Population Genetics Course

Population Genetics Exercises 1

1. A sample of 5 alleles at the *Booze-1* locus of humans was sequenced for a portion of the gene that was 500 nucleotides long. Two alleles had nucleotides that differed from the first allele in the set at position 121, one differed at positions 10 and 300, and one at position 430. Use this information to estimate the two standard measures of sequence variability, π and θ_w, using the formulae in BC's first lecture.

There are 5x4/2 = 10 pairs of distinct alleles. The mean number of pairwise differences between all possible pairs of distinct alleles is (1 + 1 + 2 + 1 + 0 + 3 + 3 + 2 + 2 + 3)/10 = 18/10 = 1.8. Hence, π is 1.8/500 = 0.0036.

The number of segregating sites in the sample is 4. The Watterson correction factor is 1 + 0.5 + 0.333 + 0.25 = 2.083, so that θ = 4/(2.083 x 500) = 0.0038.

(see section 5 of handout for lectures 1 and 2; pp.28-29 of C & C).

What might cause any disagreement between these two estimates?

The two estimates are in fact quite close, so the most likely explanation is sampling error. Deviations from the neutral equilibrium model, such as selection, might be involved if sampling error could be ruled out.

2. The time to the most common recent ancestor (MRCA) of a sample of k alleles from a population is defined as the expected time to coalescence of all the alleles in the sample. Assuming a Wright-Fisher population model, derive an expression for this quantity, using the fact that the expected time for i alleles to coalesce to $i - 1$ is $4N/[i(i - 1)$. State what the asymptotic value for this is when k is very large. How does this relate to the mean coalescent time for a pair of alleles?

The mean time taken for i alleles to coalesce to $i - 1$ alleles is $2N/[i(i-1)/2] = 4N/i(i - 1)$. The mean time to the most recent common ancestor is the sum of this from $i = k$ to $i = 2$.

We can use the relation $1/(i-1) - 1/i = 1/i(i-1)$ to express the sum of the terms in i as:

$$\sum_{i=2}^{k}\left(\frac{1}{i-1}-\frac{1}{i}\right) = -\frac{1}{k}+1 = \frac{(k-1)}{k}$$

The mean time to the most recent common ancestor is thus $4N(k-1)/k$.
This approaches $4N$ as k increases. This is twice the mean coalescent time for a pair of alleles.

3. Use the formulae given in BC's first lecture to calculate the effective populations sizes (N_e) for autosomal loci in randomly mating populations with Poisson distributions of offspring numbers, when there are 50 breeding males and 50 breeding females, and when there are 5 males and 95 females. Compare the results with the case when there are 50 individuals of each sex of breeding age, but the variance in the offspring number of males and females is 10 rather than 2. Think about why the differences between the numbers of breeding individuals and the N_e values arise.

The population size in the first case is 100; from, $1/N_e \approx 1/200 + 1/200 = 1/100$ (using the formula on slide 89), giving $N_e \approx 100$. Population size N and N_e are thus approximately the same in this case.
In the second case, $1/N_e \approx 1/380 + 1/20 = 0.0026 + 0.050 = 0.0526$, so that $N_e \approx 19$. N_e is thus much smaller than N (which is equal to 100 again), and is close to four times the number of males.
Since each individual receives one gene copy at a locus from each parent, males and female make equal contributions to the next generation. Thus, if males are much less common than females in the breeding population, half the genes in the population will be transmitted through the small number of breeding males.
In the third case, $V_k = 10$, and $N_e \approx 4 \times 100/[2 + 10] = 33$, so that $N_e \approx 50$.

The excess variance has a similar effect to a sex inequality in the numbers of breeding individuals, reflecting the fact that genes are passed to the next generation through a smaller number of parents than with purely random variation.

4. The melanic form of a species of moth with annual generations is controlled by a single semi-dominant, autosomal allele, with a selective advantage to the homozygote of 0.1 in a polluted environment. At the time of the establishment of industry in Smoketown, Scotland, the frequency of the allele for melanism was 0.001.

i. What will be the frequency of non-melanics in the Smoketown area after 120 generations?

From slide 49 of lecture 3, the rate of change of allele frequency with intermediate dominance is about one-half that for the asexual case with the same selection coefficient s, provided that selection is weak, which is reasonably accurate in this case.

Thus, we can use a modification of eqn 3. of Selection Notes 2 to obtain the frequency q_t of the allele for melanism at time t:

$$t \approx (2/s) \ln \{q_t p_0 / q_0 p_t\}$$

Thus: $q_t p_0 / q_0 p_t = \exp(st/2)$

$$q_t/p_t = (q_0/p_0) \exp(st/2)$$

In this case, $q_0 = 0.001$, $p_0 = 0.999$, $t = 120$, $s = 0.1$, $st/2 = 6$.

$$q_t/p_t = = q_t/(1 - q_t)\ (0.001/0.999)\exp(6.0) = 1.81$$

$$q_t = 0.404\ (1 - q_t);$$

so that $1.04\ q_t = 0.404$

$$q_t = 0.404/1.04 = 0.288$$

The frequency of non-melanics is the frequency of homozygotes for the non-melanic allele, which is $p_t^2 = (1 - 0.288)^2 = 0.712^2 =$

0.508.

Alternatively, you could use the formula for the haploid case with half the selection coefficient for the haploid case, $q_t/p_t = (q_0/p_0)\ (1 + 0.05)^t$.(p.1 of Selection Notes 2). It gives very similar results

ii. If no melanics had been present in the population at the time of industrialization, how many mutations to melanism would have to occur to ensure a probability of 0.99 that at least one of them would escape random loss?
(Use **Selection Notes 2** and BC's 2nd lecture to find the relevant formulae).

Using p.5 of the notes for Lectures 5-6, the probability of survival of the mutation is (approximately) $Q = 2s$, where s is the selective advantage to the mutant heterozygote NOT THAT TO THE HOMOZYGOTE: this means that $s = 0.05$, so that $Q = 0.1$.

For there to be a probability of 0.99 that that at least one mutation out of n survives, there has to be a probability 0.01 that all n go extinct. By the law of independent probabilities (or the first term of the binomial distribution), this probability is $(1 - Q)^n$.

We thus need to solve

$$(1 - Q)^n = 0.01$$

i.e. $n \ln (1-2s) = \ln (0.01)$

$$n = \ln (0.01)/ \ln (1-2s)$$

$$= - 2.996/\ln(0.9) = 43.7$$

43 mutations are thus needed.