Population Genetics Course

Population Genetics: Selection Notes 1

Single-Locus Selection Theory

1. GENERAL SELECTION EQUATION

An infinitely large, diploid, random-mating population with discrete generations and constant fitnesses is assumed. There are two alleles A_1 and A_2, with frequencies p and q (where $p + q = 1$) in a given generation. We have

Genotypes	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
Frequency at birth	p^2	$2pq$	q^2
Fitnesses	w_{11}	w_{12}	w_{22}
Frequencies After Selection	$\frac{p^2 w_{11}}{\bar{w}}$	$\frac{2pq\, w_{12}}{\bar{w}}$	$\frac{q^2 w_{22}}{\bar{w}}$

where $\bar{w} = p^2 w_{11} + 2pq\, w_{12} + q^2 w_{22}$ is the *mean fitness* of the population.

The new frequency of allele A_1 is : $p' = (p^2 w_{11} + pq\, w_{12})/\bar{w}$

The new frequency of allele A_2 is : $q' = (pq\, w_{12} + q^2 w_{22})/\bar{w}$

We can define *marginal fitnesses* for alleles A_1 and A_2 as

$$w_{1.} = p\, w_{11} + q\, w_{12} \text{ and } w_{2.} = p\, w_{12} + q\, w_{22}\text{, respectively.}$$

These measure the average fitnesses of individuals carrying the alleles in question, and we have $\bar{w} = p\, w_{1.} + q\, w_{2.}$.

The change in frequency of A_2 is $\Delta q = q' - q$; using the above equations for q' and $w_{1.}$ and $w_{2.}$, we obtain

$$\Delta q = q\,(w_{2.}/\bar{w}) - q = q\,(w_{2.} - \bar{w})/\bar{w} = q\,(w_{2.} - p\,w_{1.} - q w_{2.})/\bar{w}$$

i.e. $$\Delta q = q\,(p w_{2.} - p w_{1.})/\bar{w} = pq\,(w_{2.} - w_{1.})/\bar{w}$$

2. MODES OF SELECTION WITH CONSTANT RELATIVE FITNESSES

(i) Favoured allele A_1 is dominant, semi-dominant or recessive

The *relative fitnesses* of the three genotypes can be written as:

Genotypes	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
Relative Fitnesses	1	$1-hs$	$1-s$

s (assumed to be postive) is the *selection coefficient* for the less fit homozygote, A_2A_2, and h is the *dominance coefficient* ($0 \leq h \leq 1$).

Substituting these fitnesses into the equations for $w_{1.}$ and $w_{2.}$, we obtain

$$w_{1.} = p + q\,(1-hs) = 1 - qhs$$

$$w_{2.} = p\,(1-hs) + q\,(1-s) = 1 - phs - qs$$

so that

$$w_{1.} - w_{2.} = qs + phs - qhs = qs\,(1-h) + phs$$

Since $s > 0$, A_1 is always favoured (*directional selection*), and $w_{1.} - w_{2.} > 0$. This means that A_2 always *decreases* in frequency, regardless of the value of h.

(ii) Heterozygote advantage

In this case it is convenient to write the relative fitnesses as

Genotypes	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
Rel. Fitnesses	$1-s$	1	$1-t$

where s and t are positive.

We have: $w_{1.} = 1 - ps$ and $w_{2.} = 1 - qt$

and so $w_{2.} - w_{1.} = ps - qt$

When $qt = ps$, it is clear that Δq is zero and that there must be an *equilibrium* for the allele frequencies. At this equilibrium, we use p^* and q^* to indicate the values of the allele frequencies. From the above, $p^*s = q^*t$, so that $p^*/q^* = t/s$, which can be rearranged to give

$$p^* = t/(s + t) \quad \text{and} \quad q^* = s/(s + t)$$

If $q < q^*$, then $ps > qt$ and $\Delta q > 0$. If $q > q^*$, then $ps < qt$ and $\Delta q < 0$.

This shows that the system tends to approach the equilibrium from either side i.e. the equilibrium is *stable*.

(iii) Heterozygote disadvantage

This can be treated in the same way, except that s and t are now negative, and the equilibrium is *unstable.*

3. SELECTION WITH VARYING FITNESSES

(i) Temporal variation

Suppose that the population is exposed in successive generations to a cyclical alternation of two different sorts of environment, such that the fitnesses of the three genotypes at the locus are

Genotype	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
Env. 1	w_{11}^1	w_{12}^1	w_{22}^1
Env. 2	w_{11}^2	w_{12}^2	w_{22}^2

We can ask whether both $p = 0$ and $q = 0$ are unstable points, in which case we have a *protected polymorphism*, by asking whether both A_1 and A_2 show a net increase in frequency when rare.

Consider the case of rare A_2 in a generation when the environment is type 1; let the frequency of A_2 be $q \ll 1$. We then have

$$w_{2\cdot}^1 = pw_{12}^1 + qw_{22}^1 \approx w_{12}^1$$

$$\bar{w}^1 = p^2 w_{11}^1 + 2pqw_{12}^1 + q^2 w_{22}^1 \approx w_{11}^1$$

The new frequency of A_2 is thus

$$q' \approx q\, w_{12}^1 / w_{11}^1$$

(Terms of order q^2 have been neglected in obtaining this approximation.)

Next generation, the population is in environment 2, in which the marginal fitness of A_2 is $w_{2.}^{2}$ and the mean fitness is approximately w_{11}^{2}. Using the same principle as for the first generation, the frequency of A_2 becomes

$$q'' \approx q' \, (w_{12}^{2}/w_{11}^{2}) \approx q \, (w_{12}^{1}/w_{11}^{1}) \, (w_{12}^{2}/w_{11}^{2})$$

so that A_2 increases over the whole cycle if

$$(w_{12}^{1} w_{12}^{2}) / \, (w_{11}^{1} w_{11}^{2}) > 1$$

Similarly, A_1 increases when rare if

$$(w_{12}^{1} w_{12}^{2}) / \, (w_{22}^{1} w_{22}^{2}) > 1$$

These conditions are equivalent to the *geometric mean* over the two environments of the heterozygote fitness exceeding the geometric means for both homozygotes (the g.m. of two numbers is the square root of their product).

This can be generalized to any number of environments, or even a random sequence of environments (in general, the *geometric mean* of a set of *n* numbers is the *n* th root of their product). It is also equivalent to the arithmetic mean of the logarithm of the fitness of the heterozyotes exceeding the values for the homozygotes.

(ii) Spatially varying environments (Levene's Model)

In this model, the environment is divided up into a number of "niches"; the fitness of each individual in general depends on the niche in which it was born. Individuals spend their whole life from egg to adult in one niche, and then emerge from the niche to mate at random with adults from all niches. The *k* th niche contributes a *fixed* fraction c_k to the gene pool of the next generation. Females lay eggs randomly across niches in the next generation, so that each niche starts with the same gene frequency.

Let q be the frequency of allele A_2 before selection in a given generation. When q is small, the frequency after selection in niche k is

$$q_k' = q \, w_{2.}^{k}/\bar{w}^{k} \approx q \, w_{12}^{k}/w_{11}^{k}$$

where the fitnesses for the *k* th niche are given in the same way as in the temporal model.

The gene frequency next generation when $q \approx 0$ is thus

$$q' = \sum_k c_k q_k' \approx q \sum_k c_k (w_{12}{}^k / w_{11}{}^k)$$

so that A_2 increases if

$$\sum_k c_k (w_{12}{}^k / w_{11}{}^k) > 1$$

Similarly, A_1 increases if

$$\sum_k c_k (w_{12}{}^k / w_{22}{}^k) > 1$$