in the population is about $(1,000,000)^{-0.000945} = 0.987$. The estimated probability of population polymorphism, following the Harris definition, is $1 - (0.01)^{-0.000945} = .0044$.

These calculations lead us to a more detailed examination of the (random) number S_n of segregating sites in the sample of n genes. For the case $n = 2$, Watterson (1975) showed that

$$\text{Prob}(S_2 = s) = \left(\frac{\theta}{1 + \theta}\right)^s \frac{1}{1 + \theta}, \quad j = 0, 1, 2, \dots \quad (9.50)$$

For $s = 0$ this is $1/(1 + \theta)$. This agrees with the infinitely many alleles expression in (3.74), as it must, since if $s = 0$, the two genes sampled are of the same allelic type. From (9.50),

$$\text{E}(S_2) = \theta, \quad \text{var}(S_2) = \theta + \theta^2. \quad (9.51)$$

This value for the mean of S_2 agrees with the expression given in (9.49), found by a different approach. The variance of S_2 exceeds the mean because of correlation between sites caused by the complete linkage between sites.

The probability distribution for S_n for general values of n is more complicated. Watterson (1975) showed that S_n has the distribution of the sum of $n - 1$ independent geometric random variables Y_1, Y_2, \dots, Y_{n-1} , where

$$\text{Prob}(Y_j = i) = \left(\frac{\theta}{j + \theta}\right)^i \left(\frac{j}{j + \theta}\right), \quad i = 0, 1, 2, \dots, \quad (9.52)$$

and made the perceptive comment that Y_j is the number of mutations occurring in the ancestry of the sample during those times when the n genes in the sample have exactly $j + 1$ distinct ancestor genes. This remark, which we may take as heralding the development of coalescent theory, will be developed at length in Chapter 10.

The probability that $S_n = 0$ is the product of the probabilities that each $Y_j = 0$, and this is

$$\frac{(n - 1)!}{(\theta + 1)(\theta + 2) \cdots (\theta + n - 1)}.$$

This is identical, as it must be, to the probability, in the infinitely many alleles model, that the number of different alleles in the sample is 1 (see (3.87)).

The representation (9.52) shows that the mean of S_n , being the mean of $Y_1 + Y_2 + \cdots + Y_{n-1}$, is

$$\text{mean of } S_n = \theta g_1, \quad (9.53)$$

where g_1 is defined by

$$g_1 = \sum_{j=1}^{n-1} \frac{1}{j}. \quad (9.54)$$

Similarly, the variance of S_n is (Watterson (1975))

$$\text{var}(S_n) = g_1\theta + g_2\theta^2, \quad (9.55)$$

where g_1 is defined in (9.54), and

$$g_2 = \sum_{j=1}^{n-1} \frac{1}{j^2}. \quad (9.56)$$

The complete distribution of S_n was found by Tavaré (1984), who showed that

$$\text{Prob}(S_n = s) = \frac{n - 1}{\theta} \sum_{j=1}^{n-1} (-1)^{j-1} \binom{n-2}{j-1} \left(\frac{\theta}{j + \theta}\right)^{s+1}. \quad (9.57)$$

This probability can also be found from the recurrence relation (10.11) below, and the discussion in Chapter 10 shows why this recurrence relation holds.

Summation in (9.57) shows that the probability that $S_n \leq s$, which we denote by $F(s, \theta)$, is given by

$$F(s, \theta) = 1 - \sum_{j=1}^{n-1} (-1)^{j-1} \binom{n-1}{j} \left(\frac{\theta}{j+\theta} \right)^{s+1}. \quad (9.58)$$

We now turn to questions of statistical inference, and consider first the estimation of θ . Equation (9.53) implies that an unbiased estimator $\hat{\theta}_S$ of θ is (Ewens, (1974b), implicit in Watterson (1975)),

$$\hat{\theta}_S = \frac{S_n}{g_1}, \quad (9.59)$$

where g_1 is defined in (9.54). If s_n is the observed value of S_n derived from a particular sample, the corresponding estimate of θ is, immediately,

$$\hat{\theta}_s = \frac{s_n}{g_1}. \quad (9.60)$$

Equation (9.55) implies that the variance of $\hat{\theta}_S$ is

$$\text{var}(\hat{\theta}_S) = \frac{\theta}{g_1} + \frac{g_2 \theta^2}{g_1^2}. \quad (9.61)$$

Equation (9.49) shows that another possible estimator of θ is provided by taking all possible pairs of genes and finding the average number of sites at which any two pairs differ. Equation (9.59) provides a different estimator of θ . Both estimators are based on the assumption of selective neutrality and can be expected to differ when selection exists. This observation forms the basis of one test of selective neutrality, discussed at greater length in Chapter 11.

The expression θx^{-1} found in (9.18) applies to a sample of genes as well as to the entire population, in that for those sites segregating in the sample, the mean number of sites, in a sample of n sequences, at which there are j representatives of the mutant nucleotide is

$$\frac{\theta}{j}, \quad j = 1, 2, \dots, n-1. \quad (9.62)$$

This is the sample analogue of the population frequency spectrum (9.17), and could be called the sample frequency spectrum. Equivalently, given that a site is segregating, the probability that the mutant nucleotide appears j times ($j = 1, 2, \dots, n-1$) is

$$\frac{1}{jg_1}. \quad (9.63)$$

Passing to a continuous approximation, the mean number of sites for which the frequency of the mutant nucleotide is in the range $(x, x + \delta x)$ is

$$\frac{\theta}{x} \delta x, \quad \frac{1}{n} \leq x \leq \frac{n-1}{n}. \quad (9.64)$$

The expression (9.62) leads to what might be called the conditional mutant nucleotide frequency spectrum, given the number S_n of segregating sites and the fact that S_n/g_1 is an unbiased estimator of θ , as

$$\frac{S_n}{jg_1}. \quad (9.65)$$

We return to this result in Section 11.3.2.

In many cases the mutant nucleotide might not be distinguishable from the original nucleotide, and in these cases a more relevant calculation is that the mean number of sites at which there are j representatives of one nucleotide and $n - j$ of another is

$$\theta j^{-1} + \theta(n-j)^{-1} = \frac{n\theta}{j(n-j)}. \quad (9.66)$$

We could call (9.66) the sample frequency spectrum. The parallel conditional frequency spectrum is

$$\frac{nS_n}{j(n-j)g_1}. \quad (9.67)$$

(An obvious modification to both these formulas is needed when n is even and $j = n/2$.)

The expressions in (9.36) show that θ does not admit unbiased estimation using infinitely many alleles data. By contrast, it is clear from the above that both θ and θ^2 admit unbiased estimation in the infinitely many sites case, using estimators based on S_n and S_n^2 . This makes it all the more remarkable that, whereas the infinitely many alleles quantity K_n is a sufficient statistic for θ in that model, S_n is *not* a sufficient statistic for θ in the infinitely many sites model. This implies that in the infinitely many sites model, the data in a sample of genes beyond the information given by S_n on its own can in principle be used to provide better estimation of θ than that provided through S_n only. We return to this point below.

We make four remarks about the variance (9.61). First, with free recombination between sites and a Poisson mutation process, the variance of $\hat{\theta}_S$ is the first term on the right-hand side of (9.61). It is thus plausible that with small but nonzero recombination between sites, the variance of $\hat{\theta}_S$ is slightly less than the value given in (9.61), but on the other hand exceeds the first term on the right-hand side of (9.61).

The second comment follows from the first. The variance (9.61) of $\hat{\theta}_S$ applies for completely linked sites. Despite this, the expression (9.61) is often used for data arising from a sample of many genes, often on several

different chromosomes. The bounds just described might then be useful for such samples in assessing the variance of $\hat{\theta}_S$.

Third, the variance (9.61) is of order $1/\log n$ and is thus quite large even for large n . The same is true of the variance of the estimator (9.40). This implies that neither K_n nor S_n provides reliable estimation of θ . A variance of order $1/\log n$, rather than the classic statistical order $1/n$, arises in both cases because of the dependence between the genes in the sample arising from their common ancestry, a matter taken up in more detail in Chapter 10.

Finally, we can compare the variance (9.61) of $\hat{\theta}_S$ with the approximate infinitely many alleles mean square error (MSE) of $\hat{\theta}_K$, given in (9.40). This comparison shows that the variance of $\hat{\theta}_S$ is always less than the MSE of $\hat{\theta}_K$. While for small θ the two expressions are quite close, for $\theta = 1$ the variance of $\hat{\theta}_S$ is about 94% of the MSE of $\hat{\theta}_K$, and as θ increases, the variance of $\hat{\theta}_S$ becomes increasingly small relative to the MSE of $\hat{\theta}_K$. This confirms the general principle that estimation of θ using S_n is better than that using K_n .

Despite the fact that $\hat{\theta}_S$ provides more precise estimation of θ than does $\hat{\theta}_K$, it is in principle possible, as stated above, to employ more detailed “sites” data to find a better estimator of θ than that provided by using only S_n , which ignores aspects of these more detailed data. This matter has been discussed at length in the literature. Optimal estimation in statistics arises through the method of maximum likelihood, and thus the aim is to find the likelihood of a sample of n genes, the data in this sample involving not only the value of S_n but the complete configuration of the nucleotides at the various segregating sites.

Unfortunately an explicit expression for this likelihood depends on historical factors concerning when the various mutational events occurred, to be discussed in detail in Chapter 10. This information is of course not directly available from the data in the sample, although it might be possible to infer it from “out-group” data. The interpretation of Y_j given above is that Y_j is the number of mutations arising in the ancestry of the sample during those times when the n genes in the sample have exactly $j + 1$ distinct ancestor genes. If the Y_j were known, the likelihood of the data would be

$$\theta^{S_n} (n-1)! \prod_{j=1}^{n-1} (j+\theta)^{-(Y_j+1)}. \quad (9.68)$$

This leads to an implicit equation for the maximum likelihood estimator of θ , namely,

$$\frac{S_n}{\theta} = \sum_{j=1}^n \frac{Y_j + 1}{j + \theta} \quad (9.69)$$

(Fu and Li (1993)). Standard statistical maximum likelihood theory then shows that the variance of any unbiased estimator of θ cannot be less than

$$\frac{\theta}{\sum_{j=1}^{n-1} \frac{1}{j+\theta}},$$

and that the maximum likelihood estimator of θ using Y_1, Y_2, \dots, Y_{n-1} has a variance achieving this bound. The variance (9.61) of $\hat{\theta}_S$ exceeds this bound, and when θ is large it can significantly exceed this bound.

These calculations suggest that approaches to estimation taking historical factors into account should be useful. This implies that the use of computationally intensive methods, using the coalescent of the sample of genes (see Chapter 10), are needed. This theory was developed by Griffiths and Tavaré (1994a,b,c, 1995, 1997, 1999) and from the point of view of Markov chain Monte Carlo (MCMC) methods by Kuhner et al. (1995, 1998). The approach of Griffiths and Tavaré uses a version of importance sampling, and this observation led Stephens and Donnelly (2000) to improved estimation methods using a new importance sampling approach. The details of these procedures will be discussed in Volume II.

9.6.3 The Moran Model

As might be expected, there are many exact results available for the sample properties in the infinitely many sites Moran model. These can often be found directly from the corresponding population formulas by simply replacing $2N$ in the latter by n . For example, the mean of the number S_n of segregating sites in the sample is given by (9.24) with $2N$ replaced by n , and the comments following (9.24) continue to apply with this replacement. The probability of sample monomorphism is given by (9.25) with $2N$ replaced by n . In these formulas the definition of θ given in (9.27) is used, as it is throughout this section. Many exact sample results for the Moran model follow immediately from the population results given above.

However, other exact properties of the Moran model are not so easily obtained, and we now discuss these, beginning with the simplest case $n = 2$. The case $n = 2$ is of particular interest, since it relates to the homozygosity probability, or more precisely, since the Moran model relates to haploids, to the probability that two randomly chosen genes have identical nucleotide sequences. In the Poisson mutation input case this is found immediately from the expression (9.25) by putting $2N = 2$ to get $1/(2Ne^\nu - 2N + 1)$. This is very close to the diffusion approximation $1/(1 + \theta)$. For the general mutation input distribution model (9.16), the homozygosity probability is $q_0/(2N - (2N - 1)q_0)$.

In the case of the general mutation input distribution (9.16), the complete distribution of S_2 is that of the sum of M random variables, each having the general mutation distribution, where M itself is a random variable having

the geometric distribution

$$\text{Prob}(M = m) = \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^{m-1}, \quad m = 1, 2, 3, \dots \quad (9.70)$$

We shall see in Chapter 10 why this representation exists. As a check on (9.70), when $M = m$ the probability that $S_2 = 0$ is q_0^m . Thus the unconditional probability that $S_2 = 0$ is

$$\frac{q_0}{2N} \sum_{m=1}^{\infty} \left(q_0 \left(1 - \frac{1}{2N}\right)\right)^{m-1},$$

and this reduces immediately to the expression $q_0/(2N - (2N - 1)q_0)$ given above.

We now turn to the distribution of S_n for arbitrary n . For the Poisson input mutation case, Watterson (1975) found a probability generating function for this distribution, which implies that it can be represented as the distribution of the sum of $n - 1$ random variables Y_1, Y_2, \dots, Y_{n-1} , parallel to those surrounding (9.52). In this case, the random variable Y_j has the distribution of the sum of M_j independent random variables, each having the Poisson input mutation distribution with mean ν . Here M_j is itself a random variable having a geometric probability distribution generalizing (9.70), namely,

$$\text{Prob}(M_j = m) = \frac{j}{2N} \left(1 - \frac{j}{2N}\right)^{m-1}, \quad m = 1, 2, 3, \dots \quad (9.71)$$

This representation leads to a form for the distribution of S_n that is more complicated than the Wright–Fisher diffusion approximation (9.57). It does however allow an exact calculation for the variance of S_n for the Poisson mutation case, namely

$$\text{var}(S_n) = g_1\theta + g_2\theta^2 - \frac{g_1\theta^2}{2N}, \quad (9.72)$$

with θ defined as $2N\nu$, and with g_1 and g_2 defined respectively in (9.54) and (9.56).

The geometric mutation input distribution (9.28) was introduced in Section 9.4.3. This input distribution has the further interesting property that, perhaps uniquely, it allows an explicit form for the distribution of S_n . This distribution is again represented as that of the sum of random variables Y_1, Y_2, \dots, Y_{n-1} , whose distribution is given exactly by (9.52), with θ defined by (9.27). It follows from this that the distribution of S_n is given exactly by (9.57). This implies that the variance of S_n is given exactly by (9.55), a slightly different value from the Poisson input mutation value given in (9.72). This shows that the variance of S_n depends on the nature of the input mutation process. However, the difference between the two variances is small, and it vanishes in the diffusion approximation limit.

9.7 Relation Between Infinitely Many Alleles and Infinitely Many Sites Models

It was remarked several times above that when there is no recombination within the gene of interest, the infinitely many sites model acts as an infinitely many alleles model, so that the two models share various formulas in common. However, when recombination does exist, new alleles can be created in the infinitely many sites model by recombination. This process has no analogue in the infinitely many alleles model, for which new allelic types are assumed to arise from “normal” mutational nucleotide changes. Properties of the formation of new alleles by recombination and by nucleotide mutation are different: For example, in a population containing one allelic type only, new alleles cannot be formed by recombination, whereas they can be formed by nucleotide mutation.

Since the sampling formulas (3.83)–(3.85) are used frequently in population genetics theory, it is important to ask how satisfactory they are for an infinitely many sites model with recombination. Unfortunately, in the infinitely many sites model, the distribution of the number of alleles, that is of the number of distinct nucleotide sequences, appears to be very difficult to obtain. The same is true of the distribution of their frequencies. However, it is at least clear that the view that the generation of a new allele through intracistronic recombination can be regarded for all practical purposes as a new “normal” mutation, so that (3.83)–(3.85) still apply to a close approximation, with a new definition of θ embracing the possibility of “mutation” through recombination, is not justified. The following analysis, due to Strobeck and Morgan (1978), shows this.

Strobeck and Morgan considered two sites in a gene and supposed that mutation occurs at each site at rate v , all mutations being new. A more realistic model takes into account the fact that there are only three possible “new” mutant nucleotides, but for the small values of v appropriate to nucleotide mutation rates the two models are probably reasonably close. In any event, several of the formulas given below are easily amended to the more accurate model. We denote the recombination fraction between sites by R , with ($R \ll 1$), and the population size by N , and assume that a multi-site neutral Wright–Fisher model is applicable.

We now consider four such “two-site” genes, labeled for convenience (a_1b_1) , (a_2b_2) , (a_3b_3) , and (a_4b_4) . Here a_i is the nucleotide at site 1 in gene i , and b_i is the nucleotide at site 2 in gene i . We define the symbol “ \equiv ” to denote identity of nucleotide type and define

$$F_A = \text{Prob}(a_i \equiv a_j), \quad \text{Prob}(b_i \equiv b_j) = F_B, \quad (i \neq j), \quad (9.73)$$

$$F_{AB} = \text{Prob}(a_i \equiv a_j, b_i \equiv b_j), \quad (i \neq j), \quad (9.74)$$

$$G = \text{Prob}(a_i \equiv a_j, b_i \equiv b_k), \quad (i \neq j \neq k), \quad (9.75)$$

$$G^* = \text{Prob}(a_i \equiv a_j, b_i \equiv b_l), \quad (i \neq j \neq k \leq l). \quad (9.76)$$

The formal mathematics of the evolution of the two-site system is now identical to that of the two-locus system considered in Sections 3.6 and 3.9. In particular, from (3.74), the equilibrium values of F_A and F_B are

$$\hat{F}_A = \hat{F}_B = (1 + \psi)^{-1}, \tag{9.77}$$

where $\psi = 4Nv$. (In the more accurate model allowing four nucleotides only, the equilibrium value from (3.70) with $K = 4$ is $(3 + \psi)/(3 + 4\psi)$.)

A recurrence relation analogous to (3.73) can be found for F_{AB} : This was first done, in the context of two-locus models, by Serant (1974). This recurrence relation takes into account the possibilities of no, one, or two recombination events between the sites. If terms in N^{-2} , R^2 , and v^2 are ignored, the recurrence relation is

$$F'_{AB} = (1 - 4v)\{(1 - 2R)\{(2N)^{-1} + (1 - (2N)^{-1})F_{AB}\} + RG\}. \tag{9.78}$$

Similar recurrence relations hold for G and G^* , and simultaneous equilibrium solutions of all equations may easily be found. The solution depends on the relative order of magnitude assumptions made about R and v . When $N \gg 1$ and $v = O(N^{-1})$ it is found, for example, that

$$(1) \quad R \ll v: \quad \hat{F}_{AB} \approx (1 + 2\psi)^{-1}, \tag{9.79}$$

$$(2) \quad R \approx v: \quad \hat{F}_{AB} \approx \alpha \quad (\text{see below}), \tag{9.80}$$

$$(3) \quad R \gg v: \quad \hat{F}_{AB} \approx (1 + \psi)^{-2}, \tag{9.81}$$

where

$$\alpha = \frac{2\psi^3 + \psi^2\xi + 11\psi^2 + 6\psi\xi + 2\xi^2 + 18\psi + 13\xi + 9}{(1 + \psi)\{4\psi^3 + 64\xi^2 + 2\psi\xi^2 + 20\psi^2 + 19\psi\xi + 2\xi^2 + 27\psi + 27\xi + 9\}}$$

and $\xi = 2NR$. It is clear why the values (9.79) and (9.81) arise. In (9.79) the recombination fraction is so low that the system can effectively be considered to be a one-site system with mutation rate $2v$, while in (9.81) the recombination rate is high enough so that the sites act effectively independently. These conclusions are analogous to the two interpretations of the fixation probability (3.138) for large and small NR considered in Section 9.1. Since v is a nucleotide mutation rate, of order 10^{-8} or 10^{-9} , we may expect ψ to be quite small for all populations of size 10^6 or fewer, in which case the three equilibrium values of F_{AB} are quite close. For larger values of ψ , however, this is not the case. Thus for $\psi = 4$, \hat{F}_{AB} decreases from 0.1111 at $R = 0$ to 0.0641 at $R = 4v$ and from 0.0463 at $R = 20v$ to 0.0400 when $R \gg v$. The formulas (9.79)–(9.81) were checked by simulation by Strobeck and Morgan (1978).

These simulations also allow a check to be made of the adequacy of (3.84) and (9.30) for the distribution of allele number and frequencies in the present model. Watterson (1974b) found that if (3.84) and (9.30) hold, the variance in heterozygosity will be given by

$$\text{var}(F) = \frac{2\theta}{(1 + \theta)^2(2 + \theta)(3 + \theta)}, \tag{9.82}$$

R	$v = 0.00125$			
	0	$4v$	$10v$	$20v$
\hat{F}_{AB}	0.5000	0.4758	0.4629	0.4552
θ	1.0000	1.1017	1.1603	1.1968
$\text{var}(F)$ (theoretical)	0.0417	0.0392	0.0378	0.0370
$\text{var}(F)$ (empirical)	0.0410	0.0381	0.0437	0.0391
R	$v = 0.005$			
	0	$4v$	$10v$	$20v$
\hat{F}_{AB}	0.2000	0.1477	0.1301	0.1215
θ	4.0000	6.0572	6.6864	7.2305
$\text{var}(F)$ (theoretical)	0.0076	0.0037	0.0027	0.0023
$\text{var}(F)$ (empirical)	0.0088	0.0050	0.0047	0.0051

Table 9.2. Values of \hat{F}_{AB} calculated from (9.79)–(9.81), values of θ thus calculated from (9.83), values of the variance of (F) calculated from (9.82), and empirical values of this variance (Strobeck and Morgan, 1978) for $N = 100$ and various values of R and v .

as in (5.138). This formula can be used for comparison with empirical values of the variance of (F) once an adequate definition of θ can be made. Strobeck and Morgan (1978) do this by defining θ as the solution of the equation

$$\hat{F}_{AB} = (1 + \theta)^{-1}, \quad (9.83)$$

suggested by (5.138), where \hat{F}_{AB} is given by (9.79)–(9.81). In Table 9.2 we give values of \hat{F}_{AB} , θ as computed from (9.83), the variance of (F) as computed from (9.82) and empirical values of this variance, found from simulations. The latter differ consistently from the values calculated from (9.82) for $4Nv > 1$, so we conclude, at least for these parameter values, that (3.84) and (9.30) do *not* apply for the two-site model with recombination.

It is difficult to find properties of the distribution of the number of alleles in this model theoretically. Strobeck and Morgan observe in their simulations that whereas for $R = 0$ the mean number of alleles somewhat exceeds the variance in the number of alleles, as may be deduced from (3.84), this no longer applies when $R > 0$, so that, for example, for $R = 20v$ the variance is slightly in excess of the mean for $v = 0.00125$ and more than twice the mean for $v = 0.005$. Thus (3.84) cannot hold for such values of R , and the conditional distribution (9.30), upon which some of the tests of neutrality considered in Chapter 11 are based, is also suspect. These observations confirm those made from consideration of the homozygosity.

It is clearly important to assess realistic values of the scaled parameters $\xi = 2NR$ and $\psi = 4Nv$. Since v is a nucleotide mutation rate, we may expect $v \approx 10^{-8}$ or 10^{-9} . Typical values of R are less precise: Possibly values of order 10^{-5} may be expected. These values certainly imply $R \gg v$,

and the values in Table 9.2 then suggest that (3.84) and (9.30) are in doubt if v is sufficiently large. Unfortunately, the simulation values possibly do not cover the (R, v) combinations of most importance, and extrapolation from Table 9.2 is difficult. A conservative argument is to note that the effect of recombination is certainly less for $v = 10^{-8}$ than for $v = 10^{-6}$. But for $v = 10^{-6}$, $R = 10^{-5}$, $N = 125,000$, the values in Table 9.2 suggest that (3.84) and (9.30) might apply to a reasonable approximation.

9.8 Genetic Variation Within and Between Populations

In Chapter 12 we shall examine aspects of the evolution of genetic material in different populations or even species. In this section we consider how genetic variation at the molecular level can be divided, at least approximately, into “within” and “between” population components by an analysis of variance technique. Although the approach considered has points of similarity with that of Lewontin (1973), who uses entropy measures instead of sums of squares, it is based essentially on ANOVA concepts and the work of Wright (1943, 1951, 1965a) and Nei (1973).

Suppose that a sample of n genes is taken from each of h populations and that at any chosen nucleotide site only two nucleotides are observed in the entire sample. Define y_{ij} by

$$y_{ij} = \begin{cases} +1 & \text{if the } j\text{th gene in the } i\text{th population contains nucleotide 1,} \\ 0 & \text{if the } j\text{th gene in the } i\text{th population contains nucleotide 2.} \end{cases} \quad (9.84)$$

Then the classical analysis of variances sums of squares

$$\begin{aligned} \sum \sum (y_{ij} - \bar{y}_i)^2 &= \text{within group sum of squares,} \\ n \sum (\bar{y}_i - \bar{y})^2 &= \text{between group sum of squares,} \end{aligned} \quad (9.85)$$

become, with the identification (9.84),

$$n \sum \bar{x}_i (1 - \bar{x}_i)^2 \quad \text{and} \quad n \sum (\bar{x}_i - \bar{x})^2 \quad (9.86)$$

respectively, where \bar{x}_i is the frequency of nucleotide 1 in the sample from population i , and \bar{x} is the average frequency over all samples. If σ_w^2 is the within-group variance in frequency and σ_b^2 the between-group variance, the sums of squares in (9.86) are unbiased estimators of

$$k(n-1)\sigma_w^2 \quad \text{and} \quad (k-1)\sigma_w^2 + n(k-1)\sigma_b^2, \quad (9.87)$$

respectively, so that σ_w^2 and σ_b^2 can be estimated by

$$\hat{\sigma}_w^2 = n\{k(n-1)\}^{-1} \sum \bar{x}_i (1 - \bar{x}_i) \quad (9.88)$$

and

$$\hat{\sigma}_b^2 = (k - 1)^{-1} \sum (\bar{x}_i - \bar{\bar{x}})^2 - \{k(n - 1)\}^{-1} \sum \bar{x}_i(1 - \bar{x}_i). \quad (9.89)$$

The estimator (9.88) is necessarily nonnegative, whereas the right-hand side in (9.89) can be negative: If it is, we conventionally put $\hat{\sigma}_b^2 = 0$.

A measure of within- and between-group variation can now be found by averaging $\hat{\sigma}_w^2$ and $\hat{\sigma}_b^2$ over a number of nucleotide sites: This is in effect the procedure of Lewontin (1973). However, the ability to allocate individuals to groups with high success on the basis of genetic characteristics is not incompatible with a high $\hat{\sigma}_w^2/\hat{\sigma}_b^2$ ratio, since such an allocation can take advantage of multivariate analysis of variance techniques, and does not rely on simple averaging of $\hat{\sigma}_w^2$ and $\hat{\sigma}_b^2$ values.

9.9 Age-Ordered Alleles: Frequencies and Ages

The current direction of interest in population genetics is a retrospective one, looking backward to the past rather than (as with much of the theory in this book) looking forward into the future. This change of direction is largely spurred by the large volume of genetic data now available at the molecular level and a wish to infer the forces that led to the data observed. Tests of the neutral theory, discussed in Chapter 11, form one such inferential procedure.

Far more important, however, is the retrospective process associated with the coalescent, discussed at greater length in Chapter 10. The concept of the coalescent leads naturally into a discussion of the age properties of alleles as well as a discussion of age-ordered allele frequencies. This topic has recently been reviewed by Slatkin and Rannala (2000). The discussion in this section does not aim at a general overview such as that provided by Slatkin and Rannala. Instead it is more specific, being slanted toward explaining some of the formulas in previous sections of this chapter by using age properties of alleles, and then an introduction to further explanations using coalescent theory.

The material in this section covers both sample and population formulas relating to the infinitely many alleles model. Some results are diffusion approximations, and for them the definition of θ depends on the population model implicitly discussed. Various formulas for the Moran model are exact. The concept of reversibility was introduced in Section 2.12. This concept can be used to derive age properties from the prospective theory, and vice versa. Reversibility arguments were used, for example, in deriving (5.134), and further examples of the form of argument leading to (5.134) will be given later. We shall freely use reversibility arguments throughout, relying on reversibility properties of the diffusion process and also of the Moran infinitely many alleles model.

We first discuss allelic frequencies, for which finding “age” properties amounts to finding size-biased properties. Kingman’s (1975) Poisson–Dirichlet distribution was introduced in Section 5.10. Unfortunately, this distribution is not user-friendly, as, for example, (5.130) and (5.131) imply. This makes it all the more interesting that a *size-biased* distribution closely related to it, namely the GEM distribution, named for Griffiths, (1980), Engen (1975) and McCloskey (1965), who established its salient properties, is both simple and elegant. More important, it has a central interpretation with respect to the ages of the alleles in a population. We now describe this distribution.

The ordered allelic frequencies in the population follow the Poisson–Dirichlet distribution. Suppose that a gene is taken at random from the population. The probability that this gene will be of an allelic type whose frequency in the population is x is just x . In other words, alleles are sampled by this choice in a size-biased way. It can be shown from properties of the Poisson–Dirichlet distribution that the (random) frequency of the allele determined by this randomly chosen gene is

$$f(x) = \theta(1-x)^{\theta-1}. \quad (9.90)$$

This result also follows from the frequency spectrum (3.95): The probability that there exists an allele in the population with frequency between x and $x + \delta x$, and that the gene chosen is of this allelic type, is $\theta x^{-1}(1-x)^{\theta-1}x\delta x = \theta(1-x)^{\theta-1}\delta x$. Equation (9.90) follows immediately.

Suppose now that all genes of the allelic type just chosen are removed from the population. A second gene is now drawn at random from the population and its allelic type observed. The frequency of the allelic type of this gene among the genes remaining at this stage is also given by (9.90). All genes of this second allelic type are now also removed from the population. A third gene is then drawn at random from the genes remaining, its allelic type observed, and all genes of this (third) allelic type removed from the population. This process is continued indefinitely. At any stage, the distribution of the frequency of the allelic type of any gene just drawn among the genes left when the draw takes place is given by (9.90). This leads to the following representation. Denote by w_j the original population frequency of the j th allelic type drawn. Then we can write

$$w_1 = x_1, \dots, w_j = (1-x_1)(1-x_2)\cdots(1-x_{j-1})x_j, \quad j = 2, 3, \dots, \quad (9.91)$$

where the x_j are independent random variables, each having the distribution (9.90). The random vector (w_1, w_2, \dots) then has the GEM distribution.

All the alleles in the population at any time eventually leave the population, through the joint processes of mutation and random drift, and any allele with current population frequency x survives the longest with probability x . That is, since the GEM distribution was found according to a size-biased process, it also arises when alleles are labeled according to the

length of their future persistence in the population. Reversibility arguments then show that the GEM distribution also applies when the alleles in the population are labeled by their age. In other words, the vector (w_1, w_2, \dots) can be thought of as the vector of allelic frequencies when alleles are ordered with respect to their ages in the population (with allele 1 being the oldest).

The elegance of many age-ordered formulas derives directly from the simplicity and tractability of the GEM distribution. We now give two examples. First, the GEM distribution shows immediately that the mean population frequency of the oldest allele in the population is

$$\theta \int_0^1 x(1-x)^{\theta-1} dx = \frac{1}{1+\theta}, \quad (9.92)$$

and more generally that the mean population frequency of the j th oldest allele in the population is

$$\frac{1}{1+\theta} \left(\frac{\theta}{1+\theta} \right)^{j-1}.$$

Second, the probability that a gene drawn at random from the population is of the type of the oldest allele is the mean frequency of the oldest allele, namely $1/(1+\theta)$, as just shown. More generally, the probability that n genes drawn at random from the population are all of the type of the oldest allele is

$$\theta \int_0^1 x^n (1-x)^{\theta-1} dx = \frac{n!}{(1+\theta)(2+\theta) \cdots (n+\theta)}.$$

The probability that n genes drawn at random from the population are all of the same *unspecified* allelic type is

$$\theta \int_0^1 x^{n-1} (1-x)^{\theta-1} dx = \frac{(n-1)!}{(1+\theta)(2+\theta) \cdots (n+\theta-1)},$$

in agreement with (3.87). From this, given that n genes drawn at random are all of the same allelic type, the probability that they are all of the allelic type of the oldest allele is $n/(n+\theta)$. The similarity of this expression with that deriving from a Bayesian calculation is of some interest.

The GEM distribution is, of course, a diffusion approximation, and the above results are diffusion approximations. The distribution has a number of interesting mathematical properties. It is invariant under size-biased sampling, and this property has been used by Hoppe (1987) to derive the frequency spectrum (3.95). It also has important properties with respect to the concepts of random deletions and noninterference, discussed in Section 9.5.3, which were also exploited by Hoppe (1986). These properties are perhaps of more interest in ecology than in genetics, so we do not develop them here.

It will be expected that various exact results hold for the Moran model, with θ defined as $2Nu/(1-u)$. The first of these is an exact representation of the GEM distribution, analogous to (9.91). This has been provided by Hoppe (1987). Denote by N_1, N_2, \dots the numbers of genes of the oldest, second-oldest, ... alleles in the population. Then N_1, N_2, \dots can be defined in turn by

$$N_i = 1 + M_i, \quad i = 1, 2, \dots, \tag{9.93}$$

where M_i has a binomial distribution with index $2N - N_1 - N_2 - \dots - N_{i-1} - 1$ and parameter x_i , where x_1, x_2, \dots are iid continuous random variables each having the density function (9.90). Eventually the sum $N_1 + N_2 + \dots + N_k$ reaches the value $2N$ and the process then stops, the final index k being identical to the number K_{2N} of alleles in the population.

It follows directly from this representation that the mean of N_1 is

$$1 + (2N - 1)\theta \int_0^1 x(1-x)^{\theta-1} dx = \frac{2N + \theta}{1 + \theta}.$$

The mean of the proportion $N_1/(2N)$ is $1/\{1 + (2N - 1)u\}$, which is very close to the diffusion approximation $1/\{1 + \theta\}$.

If there is only one allele in the population, so that the population is monomorphic, this allele must be the oldest one in the population. The above representation shows that the probability that the oldest allele arises $2N$ times in the population is

$$\text{Prob}(M_1 = 2N - 1) = \theta \int_0^1 x^{2N-1}(1-x)^{\theta-1} dx,$$

and this reduces to the monomorphism probability (3.99).

More generally, Kelly (1977) has shown that the complete distribution of the number of genes of the oldest allele is, for the Moran model,

$$\text{Prob}(\text{oldest allele represented by } j \text{ genes}) = \frac{\theta}{2N} \binom{2N}{j} \binom{2N + \theta - 1}{j}^{-1} \tag{9.94}$$

The case $j = 2N$ considered above is a particular example of (9.94), and the mean number $(2N + \theta)/(1 + \theta)$ follows from (9.94).

We now turn again to approximations deriving from diffusion methods. A question of some interest is to find the probability that the oldest allele in the population is also the most frequent. By reversibility arguments this is also the probability that the most frequent allele in the population will survive the longest into the future, and in turn this is the mean of the frequency of the most frequent allele. Unfortunately, the distribution of the frequency of the most frequent allele is the user-unfriendly Poisson-Dirichlet distribution, and no exact results are available. It is easy to see from the form of the Poisson-Dirichlet distribution that a lower bound for the mean frequency of the most frequent allele is $(1/2)^\theta$, which is useful

for small θ but not of much value for larger θ . Numerical calculations are given by Watterson and Guess (1977) for a range of θ values, who provide also the upper bound $1 - \theta(1 - \theta) \log 2$. For example, when $\theta = 1$ this mean is 0.624, which may be compared with the mean frequency of the oldest allele (which must be less than the mean frequency of the most frequent allele) of 0.5.

We now turn to “age” questions. Some for these follow immediately from our previous calculations. For example, the mean time for all alleles existing in the population at any time to leave the population is given in (9.5), and by reversibility this is the mean time, into the past, that the oldest of these originally arose by mutation. This is then the mean age of the oldest allele in the population, given on a “generations” basis. Since we refer to this calculation with reference to the mean age of the oldest allele in the population, we repeat it here, with this new interpretation:

$$\text{mean age of oldest allele} = \sum_{j=1}^{2N} \frac{4N}{j(j + \theta - 1)} \text{ generations.} \quad (9.95)$$

In the case $\theta = 2$, this mean age is very close to $4N - 2$, that is, to the conditional mean fixation time (5.36). The exact result corresponding to (9.95) for the Moran model is given in (9.10), or equivalently in (9.11), being almost exactly $4N^2$ birth and death events when $\theta = 2Nu/(1 - u) = 2$. This is close to the conditional mean fixation time given in (3.54), and the reason for these identities is discussed below (9.5).

In employing the argument leading to (9.95) we in effect use a result of Watterson and Guess (1977) and Kelly (1977), stating that not only the mean age of the oldest allele, but indeed the entire probability distribution of its age, is independent of its current frequency and indeed of the frequency of all alleles in the population.

We next ask, If an allele is observed in the population with frequency p , what is its mean age? By reversibility, this is the mean time $\bar{t}(p)$ that it persists in the population, and in the diffusion approximation to the Wright–Fisher model this is found immediately from (3.20) as

$$\sum_{j=1}^{\infty} \frac{4N}{j(j + \theta - 1)} \left(1 - (1 - p)^j\right). \quad (9.96)$$

This is clearly a generalization of the expression in (9.95), since if $p = 1$, only one allele arises in the population, and it must then be the oldest allele. A parallel exact calculation for the Moran model follows from the mean persistence time found eventually using (2.160) and (3.57).

A question whose answer follows from the above calculation is the following: If a gene is taken at random from the population, what is the diffusion approximation for the mean age of its allelic type? With a change of notation, the density function of the frequency p of the allelic type of the randomly chosen gene is, from (9.90), $f(p) = \theta(1 - p)^{\theta-1}$. The mean age

$\bar{t}(p)$ of an allele with frequency p is, by reversibility, given by (3.20). The required probability is

$$\theta \int_0^1 \bar{t}(p)(1-p)^{\theta-1} dp, \tag{9.97}$$

and use of (3.20) for $\bar{t}(p)$ shows that this reduces to $2/\theta$ diffusion time units, or for the Wright–Fisher model, $1/u$ generations. This conclusion may also be derived by looking backward to the past and using the coalescent arguments given in Chapter 10. However, we shall not derive it this way, since it is an immediate result. Looking backward to the past, we see that the probability that the original mutation creating the allelic type of the gene in question occurred j generations in the past is $u(1-u)^{j-1}$, $j = 1, 2, \dots$, and the mean of this (geometric) distribution is $1/u$.

An exact calculation parallel to this is possible for the Moran model, using the exact frequency spectrum (3.102) and the exact mean age deriving from (3.57). However, a direct argument parallel to that just given for the Wright–Fisher model shows that the exact mean time, measured in birth and death events, is $2N/u$.

We turn now to sample properties, which are in practice more important than population properties. The most important sample distribution concerns the frequencies of the alleles in the sample when ordered by age. This distribution was obtained by Donnelly and Tavaré (1986), who found the probability that the number K_n of alleles in the sample takes the value k , and that the age-ordered numbers of these alleles in the sample are (in age order) $n_{(1)}, n_{(2)}, \dots, n_{(k)}$. This probability is

$$\frac{\theta^k (n-1)!}{S_n(\theta) n_{(k)} (n_{(k)} + n_{(k-1)}) \cdots (n_{(k)} + n_{(k-1)} + \cdots + n_{(2)}),} \tag{9.98}$$

where $S_j(\theta)$ is defined below (3.83). This formula can be found in several ways, one being as the size-biased version of (3.88).

The expression (9.98) is exact for the Moran model with θ defined as $2Nu/(1-u)$.

Several results concerning the oldest allele in the sample can be found from this formula, or in some cases more directly by other methods. For example, the probability that the oldest allele in the sample is represented by j genes in the sample is (Kelly, (1976))

$$\frac{\theta}{n} \binom{n}{j} \binom{n+\theta-1}{j}^{-1}. \tag{9.99}$$

This is identical to the expression (9.94) if $2N$ is replaced by n in the latter.

Further results provide connections between the oldest allele in the sample and the oldest allele in the population. Some of these results are exact for a Moran model, and others are the corresponding diffusion approximations. For example, Kelly (1976) showed that in the Moran model, the

probability that the oldest allele in the population is observed at all in the sample is $n(2N + \theta)/[2N(n + \theta)]$. This is equal to 1, as it must be, when $n = 2N$, and when $n = 1$ it reduces to a result found above that a randomly selected gene is of the oldest allelic type in the population. The diffusion approximation to this probability, found by letting $N \rightarrow \infty$, is $n/(n + \theta)$.

A further result is that in the Moran model, the probability that a gene seen j times in the sample is of the oldest allelic type in the population is $j(2N + \theta)/[2N(n + \theta)]$. Letting $N \rightarrow \infty$, the diffusion approximation for this probability is $j/(n + \theta)$. When $n = j$ this is $j/(j + \theta)$, a result found above found by other methods.

Donnelly (1986)) provides further formulas extending these. He showed, for example, that the probability that the oldest allele in the population is observed j times in the sample is

$$\frac{\theta}{n + \theta} \binom{n}{j} \binom{n + \theta - 1}{j}^{-1}, \quad j = 0, 1, 2, \dots, n. \quad (9.100)$$

This is, of course, closely connected to the Kelly result (9.99). For the case $j = 0$ this probability is $\theta/(n + \theta)$, confirming the complementary probability $n/(n + \theta)$ found above. Conditional on the event that the oldest allele in the population does appear in the sample, a straightforward calculation using (9.100) shows that this conditional probability and that in (9.99) are identical.

Griffiths and Tavaré (1998) give the Laplace transform of the distribution of the age of an allele observed b times in a sample of n genes, together with a limiting Laplace transform for the case in which θ approaches 0. These results show, for the Wright–Fisher model, that the diffusion approximation for the mean age of such an allele is

$$\sum_{j=1}^{\infty} \frac{4N}{j(j-1+\theta)} \left(1 - \frac{(n-b+\theta)_{(j)}}{(n+\theta)_{(j)}} \right) \quad (9.101)$$

generations, where $a_{(j)}$ is defined as $a_{(j)} = a(a+1) \cdots (a+j-1)$. This is the sample analogue of the population expression in (9.96), and it converges to (9.96) as $n \rightarrow \infty$ with $b = np$.

In the particular case $\theta = 2$, which we have considered several times above, the expression in (9.101) simplifies to

$$\frac{4Nb}{n-b} \sum_{j=b+1}^n j^{-1}. \quad (9.102)$$

Under the limiting process $n \rightarrow \infty$ with $b = np$ this approaches the expression in (3.22). This is as expected, since when $\theta = 2$, (3.22) is by reversibility arguments also the mean age of an allele observed with frequency p in the population.

Our final calculation concerns the mean age of the oldest allele in the sample. For the Wright–Fisher model the diffusion approximation for this

mean age is

$$4N \sum_{j=1}^n \frac{1}{j(j + \theta - 1)}. \quad (9.103)$$

For the case $n = 2N$ this is the value given in (9.5), and for the case $n = 1$ it reduces to the value $1/u$ given above. The corresponding exact result for the Moran model is

$$2N(2N + \theta) \sum_{j=1}^n \frac{1}{j(j + \theta - 1)} \quad (9.104)$$

birth and death events, with (of course) θ defined as $2Nu/(1 - u)$. When $n = 1$ this reduces to the calculation $2N/u$ given above. When $n = 2N$ it is identical to (9.11) and, less obviously, to the expression given in (9.10).

The expression in (9.103) may be written equivalently as

$$\sum_{j=1}^n \frac{1}{v_j + w_j}, \quad (9.105)$$

where

$$v_j = \frac{ju}{2N}, \quad w_j = \frac{j(j-1)(1-u)}{(2N)^2}. \quad (9.106)$$

These expressions follow the pattern of (9.12) and (9.13). In Chapter 10 we shall explain why the mean age of the oldest allele in a sample can be expressed in the form defined by (9.105) and (9.106) and why the mean age of the oldest allele in the population can similarly be expressed in the form defined by (9.12) and (9.13). These are found by an analysis of the coalescent process, which so far has been kept in the background. It is therefore now time to turn to it.

10

Looking Backward in Time: The Coalescent

10.1 Introduction

It is remarkable that the elegant Watterson formulation for the probability distribution for S_n , given implicitly by (9.52), together with the perceptive remark following it, as well as the elegance and simplicity of many of the “age” formulas in Section 9.9, were not immediately seized upon and investigated at greater length immediately after they appeared to determine why formulas of these elegant forms arise. Since these formulas relate to the past history of the population, historical factors must explain them. Similarly, the unequal frequencies that tend to arise even among selectively equivalent alleles, as shown, for example, by (3.83), must be explained by historical factors: The oldest allele in a sample will tend to have a higher frequency than a newly arisen mutant allele. It fell to Kingman (1982a,b,c) to recognize the importance of these historical factors, to see that they are most simply approached by a retrospective analysis of the ancestry of the genes in a sample, to introduce the concept of the coalescent, which provides the framework for this retrospective analysis, and then to lay down the basic mathematical machinery of the coalescent process.

The idea of the coalescent was, however, “in the air” at the time: See, for example, Tajima (1983) and Hudson (1983). Nor should one fail to acknowledge the pioneering work of Malécot (1948), which introduced and exploited the concept of “looking backward in time” to derive important results in population genetics theory.

In this chapter we give a brief introduction to the main ideas of the coalescent. We focus on the simple case in which there are no complications due to selection, recombination, geographical structure, fluctuating population sizes, and so on. The coalescent also leads to significant advances in statistical inference procedures in population genetics. Again, these are not considered in detail here. Definitive reviews of the extensions to the theory needed to handle the complications discussed above, and of inference questions in the coalescent, are provided respectively by Nordborg (2001), Griffiths and Tavaré (2003), and Tavaré (2004). Our aim in this chapter is to give an overview of the more elementary properties of the coalescent process, with a focus on demonstrating how several of the formulas arrived at in Section 9.9 are more naturally arrived at by coalescent methods. A far more complete discussion of the coalescent will be given in Volume II.

10.2 Competing Poisson and Geometric Processes

It is convenient to start with two technical results, one of which will be relevant for diffusion approximations in the coalescent, while the other will be relevant for exact Moran model calculations.

We consider first a Poisson process in which events occur independently and randomly in time, with the probability of an event in $(t, t + \delta t)$ being $a\delta t$. (Here and throughout we ignore terms of order $(\delta t)^2$.) We call a the rate of the process. Standard Poisson process theory shows that the density function of the time between events, and until the first event, is $f(x) = a e^{-ax}$, and thus that the mean time until the first event, and also between events, is $1/a$.

Consider now two such processes, process (a) and process (b), with respective rates a and b . Various results follow almost immediately from standard Poisson process theory. Given that an event occurs, the probability that it arises in process (a) is $a/(a + b)$. The mean number of “process (a)” events to occur before the first “process (b)” event occurs is a/b . More generally, the probability that j “process (a)” events occur before the first “process (b)” event occurs is

$$\frac{b}{a + b} \left(\frac{a}{a + b} \right)^j, \quad j = 0, 1, \dots \quad (10.1)$$

The mean time for the first event to occur under one or the other process is $1/(a + b)$. Given that this first event occurs in process (a), the conditional mean time until this first event occurs is equal to the unconditional mean time $1/(a + b)$. The same conclusion applies if the first event occurs in process (b).

We now turn to the geometric distribution. We consider a sequence of independent trials and two events, event A and event B . The probability that one of the events A and B occurs at any trial is $a + b$. The events A

and B cannot both occur at the same trial, and given that one of these events occurs at trial i , the probability that it is an A event is $a/(a+b)$.

We are interested in the random number of trials until the first event occurs. This number is a geometric random variable taking the value i , $i = 1, 2, \dots$, with probability $(1-a-b)^{i-1}(a+b)$. The mean of this number is $1/(a+b)$. The probability that the first event to occur is an A event is $a/(a+b)$. Given that the first event to occur is an A event, the mean number of trials before the event occurs is $1/(a+b)$. In other words, this mean number of trials applies whichever event occurs first. The similarity of properties between the Poisson process and the geometric distribution is evident.

10.3 The Coalescent Process

We start by describing the coalescent as a quite abstract process, not associated with any of the specific concrete evolutionary models discussed in previous chapters, and later we will see how this process can be used to find properties of the past history of a population whose evolution is described by these models.

We consider the ancestry of a sample of n genes taken at the present time. Since our interest is in the ancestry, we consider a process moving backward in time, and introduce a notation acknowledging this. We consistently use the notation τ for a time in the past before the sample was taken, so that if $\tau_2 > \tau_1$, then τ_2 is further back in the past than is τ_1 .

We describe the common ancestry of the sample of n genes at any time τ through the concept of an equivalence class. Two genes in the sample of n are in the same equivalence class at time τ if they have a common ancestor at this time. Equivalence classes are denoted by parentheses: Thus if $n = 8$ and at time τ genes 1 and 2 have one common ancestor, genes 4 and 5 a second, and genes 6 and 7 a third, and none of the three common ancestors are identical, the equivalence classes at time τ are

$$(1, 2), \quad (3), \quad (4, 5), \quad (6, 7), \quad (8). \quad (10.2)$$

Such a time τ is shown in Figure 10.1.

We call any such set of equivalence classes an equivalence relation, and denote any such equivalence relation by a Greek letter. As two particular cases, at time $\tau = 0$ the equivalence relation is $\phi_1 = \{(1), (2), (3), (4), (5), (6), (7), (8)\}$, and at the time of the most recent common ancestor of all eight genes, the equivalence relation is $\phi_n = \{(1, 2, 3, 4, 5, 6, 7, 8)\}$. The Kingman coalescent process is a description of the details of the ancestry of the n genes moving from ϕ_1 to ϕ_n .

We now turn to a more detailed description of this process. Let ξ be some equivalence relation, and η some equivalence relations that can be found from ξ by amalgamating two of the equivalence classes in ξ . Such

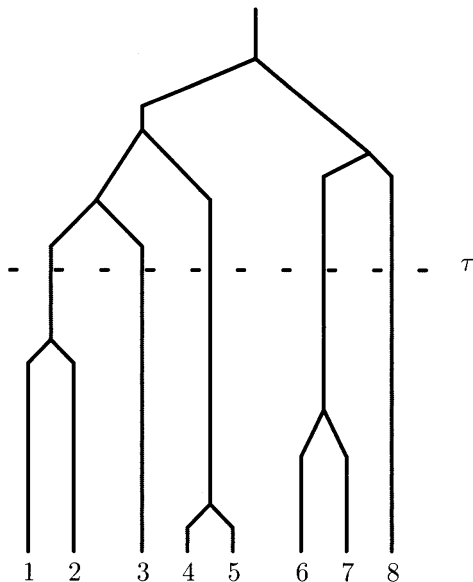


Figure 10.1. The coalescent

an amalgamation is called a coalescence, and the process of successive such amalgamations is called the coalescence process. It is assumed that, if terms of order $(\delta\tau)^2$ are ignored,

$$\text{Prob (process in } \eta \text{ at time } \tau + \delta\tau \mid \text{process in } \xi \text{ at time } \tau) = \delta\tau, \quad (10.3)$$

and if j is the number of equivalence classes in ξ ,

$$\text{Prob (process in } \xi \text{ at time } \tau + \delta\tau \mid \text{process in } \xi \text{ at time } \tau) = 1 - \frac{j(j-1)}{2} \delta\tau. \quad (10.4)$$

This might seem to be a heavy-handed description of the way in which the ancestry of a sample of genes traces back to, and coalesces at, a common ancestor. Indeed, many coalescent results can be found without the full description of the process just given. However, as we see below, the derivation of the sampling formula (3.83) requires this full description.

10.4 The Coalescent and Its Relation to Evolutionary Genetic Models

The main purpose of the coalescent is to provide results, either exact or approximate, for the evolutionary models considered so far in this book and to give a coherent framework within which to view these results.

Kingman (1982a,b,c) showed in his path-breaking work introducing the coalescent that, provided several straightforward conditions are fulfilled, the coalescent provides excellent approximations to quantities of evolutionary interest, the approximations improving as the population size increases. We discuss several of these in Section 10.5.

Kingman focused on the Cannings infinitely many alleles model outlined in Section 3.6.3, of which the Wright–Fisher model is a particular case. He considered a sequence of such models, one for each population size N . It is thus convenient to denote the (random) number of offspring genes from any one parental gene by ν_N in a population of N individuals ($2N$ genes), and to denote the variance of ν_N by σ_N^2 . Kingman then showed, under the requirements that σ_N^2 converge to a positive finite limit σ^2 as $N \rightarrow \infty$ and that the supremum of all moments of ν_N remain finite under the same limit, that the ancestral properties of a sample of fixed size n in the Cannings model converge, as $N \rightarrow \infty$, to those of the coalescent.

The Wright–Fisher model is a particular case of the Cannings model, and for it $\sigma_N^2 = 1 - (2N)^{-1}$, so that the first requirement holds, and it is equally easy to check that the second requirement holds. There are some extreme Cannings models for which one or other of the requirements listed above does not hold, but these seldom arise in practice.

Of course, any coalescent result will always be an approximation for the corresponding Wright–Fisher, or more generally Cannings, result. One reason for this is that the coalescent is a continuous-time process, while the Wright–Fisher and Cannings models are discrete-time processes. In this, the coalescent process is similar to a diffusion process, which also takes place in continuous time. The similarity goes further: In effect, coalescent results are diffusion approximation results. As with time calculations derived from diffusion processes, time calculations derived from the coalescent process must be multiplied by a scaling factor to be brought to a “generations” basis. For the Wright–Fisher model this scaling factor is $2N$.

A more important reason why coalescent results apply immediately only to samples of genes is that in the Wright–Fisher and Cannings models, several coalescent events can occur simultaneously, whereas this does not happen in the coalescent process. For a fixed sample size this becomes less and less likely in the ancestry of a sample as $N \rightarrow \infty$. In the entire population, however, simultaneous coalescences can be expected, so that coalescent results may not be taken over without further consideration to describe population properties. Despite this, we will find that some formal coalescent calculations are surprisingly accurate for population quantities. The reason for this will be given in the next section.

We will later describe a discrete-time coalescent process for the Moran model, which will allow exact calculations to be made, thus explaining the many exact “time” and “age” results for this model found in Chapter 9, holding both for a sample of genes and also for the entire set of genes in the population.

10.5 Coalescent Calculations: Wright–Fisher Models

In this section we consider various calculations arising from the coalescent process, and use them as approximations for results for the Wright–Fisher evolutionary model. Many of these will result in values agreeing with diffusion approximations found in Chapter 9, and the coalescent process often provides the simplest way of arriving at these results. This agreement confirms the claim that coalescent results are in effect diffusion approximation results for this model. We use the coalescent time scale in the calculations and then convert the results found to a Wright–Fisher time scale at the end of the analysis.

We consider first the coalescent process on its own. This process in effect consists of a sequence of $n - 1$ Poisson processes, with respective rates $j(j - 1)/2$, $j = n, n - 1, \dots, 2$, describing the Poisson process rate at which two of these classes amalgamate when there are j equivalence classes in the coalescent. Thus the rate $j(j - 1)/2$ applies when there are j ancestors of the genes in the sample for $j < n$, and the rate $n(n - 1)/2$ applies for the sample itself.

The Poisson process theory outlined in Section 10.2 shows that the time T_j to move from an ancestry consisting of j genes to one consisting of $j - 1$ genes has an exponential distribution with mean $2/\{j(j - 1)\}$. Since the total time required to go back from the contemporary sample of genes to their most recent common ancestor is the sum of the times required to go from j to $j - 1$ ancestor genes, $j = 2, 3, \dots, n$, the mean $E(T_{\text{MRCAS}})$ is, immediately,

$$E(T_{\text{MRCAS}}) = 2 \sum_{j=2}^n \frac{1}{j(j-1)} = 2 \sum_{j=1}^{n-1} \frac{1}{j(j+1)}. \quad (10.5)$$

This time is essentially 2 coalescence time units, and it requires a multiplicative scaling factor of $2N$ to convert to a “generations” basis when applied to the Wright–Fisher model.

It is clear from (10.5) that about half this mean time relates to the final coalescence of two lines of ascent into one. This observation gives some idea of the shape of the coalescent tree: The long arms tend to arise when there is a very small number of genes in the ancestry of the sample.

The times T_j , $j = 1, 2, \dots, n - 1$, are independent, so that the variance of T_{MRCAS} is the sum of the variances of the T_j . Standard calculations show that this is approximately $4\pi^2/3 - 12$, or about 1.16, (squared) time units.

The complete distribution of T_{MRCAS} is also known (Tavaré (2004)). However the expression is complicated and we do not reproduce it here, other than to note the simple inequalities

$$e^{-t} \leq \text{Prob}(T_{\text{MRCAS}} > t) \leq e^{-3t}.$$

If the above theory were to apply to the entire population of genes in a Wright–Fisher model, the mean $E(T_{\text{MRCAP}})$ of the total time to arrive at the most recent ancestor gene of all the genes in the population (MRCAP) would be

$$E(T_{\text{MRCAP}}) = 4N - 2 \quad (10.6)$$

generations. Although coalescent theory does not apply directly to the entire population, the mean number of generations given in (10.6) is correct. The reason for this is implicit in an observation made above, that the long arms in any coalescent process tend to arise when the number of genes in the ancestry of the genes considered is small, and for such small numbers the assumptions for the coalescent process hold.

The conclusion (10.6) can also be reached by reversibility arguments. We may regard the MRCAP gene as one that is certain to fix in the current generation. Given that a certain allele appears with only one representing gene, the mean number of generations until it eventually fixes the population, given that eventual fixation does occur, is $4N - 2$ generations, as is shown by (3.12). This is identical to the expression in (10.6).

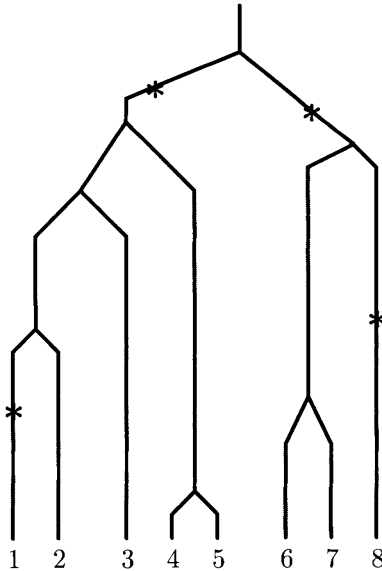


Figure 10.2. The coalescent with mutations

We now introduce mutation, and suppose that the probability that any gene mutates in the time interval $(\tau + \delta\tau, \tau)$ is $(\theta/2)\delta\tau$. All mutants are assumed to be of new allelic types. Following the coalescent paradigm, we trace back the ancestry of a sample of n genes to the mutation forming the

oldest allele in the sample. As we go backward in time along the coalescent, we shall encounter from time to time a “defining event”, taken either as a coalescence of two lines of ascent into a common ancestor or a mutation in one or other of the lines of ascent. Figure 10.2 describes such an ancestry, identical to that of Figure 10.1 but with crosses to indicate mutations.

We exclude from further tracing back any line in which a mutation occurs, since any mutation occurring further back in any such line does not appear in the sample. Thus any such line may be thought of as stopping at the mutation, as shown in Figure 10.3 (describing the same ancestry as that in Figure 10.2).

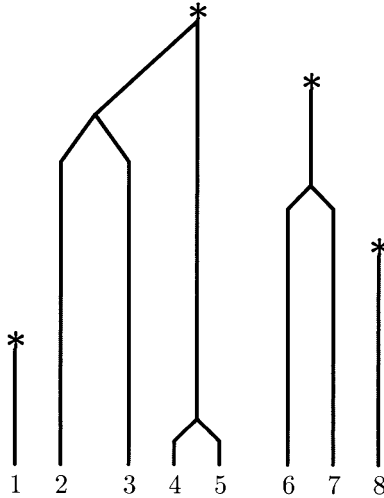


Figure 10.3. Tracing back to, and stopping at, mutational events

If at time τ there are j ancestors of the n genes in the sample, the probability that a defining event occurs in $(\tau, \tau + \delta\tau)$ is

$$\frac{1}{2}j(j-1)\delta\tau + \frac{1}{2}j\theta\delta\tau = \frac{1}{2}j(j+\theta-1)\delta\tau, \tag{10.7}$$

the first term on the left-hand side arising from the possibility of a coalescence of two lines of ascent, and the second from the possibility of a mutation.

If a defining event is a coalescence of two lines of ascent, the number of lines of ascent clearly decreases by 1. The fact that if a defining event arises from a mutation we exclude any further tracing back of the line of ascent in which the mutation arose implies that the number of lines of

ascent also decreases by 1. Thus at any defining event the number of lines of ascent considered in the tracing back process decreases by 1. Given a defining event leading to j genes in the ancestry, the Poisson process theory of Section 10.2 shows that, going backward in time, the mean time until the next defining event occurs is $2/\{j(j + \theta - 1)\}$, and that the same mean time applies when we restrict attention to those defining events determined by a mutation.

Thus starting with the original sample and continuing up the ancestry until the mutation forming the oldest allele in the sample is reached, we find that the mean age of the oldest allele in the sample is

$$2 \sum_{j=1}^n \frac{1}{j(j + \theta - 1)} \quad (10.8)$$

coalescent time units. If the value in (10.8) is multiplied by the time-scale factor $2N$, the resulting expression is identical to that in (9.103). It is interesting that this mean was found by looking backward in time, whereas (9.103) ultimately derives from a calculation looking forward in time.

This time backward until the mutation forming the oldest allele in the sample, whose mean is given in (10.8), does not necessarily trace back to, and past, the most recent common ancestor of the genes in the sample (MRCAS), and will do so only if the allelic type of the MRCAS is represented in the sample. This observation can be put in quantitative terms by comparing the MRCAS given in (10.5) to the expression in (10.8). For small θ , the age of the oldest allele will tend to exceed the time back to the MRCAS, while for large θ , the converse will tend to be the case. The case $\theta = 2$ appears to be a borderline one: For this value, the expressions in (10.5) and (10.8) differ only by a term of order n^{-2} . Thus for this value of θ , we expect the oldest allele in the sample to have arisen at about the same time as the MRCAS. It is for this reason that the value $\theta = 2$ has been used in several calculations given above.

The competing Poisson process theory of Section 10.2 shows that given that a defining event occurs with j genes present in the ancestry, the probability that this is a mutation is $\theta/(j - 1 + \theta)$. Thus the mean number of different allelic types found in the sample is

$$\sum_{j=1}^n \frac{\theta}{j - 1 + \theta},$$

and this is the value given in (3.85). The number of “mutation-caused” defining events with j genes present in the ancestry is, of course, either 0 or 1, and thus the variance of the number of different allelic types found in the sample is

$$\sum_{j=1}^n \left(\frac{\theta}{j - 1 + \theta} - \frac{\theta^2}{(j - 1 + \theta)^2} \right).$$

This expression is easily shown to be identical to the variance formula (3.86).

Even more than this can be said. The probability that exactly k of the defining events are “mutation-caused” is clearly proportional to $\theta^k / \{\theta(\theta + 1) \cdots (\theta + n - 1)\}$, the proportionality factor not depending on θ . Since this is true for all possible values of θ and since the sum of the probabilities over $k = 1, 2, \dots, n$ must be 1, the probability distribution of the number of different alleles in the sample must be given by (3.84).

The complete distribution of the allelic configuration in the sample as given in (3.83) is not so simply derived. Kingman (1982a) employed the full machinery of the coalescent process, together with a combinatorial argument considering all possible paths from ϕ_n to ϕ_1 , to derive (3.83). That is, (3.83) derives immediately from, and is best thought of as a consequence of, the coalescent properties of the ancestry of the genes in the sample. Indeed, it was in an attempt to explain the form of (3.83) through a historical argument that led Kingman to the coalescent concept (Kingman (2000)).

The sample is monomorphic if no mutants occurred in the coalescent after the original mutation for the oldest allele. Moving up the coalescent, this is the probability that all defining events before this original mutation is reached are amalgamations of lines of ascent rather than mutations. The probability of this is

$$\prod_{j=1}^{n-1} \frac{j}{(j + \theta)} = \frac{(n - 1)!}{(1 + \theta)(2 + \theta) \cdots (n - 1 + \theta)}, \quad (10.9)$$

and this agrees, as it must, with the expression in (3.87).

The results just described were found by moving up the coalescent, that is, in reverse real time, rather than down it in forward real time. The Hoppe urn process leading to the probability (9.45) in effect describes the coalescent moving forward in real time. In the genetic context the “urns” probability (9.45) is the probability that the new gene added to the ancestry of the sample as the ancestry size increases from $j - 1$ to j is a new mutant. This is identical to the corresponding probability in the coalescent arguments given above. The urn process was thought of as sampling “through space”, but we now think of it as sampling “through time”, adding new genes to the ancestry of the sample in forward time. This allows us to find all the coalescent-derived results given above. This illustrates an important property of the coalescent, that it allows both “forward” and “backward” time calculations. This is a substantial benefit, since some calculations are more easily carried out moving forward in time, and others are more easily carried out moving backward in time. It also implies that computer simulation of the coalescent process is easy. Several probability distributions relating to samples of genes, discussed in Chapter 11, are difficult to derive analytically, and are then best found, or at least approximated, by simulation using the coalescent process of the sample.

The next example concerns the Wright–Fisher infinitely many sites process. The total number of sites S_n segregating in the sample is identical to the total number of mutations down the coalescent since the MRCAS, since in the infinitely many sites model all such mutations are recorded in the sample. We consider the (random) time T_{j-1} during which there are exactly $j-1$ lines of descent to the sample. We have shown that if mutation is for the moment ignored, the mean of T_{j-1} is $2/\{j(j-1)\}$, $j = 2, 3, \dots, n$. In the Wright–Fisher infinitely many sites process the total mutation rate is $(j-1)\theta/2$ along the $j-1$ lines of descent existing during the time T_{j-1} , and this implies that the mean number of mutations to arise during this time is θ/j . Summation over the values $j = 1, 2, \dots, n-1$, gives the mean number of segregating sites given in (9.53). This justifies the perceptive comment of Watterson referred to below (9.52).

Further results follow immediately. The Poisson process equation (10.1) shows that the distribution of the number of mutations to arise between the times when there are $j-1$ ancestors of the sample and j ancestors is

$$\text{Prob}(i \text{ mutations}) = \frac{j-1}{\theta} \left(\frac{\theta}{j-1+\theta} \right)^i = Q_j(i), \quad i = 0, 1, \dots \quad (10.10)$$

The distribution of the total number of segregating sites is the distribution of the sum of $n-1$ random variables, the j th of which has the distribution given in (10.10). This confirms the distribution arising from (9.52).

The complete distribution of S_n given in (9.57) may be found (Tavaré (1984)) from (10.10) by using the recurrence relation

$$\text{Prob}(S_n = s) = \sum_{i=1}^n \text{Prob}(S_{n-1} = s-i) Q_n(i). \quad (10.11)$$

This recurrence relation shows why the distribution of S_n takes the form that it does.

10.6 Coalescent Calculations: Exact Moran Model Results

In this section we find exact results for the Moran model by a coalescent argument. The time unit used corresponds to the time between one birth and death event and the next.

As we did for the Wright–Fisher model, we first consider the coalescent process itself. Here, however, we use a coalescent theory that is not only exact, but that also applies for a sample of any size, and in particular to the entire population of genes itself. This implies that all results deriving from coalescent theory, for example the topology of the coalescent tree,

are identical to corresponding results for the exact Moran model coalescent process.

The Moran model is a birth and death process, and it is convenient to think of a gene that does not die in a birth and death event as being its own descendant after that event has take place. Consider, then, a sample of n genes, where n is not restricted to be small and could be any number up to and including the entire population size of $2N$. As we trace back the ancestry of these n genes we will encounter a sequence of coalescent events reducing the size of the ancestry to $n - 1, n - 2, \dots$ genes and eventually to one gene, the most recent common ancestor of the sample. Suppose that in this process we have just reached a time when there are exactly j genes in this ancestry. These will be “descendants” of $j - 1$ parental genes if one of these parents was chosen to reproduce and the offspring is in the ancestry of the sample of n genes. The probability of this event is $j(j - 1)/(2N)^2$. With probability $1 - j(j - 1)/(2N)^2$ the number of ancestors remains at j . It follows that, as we trace back the ancestry of the genes, the number T_j of birth and death events between the times when there are j ancestor genes and $j - 1$ ancestor genes has, exactly, a geometric distribution with parameter $j(j - 1)/(2N)^2$ and thus with mean $(2N)^2/\{j(j - 1)\}$. From this, the mean of the time T_{MRCAS} until the most recent common ancestor of all the genes in the sample is given by

$$E(T_{\text{MRCAS}}) = \sum_{j=2}^n \frac{(2N)^2}{j(j - 1)} = (2N)^2 \left(1 - \frac{1}{n}\right) \quad (10.12)$$

birth and death events. In the particular case $n = 2N$ this is

$$E(T_{\text{MRCAP}}) = 2N(2N - 1) \quad (10.13)$$

birth and death events.

Since the various T_j 's are independent, the variance of T_{MRCAP} is the sum of the variances of the T_j 's. This is

$$\text{var}(T_{\text{MRCAS}}) = \sum_{j=2}^n \frac{(2N)^4}{j^2(j - 1)^2} - \sum_{j=2}^n \frac{(2N)^2}{j(j - 1)}. \quad (10.14)$$

The complete distribution of T_{MRCAP} can be found, but the resulting expression is complicated and is not given here.

We now introduce mutation. Consider again a sample of n genes and the sequence of birth and death events that led to the formation of this sample. We again trace back the ancestry of the n genes in the sample, and consider some birth and death event when this ancestry contains $j - 1$ genes. With probability $j/2N$ the newborn created in the population at this birth and death event is in the ancestry of the sample, and with probability u is a mutant. That is, the probability that at this birth and death event a new mutant gene is added to the ancestry of the sample is $ju/(2N)$. As for the Wright–Fisher model, we trace back upward along the lines of ascent from

the sample, and do not trace back any further any line of ascent at a time when a new mutant arises in that line, so that at any mutation, the number of lines of ascent that we consider decreases by 1.

A further decrease can occur from a coalescence for which the addition of a newborn to the ancestry of the sample does not produce a mutant offspring gene. If at any time there are j lines in the ancestry, the probability of a coalescence not arising from a mutant newborn is $j(j-1)(1-u)/(2N)^2$.

It follows from the above that the number of lines of ascent from the sample will decrease from j to $j-1$ at some birth and death event with total probability

$$\frac{ju}{2N} + \frac{j(j-1)(1-u)}{(2N)^2} = \frac{2Nju + j(j-1)(1-u)}{(2N)^2}, \quad (10.15)$$

and we write the left-hand side as $v_j + w_j$, where v_j and w_j are defined in (9.106). The number of birth and death events until a decrease in the number of lines of ascent from j to $j-1$ follows a geometric distribution with parameter $v_j + w_j$. It follows from the theory of Section 10.2 that the mean number of birth and death events until the number of lines of ascent decreases from j to $j-1$ is $1/(v_j + w_j)$, and that this mean applies whatever the reason for the decrease. Tracing back to the mutation forming the oldest allele in the sample, we see that the mean age of this oldest allele is, exactly,

$$\sum_{j=1}^n \frac{1}{v_j + w_j}, \quad (10.16)$$

and this is precisely the expression (9.105).

The probability that a decrease in the number of ancestral lines from j to $j-1$, given that such a decrease occurs, is $v_j/(v_j + w_j)$, or $\theta/(j-1+\theta)$ if θ is defined as $2Nu/(1-u)$. The mean number of different alleles in the sample is thus, exactly,

$$\sum_{j=1}^n \frac{\theta}{j-1+\theta}, \quad (10.17)$$

as given by (3.85). Extending this argument as for the Wright–Fisher case, the exact distribution of the number of alleles in the sample is found to be given by (3.84), as expected.

The complete distribution of the sample allelic configuration, as with the Wright–Fisher model, requires a full description of the coalescent process. This full description is very similar to the approach of Trajstman (1974) in arriving at the exact distribution (3.97).

The argument just used, while expressed as one concerning a sample of genes, applies equally for the entire population of genes. This occurs because, even in the entire population, at most one coalescent event can occur at each birth and death event. Thus all the exact sample Moran

model results found by coalescent arguments apply for the population as a whole, with n being replaced by $2N$. This explains the identity of the form of many exact Moran model sample and population formulas, especially those in Section 9.9.

The above Moran model analysis refers to the infinitely many alleles model. A parallel analysis can be used to find the various exact infinitely many sites model results, and also to explain the identity in form between sample and population formulas. Of these, the most important is the distribution (9.57) for the number of segregating sites in the sample: This is exact in the Moran model with the definition $\theta = 2Nu/(1-u)$. If n is replaced by $2N$, the same formula gives, exactly, the distribution of the number of segregating sites in the population.

10.7 General Comments

In this brief section we make some general comments about the approximate Wright–Fisher and the exact Moran coalescent processes.

First, the coalescent concept is connected to the partition structure requirement in (9.47). This requirement was originally given as a reasonable one concerning the effects of losing one gene from a sample. But as we move backward from a sample up the coalescent, we lose genes one by one as coalescent events occur, and the consistency requirement (9.47) then states that the same sampling structure must apply at all times in the past, that is, at whatever time a sample was taken, and whatever size the sample. The noninterference requirement (9.48) has the natural interpretation in the coalescent context that if, moving forward in time, one “branch” of the coalescent is lost, the properties of the remaining branches remain unchanged except for a change in sample size.

Second, we have applied coalescent theory to provide approximate results for the Wright–Fisher model. However, the coalescent process was originally developed for the more general Cannings model, and all the approximate Wright–Fisher formulae apply with a suitable change in the definition of θ or of the time scale.

Third, the calculations for both the Wright–Fisher model and the Moran model show that the mean time until the most recent common ancestor of the genes in a sample is almost independent of the sample size, provided that this is not too small, and that about half of this mean time arises from the coalescing of the penultimate two genes in the ancestry to the final one gene. This indicates that the general shape of the coalescent tree is one for which the long branches tend to arise in the early ancestry of the sample. However, this conclusion depends critically on the assumption that the population size remains constant, and for cases of increasing population size, a quite different tree shape can be expected. Theory is now available

to handle more general cases in which the population size varies: See in particular Donnelly and Tavaré (1995) for a review of this (and other) aspects of coalescent theory.

The fourth remark follows from the previous one, and in particular from the observation concerning the shape of the coalescent tree. In the Wright–Fisher model, several coalescent results, when applied formally for the population rather than a sample, appear to be more accurate than we initially have a right to expect. As one example, the mean age of the oldest allele in the population, given by (9.5), has the same form as the coalescent-derived formula (10.8) once allowance is made for the coalescent time scale, with n replaced by $2N$ in the summation. Second, if the sample size n in (10.9) is replaced by the population number $2N$, the heuristic value given in (5.69) is obtained. But it was shown in Section 5.7 that this provides an excellent approximation to the probability of population monomorphism.

Fifth, the theory in this chapter covers only the simple cases of coalescent processes, assuming, for example, selective neutrality and a constant population size. Many extensions of the theory, covering cases involving selection, recombination, and geographical features already exist. It is not our purpose here to discuss these here, and these extensions will be discussed in detail in Volume II.

Sixth, it is clear that the coalescent lends itself to efficient simulations, either moving forward in real time (using, perhaps, the Hoppe urn), or moving backward in time. Indeed, its suitability for rapid simulation is perhaps one of its most important characteristics. In the following chapter various tests of selective neutrality will be described. The null hypothesis (neutrality) probability distribution of some of the statistics used in these tests is not easily arrived at analytically, and can often be found only empirically, using simulations based on the coalescent process.

Finally, and most important, the coalescent provides a beautiful framework in which to understand many properties of genetic populations and to arrive quickly at formulas that are less easily found by other methods. It also provides the framework for many inferential processes in population genetics. This is perhaps not surprising, since the initial motivation for its development arose from inferential questions. The influence of the coalescent on population genetics generally cannot be overestimated. Further, it provides perhaps the closest link between the merging fields of classical evolutionary population genetics and human genetics, as is discussed briefly in the following section.

10.8 The Coalescent and Human Genetics

One of the main aims of research in human genetics is to map, or locate, disease genes, using single nucleotide polymorphisms (SNPs) of known loca-

tion. More specifically, the aim is to assess whether there is any association between the two nucleotides possessed by any individual for this SNP and that individual's disease status. A typical approach is that of a case-control study, in which a contingency table is formed with the disease status (affected or not affected) of each person in the sample considered indicated by the rows of a contingency table, and the SNP characteristic of each person indicated by the columns in the table. A chi-square test is then carried out to test for association between disease status and SNP constitution.

The logic behind this approach is as follows. The SNP is assumed to be quite old and selectively neutral, whereas the original mutation causing the disease is thought to be comparatively recent, perhaps arising only a few thousand years ago. Suppose that the site of a single nucleotide polymorphism and the disease locus are on the same chromosome. Then the original mutation will have arisen on a gamete containing one of the two nucleotides segregating at the polymorphic site. At that time there is the strongest possible strong association, or linkage disequilibrium, between the nucleotides at the site and the alleles at the disease locus.

This linkage disequilibrium will break down over time, because of recombination between the site and the disease locus, following an equation of the form of (2.85), amended if necessary, using the recurrence relations (2.94), to allow for selection at the disease locus. However, if the site and the disease locus are very close, the linkage disequilibrium between the disease locus and the site will be retained for many generations, leading to an association that might be picked up in the present-day data by the chi-square procedure outlined above. On the other hand, if the disease locus and the site are not closely linked, the linkage disequilibrium between the alleles at the disease locus and the nucleotides segregating at the site will rapidly break down, and no significant association between the two loci should be observed. Because of this, the chi-square test for association in a "case-control" study is a surrogate for a linkage test.

There is a potential problem with this procedure. It was shown in Section 8.4 that linkage disequilibrium can arise with unlinked loci if the population of interest exhibits geographical structure. Thus establishing a significant association between disease status and the nucleotides carried at some site does not automatically imply that the site is linked, let alone closely linked, to the disease locus. Tests using the association concept, but which overcome this problem, have been developed in the human genetics literature. Our main interest here is not in these tests, so we do not discuss them further, other than to note the awareness in the human genetics literature of the importance of population structure.

If there was only one originating disease mutation, all disease genes in the sample trace back to it in a coalescent process. However, our observations come from the polymorphic nucleotide site, and because of recombination, the disease locus coalescent process might well differ from the coalescent process at this site. Coalescent processes of two loci with recombination,

to be discussed in Volume II, are then needed to assess the extent to which marker site data can be used to infer properties of the disease locus coalescent process.

Here we consider properties of the coalescent process at the disease locus on its own, without further reference to data arising at a site linked to the disease locus. One important feature of the form of data used for testing for linkage between a segregating site and the disease locus is that it is not obtained by random sampling: The disease gene will be at a far higher frequency in the individuals from whom the data are obtained than in the population at large. It follows from this that a conditional coalescent theory is needed rather than the unconditional theory outlined earlier in this chapter. The properties of a conditional process differ considerably from those of an unconditional process. These have been investigated by Wiuf and Donnelly (1999), and we now outline some of their results.

Suppose that there are i disease genes and $n - i$ normal genes in a sample of n genes. If only one originating disease mutation occurred in the ancestry of the sample, the i disease genes trace their ancestry back to a common ancestor gene that is not the ancestor gene of any of the $n - i$ normal genes in the sample. The disease mutation must have occurred either in that common ancestor or in some ancestor gene of it. In the latter case it must have occurred later, in real time, than the coalescence of that common ancestor and the $n - i$ normal genes in the contemporary sample.

Wiuf and Donnelly focus on the estimation of the time T back to the original disease mutation. They approach the analysis by first finding the probability $Q(i)$ that in a sample of n genes, exactly i have a most recent common ancestor gene that is not the ancestor gene of any of the remaining $n - i$ genes in the sample. (An ancestry of this form is similar to the concept of a monophyletic group, discussed in Section 12.4.) They find that

$$Q(i) = \frac{2(i-1)!(n-i)!}{(i+1)(n-1)!}. \quad (10.18)$$

This probability is very small when i is approximately half the value of n , as is likely in a sample used for testing for linkage. Nevertheless, the event that exactly i have a most recent common ancestor gene that is not the ancestor gene of any of the remaining $n - i$ genes did occur, so that calculations for the conditional coalescent process to be analyzed can be expected to differ substantially from those of a standard coalescent, where no conditioning is assumed.

The next calculation relevant to their analysis concerns the probability distribution of the random variable Y , where Y is the number of ancestor genes of the normal genes in the sample at the time of the most recent common ancestor of the disease genes in the sample. Wiuf and Donnelly

find that

$$\text{Prob}(Y = y) = \frac{\binom{n-y-2}{i-2} \binom{y+1}{2}}{\binom{n}{i+1}}. \quad (10.19)$$

This distribution has strong asymmetry properties. For example, when $n = 6$ and $i = 3$, the possible values 1, 2, and 3 for Y have respective probabilities $1/5$, $2/5$, and $2/5$. This asymmetry arises, of course, from the conditional nature of the process examined.

Wiuf and Donnelly then use these results, together with a competing Poisson process argument, to find a limiting conditional distribution for T , the limit being taken as the disease mutation rate approaches 0. It is found that as $n \rightarrow +\infty$ with $i/n = f$ fixed, the mean value of T approaches the value

$$E(T) = \frac{-2 \log f}{1 - f} \quad (10.20)$$

coalescent time units. In the case $f = 1/2$ this is about 1.38 time units, rather less than the unconditional value of about 2 time units established above for the coalescence of the ancestry of a sample of genes to their most recent common ancestor. With f replaced by $1 - p$, it is, however, identical to the conditional mean fixation time given in (5.34). This identity is not a coincidence, and it can be established by an argument using reversibility and conditional mean fixation times.

11

Looking Backward: Testing the Neutral Theory

11.1 Introduction

The coalescent theory described in the previous chapter assumes selective neutrality at the gene locus or nucleotide site considered. In this chapter we shall consider the question, May we in fact reasonably assume selective neutrality at this gene locus or nucleotide site?

The hypothesis of selective neutrality is more frequently called the “non-Darwinian” theory, and was promoted mainly by Kimura (1968). Under this theory it is claimed that, whereas the gene substitutions responsible for obviously adaptive and progressive phenomena are clearly selective, there exists a further class of gene substitutions, perhaps in number far exceeding those directed by selection, that have occurred purely by chance stochastic processes. Stochastic changes in gene frequency have been studied extensively in this book, and they can certainly lead to substitutions in which the replacing gene has no selective advantage over the replaced gene. A better name for the theory would thus be the “extra-Darwinian” theory, although here we adhere to the standard expression given above.

In a broader sense, the theory asserts that a large fraction of currently observed genetic variation between and within populations is nonselective. In this more extreme sense the theory has been described as the “neutral alleles” theory, although this term and the term “non-Darwinian” have been used interchangeably in the literature and will be so used here.

This theory has, of course, been controversial, not only among theoreticians but also among practical geneticists, and the question whether certain

specific substitutions have been neutral was argued quite strongly when the theory first appeared (cf. King and Jukes (1975), Blundell and Wood (1975), Langley and Fitch (1974), and Jukes (1976)). This controversial aspect of the theory has largely died down, perhaps because the inferential questions now of major interest often focus on comparatively short time scales, for which random changes in gene frequency are relatively more important than selective changes. Thus for these questions selective neutrality may often reasonably be assumed for many loci. On the other hand, tests for neutrality and tests associated with neutrality appear frequently in the current literature, as some of the material described below demonstrates.

Throughout this chapter, selective neutrality, the null hypothesis being tested, is assumed, so that all calculations give null hypothesis values. In broad terms the testing procedures used assess, in one way or another, whether the sample data at hand conform reasonably to what is expected under this null hypothesis. In all cases, diffusion theory approximations are used for the theoretical calculations needed. The specific evolutionary model assumed is often not important, since in particular, the conditional distributions used in tests based on the infinitely many alleles model are independent of the specific model, provided in effect that the requirements needed for coalescent theory to apply are met.

The data used to assess the non-Darwinian hypothesis are the current gene frequencies at various loci in a population, DNA and protein sequences frequencies and patterns, and differences of gene and sequence frequencies and structure between populations and species. Several reviews of statistical testing for neutrality have appeared in the literature; see, for example, Kreitman (2000) for a general review, and Fay and Wu (2001), Nielsen (2001), and Sabeti et al. (2002) for reviews focusing on genomic data. Fay and Wu also provide a substantial list of references and recommended reading. Various testing procedures use different forms of data; see, for example, Slatkin (1982), McDonald and Kreitman (1991), Hudson, Kreitman, and Aguardé (1987), Sawyer and Hartl (1992), and Li et al. (2003) for recent examples of this.

The tests that we consider focus exclusively on tests using “within population” data whose theoretical background is based on the infinitely many alleles and the infinitely many sites theory discussed in Chapter 9. Thus we do not consider procedures such as that of Sabeti et al. (2002) that are based on haplotype data and the lengths of haplotype blocks. Only comparatively simple cases will be considered in this volume. For example, no account will be taken here of geographical subdivision, so that panmixia in all populations is implicitly assumed. The extent to which subdivision must be taken into account in testing neutrality is not clear. We have seen, as a result of the calculations in (3.127) and (3.128), that only a small amount of isotropic migration is necessary for a subdivided population to act effectively as one large random-mating population. On the other hand, if

migration occurs only between adjoining subpopulations, the subdivisive structure is more important.

Stationarity is assumed throughout, as is random mating and a constant mutation rate. A constant population size, not varying over generations, is also assumed. This assumption is clearly not correct for the human population, and the shape of the coalescent tree for humans is not well described by the coalescent theory of Chapter 10. A more complete discussion of neutrality testing should take account of these factors, in particular that of a varying population size, and a much broader discussion than that provided here will be given in Volume II.

All the assumptions listed above are quite strong ones, and the theory outlined below is often applied without any substantial assessment of whether they are reasonable for the case at hand. This is unfortunate, since the tests of neutrality that we discuss are in effect tests of neutrality together with the various simplifying assumptions made in the analysis, often not known or overlooked by the investigator. As an example, one of the tests outlined below was originally put forward as a test of constancy of mutation rate, assuming neutrality, but it may equally well be used as a test of neutrality itself, assuming a constant mutation rate. Thus rejection of the neutral theory is in effect rejection of this theory together with all the assumptions, implicit and explicit, in the analysis.

Both the infinitely many alleles and the infinitely many sites models are used in the neutrality testing procedures. The latter model is, of course, appropriate for data consisting of DNA sequences. In the literature, essentially all of the neutrality testing theory depending on this model relates to the case in which only one sequence is considered, normally corresponding to one single gene. This theory depends on the general theory for completely linked sites investigated in some detail in Chapter 9, and which is employed in this chapter in Sections 11.3.2-11.3.6. However, the neutrality testing literature abounds with cases in which several genes are considered, often unlinked or essentially so. Here a revised theory, not generally considered in the literature, is needed: This theory is discussed in Section 11.3.7. Finally, the case in which the data tested relate to unlinked sites is discussed in Section 11.3.8.

Since all calculations are based on sample data, we discontinue the convention adopted in Chapter 9 of using suffixes to distinguish between population and sample variables. Thus, for example, the random variable S_n of Chapter 9 is now denoted simply by S .

11.2 Testing in the Infinitely Many Alleles Models

11.2.1 Introduction

We consider first tests based on the infinitely many alleles model described in some detail in Sections 3.6 and 9.3. This model possesses the attractive feature that an exact sampling theory is available for it: The sampling distribution formulas relevant to these tests are given in (3.83)–(3.85). The exact theory arises because the conditional distribution (9.30) deriving from (3.83)–(3.85) is free of all unknown parameters and thus can be used for an objective test of the neutrality hypothesis. Most tests that take advantage of this fact are tests of whether the observed value of some test statistic derived from the data in a sample differs significantly from its neutral theory expectation as given by the conditional distribution (9.30). Tests using the age-dependent extension of this conditional distribution are also possible, and will be discussed in Section 11.2.4.

11.2.2 The Ewens and the Watterson Tests

The first objective tests of selective neutrality based on the infinitely many alleles model were put forward by Ewens (1972) and Watterson (1977a). The broad aim of both tests was to assess whether the observed values $\{a_1, \dots, a_n\}$ in (9.30) conform reasonably to what is expected under neutrality, that is, under the formula (9.30), given the sample size n and the observed number k of alleles in the sample. It is equivalent to use the observed numbers $\{n_1, \dots, n_k\}$ defined in connection with (3.88) and to assess whether these conform reasonably to their conditional probability given n and k , namely,

$$\text{Prob}(n_1, n_2, \dots, n_k | k) = \frac{n!}{|S_n^k| k! n_1 n_2 \cdots n_k}. \quad (11.1)$$

The Ewens and the Watterson testing procedures differ only in the test statistic employed. The Ewens method used as test statistic the arbitrarily chosen information quantity $-\sum_{j=1}^k x_j \log x_j$, where $x_j = n_j/n$. The reason for this arbitrary choice was that in the standard Neyman–Pearson theory of testing statistical hypotheses, the test statistic is found by considering the ratio of the probability of the data under the null hypothesis (in this case neutrality) to the corresponding probability under the alternative hypothesis (here, selection). Since no unique selective scheme exists, no unique test statistic is found under this procedure, so that any reasonable but nevertheless arbitrary test statistic measuring genetic variation may be chosen. The statistic given above satisfies this criterion.

However, Watterson (1977a) found that for several important selective schemes and for small selective values, the conditional probability of the

observed sample numbers $\{n_1, \dots, n_k\}$, given k , is of the form

$$\text{Prob}\{n_1, \dots, n_k | k\} = \frac{n!}{|S_n^k| k! n_1 n_2 \cdots n_k} \times (1 + A\beta + 0(\beta^2)). \quad (11.2)$$

In this expression, A is defined by

$$A = n((1 + \theta)^{-1} - nf(n + \theta)^{-1})(n + 1 + \theta)^{-1}, \quad (11.3)$$

where f , the observed sample homozygosity, is defined as

$$f = \sum_{j=1}^k \frac{n_j^2}{n^2}, \quad (11.4)$$

and β is a parameter depending on the nature of the selective scheme. Since the ratio of (11.2) to the neutral value (11.1) depends on the observations, for small β at least, only through f , Watterson chose f as the appropriate test statistic. This is superior to the information statistic, and we now discuss its application to the testing procedure.

The first step is to establish what values of f will lead to rejection of the neutral hypothesis. Clearly, f will tend to be smaller under heterotic selection than under neutrality, since this form of selection will tend to equalize allele frequencies compared to the neutral case, thus decreasing f . Under a deleterious mutations model, where we expect one high-frequency “superior” allele and a collection of low-frequency deleterious alleles, f will tend to exceed its neutral theory value. Thus neutrality is rejected in favor of a heterotic scheme if f is “too small” and in favor of a deleterious alleles scheme if f is “too large”.

To determine how large or small f must be before neutrality is rejected, it is necessary to find its neutral theory probability distribution. This may be found in principle from (11.1). In practice, difficulties arise with the mathematical calculations because of the form of the distribution (11.1), and other procedures are needed.

For any observed data set $\{n_1, \dots, n_k\}$, a computer-intensive exact approach proceeds by taking n and k as given, and summing the probabilities in (11.1) over all those n_1, n_2, \dots, n_k combinations that lead to a value of f more extreme than that determined by the data. This procedure is increasingly practicable with present-day computers, but will still be difficult in practice if an extremely large number of sample points is involved.

An approximate approach is to use a computer simulation to draw a large number of random samples from the distribution in (11.1): Efficient ways of doing this are given by Watterson (1978). If a sufficiently large number of such samples is drawn, a reliable empirical estimate can be made of various significance level points. This has been done by Watterson (1978, Table 1), using a method close to that of Hoppe’s urn (discussed in Section 9.5.3), and his table, expanded by further simulations of Anderson (1978), is given in Appendix B. Use of the table in Appendix B, with interpolation

Species	n	k	n_1	n_2	n_3	n_4	n_5	n_6	n_7
<i>willistoni</i>	582	7	559	11	7	2	1	1	1
<i>tropicalis</i>	298	7	234	52	4	4	2	1	1
<i>equinoxialis</i>	376	5	361	5	4	3	3		
<i>simulans</i>	308	7	91	76	70	57	12	1	1

Table 11.1. *Drosophila* sample data

Species	f	$E(f)$	$\text{var}(f)$	P	P_{sim}
<i>willistoni</i>	0.9230	0.4777	0.0295	0.007	0.009
<i>tropicalis</i>	0.6475	0.4434	0.0253	0.130	0.134
<i>equinoxialis</i>	0.9222	0.5654	0.0343	0.036	0.044
<i>simulans</i>	0.2356	0.4452	0.0255		0.044

Table 11.2. Sample statistics, means, variances, and probabilities for the data of Table 11.1.

for values of k and n not listed, gives probably the most direct and useful test of selective neutrality using f . Examples of computing significance levels by both the exact method and the simulation method are given in Watterson (1978, Table 4).

The simulation method allows calculation of tables of $E(f|k)$ and $\text{var}(f|k)$ for various k and n values, and some representative values are given in Appendix C. We discuss a possible use of these mean and variance values below.

We illustrate the above methods of testing neutrality by applying them to particular data. The data concern numbers and frequencies of different alleles at the Esterase-2 locus in various *Drosophila* species and are quoted by Ewens (1974b) and Watterson (1977a). Since the data were obtained by electrophoresis, it is quite possible that the infinitely many alleles model is not appropriate for them, so that the calculations and the analysis given here are for illustrative purposes only.

The data are displayed in Table 11.1. For each set of data we compute f , the observed homozygosity. Then the exact neutral theory probability P (given in Table 11.2) that the homozygosity is more extreme than its observed value may be calculated (except for the *D. simulans* case where the computations are prohibitive). The simulated probabilities P_{sim} are also given in Table 11.2; these are in reasonable agreement with the exact values, so that some confidence can be placed in the values listed in Appendix B, which were found by simulation. The conclusion that we draw is that significant evidence of selection appears to exist in all species except *D. tropicalis*.

We conclude with three general remarks about the Watterson testing procedure. First, the procedure tests for selective neutrality for one gene

locus in one species or location only. If data from several different locations or species are available, it may be judged inappropriate to carry out a formal testing procedure for each data set as has been done above. In such a case it might be preferred to calculate, for each data set, the index function

$$\frac{f - E(f|k, n)}{\sqrt{\text{var}(f|k, n)}}, \quad (11.5)$$

which measures the deviation of f from its neutral theory mean in standard deviation units. A visual comparison of this index function for all the data at hand might provide useful evidence on the neutrality question. One problem with this procedure is that the distribution of f is not close to the normal distribution (see, for example, Figure 1 in Watterson (1978)), so that the usual two standard deviation limits, arrived at ultimately from the normal distribution, may not be of much value. The values of the index function for the four species of Table 11.1 are 2.59, 1.28, 1.93, and -1.31 respectively. These values agree reasonably with the probability levels in Table 11.2 except for the last one: It is clear that values of f falling short of the mean are significant at a smaller number of standard deviations than those in excess of the mean. This is because of the skewness to the right of the distribution of f .

Second, the Watterson test, as with all tests of selective neutrality, is not a powerful one. This lack of power arises from the association between the genes in the data due to their common evolutionary history. As a result the tests might not reject neutral hypothesis even when appreciable selection exists, especially when the sample size is small.

Finally, the Watterson procedure is framed above as a test of whether the observed value of f conforms reasonably to its conditional null hypothesis distribution, given the observed value k of the number of alleles, and the observed value n of the number of genes, in the sample. An equivalent procedure is to compare the estimates of θ derived from (9.32) based on k , and the estimate of θ derived from (9.35) based on f . If these are compatible, the neutral theory is accepted. Suppose, for example, that $n = 200$ and $k = 10$. Then the estimate of θ based on the value $k = 10$ is found from (9.32) to be 2.065. If the same estimate arises from the sample homozygosity, then from (9.35) the sample homozygosity f would be 0.326. This is well within a 95% probability range for f when $n = 200$, $k = 10$ (see Appendix B), and is very close to mean of f , given $n = 200$, $k = 10$ (see Appendix C). Thus with $k = 10$, $n = 200$, and $f = 0.326$ the neutral theory would be accepted, and this conclusion agrees with that reached under the Watterson testing procedure.

We mention this alternative way of viewing the testing procedure since it is very similar in spirit to the Tajima (1989) testing procedure using DNA sequence data and the infinitely many sites model, discussed in Section 11.3.

11.2.3 Procedures Based on the Conditional Sample Frequency Spectrum

In this section we outline two procedures based on the sample “frequency spectrum”. We have defined A_i as the (random) number of alleles in the sample that are represented by exactly i genes. For given k and n , the mean value of A_i can be found directly from (9.30); it is given by

$$E(A_i|k, n) = \frac{n!}{i(n-i)!} \frac{|S_{k-1}^{n-i}|}{|S_k^n|}. \tag{11.6}$$

In this formula the S_j^i are values of Stirling numbers of the first kind as discussed after (3.84). The array of the $E(A_i|k, n)$ values for $i = 1, 2, \dots, n$ is the sample conditional mean frequency spectrum, and the corresponding array of observed values a_i is the observed conditional frequency spectrum. The first approach to assessing neutrality that we outline is an informal one, consisting of a simple visual comparison of the observed and the expected sample frequency spectra. Coyne (1976 provides an illustration of this approach. In Coyne’s data, $n = 21$, $k = 10$, and

$$n_1 = n_2 = \dots = n_9 = 1, \quad n_{10} = 12.$$

Direct use of (3.57) shows that given that $k = 10$ and $n = 21$,

$$E(A_i | k = 10, n = 21) = \frac{21!}{i(21-i)!} \frac{|S_9^{21-i}|}{|S_{10}^{21}|}, \tag{11.7}$$

and this may be evaluated for $i = 1, 2, \dots, 12$, the only possible values in this case. A comparison of the observed a_i values and the expected values calculated from (11.7) is given in Table 11.3. It appears very difficult to maintain the neutral theory in the light of this comparison.

a_i	i											
	1	2	3	4	5	6	7	8	9	10	11	12
E	5.2	2.1	1.1	0.7	0.4	0.2	0.1	0.1	0.0	0.0	0.0	0.0
O	9	0	0	0	0	0	0	0	0	0.0	0.0	1

Table 11.3. Comparison of expected (E) and observed (O) sample frequency spectra: Data of Coyne (1976). See text for details.

A second approach (Ewens, (1973)) provides a formal test of hypothesis, but focuses only on the number A_1 of singleton alleles in the sample. This procedure originally assumed selective neutrality and was used to test for a recent increase in the mutation rate. However, it may equally well be used as a test of neutrality itself if a constant mutation rate is assumed, especially for any test in which the alternative selective hypothesis of interest would lead to a large number of singleton alleles. The procedure may be

generalized by using as test statistic the total number of singleton, doubleton, ..., j -ton alleles, leading to a test in which the selective alternative implies a significantly large number of low-frequency alleles. A parallel procedure, using the frequency of the most frequent allele in the data, may also be used.

We describe here only the test based on the number A_1 of singleton alleles. The total number k of alleles in the sample is taken as given, and the test is based on the neutral theory conditional distribution of A_1 , given k and n . (It is assumed, as is always the case in practice, that n strictly exceeds k .) This conditional distribution is independent of θ and is found (Ewens (1973)) from (9.30) to be

$$\text{Prob}(A_1 = a|k, n) = \sum_{j=a}^{k-1} (-1)^{j-a} \frac{|S_{k-1}^n|}{a!(j-a)!|S_k^n|}. \quad (11.8)$$

Here S_i^j is a Stirling number as discussed above. The conditional mean of A_1 is $|S_{k-1}^n|/|S_k^n|$, and the distribution (11.8) is approximately Poisson, with this mean. This observation enables a rapid approximate assessment of whether the number of singleton alleles is a significantly large one, assuming selective neutrality.

11.2.4 Age-Dependent Tests

The Watterson test of neutrality described above ultimately depends on the sampling distribution (3.83) as its basis. This distribution treats all alleles on an equal footing, and does not, for example, use age-order information about alleles, even if this information is available. This is confirmed by observing that a test statistic equivalent to the Watterson statistic f (defined in (11.4)) is the variance-like quantity

$$f^* = \sum_{i=1}^k \left(n_i - \frac{n}{k} \right)^2,$$

which is a linear function of f and whose significance points are the same linear function as those of f . The fact that the Watterson test treats all alleles on an equal footing is evident from the definition of f^* . For given n and k this statistic in effect compares each n_j with the conditional mean n/k , the same mean being used for each n_j .

However, it is well known that the probability distribution (3.83) predicts rather unequal numbers of the n_j values. The reason for this is that different alleles in a population at any time entered that population at various times in the past, and an "old" allele has a greater chance of attaining a high frequency than a "new" allele. This raises the possibility that if the age order of the alleles in the sample is known, a procedure using this age

order, based on the age-ordered analogue (9.98) of (3.83), can be used to arrive at a testing procedure more powerful than Watterson's.

The observed number of alleles k in a sample of n genes is a sufficient statistic for θ in the age-ordered distribution (9.98), so that the conditional distribution of the age-ordered allele numbers $n_{(1)}, n_{(2)}, \dots, n_{(k)}$, given n and k , is

$$\frac{(n-1)!}{|S_n^k| n_{(k)}(n_{(k)} + n_{(k-1)}) \cdots (n_{(k)} + n_{(k-1)} + \cdots + n_{(2)})}. \quad (11.9)$$

This implies that an exact test of the neutrality theory is possible, using the observed values of the age-ordered alleles.

This possibility was investigated in detail by Tavaré et al. (1989), who considered a number of possible test statistics, each a function of $n_{(1)}, n_{(2)}, \dots, n_{(k)}$. Perhaps the most natural of these is the "age-ordered" analogue of f^* , namely,

$$f_a^* = \sum_{i=1}^k \left(n_{(i)} - E(n_{(i)}) \right)^2,$$

where $E(n_{(i)})$ can be found from (11.9). Perhaps surprisingly, it is found that f_a^* has poor properties as a test statistic of neutrality, and further, that none of the age-ordered test statistics considered performed better than the Watterson statistic, which does not use age-order information. Thus, for this model, age-order information does not appear to be useful in testing for selective neutrality.

11.3 Testing in the Infinitely Many Sites Models

11.3.1 Introduction

Since the complete nucleotide sequences of genes are now available in large numbers, and since these data represent an ultimate state of knowledge of the gene, tests of neutrality based on infinitely many alleles theory are increasingly only of historical interest, and it is natural to focus instead on nucleotide sequence data for testing neutrality.

The nature of the testing procedure depends on the nature of the sequence data used in the test. In some cases the data consist of a sample of n DNA sequences, each being the sequence of one single gene. The theory for such data assumes that there is no recombination between the sites in the gene considered. The discussion in Sections 11.3.3-11.3.6 considers this case. Thus the sampling theory in these sections needed for developing tests of neutrality is based largely on the results of Section 9.6, for which completely linked segregating sites are assumed.

In other cases the data analyzed consist of DNA sequences from several unlinked genes, with several segregating sites arising within each gene. This

case is discussed in Section 11.3.7. For data of this form the theory of Section 9.6 is needed for those sites within any one gene, but further theory is needed for the “between unlinked genes” aspect of the data.

An even more extreme case arises with data from segregating sites that are all unlinked: This case is discussed in Section 11.3.8. Here the theory of Section 9.6 is not needed.

As for tests using infinitely many alleles theory, discussed above, it is assumed in all the calculations in this section that selective neutrality holds, so that these can be thought of as “null hypothesis” calculations. The notation of Section 9.6 is used throughout this section for these calculations.

It is appropriate to discuss here the broad nature of the testing procedures described below, at least those used in Sections 11.3.3 – 11.3.6, for which segregating sites with no recombination are assumed. It was remarked in Section 9.6 for such sites that when selective neutrality holds, the number S of sites segregating in the sample is not a sufficient statistic for the central parameter θ describing the stochastic behavior of the evolution of these sequences. Indeed, there is no simple nontrivial sufficient statistic for θ for this case. This implies that no direct analogue of the exact infinitely many alleles tests considered in Section 11.2.2 is possible.

On the other hand, in the infinitely many sites model there are several unbiased estimators of θ when neutrality holds, as discussed below in Section 11.3.2. The basic idea behind all of the tests described in Sections 11.3.3-11.3.6 is to form a statistic whose numerator is the difference between two such unbiased estimators and whose denominator is an estimate of the standard deviation of this difference. Although under neutrality these two observed values of these estimators should tend to be close, since they are both unbiased estimators of the same quantity, under selection they should tend to differ, since the estimators on which they are based tend to differ under selection, and in predictable ways. Thus values of the statistic formed sufficiently far from zero lead to rejection of the neutrality hypothesis. To find the sampling properties of these statistics it is necessary first to discuss properties of the various unbiased estimators of θ used in them.

11.3.2 *Estimators of θ*

In this section we consider properties of four statistics that in the neutral case are all unbiased estimators of the parameter θ . The theory considered in this section concerns only the case of completely linked segregating sites.

The first unbiased estimator of θ that we consider is that based on the number S of segregating sites, namely, the estimator $\hat{\theta}_S$ given in (9.59). This estimator was discussed in some detail in Section 9.6.2: In particular, the variance of this estimator is given in (9.61) for the completely linked sites case.

The second unbiased estimator is based on (9.49). Suppose that the nucleotide sequences of genes i and j in the sample are compared and differ at

some random number $T(i, j)$ of sites. Then $T(i, j)$ is an unbiased estimator of θ . It is natural to consider all $\binom{n}{2}$ possible comparisons of two nucleotide sequences in the sample and to form the statistic

$$T = \frac{\sum_{i < j} T(i, j)}{\binom{n}{2}}. \tag{11.10}$$

Since this is also an unbiased estimator of θ , we think of it as forming the unbiased estimator $\hat{\theta}_T$, defined by

$$\hat{\theta}_T = \frac{\sum_{i < j} T(i, j)}{\binom{n}{2}}. \tag{11.11}$$

This estimator of θ was proposed by Tajima (1983). It is a poor estimator of θ in that its variance, namely,

$$\frac{n+1}{3(n-1)}\theta + \frac{2(n^2+n+3)}{9n(n-1)}\theta^2 = b_1\theta + b_2\theta^2, \tag{11.12}$$

does not approach 0 as the sample size n increases. However, our interest here in this estimator is that it forms part of a hypothesis testing procedure, and not as a possible estimator of θ .

The third unbiased estimator of θ follows from (9.66). This equation shows that the mean number of “singleton” sites, that is, sites where one nucleotide arises once and another $n-1$ times, is $n\theta/(n-1)$. If M is the observed number of such sites, then clearly,

$$\hat{\theta}_M = \frac{(n-1)M}{n} \tag{11.13}$$

is an unbiased estimator of θ . The variance of M is

$$\frac{n}{n-1}\theta + \left(\frac{2g_1}{n-1} - \frac{1}{(n-1)^2} \right) \theta^2 \tag{11.14}$$

(Fu and Li, 1993), where g_1 is defined in (9.54). This implies that the variance of $\hat{\theta}_M$ is

$$\frac{n-1}{n}\theta + \left(\frac{2(n-1)g_1}{n^2} - \frac{1}{n^2} \right) \theta^2. \tag{11.15}$$

The fourth unbiased estimator of θ was proposed by Fay and Wu (2000). This estimator is based on the assumption that of two segregating nucleotides at any site, the mutant nucleotide can be recognized. The conditional mutant frequency spectrum (9.65) shows that if there are j representatives of this mutant nucleotide at a given site, the mean of j^2 is $n(n-1)/2g_1$. The mean number of segregating sites is $g_1\theta$. We define U as the sum of the squares of the various observed numbers of the mutant nucleotide at the various segregating sites observed, summed over all segregating sites. It then follows that the mean value of U is $n(n-1)\theta/2$. This

leads to the unbiased estimator $\hat{\theta}_H$, defined by

$$\hat{\theta}_H = \frac{U}{\binom{n}{2}}. \quad (11.16)$$

These four estimators have been used to form tests for neutrality, as described in Sections 11.3.3, 11.3.4, and 11.3.5 below, and their properties as estimators of θ are a central part of the test procedures described. It was observed in Section 9.6.2 that estimators of θ having better properties than these are possible if historical information is available, or can be unambiguously inferred, concerning the evolutionary process leading to the data observed. This fact suggests that better test procedures might be possible if historical information can be employed. This matter is discussed further in Section 11.3.4.

11.3.3 The Tajima Test

It was observed at the end of Section 11.3 that most tests of neutrality using data from completely linked segregating sites depend on the difference of two unbiased estimators of θ . By far the most frequently used test based on such a difference is that devised by Tajima (1989), which compares the values of $\hat{\theta}_T$ and $\hat{\theta}_S$, defined respectively in (9.59) and (11.11). Specifically, the procedure is carried out in terms of the statistic D , defined by

$$D = \frac{\hat{\theta}_T - \hat{\theta}_S}{\sqrt{\hat{V}}}, \quad (11.17)$$

where \hat{V} is an unbiased estimate of the variance of $\hat{\theta}_T - \hat{\theta}_S$ and is defined in (11.19) below. Tajima (1989) showed, by using adroit coalescent arguments, that the variance V of $\hat{\theta}_T - \hat{\theta}_S$ is

$$V = c_1\theta + c_2\theta^2, \quad (11.18)$$

where

$$c_1 = b_1 - \frac{1}{g_1}, \quad c_2 = b_2 - \frac{n+2}{a_1n} + \frac{g_2}{g_1^2},$$

with g_1 and g_2 defined respectively in (9.54) and (9.56) and b_1 and b_2 defined implicitly in (11.12). Since this variance depends on θ , any estimate of this variance depends on a choice of an estimate of θ .

The variance of the estimator $\hat{\theta}_S$ decreases to 0 as the sample size increases (although the decrease is very slow), so the Tajima procedure is to estimate the variance of $\hat{\theta}_T - \hat{\theta}_S$ by the function of S that provides an unbiased estimator of the variance (11.18). Elementary statistical theory shows that this function is

$$\hat{V} = \frac{c_1 S}{g_1} + \frac{c_2 S(S-1)}{g_1^2 + g_2}. \quad (11.19)$$

This is then used in the D statistic given in (11.17) above.

The next problem is to find the null hypothesis distribution of D . Although D is broadly similar in form to a z -score, it does not have a normal distribution and its mean is not zero, nor is its variance 1, since the denominator of D involves a variance estimate rather than a known variance. Further, the distribution of D depends on the value of θ , which is in practice unknown. Thus there is no null hypothesis distribution of D invariant over all θ values.

The Tajima procedure approximates the null hypothesis distribution of D in the following way. First, the smallest value that D can take arises when there is a singleton nucleotide at each site segregating. In this case $\hat{\theta}_T$ is $2S/n$, and the numerator in D is then $\{(2/n) - (1/g_1)\}S$. In this case the value of D approaches a , defined by

$$a = \frac{\{(2/n) - (1/g_1)\} \sqrt{g_1^2 + g_2}}{\sqrt{c_2}}, \quad (11.20)$$

as the value of S approaches infinity.

The largest value that D can take arises when there are $n/2$ nucleotides of one type and $n/2$ nucleotides of another type at each site (for n even) or when there are $(n-1)/2$ nucleotides of one type and $(n+1)/2$ nucleotides of another type at each site (for n odd). In this case the value of D approaches b , defined by

$$b = \frac{\{(n/2(n-1)) - (1/a_1)\} \sqrt{g_1^2 + g_2}}{\sqrt{c_2}} \quad (11.21)$$

when n is even and the value of S approaches infinity. A similar formula applies when n is odd.

Second, it is assumed, as an approximation, that the mean of D is 0 and the variance of D is 1. Finally, it is also assumed that the density function of D is the generalized beta distribution over the range (a, b) , defined by

$$f(D) = \frac{\Gamma(\alpha + \beta)(b - D)^{\alpha-1}(D - a)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)(b - a)^{\alpha+\beta-1}}, \quad (11.22)$$

with the parameters α and β chosen so that the mean of D is indeed 0 and the variance of D is indeed 1. This leads to the choice

$$\alpha = -\frac{(1 + ab)b}{b - a}, \quad \beta = \frac{(1 + ab)a}{b - a}.$$

This approximate null hypothesis distribution is then used to assess whether any observed value of D is significant.

The various approximations listed above deserve further comment. First, the use of an asymptotically large value of S in the computation of a and b is questionable, since the segregating sites are all assumed to arise within the same gene. Second, the mean of D is known not to be 0, and the variance of D is known not to be 1. Finally, the beta distribution is used

as a mathematical convenience rather than because it follows from any theoretical considerations. These comments imply that the adequacy of the distribution (11.22) as the null hypothesis distribution of D has to be examined.

Tajima (1989) investigated the implications of the various approximations listed above for a range of values of n and θ by simulating the distribution of D when neutrality holds. These simulations show that the null hypothesis mean of D is negative (but generally small) and the null hypothesis variance of D less than 1, typically being in the range 0.72 to 0.98 for the values of n and θ considered ($n = 5, 10, 20, 30, \theta = 1, 10, 100$). The beta distribution approximation is not accurate for very small values of n for the case $\theta = 1$, but appears to be far more accurate for larger values of n . This latter result is a desirable feature of the procedure since sample sizes less than 20 cannot be expected to provide a test of neutrality with any significant power.

The problem of finding more accurate significance points of D than are provided by (11.22) was also addressed by Fu and Li (1993), who used simulations with known values of θ and n to find significance points of D empirically. The values of θ considered ranged from 2 to 20, and the most extreme critical value for this range of values was chosen. This approach suffers from the problem that this range of values of θ might not correspond to the value of θ appropriate to the data at hand.

Simonsen et al. (1995) conducted a detailed examination of the accuracy of (11.22) as the null hypothesis distribution of D . Perhaps the main conclusion that they found is that the critical significance points found from (11.22) are often too conservative. While a conservative test is less likely to reject the null hypothesis incorrectly, it necessarily involves a loss of power, so that in this case the Tajima procedure might lead to acceptance of the null hypothesis of neutrality when in fact significant selection exists. This observation agrees with the fact that the true variance of D is less than 1, the value assumed in the distribution (11.22).

Simonsen et al. (1995) considered an approach that does not depend on an arbitrary range of θ values, as does that of Fu and Li (1993). In this approach a $1 - \beta$ confidence interval (θ_L, θ_U) for θ is found from the cumulative form of the Tavaré (1984) distribution for S given in (9.58). The value of β is chosen to be less than the Type I error α used for the test of hypothesis of neutrality. Standard statistical theory shows that the $1 - \beta$ confidence interval (θ_L, θ_U) for θ is found by solving the equations

$$F(s - 1, \theta_L) = 1 - \beta/2, \quad F(s, \theta_U) = \beta/2 \quad (11.23)$$

(Simonsen et al. (1995)), with $F(s, \theta)$ defined in (9.58). They then considered a grid of values of θ in this confidence interval and estimated the $\alpha - \beta$ significance points of the Tajima statistic for each value of θ in this grid, and then used the maximum upper and the minimum lower significance points taken over all values of θ considered. Statistical theory (Berger and

Boos, (1994)) shows that this procedure gives a test of hypothesis of neutrality with Type I error α . Simonsen et al. (1995) then used this procedure to arrive at adjusted significance points for the Tajima test. These imply a slightly less conservative procedure than the original Tajima test.

The Tajima procedure applies when complete linkage between sites obtains, since the coalescent theory used to find the variance of $\hat{\theta}_T - \hat{\theta}_S$ given in (11.18) assumes such complete linkage. Under complete linkage the infinitely many sites model reduces to the infinitely many alleles model, so that the infinitely many alleles testing theory of Section 11.2 may in principle be applied, where distinct alleles are now distinct nucleotide sequences. Because the Tajima procedure makes use of the actual DNA sequences it may be expected to provide a more efficient testing procedure than that based on the infinitely many alleles theory.

11.3.4 Other “Tajima-like” Testing Procedures

The numerator of the Tajima test statistic is $\hat{\theta}_T - \hat{\theta}_S$. It is also possible to form test statistics whose numerators are $\hat{\theta}_T - \hat{\theta}_M$ and $\hat{\theta}_S - \hat{\theta}_M$, where $\hat{\theta}_M$ is defined in (11.13). All three of these differences have a variance of the form $A\theta + B\theta^2$, for some constants A and B depending only on n and the difference in question. The variances of $\hat{\theta}_S$, $\hat{\theta}_T$, and $\hat{\theta}_M$ are also quadratic functions of θ (see (9.61), (11.12), and (11.15)), and thus any one of $\hat{\theta}_S$, $\hat{\theta}_T$, and $\hat{\theta}_M$ can be used to give an unbiased estimate the variance of any of the three differences given above. This implies that there are nine “Tajima-like” test statistics possible, of which the Tajima statistic described above is one.

The properties of these nine test statistics have been investigated by Simonsen et al. (1995). They all have the properties that their null hypothesis distributions depend on θ , and even if θ were known, and that they all have complicated distributions that are best approached through simulation. The broad conclusion of the investigations of Simonsen et al. is that the Tajima statistic has the best operating characteristics of all nine statistics. Because of this, we do not consider the remaining eight procedures further.

Of the above nine test statistics, three use a variance estimate based on S , and these three are the natural ones to investigate in more detail. The Tajima statistic (11.17) is one of these three. The other two, with numerators $\hat{\theta}_S - \hat{\theta}_M$ and $\hat{\theta}_T - \hat{\theta}_M$, denoted respectively by D^* and F^* , were proposed by Fu and Li (1993) as possible test statistics of the neutrality hypothesis. Fu and Li claim that these testing procedures are likely to be more powerful than the Tajima procedure, a claim contested by Tajima (1997). This matter will require more analysis before a resolution of this point can be reached.

A comparison not discussed above is that between $\hat{\theta}_T$ and $\hat{\theta}_H$, defined in (11.16). The difference between these two estimators has been used as the basis of a test of neutrality, designed specifically to test against an alternative hypothesis of a selective sweep at a locus close to the locus under consideration. This test is considered below, in Section 11.3.5.

It was remarked in Sections 9.6.2 and 11.3.2 that the estimators of θ used in the various tests discussed above are not in principle optimum ones, and thus have larger variances than an estimator using historical information. This suggests that sharper tests of neutrality might be available if estimators that use historical information were employed. On the other hand, it was shown in Section 11.2.4 that in the infinitely many alleles case, tests of neutrality using age-order information do not perform better than tests that do not use this information.

11.3.5 Testing for the Signature of a Selective Sweep

The statistic used in any hypothesis testing procedure is in principle chosen so as to maximize the probability of rejecting the null hypothesis (in this case neutrality) in favor of whichever selective alternative is of interest. Evidence for this selective alternative is provided by some specific “signature” in the data. In this section we consider aspects of tests based on a neutral locus signature suggesting a recent selective sweep at some selected locus linked to this neutral locus, carrying with it the frequency of various alleles at the neutral locus.

We assume a sample of n DNA sequences corresponding to n genes at a selectively neutral locus. The tests we consider are all based on the assumption that of two nucleotides segregating in the sample at any site in the neutral gene, the mutant (or derived) nucleotide can be recognized. With no recent selective sweep at a locus linked to the neutral locus under consideration, the mean number of sites at which there are j representatives of the derived nucleotide is θ/j , as given by (9.62). Correspondingly, the mean number of sites at which the mutant nucleotide assumes a frequency in $(x, x + \delta x)$ is given, in the continuous approximation, by (9.64).

Suppose now that a favored new mutant A arises at a locus closely linked to the neutral locus, and increases in frequency to 1 in a comparatively brief selective sweep. After the selective sweep has concluded, the frequency of the mutant nucleotide will tend to be high for those mutant nucleotides “hitchhiking” with the favored allele at the selective locus, or to be low for those mutant nucleotides not hitchhiking with the favored allele. The probability that any given mutant hitchhikes is the probability that the favored mutant arises on a gamete containing the mutant nucleotide, and this is the frequency x of the mutant nucleotide before the selective sweep. The probability of this frequency x is proportional to x^{-1} , as shown by the mutant nucleotide frequency spectrum (9.64). This will lead to a population frequency spectrum after the selective sweep different from the expression

$\phi(x) = \theta/x$ given in (9.18). This new frequency depends on the Maynard Smith and Haigh (1974) quantity c defined in (6.98), and is given by Fay and Wu (2000) as

$$\phi(x) = \frac{\theta}{x} - \frac{\theta}{c}, \quad \frac{1}{2N} \leq x < c, \quad (11.24)$$

$$\phi(x) = \frac{\theta}{c}, \quad 1 - c < x \leq 1 - \frac{1}{2N}, \quad (11.25)$$

$$\phi(x) = 0, \quad c \leq x \leq 1 - c. \quad (11.26)$$

Correspondingly, the selective sweep will tend to lead to an estimator θ_H of θ based on this new frequency spectrum that will tend to be different from that based on the expression in (11.16). Fay and Wu (2000) then form a test for such a recent sweep based on the difference H of the estimators $\hat{\theta}_H$ and $\hat{\theta}_T$, using as null hypothesis the quantity H , defined by

$$H = \frac{\hat{\theta}_H - \hat{\theta}_T}{\sqrt{\hat{V}}}, \quad (11.27)$$

where \hat{V} is an unbiased estimator of the variance of $\hat{\theta}_H - \hat{\theta}_T$. The null hypothesis to be tested is that no selective sweep such as that described has recently occurred, and the null hypothesis distribution of H was found by Fay and Wu by simulation, using the coalescent. This allows an assessment, for any particular case, of whether evidence exists for a recent hitchhiking event at a selected locus close to the neutral locus considered. Power properties of this procedure are investigated by Przeworski (2002). A similar procedure, again using (11.24), is provided by Kim and Stephan (2002).

The Fay and Wu (2000) procedure is an example of a procedure using as test statistic a quantity of the form

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\hat{V}}}, \quad (11.28)$$

where \hat{V} is an estimate of the variance of $\hat{\theta}_1 - \hat{\theta}_2$. In the Fay and Wu procedure $\hat{\theta}_1 = \hat{\theta}_H$, $\hat{\theta}_2 = \hat{\theta}_T$ and \hat{V} is an estimate of the variance of $\hat{\theta}_H - \hat{\theta}_T$. The Tajima statistic (11.17) is also a case of a statistic of the form given in (11.28), with (in that case) $\hat{\theta}_1 = \hat{\theta}_T$, $\hat{\theta}_2 = \hat{\theta}_S$. Fu (1997) considered a variety of test statistics of the form given in (11.28). He focused attention on those cases in which $\hat{\theta}_1$ and $\hat{\theta}_2$ are linear functions of the form

$$\hat{\theta}_1 = \sum \alpha_j X_j, \quad \hat{\theta}_2 = \sum \beta_j Y_j, \quad (11.29)$$

where X_j is the number of segregating sites in the sample for which there are j representatives of the mutant nucleotide (assuming that this can be recognized) and Y_j is the number of segregating sites in the sample for which there are j representatives of either the mutant or the original nucleotide. The means of X_j and Y_j are given by (9.62) and (9.66). The constants $\{\alpha_j\}$ and $\{\beta_j\}$ are required to be chosen so that the neutral theory stationary

means of $\hat{\theta}_1$ and $\hat{\theta}_2$ are both equal to θ , so that the neutral theory stationary mean value of the numerator in (11.28) is 0. The estimators $\hat{\theta}_S$, $\hat{\theta}_H$, and $\hat{\theta}_T$ considered above satisfy these requirements, so that the Fay and Wu and the Tajima statistics are of the required form.

Fu (1997) then considered various statistics of the form of (11.28), assessing their properties as test statistics for testing for hitchhiking from a selective sweep at a linked selected locus and also for testing for recent population growth. Once again, the coalescent is used to find empirical significance points of these statistics, leading to a comparison of their power properties for these tests.

If a hitchhiking event did in fact occur, can we estimate the time since it concluded? Perlitz and Stephan (1997) suggest an approach to estimating this time that depends on the assumption that θ is known, or at least can be estimated or assumed to lie in some interval of values. The mean number of segregating sites in a sample of n genes if there was no selective sweep in the recent past is given by the stationary value (9.53). One possible explanation for observing a number of segregating sites less than this mean is that a hitchhiking event concluded recently in the past and that the actual number of segregating sites has not had time to achieve a value close to its stationary value. Perlitz and Stephan (1997) find an expression for the mean number of segregating sites in the sample of n genes, given that a hitchhiking event occurred at time t in the past. This is a monotonically increasing function of t , as would be expected, and by equating this expected value to the observed number of segregating sites in the sample, an estimate of t may be found.

11.3.6 *Combining Infinitely Many Alleles and Infinitely Many Sites Approaches*

Strobeck (1987) proposed a test for population subdivision in the presence of neutrality that may equally be used as a test for neutrality in a random-mating population. The probability distribution of the number K of allelic types in the infinitely many alleles model is given by (3.84), and thus

$$\text{Prob}(K \leq k) = \frac{\sum_{i=1}^k |S_n^i| \theta^i}{S_n(\theta)}. \quad (11.30)$$

We denote the left-hand side in (11.30) by $T(K)$. If K were a continuous random variable, statistical theory would show that $T(K)$ has a uniform distribution in $(0, 1)$, so that, for example, for any value α in $(0, 1)$, $\text{Prob}(T(K) \leq \alpha) = \alpha$.

The concept behind the Strobeck procedure is that in subdivided populations, the value of the infinitely many alleles estimate $\hat{\theta}_k$ given in (9.32) should differ from the value of the Tajima infinitely many sites estimate $\hat{\theta}_T$. Strobeck then suggested that in the neutral case, a suitable statistic to

test for population subdivision is

$$\frac{\sum_{i=1}^k |S_n^i| \hat{\theta}_T^i}{S_n(\hat{\theta}_T)}, \quad (11.31)$$

and that the random-mating hypothesis be rejected (with Type I error α) if the value of this statistic is less than α . The corresponding procedure, in a random-mating population, would be to reject neutrality in favor of a selective alternative if the statistic were less than α .

Fu (1996) showed that this procedure does not provide a test with Type I error α . Nevertheless, he used it as a basis for a test of neutrality using as test statistic the quantity W , defined by

$$W = \frac{\sum_{i=1}^k |S_n^i| \hat{\theta}_s^i}{S_n(\hat{\theta}_s)}, \quad (11.32)$$

where $\hat{\theta}_s$ is the estimate of θ given by (9.60), found from the number of segregating sites in the sample.

The statistic W differs from the Strobeck statistic only in the estimate of θ used. Just as the Strobeck statistic does not have an approximately uniform distribution under neutrality, so also W does not have this distribution under neutrality. Recognizing this, Fu found an approximate neutral theory distribution for W using a logistic regression technique. We do not enter into the details, since complications arise (as for all infinitely many sites tests) because the distribution of W depends on θ , so that procedures similar to those carried out by Simonsen et al. (1995) described in Section 11.3.3 are needed.

Fu (1997) has described a procedure using a statistic similar to the Strobeck statistic (11.31), but defined instead by

$$Q = \frac{\sum_{i=k}^n |S_n^i| \hat{\theta}_T^i}{S_n(\hat{\theta}_T)}. \quad (11.33)$$

If there are many rare alleles in the sample, $\hat{\theta}_T$ will tend to be less than $\hat{\theta}_S$, and as a result, Q will tend to be small and F , defined as $\log(Q/(1-Q))$, will tend to be large and negative. Fu (1997) therefore chose F as an appropriate test statistic, aimed specifically at testing for a significantly large number of low-frequency alleles. (This procedure thus has the same aim as that discussed at the end of Section 11.2.3.) One purpose of this test procedure is as a test hitchhiking alternative to that discussed in the previous section.

11.3.7 Data from Several Unlinked Loci

All the procedures described in Sections 11.3.3-11.3.6 assume data from completely linked sites. In practice, data are often analyzed for several

genes, often unlinked, with several sites segregating within each gene. The data of Takano et al. (1991), for example, analyzed by Tajima (1997), are of this type. These data refer to four essentially unlinked genes (*Adh*, *Amy*, *Pu*, and *Gpdh*) in two populations (north and south) of *Drosophila melanogaster*. Sites within each gene may be taken as completely linked, but sites in different genes may be taken to be unlinked and to have independent evolutionary behavior.

It is necessary, for data of this form, to consider further properties of the three estimators $\hat{\theta}_S$, $\hat{\theta}_T$, and $\hat{\theta}_M$ beyond those discussed in Section 11.3.2, and as a result to find the way in which the neutral hypothesis is to be tested for data from several unlinked genes.

Tajima (1997) considered both these questions. When data arise from several genes, the definition of θ now involves the total mutation rate taken over all sites in all genes considered. We call this θ_{sum} . This is the sum of the various individual θ values for the separate genes, so it is appropriate to simply sum the individual gene estimators of θ to obtain an estimator of θ_{sum} . This can be done for all of the estimators of θ considered in Section 11.3.2.

For data deriving only from any one gene and location, the variance formulas for the three “single gene” estimators given in (9.61), (11.12), and (11.15) are appropriate, since these variances are calculated under the assumption of completely linked sites. However, these variance formulas are not appropriate for estimators of θ_{sum} , since the complete linkage assumption under which they are derived is no longer appropriate. Tajima (1997) correctly uses the result that the variance of any estimator of θ_{sum} is the sum of the variances of the “separate gene” estimators of the “separate gene” θ values, each of which is given by the theory in Section 11.3.2.

As far as the data of Takano et al. (1991) are concerned, it is interesting that even for the same gene at the same location, the numerical values of the estimates of θ often disagree considerably. This might arise because of random fluctuations arising in a neutral case from the very small sample size ($n = 43$) or because, in a selective case, selection affects the three estimators differently.

This comment leads to a discussion of the revisions needed to the tests of neutrality discussed in Sections 11.3.3 and 11.3.4, and in particular, to revisions needed for the calculation of the statistics D , D^* , and F^* . The tests described above are not directly appropriate for data pooled over several genes, since the variance formulas assumed in the statistics are no longer correct when data from unlinked sites are considered. However, they are easily amended, in the following way.

The numerator in each of the revised statistics following the general form of D , D^* , and F^* can be written as $\sum(\hat{\theta}_i - \hat{\theta}_j)$, where $\hat{\theta}_i$ and $\hat{\theta}_j$ are two different estimators of θ found from the segregating sites within one gene, and the sum is taken over all genes in the sample. Because the different genes in the sample are assumed to be unlinked and thus to evolve

independently, the variance of any such sum is the sum, over genes, of the variances of the individual gene $\hat{\theta}_i - \hat{\theta}_j$ values. These variances are found from the single gene theory of Section 11.3.2. Denoting the sum of the corresponding variance estimates by V_{ij} , we obtain “Tajima-like” statistics of the form

$$\frac{\sum(\hat{\theta}_i - \hat{\theta}_j)}{\sqrt{V_{ij}}}. \quad (11.34)$$

As an approximation, the distribution of this statistic can be taken as being close to that of the approximate distribution discussed above for the Tajima statistic.

Although it is not explicitly stated, this appears to be the procedure adopted by Tajima (1997), since his calculated values of test statistics of the form (11.34) agree well with those deriving from (11.34).

For the data of Takano et al. (1991), the values of the three test statistics given by (11.34), for the choices $ij = TS$, $ij = TM$, and $ij = SM$, usually agree in sign but often disagree in numerical value. Further, they usually agree in sign for the four different genes considered. As was the case for the estimation of θ , this might arise because of random effects, given the very small sample size, or because, in a selective case, the different statistics are sensitive to different forms of selection.

The fact that the values of the test statistics usually agree in sign for the four genes considered raises an important point. It has been remarked several times above that tests of neutrality are in effect tests of neutrality together with the various often implicit assumptions made in the testing procedure. One of the latter assumptions, for example, is that there have not been any recent population size bottlenecks in the recent past. The effect of a recent bottleneck mimics the effect of selection. Thus if the values of a test statistic for selection show a consistent deviation from zero across a number of different genes, a plausible explanation is that the deviation is caused by a bottleneck, affecting all genes equally, rather than by selection. Testing for this form of explanation rather than selection will be discussed in Volume II.

A further consideration associated with this is that the procedure using a test statistic of the form of (11.34) tests for overall selection over all gene loci considered. This might, however, not be an interesting test to perform. Further, an overall test such as this might mask selection if selection does act at the different gene loci, but causes negative values of $\hat{\theta}_i - \hat{\theta}_j$ at some loci and positive values at other loci. It is clear from this and the discussion of the previous paragraph that testing procedures using several gene loci together must be conducted with some care.

11.3.8 Data from Unlinked Sites

A situation even more extreme than that considered in Section 11.3.7 arises when the data arise from a number of unlinked segregating sites. While it is possible to extend the theory of Section 11.3.7 and to devise further “Tajima-like” testing procedures, for the case of unlinked sites a better approach is possible. For this case the number of segregating sites S is a sufficient statistic for θ , and thus a conditional test, similar in spirit to those used for infinitely many alleles data, can be used. In this case the test statistic may be taken as some function of the frequencies of the two nucleotides at any segregating site.

We assume that there are n nucleotides at each of s segregating sites in the sample. The probability distribution (9.63) of the number j of times that the mutant nucleotide is observed at any one of these sites shows that the mean and variance of the total number of times $H = \sum j$ that the mutant nucleotide arises at the various sites are, respectively,

$$\mu_H = \frac{s(n-1)}{g_1} \quad \text{and} \quad \sigma_H^2 = \frac{sn(n-1)}{2g_1} - \left(\frac{s(n-1)}{g_1} \right)^2. \quad (11.35)$$

If the mutant nucleotide at each site can be recognized, a z -like statistic of the form $(h - \mu_H)/\sigma_H$ can be formed and used to test for neutrality, where h is the observed value of H .

If it is unknown which of two segregating nucleotides is the mutant, it is necessary to amend this procedure and use as test statistic a quantity that remains unaltered if j is replaced by $n - j$. One possibility is to replace the statistic H by the statistic $K = \sum j(n - j)$, the sum being taken over all the s sites segregating in the sample. The mean and variance of this statistic are

$$\mu_K = \frac{sn(n-1)}{2g_1} \quad \text{and} \quad \sigma_K^2 = \frac{sn^2(n^2-1)}{12g_1} - \left(\frac{sn(n-1)}{2g_1} \right)^2. \quad (11.36)$$

Once again, a z -like statistic can be formed and used to test for neutrality.

The diffusion approximation for this procedure is described by Ewens, (1979, Section 9.8). In the diffusion approximation, the distribution of j/n may be written as

$$f(x) = x^{-1} \{\log(n-1)\}^{-1}, \quad n^{-1} \leq x \leq 1 - n^{-1}. \quad (11.37)$$

If it is unknown which of two segregating nucleotides is the mutant, it is necessary to use as test statistic some function that remains unaltered if x is replaced by $1 - x$. The most convenient such statistic is

$$w = \frac{|\log\{(1-x)/x\}|}{\log(n-1)}. \quad (11.38)$$

Under the approximation made above, this statistic has a uniform distribution $(0, 1)$ under the hypothesis of selective neutrality. Alternatively, under

neutrality,

$$y = -2 \log w \quad (11.39)$$

has a chi-square distribution with 2 degrees of freedom. If we are interested in heterotic selection, we reject the neutrality hypothesis for significantly large values of y . A suitable test statistic of neutrality against a heterosis alternative would be $\sum y_i$, which under neutrality has a chi-square distribution with $2s$ degrees of freedom, where s is the observed number of sites segregating in the sample.

With data such as those of Takano et al. (1991), for which sites within any one gene can be taken as completely linked and sites in different genes can be taken as unlinked, this procedure is not valid since the site-to-site independence assumption implicit in it does not then apply. One approach to this problem is to continue to use $\sum y_i$ as test statistic, but to find its null hypothesis distribution by simulation, using (as for several of the statistics considered above) a simulated coalescent process.

An informal procedure for assessing neutrality parallel to the method of Coyne (1976) discussed in Section 11.2.3 is possible whether sites are linked or unlinked. In this procedure a comparison is made between the observed number of sites at which there are j nucleotides of one type and $n - j$ of another with values given by the conditional frequency spectrum (9.67), the conditioning being on the observed value s of S . In this comparison allowance must be made for the fact that the sum of the terms in this conditional frequency spectrum is $2s$, since it is not known at each site which is the “original” nucleotide and which is the mutant. An example of the comparison of the observed conditional frequency spectrum and the neutral theory expected conditional frequency spectrum for the data of Takano et al. (1991) is given in Figures 3 and 4 of Tajima (1997). This shows in a useful visual way how the observed values of J differ from their neutral theory mean values.

11.3.9 Tests Based on Historical Features

It was mentioned at the end of Section 9.6.2 that procedures for estimating θ in the infinitely many sites model that rely on historical features are currently under intense investigation. These procedures were to some extent motivated by the corresponding attempt to devise tests of neutrality based on historical features, using the mechanism of the coalescent, initiated by Fu and Li (1993) and Fu (1996). These tests are more complicated than those described above, and research on them still continues, so the details of these procedures will be discussed in Volume II.

12

Looking Backward in Time: Population and Species Comparisons

12.1 Introduction

Perhaps the best-known retrospective activity in evolutionary genetics is the reconstruction (more accurately, estimation) of the phylogenetic tree of a collection of contemporary populations or species, given genetic data from these populations or species. In this chapter we consider stochastic processes describing, with greater or lesser accuracy, the evolution of the genetic constitution of several populations or species, all descended from a common ancestor population or species, in order to carry out this estimation procedure. We shall use the expression “different population” in this analysis, taking this to mean different species if appropriate.

In this activity we consider a far longer time scale than that considered in previous chapters. For example, we have previously considered aspects of the time until one allele substitutes for another in some population. In the phylogenetic tree estimation process we suppose, because of the far longer time scale considered, that these substitutions are in effect instantaneous.

The data used for the phylogenetic tree estimation normally consists of DNA sequences, so the analysis in this chapter is based on the infinitely many sites model appropriate for these sequences. In previous chapters we have examined aspects of the nature of the variation of DNA sequences within a population. However, there is comparatively little variation at the nucleotide level between members of the same population. For example, two randomly chosen humans typically have different nucleotides at only one site in about 500 to 1000. To a sufficient level of approximation it

is reasonable to assume, for the great majority of sites, that a single nucleotide predominates in the population. (If this were not so, the concept of a paradigm “human genome” for our own species would be meaningless.) Thus in the analysis below we assume a situation close to genetic uniformity within any population, and then use expressions like “the nucleotide at a given site in a population” rather than the more precise “the predominant nucleotide at a given site in a population”.

Several popular phylogenetic tree estimation processes are purely algorithmic. That is, they start with DNA sequences from the populations of interest and by purely algorithmic processes estimate a phylogenetic tree from these sequences. The neighbor-joining and parsimony processes are two frequently used algorithmic procedures. The mechanistic aspect of these processes often leads to the expression “tree reconstruction” rather than the more correct “tree estimation”, the latter expression recognizing the many stochastic factors involved in evolution and the sampling process leading to the data analyzed. Our focus in this chapter is on these stochastic factors.

Some of the algorithmic processes employed for tree estimation are based on the concept of a “genetic distance” between two populations. The recognition of the stochastic nature of evolution implies that the construction of such a distance is not straightforward. This matter is discussed further below.

We shall describe the effectively instantaneous change in frequency of a nucleotide from a value close to 0 to a value close to 1 as the *substitution* of one nucleotide by another, meaning more precisely the substitution of the predominant nucleotide by another in the population of interest. The time unit chosen to evaluate the properties of this substitution process is arbitrary, but is often large, perhaps on the order of hundreds of thousands of generations. Our initial analysis focuses on just one nucleotide site, and in particular on the nucleotide at this site that is predominant in the population of interest.

The analysis uses the theory of finite Markov chains, outlined above in Section 2.12, described in that section in terms of abstract “states” $E_1, E_2, E_3, \dots, E_s$. In our case $s = 4$, and the states E_1, E_2, E_3, E_4 are identified with the events that in the population of interest, the predominant nucleotide at the site in question is a, g, c , and t , respectively. Thus the Markov chain process is, for example, in state E_2 at some given time if in the population considered, the predominant nucleotide at the site of interest is g at that time. If unit time in the Markov chain is taken as, for example, 500,000 generations, a change from state E_3 to state E_1 in one time unit means the substitution of the nucleotide c by the nucleotide a after a period of 500,000 generations. If unit time is taken as, for example, 500,000 generations, a period of time n implies a period of $500,000n$ generations. During this time it is possible that for certain time periods various other

states were occupied, that is, that other nucleotides were predominant in the population.

The estimation of a phylogenetic tree is based on the comparison of genetic data from a number of contemporary populations. Various statistical procedures used in this comparison might require tracing up the tree of evolution from one population to a common ancestor and then down the tree to another population. If the stochastic process assumed for tracing upward is to be the same as that for tracing downward, the stochastic process describing the genetic evolution within each population must be reversible. We therefore start by discussing the reversibility criterion in the context of evolutionary models, focusing on DNA substitutions and the 4×4 Markov chain transition matrices used to describe these substitutions.

12.1.1 The Reversibility Criterion

The criterion of reversibility of a Markov chain was discussed in Section 2.12. Reversibility applies only to Markov chains with a stationary distribution, and the criterion that a Markov chain with stationary distribution ϕ be reversible is given in (2.164).

An arbitrary 4×4 transition matrix has twelve free parameters, namely three free transition probabilities in each of the four rows of the transition matrix. (The fourth transition probability in each row is determined by the remaining three.) However, another parameterization, using a different set of the twelve free parameters, is more useful for us in investigating the reversibility requirement. This parameterization was given by Tavaré (1986), and under this parameterization the 4×4 transition matrix is written in the form

$$\begin{bmatrix} 1 - uW & uA\phi_2 & uB\phi_3 & uC\phi_4 \\ uD\phi_1 & 1 - uX & uE\phi_3 & uF\phi_4 \\ uG\phi_1 & uH\phi_2 & 1 - uY & uI\phi_4 \\ uJ\phi_1 & uK\phi_2 & uL\phi_3 & 1 - uZ \end{bmatrix}. \quad (12.1)$$

Here A, B, \dots, L are the twelve free parameters, $(\phi_1, \phi_2, \phi_3, \phi_4)$ is the stationary distribution of the Markov chain, and

$$W = A\phi_2 + B\phi_3 + C\phi_4, \quad X = D\phi_1 + E\phi_3 + F\phi_4,$$

$$Y = G\phi_1 + H\phi_2 + I\phi_4, \quad Z = J\phi_1 + K\phi_2 + L\phi_3.$$

The necessary and sufficient condition for the Markov chain with transition matrix written in the form (12.1) to be reversible is that the equations

$$A = D, \quad B = G, \quad C = J, \quad E = H, \quad F = K, \quad I = L \quad (12.2)$$

all be satisfied. When this requirement is satisfied, the model has six free parameters, which can be taken as $A, B, C, E, F,$ and $I,$ so one can think of

paying for reversibility by losing the choice of six of the twelve parameters in the 4×4 transition matrix.

It is easily checked that when the conditions (12.2) hold, $(\phi_1, \phi_2, \phi_3, \phi_4)$ is indeed the stationary distribution of the model (12.1).

12.2 Various Evolutionary Models

12.2.1 The Jukes–Cantor Model

The simplest (and earliest) model of nucleotide substitution is the Jukes–Cantor model (Jukes and Cantor (1969)). Using the convention of the states of a Markov chain given in Section 12.1, the transition matrix P for this model is given by

$$P = \begin{bmatrix} 1 - 3\alpha & \alpha & \alpha & \alpha \\ \alpha & 1 - 3\alpha & \alpha & \alpha \\ \alpha & \alpha & 1 - 3\alpha & \alpha \\ \alpha & \alpha & \alpha & 1 - 3\alpha \end{bmatrix}. \quad (12.3)$$

Thus in the Jukes–Cantor model it is assumed that whatever the nucleotide in the population is at any time, the three other nucleotides are equally likely to substitute for it. The model therefore possesses an unrealistic assumption of symmetry, and thus may not reasonably be used as an accurate evolutionary model. We discuss more realistic models below. However, several formulas used in phylogeny theory are based on the Jukes–Cantor model, often without explicit recognition of this fact, so we now discuss the properties of this model in perhaps greater detail than its intrinsic value warrants.

In the model (12.3) α is a parameter depending on the time scale chosen: If unit time were chosen as 500,000 generations, α would take a value smaller than it would if unit time were chosen as 1,000,000 generations. Whatever time scale is chosen, it is clearly necessary that α be less than $\frac{1}{3}$.

Elementary Markov chain theory shows that the stationary distribution $\phi = (\phi_1, \phi_2, \phi_3, \phi_4)'$ for this model, defined in (2.157), is the uniform distribution

$$(\phi_1, \phi_2, \phi_3, \phi_4)' = (.25, .25, .25, .25)', \quad (12.4)$$

as might be expected from the symmetry of the model. The results of Section 12.1.1 then show that this model is reversible.

It is straightforward to show for this model that whatever the predominant nucleotide in the population is at time 0, the probability that this is also the predominant nucleotide at time n is

$$\frac{1}{4} + \frac{3}{4} (1 - 4\alpha)^n, \quad (12.5)$$

and the probability that some other specified nucleotide is the predominant nucleotide at time n is

$$\frac{1}{4} - \frac{1}{4} (1 - 4\alpha)^n. \quad (12.6)$$

12.2.2 The Kimura Model and Its Generalizations

The highly symmetric assumptions implicit in the Jukes–Cantor model are not realistic. A *transition*, that is, the replacement of one purine by the other (for example, of a by g) or of one pyrimidine by the other, is in practice more likely than a *transversion*, that is, the replacement of a purine by a pyrimidine or of a pyrimidine by a purine. Kimura (1980) proposed a (continuous-time) two-parameter model to allow for this. The transition matrix P for the discrete-time version of this model, with the ordering of states given in Section 12.1, is

$$\begin{bmatrix} 1 - \alpha - 2\beta & \alpha & \beta & \beta \\ \alpha & 1 - \alpha - 2\beta & \beta & \beta \\ \beta & \beta & 1 - \alpha - 2\beta & \alpha \\ \beta & \beta & \alpha & 1 - \alpha - 2\beta \end{bmatrix}. \quad (12.7)$$

Here α is the probability of a transition in one time unit, while β is the probability that a purine is substituted by a nominated pyrimidine in one time unit and is also the probability that a pyrimidine is substituted by a nominated purine in one time unit. It is, of course, required that $\alpha + 2\beta < 1$.

The stationary distribution for this model is easily shown to be the discrete uniform distribution given in (12.4), and from this the results of Section 12.1.1 show that the model is reversible.

It can also be shown for this model that whatever the predominant nucleotide at time 0 at any site, the probability that this is also the predominant nucleotide at time n is

$$\frac{1}{4} + \frac{1}{4} (1 - 4\beta)^n + \frac{1}{2} (1 - 2(\alpha + \beta))^n. \quad (12.8)$$

If the initial nucleotide is a purine, the probability that at time n the predominant nucleotide is the other purine is

$$\frac{1}{4} + \frac{1}{4} (1 - 4\beta)^n - \frac{1}{2} (1 - 2(\alpha + \beta))^n. \quad (12.9)$$

A parallel remark holds for pyrimidines. The probability that after n time units a purine has been substituted by a specific pyrimidine is

$$\frac{1}{4} - \frac{1}{4} (1 - 4\beta)^n, \quad (12.10)$$

and the probability that it has been replaced by one or the other pyrimidine is

$$\frac{1}{2} - \frac{1}{2} (1 - 4\beta)^n. \quad (12.11)$$

A parallel remark holds for the replacement of a pyrimidine by a purine.

Although the Kimura model is more realistic than the Jukes–Cantor model, it still does not provide a satisfactory evolutionary model. Apart from the assumptions implied in the form of the transition matrix P for the model, the uniform stationary distribution implied by the model is not realistic. There are various increasingly realistic, but at the same time increasingly complex, generalizations of this model in the literature, leading, for example, to models with nonuniform stationary distributions, but the increasing complexity implies that the criterion of reversibility for a complex model is less likely to hold. This means that in practice, a compromise must be reached, in the modeling process, between a simple model allowing a tractable analysis and satisfying the reversibility criterion, and a more realistic model that is difficult to analyze and might not be reversible.

One model more complex than the Kimura model is that of Blaisdell (1985), which allows different within-transition and within-transversion rates. The transition matrix P for this model is

$$\begin{bmatrix} 1 - \alpha - 2\gamma & \alpha & \gamma & \gamma \\ \beta & 1 - \beta - 2\gamma & \gamma & \gamma \\ \delta & \delta & 1 - \beta - 2\delta & \beta \\ \delta & \delta & \alpha & 1 - \alpha - 2\delta \end{bmatrix}. \quad (12.12)$$

The stationary distribution of this Markov chain is found to be

$$\left(\frac{\delta(\beta+\gamma)}{\theta(\alpha+\beta+2\gamma)}, \frac{\delta(\alpha+\gamma)}{\theta(\alpha+\beta+2\gamma)}, \frac{\gamma(\alpha+\delta)}{\theta(\alpha+\beta+2\delta)}, \frac{\gamma(\beta+\delta)}{\theta(\alpha+\beta+2\delta)} \right)', \quad (12.13)$$

where $\theta = \gamma + \delta$. This stationary distribution and the elements in the transition matrix (12.12) show that this model is not reversible. On the other hand, the elements in the stationary distribution can now all be different from one another, a property not enjoyed by the Jukes–Cantor and Kimura models discussed above.

12.2.3 The Felsenstein Models

A form of generalization of the Jukes–Cantor model different from those considered above was introduced by Felsenstein (1981), whose notation we adopt here. In these models the probability of substitution of any nucleotide by another is proportional to the stationary probability of the substituting

nucleotide. This implies a transition matrix P of the form

$$\begin{bmatrix} 1 - u + u\phi_1 & u\phi_2 & u\phi_3 & u\phi_4 \\ u\phi_1 & 1 - u + u\phi_2 & u\phi_3 & u\phi_4 \\ u\phi_1 & u\phi_2 & 1 - u + u\phi_3 & u\phi_4 \\ u\phi_1 & u\phi_2 & u\phi_3 & 1 - u + u\phi_4 \end{bmatrix}, \quad (12.14)$$

where $(\phi_1, \phi_2, \phi_3, \phi_4)'$ is the stationary distribution and u is a parameter of the model. (It is easily checked that the stationary distribution for the model defined by (12.14) is indeed $(\phi_1, \phi_2, \phi_3, \phi_4)'$.)

A second Felsenstein model (Felsenstein and Churchill (1996); see also Kishino and Hasegawa (1989)) is more general than that given by (12.14), and is important because it is the evolutionary model used in the PHYLIP phylogenetic tree estimation package. This model has a transition matrix similar to that of (12.14), except that the upper-left 2×2 submatrix of the 4×4 matrix in (12.14) is replaced by

$$\begin{bmatrix} 1 - u + u\phi_1 - \frac{uK\phi_2}{\phi_1 + \phi_2} & u\phi_2 + \frac{uK\phi_2}{\phi_1 + \phi_2} \\ u\phi_1 + \frac{uK\phi_1}{\phi_1 + \phi_2} & 1 - u + u\phi_2 - \frac{uK\phi_1}{\phi_1 + \phi_2} \end{bmatrix}, \quad (12.15)$$

and the lower-right 2×2 submatrix of the 4×4 matrix in (12.14) is replaced by

$$\begin{bmatrix} 1 - u + u\phi_3 - \frac{uK\phi_4}{\phi_3 + \phi_4} & u\phi_4 + \frac{uK\phi_4}{\phi_3 + \phi_4} \\ u\phi_3 + \frac{uK\phi_3}{\phi_3 + \phi_4} & 1 - u + u\phi_4 - \frac{uK\phi_3}{\phi_3 + \phi_4} \end{bmatrix}. \quad (12.16)$$

The transition matrix defined jointly by (12.14), (12.15) and (12.16), as with the simpler model (12.14), has stationary distribution $(\phi_1, \phi_2, \phi_3, \phi_4)$. From this it is easily shown that the model is reversible. The quantity K is positive and is a further parameter of the model: Larger values of K increase transition substitution rates compared to those in the model (12.14).

Although the model (12.14) generalizes the Jukes–Cantor model, to which it reduces if $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 1/4$, it does not generalize the Kimura two-parameter model (12.7). On the other hand, the model defined jointly by (12.14), (12.15), and (12.16) does generalize the Kimura two-parameter model, reducing to that model when the stationary distribution is uniform. (This requires the identifications of the parameters α and β in the Kimura model with $u(2K + 1)/4$ and $u/4$, respectively.) It also, of course, generalizes the model (12.14), to which it reduces when $K = 0$.

A model rather similar to that defined jointly by (12.14), (12.15), and (12.16) was introduced by Hasegawa et al. (1985). In this model the transition probability matrix P is of the form

$$\begin{bmatrix} 1 - u\phi_2 - v\phi_A & u\phi_2 & v\phi_3 & v\phi_4 \\ u\phi_1 & 1 - u\phi_1 - v\phi_A & v\phi_3 & v\phi_4 \\ v\phi_1 & v\phi_2 & 1 - u\phi_4 - v\phi_B & u\phi_4 \\ v\phi_1 & v\phi_2 & u\phi_3 & 1 - u\phi_3 - v\phi_B \end{bmatrix}, \quad (12.17)$$

where $\phi_A = \phi_c + \phi_t$, $\phi_B = \phi_a + \phi_g$. This model is an amalgam of the Kimura model (12.7) and the simpler Felsenstein model (12.14), and includes these as particular cases. The stationary distribution in this model is $(\phi_1, \phi_2, \phi_3, \phi_4)'$, as the notation anticipates, and this model is also reversible.

12.3 Some Implications

12.3.1 Introduction

In this section we consider some of the immediate implications of the calculations given in the previous section for the Jukes–Cantor and the Kimura models.

Before doing so, we observe that most calculations used for phylogenetic estimation use continuous-time Markov processes rather than the discrete-time Markov chains considered above. We therefore list here the continuous-time analogues of relevant calculations for the Jukes–Cantor and the Kimura models, and use these in the discussion in this section. Specifically, the continuous-time analogues of the Jukes–Cantor model expressions in (12.5) and (12.6) are

$$q_1 = \frac{1}{4} + \frac{3}{4} e^{-4\alpha t} \quad \text{and} \quad q_2 = \frac{1}{4} - \frac{1}{4} e^{-4\alpha t}, \quad (12.18)$$

respectively, and the continuous-time analogues of the Kimura model expressions (12.8), (12.9), and (12.11) are

$$q_3 = \frac{1}{4} + \frac{1}{4} e^{-4\beta t} + \frac{1}{2} e^{-2(\alpha+\beta)t}, \quad q_4 = \frac{1}{4} + \frac{1}{4} e^{-4\beta t} - \frac{1}{2} e^{-2(\alpha+\beta)t}, \quad (12.19)$$

and

$$q_5 = \frac{1}{2} - \frac{1}{2} e^{-4\beta t}, \quad (12.20)$$

respectively.

12.3.2 The Jukes–Cantor Model

We start by considering various implications of the Jukes–Cantor expressions (12.18). Suppose that two populations split at time t in the past. Then the same nucleotide type arises at a given site in both contemporary populations if they are both copies of the ancestral nucleotide at the time of the split (probability q_1^2) or if they are both copies of some other nucleotide (probability $3q_2^2$). The total probability r_1 is then given by

$$r_1 = q_1^2 + 3q_2^2 = \frac{1}{4} + \frac{3}{4} e^{-8\alpha t}. \quad (12.21)$$

This is the analogue of q_1 . Similarly, the analogue of q_2 is

$$r_2 = \frac{1}{4} - \frac{1}{4} e^{-8\alpha t}. \quad (12.22)$$

These results can be found immediately by reversibility, considering the stochastic process going up one line of ascent from one of the two populations to the common ancestor, and then down the other line of descent to the other population. Thus the probabilities r_1 and r_2 can be found from q_1 and q_2 , respectively, simply by replacing t by $2t$, and this gives the expressions in (12.21) and (12.22).

We write $p = 1 - r_1 = 3r_2$ as the probability that the two nucleotides differ, so that

$$p = \frac{3}{4} (1 - e^{-8\alpha t}). \quad (12.23)$$

From this,

$$\alpha t = -\frac{1}{8} \log \left(1 - \frac{4}{3} p \right). \quad (12.24)$$

The probability p can be estimated unbiasedly by the proportion \hat{p} of nucleotide sites at which the two populations being compared differ in their respective homologous DNA sequences sampled. Common practice is, then, to estimate αt by $\widehat{\alpha t}$, defined by

$$\widehat{\alpha t} = -\frac{1}{8} \log \left(1 - \frac{4}{3} \hat{p} \right). \quad (12.25)$$

If an extrinsic estimate of α is available, this gives an estimate of the time t since the initial split of the two populations.

This procedure gives a biased estimator of t , and indeed, the estimator is not even defined if $p \geq 3/4$. It also depends on the unrealistic Jukes–Cantor model. Despite this, this estimator appears, often uncritically, in the literature. Since this time t is a critical feature of a phylogenetic tree estimation, we can then expect biased estimation of phylogenetic trees if the estimator, or its generalizations when many populations are considered, is used.

The estimator (12.25) does have one interesting property. Write $\lambda = 3\alpha$ as the “total” substitution rate and suppose that the estimator $\hat{p} = D/N$ is derived from the comparison of N nucleotide sites in one population and the corresponding N sites in the other population, where D is the number of sites for which the two sequences compared differ. Then the total mean number of substitutions down both lines of descent from the initial ancestor population is $\nu = 2N\lambda t = 6N\alpha t$. This would then be estimated, from (12.25), by

$$\hat{\nu} = -\frac{3N}{4} \log \left(1 - \frac{4}{3} \hat{p} \right). \quad (12.26)$$

If \hat{p} is small, a Taylor series approximation gives

$$\hat{\nu} \approx N\hat{p} + \frac{2}{3} N\hat{p}^2 = D + \frac{2D^2}{3N}. \quad (12.27)$$

This implies that the estimated number of substitutions is somewhat larger than the observed number, the difference arising from the estimated number of sites at which either the same substitution arose in the two populations together with the estimated number of sites, or at which two or more substitutions occurred down one or other line of descent, concluding with the same nucleotide. If, for example, $N = 3,000$, $D = 300$, the approximation (12.27) leads to an estimated total of 320 substitutions, 20 of which are estimated not to be observed in the contemporary sample.

As remarked above, several tree estimation algorithms depend on the use of some measure of genetic distance between two populations. The argument given above implies that $\hat{\nu}$ forms a better measure of genetic distance between the two populations than does the count D of sites at which different nucleotides are observed in the two populations.

There are many further implicit assumptions made when α is estimated from data from many sites. One of these is that the value of α is the same at all sites. This assumption is undoubtedly untrue, and with site to site variation in the value of α , (12.25) gives an underestimate of $\bar{\lambda}t$ (Nei (1975, pp. 225–226)), and thus of the genetic distance between the two populations.

The concept of a distance as just described assumes genetic uniformity within populations. A more general definition of distance allows for nucleotide segregation within populations. Here various measures (Sokal and Sneath (1963), Rogers (1972), Hedrick (1971), Cavalli-Sforza and Edwards (1967), and Nei (1972)) have been proposed for various purposes. For evolutionary considerations we require a measure that, if substitutions occur at a constant rate, is proportional to the time t between the splitting of the two populations considered. Nei (1976) showed by computer simulation in the infinitely many alleles case that the expected value of his genetic distance measure D_N increases almost linearly with time, and that this property is not shared by the other distance measures above. The infinitely many alleles model is not appropriate for nucleotide sequence data, so it does not necessarily follow that D_N has this linearity property for these data. It is therefore useful to examine properties of D_N for the infinitely many sites model. We now do this, assuming that the Jukes–Cantor model holds.

We consider two populations that split at time t in the past. Suppose that at a given site, the frequencies of the four nucleotides in one population are given as x_1 , x_2 , x_3 , and x_4 . Suppose that y_1 , y_2 , y_3 , and y_4 are the corresponding (random) frequencies in the other population. For this case

D_N is defined as

$$D_N = -\log \left(\frac{\sum \sum_{i=1}^4 x_i y_i}{\sqrt{\sum \sum_{i=1}^4 x_i^2 \sum \sum_{i=1}^4 y_i^2}} \right), \quad (12.28)$$

the three outer sums being taken over all sites considered.

Now, $E(y_i | x_i) = x_i r_1 + (1 - x_i) r_2$, so that from (12.21) and (12.22),

$$E(y_i | x_i) = \frac{1}{4}(1 - e^{-8\alpha t}) + x_i e^{-8\alpha t}. \quad (12.29)$$

From this,

$$E \sum x_i y_i = \frac{1}{4}(1 - e^{-8\alpha t}) + e^{-8\alpha t} E \sum x_i^2. \quad (12.30)$$

Since we assume essential homogeneity within any population, we make the approximations $E \sum x_i^2 \approx E \sum y_i^2 \approx 1$. Inserting these approximate values in (12.28) and (12.30), and assuming a large number of nucleotide sites examined, so that random sampling effects can be ignored, we obtain

$$D_N \approx -\log \left(\frac{1}{4} + \frac{3}{4} e^{-8\alpha t} \right). \quad (12.31)$$

(The right-hand side in this expression could also be obtained directly from (12.21).) If terms of order $(\alpha t)^3$ are ignored, D_N is approximately $6\alpha t(1 - \alpha t)$, and is thus essentially a linear function of t only when $\alpha t < 0.1$.

12.3.3 The Kimura Model

If the Kimura evolutionary model is assumed, a set of calculations similar to those leading to (12.26) and (12.27) can be made. In this case both the parameters αt and βt are estimated, the data used being the numbers of transitional and transversional differences observed in the data. Suppose that at any time the predominant nucleotide is a specified purine (respectively pyrimidine). We must first find the probability p_1 that at time t later the predominant nucleotide is the other purine (respectively pyrimidine). This probability is given by

$$p_1 = \frac{1}{4} + \frac{1}{4} e^{-4\beta t} - \frac{1}{2} e^{-2(\alpha+\beta)t}. \quad (12.32)$$

If at any time the predominant nucleotide is a specified purine (respectively pyrimidine), then the probability p_2 that at time t later the predominant nucleotide is one or other pyrimidine (respectively purine) is

$$p_2 = \frac{1}{2} - \frac{1}{2} e^{-4\beta t}. \quad (12.33)$$

These probabilities may be estimated by the respective sample proportions $\hat{p}_1 = n_1/N$ and $\hat{p}_2 = n_2/N$, where in a sample of N sites there are n_1 sites at which one purine (pyrimidine) arises at some site in the sequence

of one population and the other pyrimidine (purine) arises at that site in the sequence of the other population, and n_2 sites at which a purine occurs in one sequence and a pyrimidine in the other.

From these equations αt and βt may be estimated by solving the equations

$$\hat{p}_1 = \frac{1}{4} + \frac{1}{4}e^{-4\hat{\beta}t} - \frac{1}{2}e^{-2(\hat{\alpha}t + \hat{\beta}t)} \quad (12.34)$$

and

$$\hat{p}_2 = \frac{1}{2} - \frac{1}{2}e^{-4\hat{\beta}t}. \quad (12.35)$$

These estimation procedures are subject to the same qualifications as were made for the parallel Jukes–Cantor model estimation procedure.

The mean rate at which substitutions of one form or another arise is $\alpha + 2\beta$, and from this it is found that the mean number of substitutions in the evolution of the two populations since their common ancestor is $\nu = 2N(\alpha + 2\beta)t$. This can be estimated from the estimates of αt and βt implicit in (12.34) and (12.35). The result is that the estimate of ν is

$$\hat{\nu} = \frac{1}{2}N \log(1 - 2\hat{p}_1 - \hat{p}_2) + \frac{1}{4}N \log(1 - 2\hat{p}_2). \quad (12.36)$$

Suppose that in a sample of $N = 3000$ sites there are 210 transitional and 90 transversional differences. Then $\hat{p}_1 = 0.07$ and $\hat{p}_2 = 0.03$. Then (12.36) leads to an estimate of ν of 326. This exceeds the observed number 300 of nucleotide differences, indicating that it is estimated that 26 substitutions are estimated not to be “observed” in the contemporary sample.

More important, it differs from the estimated value of 320 that would arise if the Jukes–Cantor estimation procedure, which does not distinguish between transitional and transversional substitutions, had been used. This implies that if the Kimura model faithfully describes the evolutionary process, a bias will arise in the estimation of the genetic distance between the two populations if the simple Jukes–Cantor model is used for the distance estimation procedure. In practice the Kimura model is itself over-simplified, and even greater biases may be expected if the Jukes–Cantor model is used when a far more complex model is appropriate. This is in addition to the bias inherent in the Jukes–Cantor model itself, described below (12.25). The same conclusion applies for any comparatively simple evolutionary model. Thus phylogenetic trees estimated from simple stochastic models must be viewed with much caution.

12.4 Statistical Procedures

The fact that one is only estimating a phylogenetic tree from contemporary data, rather than constructing it without error, implies that many statisti-

cal issues arise, mainly questions of hypothesis testing. One frequently used test statistic is the well-known $-2 \log \lambda$ statistic, where λ is a ratio of (maximum) likelihoods, one calculated under the null hypothesis and one under the alternative hypothesis. When various criteria are satisfied, $-2 \log \lambda$ has an approximate chi-square distribution when the null hypothesis is true.

One of these criteria is that the null hypothesis be a particular case of the alternative hypothesis. Another is that the hypothesis testing procedure must relate to the value (or values) of some parameter (or parameters) that can take continuous real number values only. There are further criteria also, but we do not discuss them here.

One procedure that is frequently carried out is to test whether some more complex evolutionary model explains the data better than a simpler evolutionary model. Although neither model can be accepted as giving a reasonable description of evolution, we illustrate the problems involved with such a procedure by discussing the test of whether the Kimura two-parameter model explains the data better than the Jukes–Cantor model. In statistical terms, this is a test of the null hypothesis that the parameter β in (12.7) is equal to the parameter α against the alternative that allows the two parameters to take any values.

There are two aspects of this test that deserve discussion. The first is that the null hypothesis (that the Jukes–Cantor model holds) is a particular case of, or is “nested within”, the alternative hypothesis (that the Kimura two-parameter model holds). Thus the first criterion listed above for the use of the $-2 \log \lambda$ approach is satisfied. The second also appears, at first sight, to be satisfied, but if the topology of the phylogenetic tree, which can loosely be thought of as a parameter, is estimated in the procedure, then the second criterion is not satisfied, since the shape of the phylogenetic tree is not a real number.

Even if the shape of the phylogenetic tree is given a priori, this last problem still arises. Part of the estimation procedure is to estimate various DNA sequences at the internal nodes of the phylogenetic tree, and these sequences are not real numbers.

As another problem, suppose that the null hypothesis is the Jukes–Cantor model and the alternative hypothesis is the simple Felsenstein model (12.14) with stationary probability values equal to the observed values in the data. Then neither model is nested within the other and there is no theoretical support for the claim that the null hypothesis distribution of $-2 \log \lambda$ is chi-square. Whelan and Goldman (1999) show that the null hypothesis distribution of $-2 \log \lambda$ is indeed not close to a chi-square in this case, and that in fact, negative values of $-2 \log \lambda$ can arise, an impossibility for a random variable truly having a chi-square distribution.

A third problem concerns testing for a monophyletic group, or clade. A collection of species derived from some internal node in the phylogenetic tree is called a monophyletic group if no other species descends from this node. It is often of interest to test whether some group of species of special

interest forms a monophyletic group. Here the null hypothesis is that this collection of species does indeed form a monophyletic group. The maximum of the likelihood of the data under this hypothesis can in principle be formed as well as the maximum of the likelihood of the data when no monophyly claim is made. A value of $-2 \log \lambda$ can then, in principle, be calculated, but this does not have a chi-square distribution under the null hypothesis, since the shape of the phylogenetic tree is estimated as part of both likelihood procedures.

Apart from these and other statistical issues, there are very difficult problems of computing to be overcome when likelihood calculations are made. We do not discuss these, or further statistical problems, here: An excellent summary of these matters is given by Goldman, Anderson and Rodrigo (2000).

Appendix A

Eigenvalue Calculations

Let X_t ($t = 0, 1, 2, \dots$) be a (possibly vector) Markovian random variable with state space $\{0, 1, \dots, M\}$ and transition matrix P . Suppose that $p_{00} = p_{MM} = 1$, that the states $\{1, 2, \dots, M - 1\}$ are transient, and that there exists an integer m such that $p_{ij}^{(m)} > 0$ for $1 \leq i \leq M - 1$ and all j .

Suppose further that a function $f(X)$ exists such that $f(0) = f(M) = 0$, $f(i) > 0$ otherwise, and for which

$$E\{f(X_{t+1}) \mid X_t\} = \lambda_2 f(X_t) \tag{A.1}$$

for some constant λ_2 . Then λ_2 is real and positive and is the leading nonunit eigenvalue of P .

The proof is almost immediate. The matrix P has two unit eigenvalues, and if the first and last rows and columns of P are removed, the remaining eigenvalues of P are those of the resultant matrix Q . Denoting $(f(1), \dots, f(M - 1))$ by \mathbf{f}' , we see that (A.1) and the assumption that $f(0) = f(M) = 0$ show that

$$Q\mathbf{f} = \lambda_2\mathbf{f}.$$

Since the matrix Q satisfies the conditions of Theorem 2.2 of Karlin and Taylor (1975, p. 545), the Frobenius theory of their Theorem 2.1 proves the desired result.

Appendix B

Significance Levels for \hat{F}

Empirical¹ significance levels (2.5%, 5%, 97.5%) of the test statistic \hat{F} for given values of k and n . "N.S." means significance is not possible at the probability level indicated.

n	Prob	k							
		3	5	7	10	15	20	25	30
100	2.5%	0.36	0.27	0.20	0.15	0.11	0.08	0.06	0.05
	5%	0.40	0.29	0.21	0.16	0.11	0.08	0.07	0.05
	97.5%	N.S.	0.87	0.71	0.48	0.33	0.22	0.15	0.12
200	2.5%	0.37	0.28	0.22	0.17	0.12	0.09	0.08	0.06
	5%	0.41	0.30	0.23	0.18	0.13	0.10	0.08	0.07
	97.5%	N.S.	0.89	0.78	0.63	0.41	0.29	0.23	0.17
300	2.5%	0.38	0.29	0.23	0.17	0.12	0.10	0.08	0.07
	5%	0.43	0.31	0.24	0.19	0.13	0.11	0.08	0.07
	97.5%	N.S.	0.93	0.83	0.68	0.48	0.34	0.26	0.20
400	2.5%	0.41	0.29	0.23	0.17	0.13	0.10	0.08	0.07
	5%	0.45	0.31	0.25	0.19	0.14	0.11	0.09	0.08
	97.5%	0.99	0.93	0.86	0.71	0.51	0.35	0.28	0.21
500	2.5%	0.40	0.28	0.24	0.18	0.13	0.11	0.09	0.07
	5%	0.45	0.31	0.25	0.20	0.15	0.11	0.09	0.08
	97.5%	0.99	0.93	0.86	0.74	0.52	0.41	0.31	0.24

¹Based on 1,000 independent drawings for each (k, n) combination from the distribution (9.30). Values by kind courtesy of R. Anderson.

Appendix C

Means and Variances of \hat{F}

Values of $E(\hat{F} | k)$ and $\text{var}(\hat{F} | k)$ for various k, n values. Values by kind courtesy of R. Anderson.

		k							
		3	5	7	10	15	20	25	30
		$E(\hat{F} k)$							
n	100	0.671	0.490	0.376	0.271	0.176	0.125	0.094	0.073
	200	0.705	0.532	0.421	0.313	0.212	0.156	0.120	0.096
	300	0.722	0.554	0.444	0.336	0.232	0.173	0.135	0.110
	400	0.732	0.568	0.459	0.351	0.245	0.185	0.146	0.119
	500	0.740	0.579	0.470	0.362	0.255	0.193	0.153	0.126
		$\text{var}(\hat{F} k)$							
n	100	0.0325	0.0254	0.0169	0.0089	0.0033	0.0013	0.0006	0.0003
	200	0.0350	0.0306	0.0224	0.0133	0.0058	0.0028	0.0014	0.0008
	300	0.0359	0.0331	0.0253	0.0159	0.0075	0.0038	0.0021	0.0012
	400	0.0364	0.0346	0.0272	0.0176	0.0087	0.0046	0.0026	0.0015
	500	0.0366	0.0356	0.0286	0.0190	0.0096	0.0052	0.0030	0.0018

References

- Abramowitz, M., Stegun, I.A.: *Handbook of Mathematical Functions*. New York: Dover Publ. Inc., 1965.
- Akin, E.: Cycling in simple genetic systems. *J. Math. Biol.* **13**, 305–324 (1982).
- Anderson, R.: Unpublished M.Sc. thesis, Monash University (1978).
- Anderson, W.W.: Genetic equilibrium and population growth under density-regulated selection. *Amer. Nat.* **105**, 489–498 (1971).
- Avery, P.J.: The effect of random selection coefficients on populations of finite size – some particular models. *Gen. Res. Camb.* **29**, 97–112 (1977).
- Avery, P.J., Hill, W.G.: The effect of linkage disequilibrium on the genetic variance of a quantitative trait. *Adv. App. Prob.* **10**, 4–6 (1978).
- Barker, W.C., Ketcham, L.K., Dayhoff, M.O.: A comprehensive examination of protein sequences for evidence of internal gene duplication. *J. Mol. Evol.* **10**, 265–281 (1978).
- Bennett, J.H.: On the theory of random mating. *Ann. Eugenics* **18**, 311–317 (1954).
- Bennett, J.H.: Selectively balanced polymorphism at a sex-linked locus. *Nature* **180**, 1363–1364 (1957).
- Berger, E.M.: Pattern and chance in the use of the genetic code. *J. Mol. Evol.* **10**, 319–323 (1978).
- Berger, R.L., Boos, D.D.: P -values maximized over a confidence set for a nuisance parameter. *J. Amer. Stat. Assoc.* **89**, 1012–1016 (1995).
- Blaisdell, B.E.: A method of estimating from two aligned present-day DNA sequences their ancestral composition and subsequent rates of substitution, possibly different in the two lineages, corrected for multiple and parallel substitutions at the same site. *J. Molec. Evol.* **22**, 69–81 (1985).
- Blundell, T.L., Wood, S.P.: Is the evolution of insulin Darwinian or due to selectively neutral mutation? *Nature* **257**, 197–198 (1975).

- Bodmer, W.F.: Differential fertility in population genetics models. *Genetics* **51**, 411–424 (1965).
- Bodmer, W.F., Edwards, A.W.F.: Natural selection and the sex ratio. *Ann. Hum. Genet.* **24**, 239–244 (1960).
- Bodmer, W.F., Felsenstein, J.: Linkage and selection — theoretical analysis of the deterministic two locus random mating model. *Genetics* **57**, 237–265 (1967).
- Bulmer, M.G.: The effect of selection on genetic variability: a simulation study. *Gen. Res.* **28**, 101–117 (1976).
- Bürger, R.: (2000) *The Mathematical Theory of Recombination, Selection, and Mutation*. New York: Wiley, (2000).
- Burt, C.: Quantitative genetics in psychology. *Brit. Jour. of Math. & Stat. Psych.* **24**, 1–21 (1971).
- Cannings, C.: Equilibrium, convergence and stability at a sex-linked locus under natural selection. *Genetics* **56**, 613–618 (1967).
- Cannings, C.: Equilibrium under selection at a multi-allelic sex-linked locus. *Biometrics* **24**, 187–189 (1968).
- Cannings, C.: The latent roots of certain Markov chains arising in genetics: a new approach 1. Haploid models. *Adv. Appl. Prob.* **6**, 260–290 (1974).
- Castle, W.E.: The laws of Galton and Mendel and some laws governing race improvement by selection. *Proc. Amer. Acad. Arts and Sci.* **39**, 233–242 (1903).
- Cavalli-Sforza, L.L., Edwards, A.W.F.: Phylogenetic analysis: models and estimation procedures. *Amer. J. Hum. Genet.* **19**, 233–257 (1974).
- Charlesworth, B.: Selection in populations with overlapping generations. I. The use of Malthusian parameters in population genetics. *Theory Pop. Biol.* **1**, 352–370 (1970).
- Charlesworth, B.: Selection in density-regulated populations. *Ecology* **52**, 469–474 (1971).
- Charlesworth, B.: Selection in populations with overlapping generations. III. Conditions for genetic equilibrium. *Theoret. Pop. Biol.* **3**, 377–395 (1972).
- Charlesworth, B.: Selection in populations with overlapping generations. V. Natural selection and life histories. *Amer. Natur.* **107**, 303–311 (1973).
- Charlesworth, B.: Selection in populations with overlapping generations. VI. Rates of change of gene frequency and population growth rate. *Theoret. Pop. Biol.* **6**, 108–133 (1974).
- Charlesworth, B.: Natural selection in age-structured populations. *Lectures on Mathematics in the Life Sciences* **8**, 69–87 (1976).
- Christiansen, F.B., Frydenberg, O.: Geographical patterns of four polymorphisms in *viviparus* as evidence of selection. *Genetics* **77**, 765–770 (1974).
- Christiansen, F.B.: *Population Genetics of Multiple Loci*. New York, Wiley (2000).
- Cockerham, C.C.: An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882 (1954).
- Cockerham, C.C.: Effects on linkage on the covariances between relatives. *Genetics* **41**, 138–141 (1956).
- Cockerham, C.C.: Higher order probability functions of identity of alleles by descent. *Genetics* **69**, 235–246 (1971).

- Cockerham, C.C., Weir, B.S.: Descent measures for two loci with some applications. *Theoret. Pop. Biol.* **4**, 300–330 (1973).
- Conley, C.C.: Unpublished lecture notes. University of Wisconsin, Madison (1972).
- Cornish-Bowden, A., Marson, A.: Evolution of the non-randomness of protein composition. *J. Mol. Evol.* **10**, 231–240 (1977).
- Coyne, J.A.: Lack of genetic similarity between two sibling species of *Drosophila* as revealed by varied techniques. *Genetics* **84**, 593–607 (1976).
- Crow, J.F., Felsenstein, J.: The effect of assortative mating on the genetic composition of a population. *Eugenics Quart.* **15**, 85–97 (1968).
- Crow, J.F., Kimura, M.: Evolution in sexual and asexual populations. *Am. Nat.* **99**, 439–450 (1965).
- Crow, J.F., Kimura, M.: Evolution in sexual and asexual populations: a reply. *Am. Nat.* **103**, 89–91 (1969).
- Crow, J.F., Kimura, M.: *An Introduction to Population Genetics Theory*. New York: Harper and Row, (1970).
- Crow, J.F., Kimura, M.: The effective number of a population with overlapping generations — a correction and further discussion. *Amer. J. Hum. Genet.* **24**, 1–10 (1972).
- Dayhoff, M.: *Atlas of Protein Sequence and Structure*, Vol. 5. Washington, D.C.: Nat. Biomed. Res. Foundn., (1972).
- Demetrius, L.: Primitivity conditions for growth matrices. *Math. Biosci.* **12**, 53–58 (1971).
- Demetrius, L.: Demographic parameters and neutral selection. *Proc. Nat. Acad. Sci.* **71**, 4645–4647 (1974).
- Demetrius, L.: Reproductive strategies and natural selection. *Am. Nat.* **109**, 243–249 (1975).
- Demetrius, L.: Measures of variability in age-structured populations. *Journ. Theoret. Biol.* **63**, 397–404 (1976).
- Demetrius, L.: Measures of fitness and demographic stability. *Proc. Nat. Acad. Acad. Sci.* **74**, 384–386 (1977).
- Dobzansky, T. *Genetics of the Evolutionary Process*. New York: Columbia University Press 1970.
- Donnelly, P.J.: Partition structure, Polya urns, the Ewens sampling formula, and the ages of alleles. *Theoret. Pop. Biol.* **30**, 271–288 (1986).
- Donnelly, P.J., Tavaré, S.: The ages of alleles and a coalescent. *Adv. Appl. Prob.* **18**, 1–19 (1986).
- Donnelly, P.J., Tavaré, S.: Coalescents and genealogical structure under neutrality. In: *Annual Review of Genetics*, Campbell, A., Anderson, W., Jones, E. (eds.), pp 401–421. Palo Alto, Annual Reviews Inc., (1995).
- Donnelly, P.J., Tavaré, S.: *Progress in Population Genetics and Human Evolution*. New York: Springer, (1997).
- Eaves, L.J., Last, K., Martin, N.G., Jinks, J.L.: A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *Br. J. Math. Statist. Psychol.* **30**, 1–42 (1977).
- Edwards, A.W.F.: The population genetics of “sex-ratio” in *Drosophila pseudoobscura*. *Heredity* **16**, 291–304 (1961).
- Edwards, A.W.F.: On Kimura’s maximum principle in the genetical theory of natural selection. *Adv. Appl. Prob.* **3**, 1–3 (1974).

- Elliott, J.: Eigenfunction expansion associated with singular differential operators. *Trans. Amer. Math. Soc.* **78**, 406–425 (1955).
- Engen, S.: A note on the geometric series as a species frequency model. *Biometrika* **62**, 694–699 (1975).
- Epperson, B.K.: *Geographical Genetics*. Princeton: Princeton University Press (2003).
- Eshel, I.: Selection on sex ratio and the evolution of sex-determination. *Heredity* **34**, 351–361 (1975).
- Eshel, I., Feldman, M.W.: On the evolution effects of recombination. *Theoret. Pop. Biol.* **1**, 88–100 (1970).
- Ethier, S.N., Norman, M. Frank: An error estimate for the diffusion approximation to the Wright–Fisher model. *Proc. Nat. Sci.* **74**, 5096–5098 (1977).
- Ewens, W.J.: The pseudo-transient distribution and its uses in genetics. *J. Appl. Prob.* **1**, 141–156 (1964).
- Ewens, W.J.: A note on the mathematical theory of the evolution of dominance. *Amer. Nat.* **101**, 35–40 (1967).
- Ewens, W.J.: A generalized fundamental theorem of natural selection. *Genetics* **63**, 531–537 (1969a).
- Ewens, W.J.: Mean fitness increases when fitnesses are additive. *Nature* **221**, 1076 (1969b).
- Ewens, W.J.: *Population Genetics*. London: Methuen, (1969c).
- Ewens, W.J.: The transient behavior of stochastic processes, with applications in the natural sciences. Bull. 36th Session I.S.I., 603–622 (1969d).
- Ewens, W.J.: Remarks on the substitutional load. *Theoret. Pop. Biol.* **1**, 129–139 (1970).
- Ewens, W.J.: The sampling theory of selectively neutral alleles. *Theoret. Pop. Biol.* **3**, 87–112 (1972).
- Ewens, W.J.: Testing for increased mutation rate for neutral alleles. *Theoret. Pop. Biol.* **4**, 251–259 (1973).
- Ewens, W.J.: Mathematical and statistical problems arising in the non-Darwinian theory. *Lectures on Mathematics in the Life Sciences* **7**, 25–42 (1974a).
- Ewens, W.J.: A note on the sampling theory for infinite alleles and infinite sites models. *Theoret. Pop. Biol.* **6**, 143–148, (1974b).
- Ewens, W.J.: Remarks on the evolutionary effect of natural selection. *Genetics* **83**, 601–607 (1976).
- Ewens, W.J.: The effective population size in the presence of catastrophes. In: *Mathematical Evolutionary Theory*, Feldman, M. (ed.), pp. 9–26. Princeton: Princeton University Press, (1989).
- Ewens, W.J.: An optimizing principle of natural selection in evolutionary population genetics. *Theoret. Pop. Biol.* **42**, 333–346, (1992).
- Ewens, W.J., Feldman, M.W.: The theoretical assessment of selective neutrality. In: *Population Genetics and Ecology*, Karlin, S., Nevo, E. (eds.), pp. 303–337. New York: Academic Press, 1976.
- Ewens, W.J., Gillespie, J.H.: Some simulation results for the neutral allele model, with interpretations. *Theoret. Pop. Biol.* **6**, 35–57, 1974.
- Ewens, W.J., Kirby, K.: The eigenvalues of the neutral alleles processes. *Theoret. Pop. Biol.* **7**, 212–220 (1975).

- Ewens, W.J., Thomson, G.: Heterozygote selective advantage. *Ann. Hum. Gen.* **33**, 365–376 (1970).
- Ewens, W.J., Thomson, G.: Properties of equilibria in multilocus genetic systems. *Genetics* **87**, 807–819 (1977).
- Fay, J.C., Wu, C.-I.: Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
- Fay, J.C., Wu, C.-I.: The neutral theory in the genomic era. *Curr. Opin. Gen. Dev.* **11**, 642–646 (2002).
- Feldman, M.W.: On the offspring number distribution in a genetic population. *J. Appl. Prob.* **3**, 129–141 (1966).
- Feldman, M.W.: Selection for linkage modification - I. Random mating populations. *Theoret. Pop. Biol.* **3**, 324–346 (1972).
- Feldman, M.W., Christiansen, F.G.: The effect of population subdivision on two loci without selection. *Genet. Res. Camb.* **25**, 151–162 (1975).
- Feldman, M.W., Cavalli-Sforza, L.L.: Darwinian selection and “altruism”. *Theoret. Pop. Biol.* **14**, 268–280 (1978).
- Feldman, M.W., Cavalli-Sforza, L.L.: Further remarks on Darwinian selection and “altruism”. *Theoret. Pop. Biol.* **19**, 251–260 (1981).
- Feldman, M.W., Franklin, I., Thomson, G.J.: Selection in complex genetic system I. the symmetric equilibria of the three-locus symmetric viability model. *Genetics* **76**, 135–162 (1974).
- Feldman, M.W., Karlin, S.: The evolution of dominance - a direct approach through the theory of linkage and selection. *Theoret. Pop. Biol.* **2**, 482–492 (1971).
- Feldman, M.W., Krakauer, J.: *Genetic Modification and Modifier Polymorphisms. Population Genetics and Ecology*. Karlin, S., Nevo, E. (eds.), pp. 547–583. New York: Academic Press, 1976.
- Feldman, M.W., Lewontin, R.C., Franklin, I.R., Christiansen, F.B.: Selection in complex genetic systems III. An effect of allele multiplicity with two loci. *Genetics* **79**, 333–347 (1975).
- Feller, W.: Diffusion processes in genetics. In *Proc. 2nd Berkeley Symp. on Math. Stat. and Prob.* Neyman, J. (ed.), pp. 227–246. Berkeley: University of California Press, 1951.
- Feller, W.: Diffusion processes in one dimension. *Trans. Amer. Math. Soc.* **77**, 1–31 (1954).
- Felsenstein, J.: Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* **68**, 581–597 (1971).
- Felsenstein, J.: Maximum likelihood and minimum-steps method for estimating evolutionary trees from data on discrete characters. *Sys. Zoology* **22**, No. 3, 240–249 (1973).
- Felsenstein, J.: The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
- Felsenstein, J., Yokoyama, S.: The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**, 845–859 (1976).
- Felsenstein, J.: Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Molec. Evol.* **17**, 368–376 (1981).
- Felsenstein, J., and Churchill, G.A.: A hidden Markov model approach to variation among sites in rate of evolution. *Mol. Biol. Evol.* **13**, 93–104 (1996).

- Fenchel, T.M., Christiansen, F.B.: Measuring selection in natural populations. In: *Lectures Notes in Biomathematics* **19**, Christiansen, F.B., Fenchel, T.M. (eds.), Berlin: Springer, 1977.
- Fisher, R.A.: The correlation between relatives on the supposition of Mendelian inheritance. *Trans. of the Roy. Soc. of Edinburgh* **52**, 399–433 (1918).
- Fisher, R.A.: On the dominance ratio. *Proc. Roy. Soc. Edin.* **42**, 321–341 (1922).
- Fisher, R.A.: The arrangements of field experiments. *Jour. of the Min. of Agr. of Great Brit.* **33**, 503–513 (1926).
- Fisher, R.A.: The possible modification of the response of the wild type to recurrent mutation. *Amer. Nat.* **62**, 115–226 (1928a).
- Fisher, R.A.: Two further notes on the origin of dominance. *Amer. Nat.* **62**, 571–574 (1928b).
- Fisher, R.A.: The evolution of dominance - reply to Professor Sewall Wright. *Amer. Nat.* **63**, 553–556 (1929).
- Fisher, R.A.: *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, (1930a).
- Fisher, R.A.: The evolution of dominance in certain polymorphic species. *Amer. Nat.* **64**, 385–406 (1930b).
- Fisher, R.A.: The evolution of dominance. *Biol. Reviews* **6**, 345–368 (1931).
- Fisher, R.A.: Prof. Wright on the theory of dominance. *Amer. Nat.* **68**, 370–374 (1934).
- Fisher, R.A.: The wave of advance of advantageous genes. *Ann. Eug.* **7**, 355–369 (1937).
- Fisher, R.A.: Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63 (1941).
- Fisher, R.A.: Population genetics. The Croonian lecture. *Proc. Roy. Soc. B.* **141**, 510–523 (1953).
- Fisher, R.A.: *The Genetical Theory of Natural Selection* (second revised edit.). New York: Dover, (1958).
- Fisher, R.A.: Heredity (Address to Camb. Uni. Eug. Soc.) *Notes and Records of the Roy. Soc. of London* **31**, 155–162 (1976).
- Fitch, W.M.: Toward defining the course of evolution: minimum change for a specific tree topology. *Sys. Zoology* **20**, No. 4, 406–416 (1971).
- Fitch, W.M.: Margoliash, E.: Construction of phylogenetic trees. *Science* **155**, 279–284 (1967).
- Franklin, I.R., Feldman, M.W.: Two loci with two alleles: linkage equilibrium and linkage disequilibrium can be simultaneously stable. *Theoret. Pop. Biol.* **12**, 95–113 (1977).
- Franklin, I.R., Lewontin, R.C.: Is the gene the unit of selection? *Genetics* **65**, 701–734 (1970).
- Freedman, D.: *Brownian Motion and Diffusions*. San Francisco: Holden-Day, (1971).
- Freeling, M.: Allelic variation at the level of intragenic recombination. *Genetics* **89**, 211–224 (1978).
- Frydenberg, O.: Population studies of a lethal mutant in *Drosophila melanogaster*. I. Behavior in populations with discrete generations. *Hereditas* **50**, 89–116 (1963).
- Fu, Y.-X.: New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**, 557–570 (1996).

- Fu, Y.-X.: Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
- Fu, Y.-X., Li, W.-H.: Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).
- Fu, Y.-X., Li, W.-H.: Coalescing into the 21st century: An overview and prospects of coalescent theory. *Theoret. Pop. Biol.* **56**, 1–10 (1999).
- Gallais, A.: Covariances entre apparentés quelconques avec linkage et épistasie. I-Expression générale. *Ann. Génét. Sél. Anim.* **2**, 281–310 (1970).
- Gallais, A.: Covariances between arbitrary relatives with linkage disequilibrium. *Biometrics* **30**, 429–446 (1974).
- Geiringer, H.: On the probability theory of linkage in Mendelian heredity. *Ann. Math. Statist.* **15**, 15–50 (1944).
- Gillespie, J.H.: Polymorphism in patchy environments. *Amer. Nat.* **108**, 145–151 (1974a).
- Gillespie, J.H.: The role of environmental grain in the maintenance of genetic variation. *Amer. Nat.* **108**, 831–836 (1974b).
- Gillespie, J.H.: The role of migration in the genetic structure of populations in temporally and spatially varying environments. II. Island models. *Theoret. Pop. Biol.* **10**, 227–238 (1976a).
- Gillespie, J.H.: A general model to account for enzyme variation in natural populations. II. Characterization of the fitness function. *Amer. Nat.* **100**, 809–821 (1976b).
- Gillespie, J.H.: A general model to account for enzyme variation in natural populations. III. Multiple Alleles. *Evolution* **31**, 85–90 (1977a).
- Gillespie, J.H.: A general model to account for enzyme variation in natural populations. IV. The quantitative genetics of fitness traits. In: *Lecture Notes in Biomathematics* **19**. Christiansen, F.B., Fenchel, T.M. (eds.), pp. 301–314. Berlin: Springer, (1977b).
- Gillespie, J.H.: A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. *Theoret. Pop. Biol.* **14**, 1–45 (1978).
- Gillespie, J.H.; Langley, C.H.: A general model to account for enzyme variation in natural populations. *Genetics* **76**, 837–848 (1974).
- Gladstien, K.: The characteristic values and vectors for a class of stochastic matrices arising in genetics. *SIAM J. of Appl. Math.* **34**, 630–642 (1978).
- Goldberger, A.S.: Models and methods in I.Q. debate: Part 1. University of Wisconsin, SSRI Series, (1978a).
- Goldberger, A.S.: Pitfalls in the resolution of IQ inheritance. In: *Genetic Epidemiology*. Morton, N.E., Yung, C.S. (eds.). New York: Academic Press, (1978b).
- Goldman, N., Anderson, J.P., Rodrigo, A.G.: Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**, 652–670 (2000).
- Grafen, A.: Natural selection, kin selection, and group selection. In *Behavioral Ecology: An Evolutionary Approach*, Krebs, J.R., Davies, N.B. (eds.). Oxford: Blackwell, (1984).
- Griffiths, R.C.: Exact sampling distributions from the infinite neutral alleles model. *Adv. Appl. Prob.* **11**, 326–354, (1979a).
- Griffiths, R.C.: A transition density expansion for a multi-allele diffusion model. *Adv. Appl. Prob.* **11**, 310–325, (1979b).

- Griffiths, R.C.: On the distribution of allele frequency in a diffusion model. *Theoret. Pop. Biol.* **15**, 140–158 (1979c).
- Griffiths, R.C. unpublished notes, (1980).
- Griffiths, R.C. personal communication, (2003).
- Griffiths, R.C., Tavaré, S.: Simulating probability distributions in the coalescent. *Theoret. Pop. Biol.* **46**, 131–159, (1994a).
- Griffiths, R.C., Tavaré, S.: Ancestral inference in population genetics. *Statist. Sci* **9**, 307–319, (1994b).
- Griffiths, R.C., Tavaré, S.: Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. B* **344**, 403–410, (1994c).
- Griffiths, R.C., Tavaré, S.: Unrooted tree probabilities in the infinitely-many-sites model. *Math. Biosci.* **127**, 77–98, (1995).
- Griffiths, R.C., Tavaré, S.: Computational methods for the coalescent. *IMA Vol. Math. Applic.* **87**, 165–182, (1997).
- Griffiths, R.C., Tavaré, S.: The age of a mutation in a general coalescent tree. *Stochastic Models* **14** 273–298 (1998).
- Griffiths, R.C., Tavaré, S.: The ages of mutations in gene trees. *Ann. Appl. Probab.* **9**, 567–590, (1999).
- Griffiths, R.C., Tavaré, S.: The genealogy of a neutral mutation. In *Highly Structured Stochastic Systems*. Green, P., Hjort, N., Richardson, S. (eds.), 393–412 (2003).
- Hadeler, K.P., Liberman, U.: Selection models with fertility differences. *J. Math. Biol.* **2**, 19–32 (1975).
- Haigh, J., Maynard Smith, J.: The hitch-hiking effect — a reply. *Genet. Res. Camb.* **27**, 85–87 (1976).
- Haldane, J.B.S.: A mathematical theory of natural and artificial selections, Part II, the influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of Mendelian populations, and on natural selection. *Proc. Camb. Phil. Soc., Biol. Sci.* **1**, 158–163 (1924).
- Haldane, J.B.S.: A mathematical theory of natural and artificial selection, Part III, *Proc. Camb. Phil. Soc.* **23**, 363–372 (1926).
- Haldane, J.B.S.: A mathematical theory of natural and artificial selection, Part IV, *Proc. Camb. Phil. Soc.* **23**, 607–615 (1927a).
- Haldane, J.B.S.: A mathematical theory of natural and artificial selection, V, selection and mutation. *Proc. Camb. Phil. Soc.* **23**, 838–844 (1927b).
- Haldane, J.B.S.: A mathematical theory of natural and artificial selection, VII, selection intensity as a function of mortality rate. *Proc. Camb. Phil. Soc.* **27**, 131–135 (1930a).
- Haldane, J.B.S.: A mathematical theory of natural and artificial selection, VI, isolation. *Proc. Camb. Phil. Soc.* **26**, 220–230 (1930b).
- Haldane, J.B.S.: *The Causes of Evolution*. London: Longmans Green, 1932a.
- Haldane, J.B.S.: A mathematical theory of natural and artificial selection, Part IX, rapid selection. *Proc. Camb. Phil. Soc.* **28**, 244–248 (1932b).
- Haldane, J.B.S.: The effect of variation on fitness. *Amer. Nat.* **71**, 337–349 (1937).
- Haldane, J.B.S.: The cost of natural selection. *J. Gen.* **55**, 511–524 (1957).
- Haldane, J.B.S.: More precise expressions for the cost of natural selection. *J. Gen.* **57**, 351–360 (1961).
- Hamilton, W.D.: The genetical evolution of social behavior. *Jour. of Theoret. Biol.* **7**, 1–52 (1964).

- Hamilton, W.D.: *Narrow Roads of Gene Land. Vol. I. Evolution of Social Behavior*. Basingstoke: Freeman/Macmillan, (1996).
- Hammerstein, P.: Darwinian adaptation, population genetics and the streetcar theory of evolution. *J. Math. Biol.* **34**, 511–532 (1996).
- Hardy, G.H.: Mendelian proportions in a mixed population. *Science* **28**, 49–50 (1908).
- Harris, H.: *The Principles of Human Biochemical Genetics* (third revised edition). Amsterdam: Elsevier, (1980).
- Hasegawa, M., Kishino, H., Yano, T.: Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
- Hedrick, P.W.: A new approach to measuring genetic similarity. *Evolution* **25**, 276–280 (1971).
- Hill, W.G.: Effective size of populations with overlapping generations. *Theoret. Pop. Biol.* **3**, 278–289 (1972).
- Hill, W.G., Robertson, A.: The effect of linkage on limits to artificial selection. *Genet. Res. Camb.* **8**, 269–294 (1966).
- Hill, W.G., Robertson, A.: Linkage disequilibrium in finite populations. *Theoret. and Appl. Gen.* **38**, 226–231 (1968).
- Hoppe, F.: Polya-like urns and the Ewens sampling formula. *J. Math. Biol.* **20**, 91–99 (1984).
- Hoppe, F.: Size-biased sampling of Poisson–Dirichlet samples with an application to partition structures in population genetics. *J. Appl. Prob.* **23**, 1008–1012 (1986).
- Hoppe, F.: The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.* **25**, 123–159 (1987).
- Hoppensteadt, F.C.: A slow selection analysis of two-locus, two-allele traits. *Theoret. Pop. Biol.* **9**, 68–81 (1976).
- Hudson, R.R.: Testing the constant rate neutral allele model with protein sequence data. *Evolution* **37**, 203–217 (1983).
- Hudson, R.R., Kreitman M., Aguardé, M.: A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
- Ito, K., McKean, H.P.: *Diffusion Processes and Their Sample Paths*. Berlin: Springer, (1965).
- Jayakar, S.D.: A mathematical model for interaction of gene frequencies in a parasite and its host. *Theoret. Pop. Biol.* **1**, 140–164 (1970).
- Jinks, J.L., Eaves, L.J.: IQ and inequality. *Nature* **248**, 287–289 (1974).
- Jukes, T.H.: Evolutionary changes in insulin. *Nature* **259**, 250–251 (1976).
- Jukes, T.H., Cantor, C.R.: Evolution of protein molecules. In: *Mammalian Protein Metabolism*, Munro, H.N. (ed.), pp. 21–132. New York, Academic Press, (1969).
- Karlin, S.: Rates of approach to homozygosity for finite stochastic models with variable population size. *Amer. Nat.* **102**, 443–455 (1968).
- Karlin, S.: Sex and infinity: a mathematical analysis of the advantages and disadvantages of recombination. In: *The Mathematical Theory of the Dynamics of Biological Populations*. Bartlett, M.S., Hiorns, R.W., (eds.), pp. 155–194. London and New York: Academic Press, (1973).
- Karlin, S.: General two-locus selection models: some objectives, results and interpretations. *Theoret. Pop. Biol.* **7**, 364–398 (1975).

- Karlin, S.: Selection with many loci and possible relations to quantitative genetics. In: *Proc. Int. Conf. Quart. Gen.* Pollak, K., Kempthorne, O., Bailey, T., (eds.), pp. 207–226. Ames: Iowa State Univ. Press, 1977a.
- Karlin, S.: Gene frequency patterns in the Levene subdivided population model. *Theoret. Pop. Biol.* **11**, 356–385 (1977b).
- Karlin, S., Carmelli, D.: Numerical studies on two-loci selection models with general viabilities. *Theoret. Pop. Biol.* **7**, 399–421 (1975).
- Karlin, S., Feldman, M.W.: Linkage and selection — two locus symmetric viability model. *J. Appl. Prob.* **1**, 39–71 (1970a).
- Karlin, S., Feldman, M.W.: Linkage and selection — two locus additive viability model. *J. Appl. Prob.* **7**, 262–271 (1970b).
- Karlin, S., Liberman, U.: Random temporal variation in selection intensities: case of large population size. *Theoret. Pop. Biol.* **6**, 355–382 (1974).
- Karlin, S., Liberman, U.: Representation of non-epistatic selection models and analysis of multilocus Hardy–Weinberg equilibrium configurations. *J. Math. Biol.* **7**, 353–374 (1979).
- Karlin, S., McGregor, J.L.: On some stochastic models in genetics. In: *Stoch. Models in Med. and Biol.* Gurland, J., (ed.), pp. 245–279. Madison: University of Wisconsin Press, 1964.
- Karlin, S., McGregor, J.L.: Direct product branching processes and related induced Markoff chains. I. Calculations of rates of approach to homozygosity. In: *Bernoulli (1723), Bayes (1773), Laplace (1813): Anniv. Vol.*, LeCam, L., Neyman, J., (eds.), pp. 111–145. Berlin, Heidelberg, New York: Springer, 1965.
- Karlin, S., McGregor, J.L.: The number of mutants forms maintained in a population. *Proc. Fifth Berk. Symp. of Math. Stat. and Prob.* **4**, 403–414 (1966).
- Karlin, S., McGregor, J.L.: Rates and probabilities of fixation for two locus random mating finite populations without selection. *Genetics* **58**, 141–159 (1968).
- Karlin, S., McGregor, J.L.: Addendum to a paper of W. Ewens. *Theoret. Pop. Biol.* **3**, 113–116 (1972).
- Karlin, S., McGregor, J.L.: Towards a theory of the evolution of modifier genes. *Theoret. Pop. Biol.* **5**, 59–105 (1974).
- Karlin, S., Taylor, H.: *A First Course in Stochastic Processes*. New York: Academic Press, 1975.
- Keeler, C.: Some oddities in the delayed appreciation of “Castle’s Law.” *J. Heredity* **59**, 110–112 (1968).
- Kelly, F.P.: On stochastic population models in genetics. *J. Appl. Prob.* **13**, 127–131 (1976).
- Kelly, F.P.: Exact results for the Moran neutral allele model. *J. Appl. Prob.* **9**, 197–201 (1977).
- Kemeny, J.G., Snell, J.L.: *Finite Markov Chains*. New York: Van Nostrand, (1960).
- Kempthorne, O.: *The Design and Analysis of Experiments*. New York: John Wiley & Sons, (1952).
- Kempthorne, O.: The correlation between relatives in a random mating population. *Proc. Roy. Soc. B.* **143**, 103–113 (1954).
- Kempthorne, O.: The theoretical values of correlations between relatives in random mating populations. *Genetics* **40**, 153–167 (1955).

- Kempthorne, O.: *An Introduction to Genetic Statistics*. New York: Wiley, (1957).
- Kempthorne, O., Pollak, E.: Concepts of fitness in Mendelian populations. *Genetics* **64**, 125–145 (1970).
- Kidwell, J.F., Clegg, M.T., Steward, F.M., Prout, T.: Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics* **85**, 171–183 (1977).
- Kim, Y., Stephan, W.: Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**, 765–777 (2002).
- Kimura, M.: Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics* **39**, 280–295 (1954).
- Kimura, M.: Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci.* **41**, 144–150 (1955a).
- Kimura, M.: Random drift in a multi-allelic locus. *Evolution* **9**, 419–435 (1955b).
- Kimura, M.: Stochastic processes and distribution of gene frequencies under natural selections. *Cold Spring Harbor on Quant. Biol.* **20**, 33–53 (1955c).
- Kimura, M.: Random genetic drift in a tri-allelic locus — exact solution with a continuous model. *Biometrics* **12**, 57–66 (1956a).
- Kimura, M.: A model of a genetic system which leads to closer linkage under natural selection. *Evolution* **10**, 278–287 (1956b).
- Kimura, M.: Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**, 882–901 (1957).
- Kimura, M.: On the change of population fitness by natural selection. *Heredity* **12**, 145–167 (1958).
- Kimura, M.: A probability method for treating inbreeding systems especially with linked genes. *Biometrics* **19**, 1–17 (1963).
- Kimura, M.: Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* **52**, 875–890 (1965).
- Kimura, M.: On the evolutionary adjustment of spontaneous mutation rates. *Genet. Res.* **9**, 25–34 (1967).
- Kimura, M.: Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
- Kimura, M.: The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutation. *Genetics* **61**, 893 (1969).
- Kimura, M.: The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population. *Gen. Res.* **15**, 131–134 (1970).
- Kimura, M.: Theoretical foundations of population genetics at the molecular level. *Theoret. Pop. Biol.* **2**, 174–208 (1971).
- Kimura, M.: Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267–275 (1977).
- Kimura, M.: A simple method for estimating evolutionary rate in a finite population due to mutational production of neutral and nearly neutral base substitution through comparative studies of nucleotide sequences. *J. Molec. Biol.* **16**, 111–120 (1980).
- Kimura, M., Crow, J.F.: The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738 (1964).
- Kimura, M., Ohta, T.: *Theoretical Aspects of Population Genetics*. Princeton: Princeton University Press, 1971a.

- Kimura, M., Ohta, T.: On the rate of molecular evolution. *J. Mol. Evol.* **1**, 1 (1971b).
- Kimura, M., Ohta, T.: The age of a neutral mutant persisting in a finite population. *Genetics* **75**, 199–212 (1973).
- King, J.L., Jukes, T.H.: Evolutionary loss of ascorbic acid synthesizing ability. *J. Hum. Evol.* **4**, 85–88 (1975).
- Kingman, J.F.C.: A matrix inequality. *Quart. J. Math.* **12**, 78–80 (1961a).
- Kingman, J.F.C.: A mathematical problem in population genetics. *Proc. Camb. Phil. Soc.* **57**, 574–582 (1961b).
- Kingman, J.F.C.: Random discrete distributions. *J. Roy. Stat. Soc. B.* **37**, 1–22 (1975).
- Kingman, J.F.C.: Coherent random walks arising in some genetical models. *Proc. R. Soc. Lond. A* **351**, 19–31 (1976).
- Kingman, J.F.C.: A note on multi-dimensional models of neutral mutation. *Theoret. Pop. Biol.* **11**, 285–290 (1977a).
- Kingman, J.F.C.: The population structure associated with the Ewens sampling formula. *Theoret. Pop. Biol.* **11**, 274–283 (1977b).
- Kingman, J.F.C.: Random partitions in population genetics. *Proc. Roy. Soc. London Ser. A* **361**, 1–20 (1978).
- Kingman, J.F.C.: The coalescent. *Stoch. proc. Applns.* **13**, 235–248, (1982a).
- Kingman, J.F.C.: Exchangeability and the evolution of large populations. pp 97–112 in *Exchangeability in Probability and Statistics*, (Koch, G. and Spizzichino, F. eds.). Amsterdam, North-Holland, (1982b).
- Kingman, J.F.C.: On the genealogy of large populations. *J. Appl. Prob.* **19A**, 27–43 (1982c).
- Kingman, J.F.C.: Origins of the coalescent: 1974–1982. *Genetics* **156**, 1461–1463 (2000).
- Kishino, H., Hasegawa, M.: Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in *Hominoidea*. *J. Molec. Evol.* **29**, 170–179 (1989).
- Kojima, K., Kelleher, T.M.: Changes of mean fitness in random-mating populations when epistasis and linkage are present. *Genetics* **46**, 527–540 (1961).
- Kolman, W.: The mechanism of natural selection for the sex ratio. *Amer. Natur.* **94**, 373–377 (1960).
- Kreitman, M.: Methods to detect selection in populations with applications to the human. *Ann. Rev. Genomics Hum. Genet.* **1**, 539–559 (2000).
- Kuhner, M.K., Yamato, J., Felsenstein, J.: Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140**, 1421–1430, (1995).
- Kuhner, M.K., Yamato, J., Felsenstein, J.: Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**, 429–434, (1998).
- Langley, C.H., Fitch, W.M.: An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**, 161–177 (1974).
- Last, K.: Genetical aspects of human behaviour. Unpublished Ph.D. thesis. Dept. of Genetics, University of Birmingham, England, (1976).
- Lessard, S.: Fisher's fundamental theorem of natural selection revisited. *Theoret. Pop. Biol.* **52**, 119–136 (1997).

- Lessard, S., Castilloux, A.-M.: The fundamental theorem of natural selection in Ewens' sense: case of fertility selection. *Genetics* **141**, 733–742 (1995)
- Levene, H.: Genetic equilibrium when more than one ecological niche is available. *Amer. Nat.* **87**, 331–333 (1953).
- Levikson, B.: The age distribution of Markov processes. *J. Appl. Prob.* **14**, 492–506 (1977).
- Levikson, B., Karlin, S.: Random temporal variation in selection intensities acting on infinite diploid populations: diffusion method analysis. *Theoret. Pop. Biol.* **8**, 292–300 (1975).
- Lewontin, R.C.: The interaction of selection and linkage I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1964a).
- Lewontin, R.C.: The interaction of selection and linkage II. Optimum models. *Genetics* **50**, 757–782 (1964b).
- Lewontin, R.C.: The apportionment of human diversity. *Evol. Biol.* **6**, 381–398 (1973).
- Lewontin, R.C.: *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press, (1974).
- Lewontin, R.C., Kojima, K.: The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472 (1960).
- Li, H., Zhang, Y., Zhang, Y.-P., Fu, Y.X.: Neutrality tests using DNA polymorphism from multiple samples. *Genetics* **163**, 1147–1151 (2003).
- Li, W.-H.: A mixed model of mutation for electrophoretic identity of proteins within and between populations. *Genetics* **83**, 423–432 (1976).
- Li, W.-H.: Maintenance of genetic variability under mutation and selection pressures in a finite population. *Proc. Nat. Acad. Sci.* **74**, 2509–2513 (1977).
- Li, W.-H., Nei, M.: Stable linkage of disequilibrium without epistasis in subdivided populations. *Theoret. Pop. Biol.* **6**, 173–183 (1974).
- Li, W.-H., Nei, M.: Persistence of common alleles in two related populations or species. *Genetics* **86**, 901–914 (1977).
- Littler, R.A.: Linkage disequilibrium in two-locus, finite, random mating models without selection or mutation. *Theoret. Pop. Biol.* **4**, 259–275 (1973).
- Littler, R.A.: Loss of variability at one locus in a finite population. *Math. Bio.* **25**, 151–163 (1975).
- Littler, R.A., Fackerell, E.D.: Transition densities for neutral multi-allele diffusion models. *Biometrics* **31**, 117–123 (1975).
- Løvtrup, S. *Darwinism: The Refutation of a Myth*. London: Croon Helm, 1987.
- Lynch, M., Walsh, B.: *Genetics and Analysis of Quantitative Traits*. Sunderland MA: Sinauer (1998).
- Lyubich, Y.I.: *Mathematical Structures in Population Genetics*. New York, Springer (1992).
- Malécot, G.: *Les Mathématiques de l'Hérédité*. Paris: Masson, (1948).
- Mandl, P.: *Analytical Treatment of One-Dimensional Markov Processes*. Berlin: Springer, (1968).
- Marrow, P., Johnstone, R.A., Hurst, L.D.: Riding the evolutionary streetcar: where population genetics and game theory meet. *TREE* **11**, 445–446 (1996).
- Maruyama, T.: On the fixation probability of mutant genes in a subdivided population. *Gen. Res.* **15**, 221–226 (1970).
- Maruyama, T.: An invariance property of a structured population. *Gen. Res.* **18**, 81–84 (1971).

- Maruyama, T.: The rate of decay of genetic variability in a geographically structured finite population. *Math. Biosci.* **14**, 325–335 (1972).
- Maruyama, T.: The age of an allele in a finite population. *Genetics. Res. Camb.* **23**, 137–143 (1974).
- Maruyama, T.: *Stochastic Problems in Population Genetics*. Lecture Notes in Biomathematics **17**. Berlin: Springer, 1977.
- Maruyama, T., Kimura, M., A note on the speed of gene frequency changes in reverse directions in a finite population. *Evolution* **28**, 161–163 (1974).
- Maruyama, T., Yamazaki, T.: Analysis of heterozygosity in regard to the neutrality theory of protein polymorphisms, *J. Mol. Evol.* **4**, 195 (1974).
- Mather, K., Jinks, J.L.: *Introduction to Biometrical Genetics*. London: Chapman & Hall, 1977.
- Matessi, C., Jayakar, S.D.: Conditions for the evolution of altruism under Darwinian selection. *Theoret. Pop. Biol.* **9**, 360–387 (1976).
- May, R.M.: *Stability and Complexity in Model Ecosystems*. (Second ed.), Princeton: Princeton University Press, (1975).
- May, R.M.: *Theoretical Ecology*. Philadelphia: W.B. Saunders, (1976).
- Maynard Smith, J.: Evolution in sexual and asexual populations. *Am. Nat.* **102**, 469–473 (1968).
- Maynard Smith, J.: What use is sex? *J. Theoret. Biol.* **30**, 319–335 (1971).
- Maynard Smith, J., Haigh, J.: The hitch-hiking effect of a favourable gene. *Genet. Res. Camb.* **23**, 23–35 (1974).
- McCloskey, J.W.: A model for the distribution of individuals by species in an environment. Unpublished PhD. thesis, Michigan State University, (1965).
- McDonald, J.H., Kreitman, M.: Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **354**, 652–654 (1991).
- McKean, H.P.: Elementary solutions for certain parabolic partial differential equations. *Trans. Amer. Math. Soc.* **82**, 519–548 (1956).
- McNaughton, S.J.: Natural selection at the enzyme level. *Am. Nat.* **108**, 616 (1974).
- Miller, G.F.: The evaluation of eigenvalues of a differential equation arising in a problem in genetics. *Proc. Camb. Phil. Soc.* **58**, 588–593 (1962).
- Moran, P.A.P.: Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**, 60–71 (1958).
- Moran, P.A.P.: *The Statistical Processes of Evolutionary Theory*, Oxford: Clarendon Press, (1962).
- Moran, P.A.P.: On the theory of selection dependent on two loci. *Ann. Hum. Genet.* **32**, 183–190, (1968).
- Moran, P.A.P.: Wandering distribution and the electrophoretic profile. *Theoret. Pop. Biol.* **8**, 318–330 (1975).
- Moran, P.A.P.: Wandering distributions and the electrophoretic profile II. *Theoret. Pop. Biol.* **10**, 145–149 (1976).
- Moran, P.A.P., Watterson, G.A.: The genetic effects of family structure in natural populations. *Aust. J. Biol. Sci.* **12**, 1–15 (1958).
- Muller, H.J.: Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
- Nagylaki, T.: The moments of stochastic integrals and the distribution of sojourn times. *Proc. Nat. Acad. Sci.* **71**, 746–749 (1974a).
- Nagylaki, T.: The decay of genetic variability in geographically structured populations. *Proc. Nat. Acad. Sci.* **71**, 2932–2936 (1974b).

- Nagylaki, T.: Continuous selective models with mutation and migration. *Theoret. Pop. Biol.* **5**, 284–295 (1974c).
- Nagylaki, T.: Conditions for the existence of clines. *Genetics* **80**, 595–615 (1975).
- Nagylaki, T.: The evolution of one- and two-locus systems. *Genetics* **83**, 583–600 (1976).
- Nagylaki, T.: The evolution of one- and two-locus systems. II. *Genetics* **85**, 347–354 (1977a).
- Nagylaki, T.: *Selection in One- and Two-Locus Systems*. Lecture Notes in Biomathematics. Berlin: Springer, (1977b).
- Nagylaki, T.: Decay of genetic variability in geographically structured populations. *Proc. Nat. Acad. Sci.* **74**, 2523–2525 (1977c).
- Nagylaki, T.: The correlation between relatives with assortative mating. *Ann. Hum. Genet.* **42**, 131–137 (1978).
- Nagylaki, T.: The island model with stochastic migration. *Genetics* **91**, 163–176 (1979).
- Nagylaki, T.: Clines with asymmetric migration. *Genetics* **88**, 813–827 (1978d).
- Nagylaki, T.: *Theoretical Population Genetics*. Berlin: Springer-Verlag, (1992).
- Nagylaki, T., Crow, J.F.: Continuous selection models. *Theoret. Pop. Biol.* **5**, 257–283 (1974).
- Nei, M.: Effective population size when fertility is inherited. *Genet. Res.* **8**, 257–260 (1966).
- Nei, M.: Modification of linkage intensity by natural selection. *Genetics* **57**, 625–641 (1967).
- Nei, M.: Linkage modification and sex difference in recombination. *Genetics* **63**, 681–699 (1969).
- Nei, M.: Genetic distance between populations. *Amer. Nat.* **106**, 283–292 (1972).
- Nei, M.: Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci.* **70**, No. 12, Part I, 3321–3323 (1973).
- Nei, M.: *Molecular Population Genetics and Evolution*. Oxford: North-Holland Publishing Co., (1975).
- Nei, M.: Mathematical models of speciation and genetic distance. In: *Population Genetics and Ecology*, Karlin, E., Nevo, E., (eds.), pp. 723–765. New York: Academic Press, (1976).
- Neilsen, R.: Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**, 641–647 (2001).
- Nordborg, M.: Coalescent theory. In *Handbook of Statistical Genetics*, Balding, D.J., Bishop, T., Cannings, C. (eds.), pp. 179–212. New York, Wiley, (2001).
- Norman, M.F.: A central limit theorem for Markov processes that move by small steps. *Annals of Prob.* **2**, 1065–1074 (1974).
- Norman, M.F.: Approximation of stochastic processes by Gaussian diffusions, and applications to Wright–Fisher genetic models. *SIAM J. Appl. Math.* **29**, 225–242 (1975a).
- Norman, M.F.: Limit theorems for stationary distributions. *Adv. Appl. Prob.* **7**, 561–575 (1975b).
- Norman, M.F.: Diffusion approximation of non-Markovian processes. *The Annals of Prob.* **3**, 358–364 (1975c).
- Norman, M.F.: Personal communication, 1978.
- Norton, H..J.: Natural selection and Mendelian variation. *Proc. London Math. Soc.* (series 2) **28**, 1–45 (1928).

- Ohta, T.: Fixation probability of a mutant influenced by random fluctuation of selection intensity. *Gen. Res.* **19**, 33–38 (1972).
- Ohta, T.: Mutational pressure as main cause of molecular evolution. *Nature* **252**, 351–354 (1974).
- Ohta, T.: Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theoret. Pop. Biol.* **10**, 254–275 (1976).
- Ohta, T., Kimura, M.: Linkage disequilibrium due to random genetic drift. *Gen. Res.* **13**, 47–55 (1969a).
- Ohta, T., Kimura, M.: Linkage disequilibrium at steady state determined by random random genetic drift and recurrent mutation. *Genetics* **63**, 229–238 (1969b).
- Ohta, T., Kimura, M.: Development of associative overdominance through linkage disequilibrium in finite populations. *Gen. Res.* **16**, 165–177 (1970).
- Ohta, T., Kimura, M.: A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Gen. Res. Camb.* **22**, 201–204 (1973).
- Ohta, T., Kimura, M.: The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Gen. Res. Camb.* **25**, 313–326 (1975).
- Ohta, T., Kimura, M.: Hitch-hiking effect — counter reply. *Gen. Res. Camb.* **28**, 307–308 (1976).
- Olby, R.C.: Francis Galton's derivation of Mendelian ratios in 1875. *Heredity* **20**, 636–638 (1965).
- Pearson, E.S., Hartley, H.O.: *Biometrika Tables for Statisticians*. Cambridge: Cambridge University Press, 1962.
- Pearson, K.: Mathematical contributions to the theory of evolution. XII. on a generalized theory of alternative inheritance, with special reference to Mendel's laws. *Phil. Trans. of the Roy. Soc. A.* **203**, 53–86 (1904).
- Perlitz, M., Stephan, W.: The mean and variance of the number of segregating sites since the last hitchhiking event. *J. Math. Biol.* **36**, 1–23 (1997).
- Pielou, E.C.: *Populations and Community Ecology: Principles and Methods*. New York: Gordon and Breach, (1974).
- Pollak, E.: The effective population size of some age-structured populations. *Math. Biosci.* **168**, 39–56 (2000).
- Price, G.R.: Fisher's "Fundamental Theorem" made clear. *Ann. Hum. Genet.* **36**, 129–140, (1972).
- Przeworski, M.: The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189 (2002).
- Prohorov, Y., Rozanov, Y.: *Probability Theory*. Berlin: Springer, (1969).
- Provine, W.B.: *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press, (1971).
- Punnett, E.C.: Eliminating feeble-mindedness. *J. Heredity* **8**, 464–465 (1917).
- Robertson, A.: Selection for heterozygotes in small populations. *Genetics* **47**, 1291–1300 (1962).
- Robertson, A.: Artificial selection in plants and animals. *Proc. Roy. Soc. B.* **164**, 341–349 (1966).
- Robertson, A.: The spectrum of genetic variation. In: *Population Biology and Evolution*. By Lewontin, R.C. (ed.), pp. 5–16. Syracuse: Syracuse University Press, (1968).

- Rogers, J.S.: Measures of genetic similarity and genetic distance. *Studies in Genetics VII* (Univ. Texas Publ. No. 7213), 145–153 (1972).
- Roughgarden, J.: Resource partitioning among competing species — a coevolutionary approach. *Theoret. Pop. Biol.* **3**, 388–424 (1976).
- Roughgarden, J.: Coevolution in ecological systems: results from “loop analysis” for purely density-dependent coevolution. In: *Lecture Notes in Biomathematics* **19**, Christiansen, F.G., Fenchel, T.M. (ed.). Berlin: Springer, (1977).
- Roux, C.Z.: Hardy–Weinberg equilibria in random mating populations. *Theoret. Pop. Biol.* **5**, 393–416 (1974).
- Sabeti, P.C., Reich, D.E., Higgins, J.M. et al.: Detecting recent positive selection in the human genome from haplotype structure. *Nature* **24**, 832–837 (2002).
- Sawyer, S.A.: On the past history of an allele now known to have frequency p . *J. Appl. Prob.* **14**, 439–450 (1977).
- Sawyer, S.A., Hartl, D.L.: Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176 (1992).
- Schaeffer, H.E., Johnson, F.M.: Isozyme allelic frequencies related to selection and gene-flow hypothesis. *Genetics* **77**, 163 (1974).
- Schnell, F.W.: The covariance between relatives in the presence of linkage. In: *Statistical Genetics and Plant Breeding*. Nat. Acad. Sci. Nat. Res. Council. Publication **982**, 468–483 (1963).
- Schwartz, J.: Population genetics and sociobiology. *Persp. Biol. Med.* **45**, 224–240 (2002).
- Seneta, E.: Quasi-stationary distribution and time-reversion in genetics. *J. Roy. Stat. Soc. B.* **28**, 253–277 (1966).
- Seneta, E.: On a genetic inequality. *Biom.* **29**, 810–813 (1973).
- Serant, D.: Linkage and inbreeding coefficients in a finite random mating population. *Theoret. Pop. Biol.* **6**, 251–263 (1974).
- Serant, D., Villard, M.: Linearization of crossing-over and mutation in a finite random-mating population. *Theoret. Pop. Biol.* **3**, 249–257 (1972).
- Shahshahani, S.: A new mathematical framework for the study of linkage and selection. in *Memoirs of the American Mathematical Society* **17**. Providence, RI: Amer. Math. Soc. (1979).
- Shaw, R.F.: The theoretical genetics of the sex-ratio. *Genetics* **43**, 149–163 (1958).
- Shaw, R.F., Mohler, J.D.: The selective significance of the sex ratio. *Amer. Nat.* **87**, 337–342 (1953).
- Simonsen, K.L., Churchill, G.A., Aquadro, C.F.: Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429 (1995).
- Slatkin, M.: On treating the chromosome as the unit of selection. *Genetics* **72**, 157–168 (1972).
- Slatkin, M.: Testing neutrality in a subdivided population. *Genetics* **100**, 533–545 (1982).
- Slatkin, M., Rannala, B. Estimating allele age. *Ann. Rev. Genomics Hum. Genet.* **1**, 225–249 (2000).
- Sokal, R.R., Sneath, P.H.A.: *Principles of Numerical Taxonomy*. San Francisco: Freeman, (1963).
- Spratt, D.A.: The stability of a sex-linked allelic system. *Ann. Hum. Gen.* **22** 1–6 (1957).

- Stephens, M.: Inference under the coalescent. In *Handbook of Statistical Genetics*, Balding, D.J., Bishop, T., Cannings, C. (eds.), pp. 213–238. New York: Wiley, (2001).
- Stephens, M., Donnelly, P.J.: Inference in molecular population genetics. *J. Roy. Statist. Soc. B* **62**, 605–655, (2000).
- Strobeck, C.: The two-locus model with sex differences in recombination. *Genetics* **78**, 791–797 (1974).
- Strobeck, C.: Average number of nucleotide difference in a sample from a single population: a test for population subdivision. *Genetics* **117**, 533–545 (1987).
- Strobeck, C., Morgan, K.: The effect of intragenic recombination on the number of alleles in a finite population. *Genetics* **88**, 829–844 (1978).
- Svirezhev, Y.M.: Optimizing principles in genetics. in *Studies on Theoretical Genetics*, (V.A. Ratner, ed.), USSR Academy of Sciences, Novosibirsk. [In Russian with English summary]. (1972)
- Tajima, F.: Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).
- Tajima, F.: Statistical methods for testing the neutral mutations hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
- Tajima, F.: Estimation of the amount of DNA polymorphism and statistical tests of the neutral mutation hypothesis based on DNA polymorphism. pp. 149–164 in *Progress in Population genetics and Human Evolution*, (Donnelly, P.J., Tavaré, S., eds). New York: Springer, (1997).
- Takano, T.S., Kusakabe, S., Mukai, T.: The genetic structure of natural populations of *Drosophila melanogaster*. XXII. Comparative study of DNA polymorphisms in northern and southern natural populations. *Genetics* **129**, 753–761 (1991).
- Tanaka, K.: On limiting distributions for one-dimensional diffusion processes. *Bull. of Math. Stat.* **7**, 84–91 (1957).
- Tavaré, S.: Lines of descent and genealogical processes, and their application in population genetics models. *Theoret. Pop. Biol.* **26**, 119–164, (1984).
- Tavaré, S.: Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17**, 57–86 (1986).
- Tavaré, S.: The age of a mutant in a general coalescent tree. *Stoch. Models* **14**, 273–295 (1998).
- Tavaré, S.: Ancestral inference in population genetics. In *Proceedings of Saint Flour Summer School in Probability and Statistics*, (to appear) (2004).
- Tavaré, S., Joyce, P., Ewens, W.J.: Is knowing the age-order of alleles in a sample useful in testing for selective neutrality? *Genetics* **122**, 705–711 (1989).
- Thomson, G.: The effect of a selected locus on linked neutral loci. *Genetics* **85**, 753–788 (1977).
- Tier, C., Keller, J.B.: Asymptotic analysis of diffusion equations in population genetics. *SIAM J. Appl. Math.* **34**, 549–576 (1978).
- Trajstman, A.C.: On a conjecture of G.A. Watterson. *Adv. Appl. Prob.* **6**, 489–493 (1974).
- Uyenoyama, M.K., Bengtsson, B.O.: Towards a genetic theory for the evolution of the sex ratio. *Genetics* **93**, 721–736 (1979).
- van Aarde, I.M.R.: Covariances of relatives in random mating populations with linkage. Unpublished Ph.D. thesis. Iowa State University, Ames, 1963.

- van Aarde, I.M.R.: The effect of linkage on the mean value of inbreds derived from a random mating populations. *Genetics* **78**, 1245–1249 (1974).
- van Aarde, I.M.R.: The covariance of relatives derived from a random mating population. *Theoret. Pop. Biol.* **8**, 166–183 (1975).
- Verner, J.: Selection for sex ratio. *Amer. Nat.* **99**, 419–420 (1965).
- Veronka, R., Keller, J.: Asymptotic analysis of stochastic models in population genetics. *Math. Biosci.* **25**, 331–362 (1975).
- Wade, M.J.: A critical review of the models of group selection. *Quart. Rev. Biol.* **53**, 101–114 (1978).
- Wang, Y., Pollak, E.: The effective number of a population that varies cyclically in size. I. Discrete generations. *Math. Biosci.* **166**, 1–21 (2000).
- Watterson, G.A., Markov chains with absorbing states: a genetic example. *Ann. Math. Stat.* **32**, 716–729 (1961).
- Watterson, G.A., Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Stat.* **33**, 939–957 (1962).
- Watterson, G.A., The effect of linkage in a finite random-mating population. *Theoret. Pop. Biol.* **1**, 72–87 (1970).
- Watterson, G.A.: Errata to the effects of linkage in a finite random-mating population. *Theoret. Pop. Biol.* **3**, 117 (1972).
- Watterson, G.A.: Models for the logarithmic species abundance distributions. *Theoret. Pop. Biol.* **6**, 217–250 (1974a).
- Watterson, G.A.: The sampling theory of selectively neutral alleles. *Adv. App. Prob.* **6**, 463–488 (1974b).
- Watterson, G.A.: On the number of segregating sites in genetic models without recombination. *Theoret. Pop. Biol.* **7**, 256–276 (1975).
- Watterson, G.A.: Reversibility and the age of an allele I. Moran's infinitely many neutral alleles model. *Theoret. Pop. Biol.* **10**, 239–253 (1976a).
- Watterson, G.A.: The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Prob.* **13**, 639–651 (1976b).
- Watterson, G.A.: Heterosis or neutrality? *Genetics* **85**, 789–814 (1977a).
- Watterson, G.A.: Reversibility and the age of an allele II. two-allele models, with selection and mutation. *Theoret. Pop. Biol.* (1977b).
- Watterson, G.A.: The homozygosity test of neutrality. *Genetics* **88**, 405–417 (1978).
- Watterson, G.A.: Lines of descent and the coalescent. *Theoret. Pop. Biol.* **26**, 239–253 (1984).
- Watterson, G.A.: Guess, H.A.: Is the most frequent allele the oldest? *Theoret. Pop. Biol.* **11**, 141–160 (1977).
- Weinberg, W.: On the detection of heredity in man (in German). *Jh. Ver. Vaterl. Naturk. Wurttemb.* **64**, 368–382 (1908).
- Weir, B.S., Brown, A.H.D., Marshall, D.R.: Testing for selective neutrality of electrophoretically detectable protein polymorphisms. *Genetics* **84**, 639–659 (1976).
- Weir, B.S., Cockerham, C.C.: Group inbreeding with two linked loci. *Genetics* **63**, 711–742 (1969).
- Weir, B.S., Cockerham, C.C.: Mixed self and random mating at two loci. *Gen. Res.* **21**, 247–262 (1973).
- Weir, B.S., Cockerham, C.C.: Behavior of pairs of loci in finite moneocious populations. *Theoret. Pop. Biol.* **6**, 323–354 (1974).

- Whelan, S., Goldman, N.: Distribution of statistics used for the comparison of models of sequence evolution in phylogenetics. *Mol. Biol. Evol.* **16**, 1292–1299 (1999).
- White, M.J.D.: *Modes of Speciation*. San Francisco: Freeman, (1978).
- Williams, G.C., Mitton, J.B.: Why reproduce sexually? *J. Theoret. Biol.* **39**, 545–555 (1961).
- Wilson, E.O.: *Sociobiology: The New Synthesis*. Cambridge, MA: Belknap Press, (1975).
- Wilson, E.O.: Animal and human sociobiology. In: *The Changing Scenes in the Natural Sciences*. Goulden, C., (ed.), pp. 273–283. Philadelphia: Academy of Natural Sciences, (1977).
- Wilson, S.R.: The correlation between relatives under the multifactorial model with assortative mating. *Ann. Hum. Genet.* **37**, 189–204, 205–215 (1973).
- Wilson, S.R.: Two-sided assortative mating for a single locus. *Ann. Hum. Genet.* **40**, 225–229 (1976).
- Wiuf, C., Donnelly, P.J.: Conditional genealogies and the age of a neutral mutant. *Theoret. Pop. Biol.* **56**, 183–201 (1999).
- Wright, S.: Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics* **6**, 144–161 (1921).
- Wright, S.: The evolution of dominance. *Amer. Nat.* **63**, 556–561 (1929a).
- Wright, S.: Fisher's theory of dominance. *Amer. Nat.* **63**, 274–279 (1929b).
- Wright, S.: The genetical theory of natural selection – a review. *J. Hered.* **21**, 349–356 (1930).
- Wright, S.: Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
- Wright, S.: Physiological and evolutionary theories of dominance. *Amer. Nat.* **68**, 25–53 (1934).
- Wright, S.: The analysis of variance and the correlation between relatives with respect to deviations from an optimum. *J. Gen.* **30**, 243–256 (1935).
- Wright, S.: Isolation by distance. *Genetics* **28**, 114–138 (1943).
- Wright, S.: On the role of directed and random changes in gene frequency in the genetics of populations. *Evolution* **2**, 279–294 (1948).
- Wright, S.: Adaption and selection. In: *Genetics, Palaeontology, and Evolutions*. Simpson, G.G., Jepsen, G.L., Mayr, E., (eds.), pp. 365–389. Princeton: Princeton University Press, (1949).
- Wright, S.: The genetical structure of populations. *Ann. Eug.* **15**, 323–354 (1951).
- Wright, S.: The genetics of quantitative variability. In: *Quantitative Inheritance*. Reeve, E.C.R., Waddington, C.H. (eds.), pp. 5–41. Her Majesty's Stat. Office: London, (1952).
- Wright, S.: Modes of selection. *Amer. Nat.* **90**, 5–24 (1956).
- Wright, S.: Genetics and twentieth century Darwinism — a review and discussion. *Amer. J. Hum. Genet.* **12**, 365–372 (1960).
- Wright, S.: The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution* **19**, 395–420 (1965a).
- Wright, S.: Factor interaction and linkage in evolution. *Proc. Roy. Soc. B.* **162**, 80–104 (1965b).
- Wright, S.: Polyallelic random drift in relation to evolution. *Proc. Nat. Acad. Sci.* **55**, 1074–1081 (1966).

- Wright, S.: *Evolution and the Genetics of Populations, Vol. 1. Genetics and Biometric Foundations*. Chicago, London: University of Chicago Press, (1969a).
- Wright, S.: *Evolution and the Genetics of Populations, Vol. 2. The Theory of Gene Frequencies*. Chicago, London: University of Chicago Press, (1969b).
- Wright, S.: *Evolution and the Genetics of Populations, Vol. 3. Experimental Results and Evolutionary Deductions*. Chicago, London: University of Chicago Press, (1977).
- Wright, S.: *Evolution and the Genetics of Populations, Vol. 4. Variability Within and Among Natural Populations*. Chicago, London: University of Chicago Press, (1978).
- Yu, P.: Some host parasite genetic interaction models. *Theoret. Pop. Biol.* **3**, 347–357 (1972).
- Yule, G.U.: Mendel's laws and their probable relation to intraracial heredity. *New Phyt.* **1**, 193–207, 222–238 (1902).
- Yule, G.U.: On the theory of inheritance of quantitative compound characters on the basis of Mendel's law — a preliminary note. *Rep. of Third Int. Conf. on Gen.* 140–142 (1906).

Author Index

- Abramowitz, M, 112, 114
Aguardé, M., 347
Akin, E., 261
Anderson, J.P., 383
Anderson, R, 350
Anderson. W.W., 284
Aquadro, C.F., 360, 361, 365
Avery, P., 186, 248
- Barker, W.C., 290
Bengtsson, B.O., 277
Bennett, J.H., 48, 243
Berger, E.M., 290
Berger, R.L., 361
Blaisdell, B.E., 375
Blundell, T.L., 347
Bodmer, W.F., 54, 56, 212, 277
Boos, D.D., 361
Bulmer, M.G., 248, 274
Bürger, R., viii
Burt, C., 273
- Cannings, C., 38, 48, 95, 99, 100, 102
Cantor, C.R., 373
Carmelli, D., 202, 217, 219
Castilloux, A.-M., 67, 261
Castle, W.E., 5
- Cavalli-Sforza, L.L., 286, 379
Charlesworth, B., 282
Christiansen, F.B., viii, 239, 283
Churchill, G.A., 360, 361, 365, 376
Cockerham, C.C., 75, 77, 78, 129, 268, 270
Conley, C.C., 204
Cornish-Bowden, A., 290
Coyne, J.A., 353, 369
Crow, J.F., 35, 57, 78, 83, 128, 166, 236-239, 247, 261, 262, 272, 274, 289
- Dayhoff, M.O., 290
Demetrius, L., 282, 283
Dobzhansky, T., 79
Donnelly, P.J., viii, 314, 325, 326, 342, 344, 345
- Eaves, L.J., 273
Edwards, A.W.F., 262, 277, 379
Elliott, J., 150
Engen, S., 321
Epperson, B.K., viii, 282
Eshel, L., 238, 277
Ethier, S., 179
Ewens, W.J., 83, 113, 115, 120, 169, 177, 204, 207, 220, 224, 246,

- 247, 261, 294, 303, 311, 349,
351, 354, 355, 368
- Fackerall, E.D., 192
- Fay, J.C., 347, 357, 363
- Feldman, M.W., 103, 211, 214, 215,
218, 221, 224–227, 238, 239,
251, 286
- Feller, W., 95, 149
- Felsenstein, J., 128, 212, 239, 272,
314, 375, 376
- Fenchel, T.M., 283
- Fisher, R.A., 6, 9, 15, 16, 20, 22–25,
27–29, 33, 39, 64–66, 71, 73,
74, 137, 177, 219, 221, 223,
224, 235, 272, 280, 281, 286
- Fitch, W., 347
- Franklin, I., 85, 218, 239, 251, 252,
254, 274
- Freedman, D., 148
- Frydenberg, O., 230
- Fu, Y.-X., 347, 357, 360, 361,
363–365, 369
- Geiringer, H., 243
- Gillespie, J.H., 186–188, 303
- Gladstien, K., 101, 106
- Goldberger, A.S., 273
- Goldman, N., 382, 383
- Grafen, A., 286
- Griffiths, R.C., 96, 192, 304, 305, 314,
321, 326, 329
- Guess, H.A., 173, 196, 197, 324
- Hadeler, K.P., 55
- Haigh, J., 232, 234, 363
- Haldane, J.B.S., 12, 16, 27, 79, 82,
282, 286
- Hamilton, W.D., 286
- Hammerstein, P., 286
- Hardy, G.H., 5
- Harris, H., 291, 292
- Hartl, D.L., 347
- Hartley, H.O., 84
- Hasegawa, M., 376
- Hedrick, P.W., 379
- Higgins, J.M., 347
- Hill, W.G., 128, 129, 134, 135, 248
- Hoppe, F., 306, 307, 322, 323
- Hoppensteadt, F.C., 204
- Hudson, R.R., 328, 347
- Hurst, L.D., 286
- Ito, K., 148
- Jayakar, S.D., 283, 286
- Jinks, J.L., 273
- Johnstone, R.A., 286
- Joyce, P., 355
- Jukes, T.H., 347, 373
- Karlin, S., 103, 113, 126, 129, 132,
180, 182–184, 186, 202, 211,
214–219, 221, 226, 239, 249,
250, 252, 253, 384
- Keeler, C., 5
- Kelleher, T.M., 72, 202
- Keller, J., 162, 164
- Kelly, F.P., 295, 296, 324, 325
- Kemeny, J.G., 91
- Kempthorne, O., 54, 56, 57, 75, 267,
277
- Ketcham, L.K., 290
- Kidwell, J.F., 46
- Kim, Y., 363
- Kimura, M., 23, 35, 57, 61, 70, 72,
74, 79, 82, 83, 128, 129,
132, 135, 147, 154, 159, 164,
166, 171, 178, 181, 182, 185,
190, 202, 204, 205, 207, 210,
228–234, 236–239, 247, 261,
262, 274, 280, 288–290, 297,
346, 374
- King, J.L., 347
- Kingman, J.F.C., 49, 52, 53, 195, 307,
321, 332, 337
- Kirby, K., 115
- Kishino, H., 376
- Kojima, K., 70, 72, 202
- Kolman, W., 277
- Krakauer, J., 225, 226
- Kreitman, M., 347
- Kuhner, M.K., 314
- Kusakabe, S., 366, 367, 369
- Langley, C.H., 186, 188, 347
- Last, K., 273
- Lessard, S., 65, 67, 256, 259, 261

- Levene, H., 186
 Levikson, B., 183, 184, 189
 Lewontin, R.C., 70, 85, 239, 240, 251, 252, 254, 274, 320
 Li, H., 347
 Li, W.-H., 239, 294, 357, 360, 361, 369
 Liberman, U., 55, 182-184, 250, 252, 253
 Littler, R.A., 110, 129, 131, 132, 134, 192-194
 Løvtrup, S., ix
 Lynch, M., viii
 Lyubich, Y.I., 208, 216, 250

 Malécot, G., 267, 328
 Mandl, P., 148
 Marrow, P., 286
 Marson, A., 290
 Martin, N.G., 273
 Maruyama, T., 170, 171, 280
 Matessi, C., 286
 Mather, K., 273
 May, R.M., 283
 Maynard Smith, J., 232, 234, 237-239, 363
 McCloskey, J.W., 321
 McDonald, J.H., 347
 McGregor, J., 103, 113, 129, 132, 180, 226
 McKean, H.P., 148, 150
 Miller, G.F., 168
 Mitton, J.B., 239
 Mohler, J.D., 277, 278
 Moran, P.A.P., 103, 104, 124, 152, 212
 Morgan, K., 316-318
 Mukai, T., 366, 367, 369
 Muller, H.J., 235

 Nagylaki, T., 57, 61, 67, 78, 143, 204, 207, 272, 280-282
 Nei, M., 128, 171, 221, 224, 239, 288, 294, 319, 379
 Nielsen, R., 347
 Nordborg, M., 329
 Norman, M.F., 152, 153, 179, 180
 Norton, H.T.J., 12, 282

 Ohta, T., 82, 129, 135, 147, 154, 185, 190, 228-234, 239, 274, 294
 Olby, R.C., 2

 Pearson, E.S., 84
 Pearson, K., 5, 6
 Perlitz, M., 364
 Pielou, E.C., 283
 Pollak, E., 54, 56, 57, 127, 277
 Price, G.R., 65, 259
 Prohorov, Y., 149
 Provine, W., 2
 Przeworski, M., 363
 Punnett, E.C., 12

 Rannala, B., 320
 Reich, D.E., 347
 Robertson, A., 52, 129, 134, 135, 168, 169
 Rodrigo, A.G., 383
 Rogers, J.S., 379
 Roughgarden, J., 283-285
 Roux, C.Z., 212, 249, 251, 253
 Roychoudhury, A.K., 171
 Rozanov, Y., 149

 Sabeti, P.C., 347
 Sawyer, S.A., 191, 347
 Schnell, F.W., 268, 270
 Schwartz, J., 286
 Seneta, E., 51
 Serant, D., 129, 317
 Shahshahani, S., 261
 Shaw, R.F., 277, 278
 Simonsen, K.L., 360, 361, 365
 Slatkin, M., 243, 254, 280, 320, 347
 Sneath, P.H.A., 379
 Snell, J.L., 91
 Sokal, R.R., 379
 Sprott, D.A., 48
 Stegun, I., 112, 114
 Stephan, W., 363, 364
 Stephens, M., 314
 Strobeck, C., 239, 316-318, 364, 365
 Svirizhev, Y.M., 261, 262, 264
 Svirizhev, Y.M., 265

 Tajima, F., 328, 352, 357, 358, 360, 361, 366, 367, 369

- Takano, T.S., 366, 367, 369
Tanaka, K., 184
Tavaré, S., viii, 310, 314, 325, 326,
329, 333, 338, 342, 355, 360,
372
Taylor, H.M., 384
Thomson, G.J., 169, 220, 232, 246,
247, 251
Tier, C., 162
Trajstman, A., 118, 340

Uyenoyama, M., 277

van Aarde, I.M.R., 268, 270
Verner, J., 277
Voronka, R., 162

Wade, M.J., 286
Walsh, B., viii
Watterson, G.A., 103, 106, 129–131,
133, 152, 170, 173, 177, 178,
189, 196, 197, 199, 291, 295,
297, 299, 300, 306, 309–311,
315, 317, 324, 349–352
Weinberg, W., 5
Weir, B.S., 129, 270
Whelan, S., 382
White, M., 282
Williams, G.C., 239
Wilson, E.O., 285
Wilson, S.R., 272
Wiuf, C., 344, 345
Wood, S.P., 347
Wright, S., 20, 23–27, 33, 34, 39–41,
54, 73, 83, 178, 197, 224,
272, 274, 277, 278, 286, 319
Wu, C.-I., 347, 357, 363

Yamato, J., 314
Yano, T., 376
Yokoyama, S., 239
Yu, P., 283
Yule, G.U., 6

Zhang, Y., 347
Zhang, Y.-P., 347

Subject Index

- 2 log λ , 382, 383
- adenine, 371
- age distribution, 189
- age structure, 282
- age-ordered alleles, 320, 322, 325
- altruism, 285–287
- asymptotic expansion, 160, 162, 164
- average effect, 8, 9, 62, 63, 66, 248, 256–258, 260
 - entire genome, 255
- average excess, 62, 244, 257, 266
- Bayesian statistics, 322
- binomial distribution, 180
- bioinformatics, vii
- Biston betularia, 224
- Blaisdell model, 375
- blending theory, 2
- blind watchmaker, ix
- boundary
 - entrance, 149, 158, 173, 188
 - exit, 149, 158, 188
 - natural, 149, 150, 158, 188
 - regular, 149, 158, 173, 188
- branching processes, 5, 27–30, 103, 163
- conditional, 103
- Cannings model, 37, 38, 95, 99–104, 109–111, 117, 120–122, 291, 292, 298, 308, 332, 341
- case-control study, 343
- clines, 280, 281
- coalescent, 301, 308, 310, 314, 320, 325, 328–348, 358, 361, 363, 364, 369
- coefficient of kinship, 286
- conditional distribution, 28–30, 318
- conditional processes, 89, 94, 95, 146–148, 162, 164, 169, 170, 173, 178, 189, 190
- configurations
 - delabelled, 112
- correlation between relatives, xviii, 6–8, 11, 53, 74, 77, 266–269, 271
 - Malécot's formula, 267
 - double first cousins, 9
 - father-son, 7, 9, 53, 77, 267, 268
 - for tight linkage, 270
 - full sibs, 9, 78, 269
 - grandfather-grandson, 268
 - half sibs, 269

- uncle-nephew, 9
- covariance
 - in stationary distribution, 110
- cytosine, 371
- diffusion approximation, 93, 94, 108, 111, 118, 121, 146, 158, 176–179, 192, 291–293, 295–298, 301, 308, 314, 315, 320, 322–327, 329, 332, 333
- diffusion coefficient, 144, 147, 148, 153, 157, 171, 174, 176, 182, 184, 191, 229
- diffusion methods, 136–200, 227–230
- Dirichlet distribution, 195
- distance, genetic, 371, 379
- dominance, 13, 32, 33, 46, 165, 166, 168, 176, 226, 248, 268, 269, 272, 299
 - evolution of, 32, 33, 39, 41, 221–224
- drift coefficient, 144, 147, 148, 153, 157, 165, 170, 171, 174, 176, 182, 184, 191, 229
- effective population size, ix, 37, 116, 118–122, 127, 128, 280
 - eigenvalue, 120, 123, 125, 126
 - inbreeding, 120–123, 125, 126
 - variance, 120, 124–126, 153, 165
- eigenfunction, 138, 139, 150, 151, 161, 183, 192
- eigenfunction expansion, 139, 183
- eigenvalue, 22, 52, 53, 87, 89, 90, 93–95, 100–104, 106, 107, 109–111, 114, 115, 119, 123, 125–127, 130–132, 134, 139, 150, 151, 161, 168, 169, 192, 194, 196, 209, 212, 213, 225, 228, 229, 253, 280, 305, 306, 384
- eigenvector, 22, 87, 89, 106, 134, 151, 161
- epistasis, 83, 271, 281
- equilibrium, 113, 173, 246, 258
 - point, 15, 18, 32, 34, 46, 47, 50, 51, 53–55, 61, 70, 71, 175, 176, 186, 202, 207–209, 211, 212, 216, 226, 285, 305
 - product, 251, 253, 254
 - quasi-stable, 29
 - stable, 14, 15, 35, 46, 47, 49, 52–54, 85, 212, 215–217, 219, 225, 226, 251, 253, 254, 285
 - unstable, 14, 220, 251, 254
- equilibrium equation, 46
- equilibrium mean fitness, 16, 34, 71, 80, 212, 220
- equivalence class, 330
- equivalence relation, 330
- ergodic arguments, 24
- evolution, 1
- exchangeable process, 102, 103
- fecundity, 56, 58
- Felsenstein model, 375–377, 382
- fertility, 11, 54, 57, 67, 128
- fertility parameters, 56
- fertility selection, 85
- fitness, 3, 6, 10–12, 14, 24, 32, 34, 43, 83, 129, 156, 165, 170, 171, 182, 183, 189, 197, 198, 201, 221, 224, 225, 231, 241, 243, 245, 249, 276–277
 - additive, 187, 210, 211, 213, 217, 249
 - different between sexes, 14, 45–47
 - distribution of, 84
 - frequency-dependent, 54, 85
 - inclusive, viii, 286
 - marginal, 49, 211, 214, 220, 246, 256–258
 - mean, 16, 17, 33, 34, 49, 51, 54, 56, 202–209, 211, 213, 214, 219, 226, 235, 250, 252, 259, 261, 283
 - multilocus, 255
 - multiplicative, 83, 86, 186, 210–213, 219, 237, 238, 250, 251, 254
 - surface, 40
 - symmetric viability, 210, 214–217, 226
- fixation, 21, 29, 39, 41, 90, 211, 232, 238, 239, 249, 289

- conditional mean time for, 89, 94, 105, 146, 148, 163, 164, 169, 170
- mean time for, 21, 22, 88, 90, 93, 95, 105, 108, 115, 140–142, 144, 160, 167, 177, 186, 193, 279, 280
- probability, 21
- probability of, 21, 25, 87, 90, 92, 98, 105, 109, 132, 139, 140, 159, 165, 177, 186, 192, 280, 289, 317
- variance of time for, 88, 143, 160, 163
- frequency spectrum, 116, 198, 199, 298, 302, 304, 312, 321, 322, 353, 362, 363
 - conditional, 312, 353, 357, 369
 - exact Moran model, 118, 119, 296, 325
 - sample, 311, 353
- Frobenius theory, 384
- Fundamental Theorem of Natural Selection (FTNS), ix, 16, 18, 39, 43, 62, 64–67, 241, 255, 259–261, 275
- game theory, viii
- Gegenbauer polynomial, 159
- GEM distribution, 321–323
- gene conversion, viii
- gene-environment interaction, viii
- generating function, 28
- geographical factors, 23, 44, 67, 120, 124, 125, 127, 128, 278–282, 347
- geometric distribution, 88, 301, 310, 315, 325, 329, 330, 339, 340
- guanine, 371
- Hardy–Weinberg frequencies, 4, 6, 12, 19, 45, 48, 49, 56, 59, 61, 249
 - not holding, 18
- Hardy–Weinberg law, 3, 5, 6, 19, 22, 36
- hitchhiking, 230–235, 362–365
- Hoppe’s urn, 306–308, 337, 350
- human genetics, viii, ix, xix, 1, 18, 128, 254, 274, 279, 290, 301, 342–345
- infinitely many alleles model, 111–119, 121, 172, 173, 178, 195, 197, 199, 290–297, 301–308, 316–320, 332, 341, 347–355, 364–365, 379
- infinitely many sites model, 289, 291, 297–301, 304, 308–319, 338, 347, 348, 352, 355–370, 379
- inter-population selection, 224
- Jukes–Cantor model, 373–377, 382
- Kimura model, 374–375, 377
- kin selection, 286, 287
- Kingman distribution, 195
- Kolmogorov equation
 - backward, 137, 138, 147, 227
 - forward, 137, 138, 145, 147, 182
- linkage, 32, 224
- linkage disequilibrium, 73, 78, 85, 86, 130, 202, 218, 227, 231, 232, 237–239, 241, 243, 246, 248, 249, 251, 266, 270, 271, 274
 - coefficient of, 69, 71, 130, 134, 204, 209, 211, 213, 217, 218, 229, 230, 247, 249, 267, 270
- linkage equilibrium, 69–71, 73, 75, 82, 221, 224, 245, 268, 270–272
- load, 16, 78
 - mutational, 16
 - segregational, ix, 78, 86
 - substitutional, ix, 78, 81–83, 86
- Malthusian parameter, 58, 59, 61, 277
- Markov chain Monte Carlo (MCMC), 314
- Markov chains, 20, 21, 27, 86–92, 107, 112, 114, 115, 126, 129, 130, 136, 137, 140, 141, 144–146, 153, 156–158, 160, 163, 167, 168, 171, 176, 179, 180, 192, 227, 228, 371, 377
 - reversibility of, 372–373, 377
- Markovian random variables, 21, 92, 115, 191, 384

- joint, 124, 151, 165
- mating
 - assortative, 10, 39, 271–273
 - non-random, 6, 18, 19, 62–65, 254–260, 266, 271, 275, 347
 - random, 2, 3, 5, 6, 10, 12, 17, 18, 43, 45, 48, 49, 59, 63, 65, 67–69, 201, 226, 235, 241, 243, 245, 246, 249, 266, 275
- mean, 28
 - in stationary distribution, 110, 176
- mean fitness increase theorem (MFIT), 16, 49, 61, 62, 64, 201, 202, 216, 217, 245, 250, 274, 283
- Mendelism, 1
- modifier genes, 39, 201, 235, 236, 239
- modifier theory, 221–227
- monophyletic group, 344, 383
- Moran model, 38, 104–109, 111, 117–119, 121, 174, 291, 294–297, 300–301, 306–308, 314–315, 320, 323–327, 329, 332, 338–341
- mutation, 3, 11, 15, 16, 21, 23, 26, 30, 95, 104, 115, 135
 - multi-way, 110, 194
 - one-way, 103, 106, 171–174, 177, 189
 - two-way, 107, 156, 174–176, 185
- mutation rate, 26, 27, 32, 158, 178, 221, 224, 226, 229, 275
- mutation-selection equilibrium, 16, 32
- natural selection, 1–3, 11, 17
- neighbor-joining, 371
- neutral theory, 346–369
- non-Darwinian theory, 346, 347
- nucleotide
 - predominant, 371
- optimality principles, 261–265
- order statistics, 195–198
- Origin of Species, The, 1, 285
- overdominance, 46, 47, 166, 168–170, 175, 211, 214, 217, 220, 231, 232, 283
 - associative, 230, 231
 - induced, 220
 - marginal, 216
- parsimony, 371
- partition structure, 307, 308, 341
- phylogenetic tree, 376–378, 381, 382
- Poisson distribution, 29, 126, 297, 298
- Poisson process, 329
- Poisson–Dirichlet distribution, 195, 196, 307, 321, 323
- purine, 374, 380
- pyrimidine, 374, 380
- quasi-Hardy–Weinberg law, 61
- quasi-linkage equilibrium (QLE), 202–205, 240, 275
- quasi-Markovian random variable, 124, 152, 165
- recombination, 10, 33, 39, 67, 68, 201, 227, 250, 252, 289, 297, 299, 316–319
 - evolutionary advantage of, 235–239
 - free, 254
 - zero, 253
- recombination fraction, 34, 35, 68, 71, 129, 202, 211, 213, 215, 218, 221, 224, 225, 227, 233, 235, 250–252, 268–270, 275, 316, 317
- recombination pattern, 208, 245
- return process, 189
- reversibility, 91, 107, 153, 189, 191, 296, 320, 322–326, 345, 372, 375, 377
 - of a Markov chain, 91, 334, 372–373, 378
- saltationists, 2
- SAS-CFF model, 186
- scale function, 149, 150, 157, 165
- selection, 5, 13–17, 24, 26, 32, 70, 81, 98, 104, 108, 165–172, 181, 183, 201, 231, 232, 239, 248, 272, 274, 276, 281, 283, 285, 299, 346, 350, 369
 - intergroup, 285
 - secondary, 221, 225

- selective sweep, 234, 305, 306,
 362–364
 selfing, 226
 sex ratio, 32, 35, 45, 277, 278, 282
 sex-linked genes, 48
 size-biased sampling, 296, 321, 322,
 325
 sociobiology, xvii, xviii, 285–287
 sojourn time, 144, 146, 147
 spectral expansion, 95, 161, 168, 209
 speed measure, 150, 157, 165
 stationary distribution, 26, 90, 91,
 96–98, 107, 110, 112, 135,
 145–146, 150–154, 158, 174,
 175, 177–180, 183–185,
 187–189, 191, 194, 229,
 373–376
 of configurations, 112, 118
 Stirling number, 114, 302, 353, 354
 stochastic local stability, 185
 substitution, 371
 sufficient statistic, 302, 368

 Tajima test, 352, 358–364, 366
 Taylor series, 28, 93, 137, 176
 thymine, 371
 time reversal, 91, 153
 transition, 374, 375
 transition probability matrix, 372–376
 transversion, 374, 375

 uniform distribution
 discrete, 373, 374

 variance, 2, 22, 28
 additive, 266
 additive by additive, 77, 267–270
 additive by dominance, 77,
 267–270
 additive gametic, 74, 203, 246,
 247
 additive genetic, xix, 8–10, 17,
 39, 51, 59, 60, 62, 63, 65,
 71–74, 76, 77, 203, 207,
 246–248, 250, 256, 258, 260,
 267–270
 additive genetic, marginal, 72,
 73
 analysis of (ANOVA), 319
 between populations, 319
 dominance, 8, 9, 76, 77, 266–270
 environmental, 10, 272
 epistatic, 77
 epistatic gametic, 74
 gametic, 74
 genic, xix
 in offspring number, 122, 128
 in stationary distribution, 97,
 110, 174, 176
 of D , 134
 subdivision of, 75
 total, 7–10, 17, 75, 77, 272
 total gametic, 203
 within population, 319

 Wright–Fisher model, 21, 37, 38,
 92–102, 107, 109–122, 128,
 136, 148, 156, 161, 174, 178,
 291–295, 297–306, 308–316,
 324–327, 332–342

Interdisciplinary Applied Mathematics

1. *Gutzwiller*: Chaos in Classical and Quantum Mechanics
2. *Wiggins*: Chaotic Transport in Dynamical Systems
3. *Joseph/Renardy*: Fundamentals of Two-Fluid Dynamics:
Part I: Mathematical Theory and Applications
4. *Joseph/Renardy*: Fundamentals of Two-Fluid Dynamics:
Part II: Lubricated Transport, Drops and Miscible Liquids
5. *Seydel*: Practical Bifurcation and Stability Analysis:
From Equilibrium to Chaos
6. *Hornung*: Homogenization and Porous Media
7. *Simo/Hughes*: Computational Inelasticity
8. *Keener/Sneyd*: Mathematical Physiology
9. *Han/Reddy*: Plasticity: Mathematical Theory and Numerical Analysis
10. *Sastry*: Nonlinear Systems: Analysis, Stability, and Control
11. *McCarthy*: Geometric Design of Linkages
12. *Winfrey*: The Geometry of Biological Time (Second Edition)
13. *Bleistein/Cohen/Stockwell*: Mathematics of Multidimensional
Seismic Imaging, Migration, and Inversion
14. *Okubo/Levin*: Diffusion and Ecological Problems: Modern Perspectives
(Second Edition)
15. *Logan*: Transport Modeling in Hydrogeochemical Systems
16. *Torquato*: Random Heterogeneous Materials: Microstructure and
Macroscopic Properties
17. *Murray*: Mathematical Biology I: An Introduction (Third Edition)
18. *Murray*: Mathematical Biology II: Spatial Models and Biomedical
Applications (Third Edition)
19. *Kimmel/Axelrod*: Branching Processes in Biology
20. *Fall/Marland/Wagner/Tyson (Editors)*: Computational Cell Biology
21. *Schlick*: Molecular Modeling and Simulation: An Interdisciplinary Guide
22. *Sahimi*: Heterogeneous Materials I: Linear Transport and Optical Properties
23. *Sahimi*: Heterogeneous Materials II: Nonlinear and Breakdown Properties
and Atomistic Modeling
24. *Bloch*: Nonholonomic Mechanics and Control
25. *Beuter/Glass/Mackey/Titcombe (Editors)*: Nonlinear Dynamics in
Physiology and Medicine
26. *Ma/Soatto/Košecká/Sastry*: An Invitation to 3-D Vision: From Images to
Geometric Models
27. *Ewens*: Mathematical Population Genetics (Second Edition)