For Math 635, Spring 2012.
David F. Anderson
January 30, 2012

The Wright-Fischer Model

We consider a model from population genetics, which was first developed by Fischer, and later extended by Wright. In this model we assume the existence of N diploid (two copies of each gene) individuals. Thus, there are a total of $2N$ genes in the gene pool. We make the following assumptions:

1. The number of individuals remains constant at N from generation to generation.
2. The genes for any individual in the $(n+1)$st generation are randomly selected (with replacement) from the pool of genes in the nth generation.

Note that the last assumption allows us to disregard the individuals, and only consider the gene pool itself.

We suppose we have two alleles of the gene in question, which we denote by A and a. We let $X_n \in \{0, 1, \ldots, 2N\}$ denote the number of alleles of type A in the entire gene pool at time n. Oftentimes A is assumed to be a *mutant* that has entered the population. We are interested in the probabilities associated with *fixation*, meaning when the system is homogeneous in A, which occurs when $X_n = 2N$ and A has overtaken the population, or in a, which occurs when $X_n = 0$. Using the terminology developed so far in the class, we can set

$$\tau = \min\{n : X_n \in \{0, 2N\}\},$$

and note that τ is a stopping time with respect to the filtration $\mathcal{F}_n = \{X_0, X_1, \ldots, X_n\}$. We now want to calculate

$$p_0(j) = P\{X_\tau = 0 | X_0 = j\}.$$

Using martingale techniques to solve this problem

We consider the transition probabilities of this Markov model. Supposing that $X_n = i$, for some $i \geq 0$, what is the probability that $X_{n+1} = j$? Because of our simplifying assumptions, we see that, conditioned on X_n, the value of X_{n+1} is a binomial random variable with parameters $n = 2N$ and $p = X_n/(2N)$. Therefore,

$$P\{X_{n+1} = j | X_n = i\} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}.$$

Hence,

$$\mathbb{E}[X_{n+1} | X_n] = np = X_n,$$

and X_n is a martingale!

Now, it should be clear that $\mathbb{E}|X_{n\wedge\tau}| < 2N$, so we can apply the dominated convergence theorem so long as $n \wedge \tau \to \tau$ almost surely, which follows if $P(\tau < \infty) = 1$. We delay showing this fact until the end of this note, and will simply assume that $P(\tau < \infty) = 1$ for now.

Applying the dominated convergence theorem to $X_{n\wedge\tau}$, where we recall that $X_0 = j$, we see that

$$j = X_0 = \mathbb{E}_j X_{n\wedge\tau} \to \mathbb{E}_j X_\tau, \quad \text{as } n \to \infty.$$

Therefore,

$$j = 0P_j\{X_\tau = 0\} + 2NP_j\{X_\tau = 2N\} = 2N(1 - p_0(j)) \implies p_0(j) = 1 - \frac{j}{2N} = \frac{2N - j}{2N}.$$

Note also that

$$P_j\{X_\tau = 2N\} = 1 - p_0(j) = \frac{j}{2N}.$$

It remains to show that $P(\tau < \infty) = 1$, which we do now. Note that

$$P\{X_{n+1} \in \{1, 2, \ldots, 2N-1\} | X_n \in \{1, \ldots, 2N-1\}\} \leq 1 - \left(\frac{1}{2N}\right)^{2N} \stackrel{\text{def}}{=} p,$$

since one way to leave the set $\{1, 2, \ldots, 2N-1\}$ is the have $2N$ choices of A, which has probability at least $1/2N$. Therefore, letting $E_k = \{X_k \in \{1, 2, \ldots, 2N-1\}\}$, we see that for any initial condition X_0,

$$P\{E_n\} = P\{E_n, E_{n-1}\} + P\{E_n, E_{n-1}^c\} = P\{E_n | E_{n-1}\}P\{E_{n-1}\} \leq pP\{E_{n-1}\}.$$

Iterating shows

$$P\{E_n\} \leq p^n.$$

Therefore,

$$P\{\tau = \infty\} \leq P\{\tau > n\} = P\{E_n\} \leq p^n \to 0, \quad \text{as } n \to \infty.$$