Probabilistic Models, Homework 4, due March 29

Turn-in problems:

1(a) Consider the Wright–Fisher model with $N = 2$ (that is, $2N = 4$) with equal initial numbers (0th generation) of the two alleles A and a. Find the probability that A becomes fixed in the first generation and in the second generation, respectively (note: if it becomes fixed in the first generation, it cannot become fixed in the second).

(b) Find the probability that A is fixed in the second generation (note the difference between "is" and "becomes" fixed). Use LTP and condition on the number X_1 of As in the first generation.

(c) Find the probability that fixation of one allele occurs in the first generation.

(d) Now assume variable population size with initial size $N_0 = 2$ as in **(a)**. Do part **(c)** in the cases where population size in the first generation halves and doubles, respectively. What is the effect on the fixation probability? Does it halve/double?

(e) Compute the effective population size N_e in the two cases in part **(d)** (round to nearest integer). Compare to the average (arithmetic mean) population size.

2. In the Wright–Fisher model individuals reproduce by choosing $2N$ times with replacement, thus creating the next generation from the present. Another way of looking at this is to note that each individual has a number of offspring that is described by a random variable X. If we label the individuals $1, 2, ..., 2N$, we then have the numbers of offspring $X_1, X_2, ...X_{2N}$ where $X_1 + X_2 + \cdots + X_{2N} = 2N$.

(a) What is the distribution of X_k for a given k?

(b) Argue that the random variables $X_1, X_2, ..., X_{2N}$ have the same distribution but are *not* independent.

3. The distribution in problem **2** above corresponds to sampling with replacement. More generally, we can let the X_k be any nonnegative integer-valued random variables with the same distribution as long as they are such that $X_1 + X_2 + \cdots + X_{2N} = 2N$ (in order to keep a constant population size). As we have seen, in the Wright–Fisher model, the time T until the most recent common ancestor equals $E[T] \approx 4N$ (time unit: generations) if the sample k is reasonably large. In the more general model, it can be shown that

$$E[T] \approx \frac{4N}{\sigma^2}$$

where σ^2 is the limit of the variance in the offspring distribution as $N \to \infty$, $\sigma^2 = \lim_{N\to\infty} \mathrm{Var}[X_k]$ (from a practical point of view, σ^2 is the variance in the offspring distribution assuming N is large).

(a) Find σ^2 in the Wright–Fisher model and show that the general formula applies. Note: to do this you must have solved **2(a)** correctly, if you didn't manage to solve it, let me know and I'll give you a hint).

(b) If σ^2 increases, $E[T]$ decreases and vice versa. Give an intuitive argument why this makes sense.

(c) If $\sigma^2 = 0$, we get $E[T] = \infty$. What situation does this describe?