
Diffusion Processes

“One cannot escape the feeling that these mathematical formulas have an independent existence and an intelligence of their own.” H.R. Hertz

As we have mentioned several times, if we let the population size $N \rightarrow \infty$ then our processes become deterministic. In this section, we will see that if we let $N \rightarrow \infty$ and at the same time speed up time so that it runs at rate $O(N)$, then allele frequencies converge to limits called diffusion processes. This will allow us to obtain more detailed results about the models with selection introduced in the previous chapter. A rigorous treatment of diffusion processes requires a fair amount of mathematical sophistication and the details themselves could fill a book, see e.g., Durrett (1996). Here, we will content ourselves to state and explain the use of the main formulas useful for computation.

As an antidote to the mathematical skullduggery, we will give some anecdotes concerning the historical development of the use of diffusion process in genetics that, taking place in the 30s, 40s, and 50s, occurs in parallel to the development of a rigorous mathematical foundation for probability theory. Our first is a quote from Feller’s (1951) Berkeley Symposium paper, which began the development of the mathematical machinery for treating the convergence of Markov chains to diffusion processes:

“There exists a huge literature on the mathematical theory of evolution and statistical genetics, but existing methods and results are due almost entirely to R.A. Fisher and Sewall Wright. They have attacked individual problems with great ingenuity and an admirable resourcefulness, and had in some instances to discover for themselves isolated facts of the general theory of stochastic processes. However, as is natural with such pioneer work, it is not easy to penetrate to the mathematical core of the arguments to discover the explicit and implicit assumptions underlying the theory.”

A footnote to the first sentence of the quote says: “See Fisher (1930), Wright (1939) and Wright (1942). It is difficult to give useful references to original

papers, since these are mostly highly technical and inaccessible to nonspecialists." I am sure that many biologists have similar feelings about the mathematics literature.

7.1 Infinitesimal mean and variance

To motivate the definition of a diffusion process, we begin by recalling that a continuous-time Markov chain X_t is defined by giving the rate $q(i, j)$ at which the chain jumps from i to j . That is, if we use P_i for the distribution of the process starting from i then

$$P_i(X_s = j) = q(i, j)s + o(s)$$

where $o(s)$, pronounced "little oh of s ," is a quantity that when divided by s tends to 0 as $s \rightarrow 0$. If we define $q(i, i) = -\sum_{j \neq i} q(i, j)$ then the rows of the matrix sum to 0 and

$$P_i(X_s = i) = 1 + q(i, i)s + o(s)$$

Combining the last two formulas, it follows that if f is a bounded function then

$$E_i f(X_s) = (1 + q(i, i)s)f(i) + \sum_{j \neq i} q(i, j)s f(j) + o(s)$$

Rearranging we have

$$\frac{E_i f(X_s) - f(i)}{s} = \sum_j q(i, j)f(j) + o(1)$$

where $o(1)$ denotes a quantity that (when divided by 1) tends to 0 as $s \rightarrow 0$. Letting $s \rightarrow 0$

$$\left. \frac{d}{ds} E_i f(X_s) \right|_{s=0} = Qf(i) \tag{7.1}$$

where the right-hand side is the i th component of the product of the matrix $Q = q(i, j)$ and the vector $f(j)$. Q is called the *infinitesimal generator* of X_s .

To define diffusion processes, we will take an approach that is not intuitive, but is efficient.

Definition. A one dimensional diffusion process is a continuous Markov process with infinitesimal generator

$$Lf = \frac{1}{2}a(x)\frac{d^2}{dx^2}f + b(x)\frac{d}{dx}f$$

That is, we have $(d/dt)E_x f(X_t)|_{t=0} = Lf(x)$.

To see what this means, note that if we take $f(x) = x$ then $f'(x) = 1$ and $f''(x) = 0$ so

$$\left. \frac{d}{dt} E_x X_t \right|_{t=0} = b(x)$$

while if we fix x and define $f(y) = (y - x)^2$ then $f'(x) = 0$ and $f''(y) = 2$ so

$$\left. \frac{d}{dt} E_x (X_t - x)^2 \right|_{t=0} = a(x)$$

For this reason, $b(x)$ and $a(x)$ are called the *infinitesimal mean* and *infinitesimal variance*.

Taking $f(y) = (y - x)^4$ we have $f'(x) = 0$ and $f''(x) = 0$ so

$$\left. \frac{d}{dt} E_x (X_t - x)^4 \right|_{t=0} = 0$$

Since $(y - x)^4 \geq 0$, we have $E_x (X_t - x)^4 \geq \epsilon^4 P_x(|X_t - x| > \epsilon)$, and it follows that

$$\frac{1}{t} P_x(|X_t - x| > \epsilon) \rightarrow 0 \quad \text{as } t \rightarrow 0$$

It can be shown that this condition implies that the paths $t \rightarrow X_t$ are continuous. To see why we need this probability to be $o(t)$, recall that for continuous time Markov chains with jumps

$$P_i(X_t = j)/t \rightarrow q(i, j) \quad \text{as } t \rightarrow 0.$$

To explain the intuitive meaning of the coefficients $b(x)$ and $a(x)$ we will consider some examples.

Example 7.1. Deterministic motion. Suppose $X_0 = x$ and $dX_t/dt = b(X_t)$. A little calculus shows

$$\begin{aligned} f(X_t) - f(X_0) &= \int_0^t \frac{d}{ds} f(X_s) ds \\ &= \int_0^t f'(X_s) \frac{dX_s}{ds} ds = \int_0^t f'(X_s) b(X_s) ds \end{aligned}$$

So if f' and b are continuous

$$\frac{f(X_t) - f(x)}{t} \rightarrow f'(x)b(x)$$

i.e., $Lf(x) = b(x)f'(x)$. Thus, when $a(x) = 0$, a diffusion process reduces to a differential equation.

Notation. In the next example and in what follows $B_t = B(t)$ and the second form will often be used when t has subscripts or a complicated formula.

Example 7.2. Brownian motion. Suppose $B(0) = x$ and for $0 = t_0 < t_1 < \dots < t_n$, $B(t_1) - B(t_0)$, $B(t_2) - B(t_1)$, \dots , $B(t_n) - B(t_{n-1})$ are independent with $B(t_i) - B(t_{i-1})$ normally distributed with mean 0 and variance $\sigma^2(t_i - t_{i-1})$. Using Taylor's theorem, when t is small

$$f(B_t) - f(B_0) \approx f'(B_0)(B_t - B_0) + \frac{1}{2}f''(B_0)(B_t - B_0)^2$$

Taking expected values

$$E_x(f(B_t) - f(x)) \approx \frac{1}{2}f''(x)\sigma^2 t$$

so $Lf(x) = (\sigma^2/2)f''(x)$, i.e., $a(x) = \sigma^2$ and $b(x) = 0$. Thus, $a(x)$ measures the size of the stochastic fluctuations, or what biologists call random genetic drift.

Example 7.3. Stochastic differential equations. Let $\sigma(x) = \sqrt{a(x)}$. Intuitively, a diffusion process has for small t

$$(\star) \quad X_t - X_0 \approx b(X_0)t + \sigma(X_0)(B_t - B_0)$$

If b and σ are Lipschitz continuous, i.e., $|b(x) - b(y)| \leq K|x - y|$ and $|\sigma(x) - \sigma(y)| \leq K|x - y|$ then it can be shown that the integral equation

$$X_t - X_0 = \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s$$

has a unique solution, where the second integral is defined to be the limit of approximating sums $\sum_i \sigma(X(s_{i-1}))(B(s_i) - B(s_{i-1}))$. The formalities involved in making the last sentence precise are considerable but the intuition in (\star) is important: a diffusion process is a differential equation plus random fluctuations, which can be thought of as coming from a Brownian motion with a state dependent variance.

7.2 Examples of diffusions

In this section, we will introduce many of the examples from genetics that we will study. A formal proof of the convergence of Markov chains to limiting diffusions is somewhat complicated. Here, we will content ourselves to compute the limits of the infinitesimal mean and variance. Theoretical results which show that this is sufficient to conclude convergence can be found in Section 8.7 of Durrett (1996) or Section 7.4 of Ethier and Kurtz (1986).

Example 7.4. Wright-Fisher model with selection. There are two alleles, A and a . The fitness of A is 1 and fitness of a is $1 - s$ where $s \geq 0$. In the Wright-Fisher

model, this can be implemented by declaring that, as we build up the state at time $t + 1$ by drawing with replacement from generation t , we always accept an A that is drawn, but we keep an a with probability $1 - s$. Here, selection acts on the individual chromosomes, so, in effect, we have a population of $2N$ haploid individuals. Later we will discuss the more complicated situation of selection acting on diploids.

If the frequency of allele A in generation 0 is $X_0 = x$ then a newly drawn ball will be kept with probability $x + (1 - x)(1 - s)$ so the expected frequency in the next generation will be

$$x' = \frac{x}{x + (1 - x)(1 - s)} = \frac{x}{1 - (1 - x)s} = x + x(1 - x)s + o(s)$$

since $1/(1 - y) = 1 + y + y^2 + \dots$. The number of A 's in the next generation N_1 will be binomial($2N, x'$) so the frequency $X_1 = N_1/2N$ has

$$E(X_1 - X_0) = x(1 - x)s + o(s)$$

To take the diffusion limit, we want to write time in units of $2N$ generations, i.e., let $Y_t = X_{[2Nt]}$ where $[s]$ is the largest integer $\leq s$. Since time 1 for X corresponds to time $1/2N$ for Y , we want the change in the mean in one time step to be of order $1/2N$, so we let $\gamma = 2Ns$ and write

$$E(Y_{1/N} - Y_0) = x(1 - x)\gamma \cdot \frac{1}{2N} + o(N^{-1})$$

where $o(N^{-1})$ is a term that when divided by N^{-1} tends to 0 as $N \rightarrow \infty$.

The variance of N_1 is $2Nx'(1 - x')$, and $\text{var}(cZ) = c^2 \text{var}(Z)$ so the variance of X_1 is $x'(1 - x')/2N$. When $s = \gamma/2N$, $x' = x + o(1)$ and we have

$$\text{var}(Y_{1/N} - Y_0) = x(1 - x) \cdot \frac{1}{2N} + o(N^{-1})$$

Combining our calculations we see that the infinitesimal generator is

$$Lf = \frac{1}{2}x(1 - x)\frac{d^2}{dx^2}f + \gamma x(1 - x)\frac{d}{dx}f \quad (7.2)$$

In some papers in the biology literature time is not sped up and one sees

$$Lf = \frac{1}{4N}x(1 - x)\frac{d^2}{dx^2}f + sx(1 - x)\frac{d}{dx}f$$

Example 7.5. Wright-Fisher model with selection and mutation. As before, we have two alleles, A and a , with the fitness of A is 1 and fitness of a is $1 - s$. This time $a \rightarrow A$ with probability μ_1 and $A \rightarrow a$ with probability μ_2 . In defining our process we will suppose that selection occurs first followed by mutation. For a concrete story, suppose that the fitnesses give the relative probabilities of the two types surviving long enough to reproduce, at which point a genetic

mutation may occur. If the frequency of allele A in generation 0 is $X_0 = x$ then the frequency in the next generation will be

$$x'' = x' + \mu_1(1 - x') - \mu_2x' \quad \text{where} \quad x' = x + x(1 - x)s + o(s)$$

Letting $Y_t = X_{[2Nt]}$, $\gamma = 2Ns$, $\beta_i = 2N\mu_i$, and noting that $x' = x + o(1)$ we see that

$$E(Y_{1/N} - Y_0) = \{x(1 - x)\gamma + \beta_1(1 - x) - \beta_2x\} \cdot \frac{1}{2N} + o(N^{-1})$$

Again $x'' = x + o(1)$ so we have

$$\text{var}(Y_{1/N} - Y_0) = x(1 - x) \cdot \frac{1}{2N} + o(N^{-1})$$

Combining our calculations we see that the infinitesimal generator is

$$Lf = \frac{1}{2}x(1 - x)\frac{d^2}{dx^2}f + \{\gamma x(1 - x) + \beta_1(1 - x) - \beta_2x\}\frac{d}{dx}f \quad (7.3)$$

The source of each term is

binomial sampling	$\frac{1}{2}x(1 - x)\frac{d^2}{dx^2}f$
selection, $2Ns = \gamma$	$\gamma x(1 - x)\frac{d}{dx}f$
mutation, $2N\mu_i = \beta_i$	$\{\beta_1(1 - x) - \beta_2x\}\frac{d}{dx}f$

Example 7.6. Moran model. There are two alleles, A and a . The fitness of A is 1 and fitness of a is $1 - s$. Mutations $a \rightarrow A$ occur at rate μ_1 and $A \rightarrow a$ occur at rate μ_2 . For simplicity, we assume that mutations occur during the individual's life, not at birth, so adding the mutation rates to the transition rates from Section 6.1 we have

$$\begin{aligned} k \rightarrow k + 1 & \quad \text{at rate} \quad (2N - k) \left(\frac{k}{2N} + \mu \right) \\ k \rightarrow k - 1 & \quad \text{at rate} \quad k \left(\frac{2N - k}{2N}(1 - s) + \nu \right) \end{aligned}$$

Let X_t be the fraction of individuals with the A allele. To derive the diffusion approximation we note that if $k/2N = x$

$$\begin{aligned} \frac{d}{dt}EX_t &= \frac{1}{2N} \left[(2N - k) \left(\frac{k}{2N} + \mu_1 \right) - k \left(\frac{2N - k}{2N}(1 - s) + \mu_2 \right) \right] \\ &= (1 - x)\mu_1 - x\mu_2 + x(1 - x)s \end{aligned}$$

Letting $\beta_i = N\mu_i$, and $\gamma = Ns$ we see that the drift coefficient for the process run at rate N is

$$b(x) = (1 - x)\beta_1 - x\beta_2 + x(1 - x)\gamma$$

To compute the second order term, we note that after either an up jump or a down jump $(X_t - x)^2 = (1/2N)^2$, so

$$\frac{d}{dt}E(X_t - x)^2 = \frac{1}{(2N)^2} \left[(2N - k) \left(\frac{k}{2N} + \mu_1 \right) + k \left(\frac{2N - k}{2N}(1 - s) + \mu_2 \right) \right]$$

Since $\mu_1, \mu_2, s, \rightarrow 0$ and $k/2N = x$ we have

$$\frac{d}{dt}E(X_t - x)^2 = \frac{1}{2N} [2x(1 - x) + o(1)]$$

Thus for the process run at rate N the diffusion coefficient is

$$a(x) = x(1 - x)$$

Combining our calculations we see that the infinitesimal generator is again

$$Lf = \frac{1}{2}x(1 - x)\frac{d^2}{dx^2}f + [\gamma x(1 - x) + \beta_1(1 - x) - \beta_2x]\frac{d}{dx}f \tag{7.4}$$

In Section 1.5, we saw that the Moran model coalesces twice as fast as the Wright-Fisher model. To compensate for this, we sped up time by N rather than $2N$, in order to arrive at the same diffusion limit.

Example 7.7. General diploid selection model. We again have two alleles A and a but the fitnesses of diploid individuals are

AA	Aa	aa
$1 - s_0$	$1 - s_1$	$1 - s_2$

If the frequency of allele A in generation 0 is $X_0 = x$ then assuming random union of gametes and reasoning as in Section 6.2, the frequency in the next generation will be

$$\begin{aligned} x' &= \frac{x^2(1 - s_0) + x(1 - x)(1 - s_1)}{x^2(1 - s_0) + 2x(1 - x)(1 - s_1) + (1 - x)^2(1 - s_2)} \\ &= \frac{x - s_0x^2 - s_1x(1 - x)}{1 - x^2s_0 - 2x(1 - x)s_1 - (1 - x)^2s_2} \end{aligned}$$

Ignoring terms with s_i^2 and $s_i s_j$ the above

$$\approx x - s_0x^2 - s_1x(1 - x) + s_0x^3 + 2x^2(1 - x)s_1 + x(1 - x)^2s_2$$

A little algebra now shows

$$\begin{aligned} x' - x &\approx x(1 - x)[-s_0x - s_1(1 - 2x) + s_2(1 - x)] \\ &= x(1 - x)[s_2 - s_1 + x(2s_1 - s_0 - s_2)] \end{aligned}$$

Letting $Y_t = X_{[2Nt]}$, $\gamma_i = 2Ns_i$, $\delta = \gamma_2 - \gamma_1$ and $\eta = 2\gamma_1 - \gamma_0 - \gamma_2$ then

$$E(Y_{1/N} - Y_0) = x(1-x)[\delta + \eta x] \cdot \frac{1}{2N} + o(N^{-1})$$

Again $x' = x + o(1)$ so we have

$$\text{var}(Y_{1/N} - Y_0) = x(1-x) \cdot \frac{1}{2N} + o(N^{-1})$$

Combining our calculations we see that the infinitesimal generator is

$$Lf = \frac{1}{2}x(1-x)\frac{d^2}{dx^2}f + x(1-x)[\delta + \eta x]\frac{d}{dx}f \quad (7.5)$$

If there is mutation $a \rightarrow A$ at rate $\mu_1 - 1$ and $A \rightarrow a$ at rate μ_2 , and we let $\beta_i = 2N\mu_i$ then this adds a term of the form

$$\{\beta_1(1-x) - \beta_2x\}\frac{d}{dx}f$$

There are several important special cases

Additive selection. $s_0 = 0$, $s_1 = s$, $s_2 = 2s$, and let $\gamma = 2Ns$. $\delta = \gamma_2 - \gamma_1 = \gamma$ and $\eta = 2\gamma_1 - \gamma_0 - \gamma_2 = 0$ so

$$\begin{aligned} \delta + \eta x &= \gamma \\ b(x) &= \gamma x(1-x) \end{aligned} \quad (7.6)$$

just as in our previous Wright-Fisher model with selection.

Balancing selection. $s_1 = 0$, so $\delta = \gamma_2$ and $\eta = -(\gamma_0 + \gamma_2)$. If we let $x_0 = \gamma_2/(\gamma_0 + \gamma_2)$ then

$$\begin{aligned} \delta + \eta x &= (\gamma_0 + \gamma_2)(x_0 - x) \\ b(x) &= (\gamma_0 + \gamma_2)x(1-x)(x_0 - x) \end{aligned} \quad (7.7)$$

From this we see that the drift is < 0 for $x > x_0$ and > 0 for $x < x_0$. In the symmetric case $\gamma_0 = \gamma_2 = \gamma$ so $x_0 = 1/2$ and

$$\begin{aligned} \delta + \eta x &= \gamma(1-2x) \\ b(x) &= \gamma x(1-x)(1-2x) \end{aligned} \quad (7.8)$$

A is dominant. Aa has the same fitness as AA . $s_0 = s_1 = 0$, $s_2 = s$. $\delta = \gamma_2 - \gamma_1 = \gamma$ and $\eta = 2\gamma_1 - \gamma_0 - \gamma_2 = -\gamma$

$$\begin{aligned} \delta + \eta x &= \gamma(1-x) \\ b(x) &= \gamma x(1-x)^2 \end{aligned} \quad (7.9)$$

A is recessive. Aa has the same fitness as aa . $s_0 = 0$, $s_1 = s_2 = s$. $\delta = \gamma_2 - \gamma_1 = 0$ and $\eta = 2\gamma_1 - \gamma_0 - \gamma_2 = \gamma$.

$$\begin{aligned} \delta + \eta x &= \gamma x \\ b(x) &= \gamma x^2(1-x) \end{aligned} \quad (7.10)$$

To compare the drifts for additive selection versus dominant and recessive alleles we have graphed the solution of the differential equation $dX_t/dt = b(X_t)$ for the three drifts starting with $X_0 = 0.01$. To make the selective advantage of AA over aa the same in the three cases, we have taken $\gamma = 1, 2, 2$. Note that because of the extra factor of $(1 - x)$ the dominant case has more trouble getting to 1, while due to the extra factor of x , the recessive case has a hard time escaping from 0.

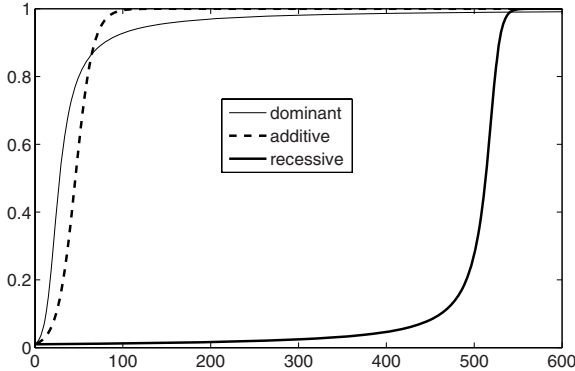


Fig. 7.1. Solution of $dX_t/dt = b(X_t)$ for the three drifts for additive selection, dominant alleles, and recessive alleles.

We can generate new examples of diffusions from old ones by

Theorem 7.1. Change of variables. *If h is increasing and has two continuous derivatives, then $Y_t = h(X_t)$ is a diffusion process with infinitesimal mean and variance*

$$\bar{a}(y) = a(x)h'(x)^2 \quad \bar{b}(y) = Lh(x)$$

where $x = f^{-1}(y)$.

Proof. By calculus

$$\frac{d}{dx}f(h) = f'(h)h' \quad \frac{d^2}{dx^2}f(h) = f''(h)(h')^2 + f'(h)h''$$

Using this in the definition of the generator

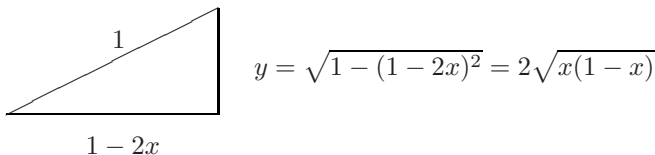
$$\begin{aligned} Lf(h) &= \frac{1}{2}a(x)[f''(h)(h')^2 + f'(h)h''] + b(x)f'(h)h' \\ &= \frac{1}{2}a(x)(h')^2 f''(h) + Lh(x)f'(h) \end{aligned}$$

which gives the result. \square

Example 7.8. Fisher's transformation. Let X_t be the Wright-Fisher model with no mutation or selection. Fisher (1922) discovered a remarkable transformation, although he did not get the answer right the first time. See pages 88-89 in Fisher (1930) and pages 119-120 in Wright (1931). Let $h(x) = \cos^{-1}(1-2x)$. This maps $[0, 1] \rightarrow [0, \pi]$. Recalling from calculus that

$$\begin{aligned} (f^{-1})'(x) &= 1/f'(f^{-1}(x)) \\ \frac{dh}{dx} &= \frac{-2}{-\sin(\cos^{-1}(1-2x))} \\ \frac{d^2h}{dx^2} &= \frac{-2 \cos(\cos^{-1}(1-2x))}{\sin^2(\cos^{-1}(1-2x))} \cdot \frac{-2}{-\sin(\cos^{-1}(1-2x))} \end{aligned}$$

To simplify the last expression we draw a picture



The last calculation shows $\sin(\cos^{-1}(1-2x)) = 2\sqrt{x(1-x)}$ so $\theta_t = h(X_t)$ has generator

$$\frac{1}{2} \frac{d^2}{d\theta^2} - \frac{1}{2} \cot(\theta) \frac{d}{d\theta}$$

The infinitesimal variance is now constant, but a drift (which Fisher missed in his first attempt) has been introduced. We leave it as an exercise for the reader to check that $\sin^{-1}(\sqrt{x})$ also results in constant variance.

7.3 Transition probabilities

In discrete time, a Markov chain is defined by giving its transition probability $p(i, j)$. For a continuous-time Markov chain or a diffusion process, the transition probability $p_t(x, y) = P(X_t = y | X_0 = x)$ must be computed by solving one of two differential equations. In the case of a continuous time Markov chain, it follows from the Markov property and the definition of the generator in (7.1) that

$$\left. \frac{d}{ds} E_i f(X_s) \right|_{s=t} = E_i Q f(X_t) \tag{7.11}$$

From this we get

$$\frac{d}{dt} \sum_j p_t(i, j) f(j) = \sum_k p_t(i, k) \sum_j q(k, j) f(j)$$

Since this holds for all f we must have

$$\frac{d}{dt}p_t(i, j) = \sum_k p_t(i, k)q(k, j) \quad (7.12)$$

or in matrix notation $(d/dt)p_t = p_t Q$. This is *Kolmogorov's forward equation*. It can also be derived by letting I be the identity matrix and using the Markov property to write

$$\begin{aligned} \frac{1}{h}(p_{t+h}(i, j) - p_t(i, j)) &= \sum_k p_t(i, k) \frac{1}{h}(p_h(k, j) - I(k, j)) \\ &\rightarrow \sum_k p_t(i, k)q(k, j) \end{aligned}$$

Here, we have broken the time interval $[0, t+h]$ into $[0, t]$ and $[t, t+h]$, with the small piece on the forward end. If we instead break it into $[0, h]$ and $[h, t+h]$ we get

$$\begin{aligned} \frac{1}{h}(p_{t+h}(i, j) - p_t(i, j)) &= \frac{1}{h} \sum_k (p_h(i, k) - I(i, k))p_t(k, j) \\ &\rightarrow \sum_k q(i, k)p_t(k, j) \end{aligned} \quad (7.13)$$

or in matrix notation $(d/dt)p_t = Qp_t$. This is *Kolmogorov's backward equation*.

Consider now a diffusion process. Imitating (7.13) we can write

$$\begin{aligned} \frac{1}{h}(p_{t+h}(x, y) - p_t(x, y)) &= \frac{1}{h} \left(\int p_h(x, z)p_t(z, y) dz - p_t(x, y) \right) \\ &= \frac{1}{h}(E_x p_t(X_h, y) - p_t(x, y)) \end{aligned}$$

If we let $f(x) = p_t(x, y)$ for fixed y then the last quantity is $(1/h)(E_x f(X_h) - f(x))$, so recalling the definition of the generator and letting $h \rightarrow 0$ we have

$$\frac{d}{dt}p_t(x, y) = \frac{1}{2}a(x)\frac{d^2}{dx^2}p_t(x, y) + b(x)\frac{d}{dx}p_t(x, y) \quad (7.14)$$

Because we broke us the interval into $[0, h]$ and $[h, t+h]$, this is *Kolmogorov's backward equation*. Another reason is that the derivatives occur in the backward variable x . To get the forward equation note that as in (7.11)

$$\left. \frac{d}{ds}E_x f(X_s) \right|_{s=t} = E_x Lf(X_t) \quad (7.15)$$

So we have

$$\begin{aligned} \frac{d}{dt} \int p_t(x, y)f(y) dy &= \int p_t(x, y)Lf(y) dy \\ &= \int p_t(x, y) \left[\frac{1}{2}a(y)\frac{d^2}{dy^2}f(y) + b(y)\frac{d}{dy}f(y) \right] dy \end{aligned}$$

To turn this into an equation for $p_t(x, y)$ we suppose that f is 0 outside $[\delta, 1-\delta]$ for some $\delta > 0$ and integrate by parts twice to get

$$\int \left[\frac{1}{2} \frac{d^2}{dy^2} (a(y)p_t(x, y)) - \frac{d}{dy} (b(y)p_t(x, y)) \right] f(y) dy$$

Since this holds for all f we have *Kolmogorov's forward equation*

$$\frac{d}{dt} p_t(x, y) = \frac{1}{2} \frac{d^2}{dy^2} (a(y)p_t(x, y)) - \frac{d}{dy} (b(y)p_t(x, y)) \quad (7.16)$$

where the derivatives occur in the forward variable y . This is not as nice as the backward equations since it does not make sense unless $b(y)$ is differentiable and $a(y)$ is twice differentiable.

(7.16) is called the Fokker-Planck equation by physicists, due to work of Fokker in 1914 and Planck in 1917. The first rigorous mathematical derivation was given by Kolmogorov in 1931. The formula makes its first appearance in the biology literature in Wright (1945).

Defining the *adjoint operator* L^*

$$L^* f = \frac{1}{2} \frac{d^2}{dy^2} (a(y)f(y)) - \frac{d}{dy} (b(y)f(y))$$

we can write the two equations as

$$\frac{d}{dt} p_t(x, y) = L_x p_t(x, y) \quad \frac{d}{dt} p_t(x, y) = L_y^* p_t(x, y)$$

where the subscript indicates the variable of $p_t(x, y)$ where the operator acts. In comparison, the two equations for continuous time Markov chains are

$$\frac{d}{dt} p_t(i, j) = \sum_k Q(i, k) p_t(k, j) \quad \frac{d}{dt} p_t(i, j) = \sum_k p_t(i, k) Q(k, j)$$

Again Q acts on different variables in the two cases, but we don't need the formalities of defining the adjoint matrix. We just shift the matrix to the other side.

Only on rare occasions can one solve the differential equations given above to determine the transition probability.

Example 7.9. Brownian motion. Suppose $a(x) = \sigma^2$, $b(x) = 0$. In this case the backward equation (7.14) is

$$\frac{d}{dt} p_t(x, y) = \frac{\sigma^2}{2} \frac{d^2}{dx^2} p_t(x, y)$$

From the definition of the process in the previous section we know that

$$\sqrt{2\pi} p_t(x, y) = (t\sigma^2)^{-1/2} e^{-(y-x)^2/2\sigma^2 t}$$

Differentiating we find

$$\begin{aligned} \frac{d}{dt} &= \left(-\frac{1}{2}t^{-3/2}\sigma^{-1} + t^{-1/2}\sigma^{-1} \frac{(y-x)^2}{2\sigma^2 t^2} \right) e^{-(y-x)^2/2\sigma^2 t} \\ \frac{d}{dx} &= t^{-1/2}\sigma^{-1} \cdot \frac{y-x}{\sigma^2 t} e^{-(y-x)^2/2\sigma^2 t} \\ \frac{d^2}{dx^2} &= \left(-t^{-3/2}\sigma^{-3} + t^{-1/2}\sigma^{-1} \frac{(y-x)^2}{\sigma^4 t^2} \right) e^{-(y-x)^2/2\sigma^2 t} \end{aligned}$$

which shows that $p_t(x, y)$ satisfies the stated differential equation.

Example 7.10. Ornstein-Uhlenbeck process. Suppose $a(x) = \sigma^2$ and $b(x) = -\alpha x$. This is a model for the velocity of a particle with a random acceleration and experiences friction forces proportional to its velocity. The transition probability $p_t(x, y)$ is a normal with mean $u(x, t) = xe^{-\alpha t}$ and variance $v(t) = \sigma^2 \int_0^t e^{-2\alpha r} dr$, so

$$\sqrt{2\pi} p_t(x, y) = v(t)^{-1/2} e^{-(y-u(x,t))^2/2v(t)}$$

Let u_x and u_t be the partial derivatives of u , and note that $u_{xx} = 0$. Differentiating we find

$$\begin{aligned} \frac{d}{dt} &= -\frac{1}{2}v(t)^{-3/2}v'(t)e^{-(y-u(x,t))^2/2v(t)} \\ &\quad + v(t)^{-1/2} \cdot \frac{y-u(x,t)}{v(t)} u_t(x, t) e^{-(y-u(x,t))^2/2v(t)} \\ &\quad + v(t)^{-1/2} \cdot \frac{(y-u(x,t))^2}{2v(t)^2} v'(t) e^{-(y-u(x,t))^2/2v(t)} \\ \frac{d}{dx} &= v(t)^{-1/2} \cdot \frac{y-u(x,t)}{v(t)} u_x(x, t) e^{-(y-u(x,t))^2/2v(t)} \\ \frac{d^2}{dx^2} &= v(t)^{-3/2} (-u_x(x, t)^2) e^{-(y-u(x,t))^2/2v(t)} \\ &\quad + v(t)^{-1/2} \frac{(y-u(x,t))^2}{v(t)^2} u_x(x, t)^2 e^{-(y-u(x,t))^2/2v(t)} \end{aligned}$$

Let f_1, f_2, \dots, f_6 denote the right hand sides. Since $v'(t) = \sigma^2 e^{-2\alpha t}$ and $u_x(x, t) = e^{-\alpha t}$, we have $f_1 = (\sigma^2/2)f_5$ and $f_3 = (\sigma^2/2)f_6$. Since $u_t(x, t) = -\alpha x e^{-\alpha t}$, we have $f_2 = -\alpha x f_4$. Combining these results we see that

$$\frac{d}{dt} p_t(x, y) = -\alpha x \frac{d}{dx} p_t(x, y) + \frac{\sigma^2}{2} \frac{d^2}{dx^2} p_t(x, y)$$

which verifies that $p_t(x, y)$ is the desired transition probability.

In most cases, one cannot find an explicit expression for the transition probability. Kimura (1955) was able to express the transition probability for

the Wright-Fisher model as an infinite series of Gegenbauer polynomials. However, for our purposes, the following qualitative result is more useful. A function $f(x)$ on $[0,1]$ is said to be Hölder continuous if there is a $\delta > 0$ and $C < \infty$ so that $|f(x) - f(y)| \leq C|x - y|^\delta$.

Theorem 7.2. *Suppose that the coefficients a and b are Hölder continuous on $[0, 1]$, and $a(x) > 0$ on $(0, 1)$. Then for any $\delta > 0$ there is an ϵ so that $p_t(x, y) \geq \epsilon$ when $x, y \in [\delta, 1 - \delta]$.*

7.4 Hitting probabilities

For our genetics models, we want to be able to compute the probability an allele becomes fixed in the population. Here, and throughout this chapter, we will first consider the analogous problems for discrete and continuous time Markov chains on $\{0, 1, \dots, 2N\}$. Let $T_k = \min\{n : X_n = k\}$ be the time of the first visit to k , and let $h(i) = P_i(T_{2N} < T_0)$. To compute $h(i)$ we note that if $0 < i < 2N$ then breaking things down according to what happens on the first step

$$h(i) = \sum_j p(i, j)h(j) \tag{7.17}$$

Introducing P for the transition matrix and I for the identity matrix, we can write (7.17) as $h = Ph$ or $(P - I)h = 0$. The second formula may look a little odd now, but soon it will seem natural. To compute $h(i)$, we first need a technical result.

Theorem 7.3. *Let $\tau = T_0 \wedge T_{2N}$. In discrete or continuous time, if it is possible to reach 0 and $2N$ from each $0 < i < 2N$ then $\sup_{0 < i < 2N} E_i\tau < \infty$.*

Proof. Our assumption implies that there are $\epsilon > 0$ and $M < \infty$ so that $P_i(\tau \leq M) \geq \epsilon$ for all $0 < i < 2N$. The Markov property implies $P_i(\tau > kM) \leq (1 - \epsilon)^k$, so

$$E_i\tau = \int_0^\infty P_i(\tau > t) dt \leq M \sum_{k=0}^\infty (1 - \epsilon)^k = M/\epsilon \quad \square$$

Theorem 7.4. *Suppose that it is possible to reach 0 and $2N$ from each $0 < i < 2N$. $h(i) = P_i(T_{2N} < T_0)$ is the unique solution of $(P - I)h = 0$ with $h(0) = 0$ and $h(2N) = 1$.*

Proof. Theorem 7.3 implies $\sup_{0 < i < 2N} E_i\tau < \infty$. (7.17) implies that $Eh(X_{n \wedge \tau})$ is constant, since for any jump that starts at a point $0 < i < 2N$, the expected value after a jump is the same as before. The irreducibility condition implies that $P_i(\tau < \infty) = 1$. Letting $n \rightarrow \infty$, which can be justified since h is a bounded function,

$$h(i) = E_i h(X_\tau) = P_i(T_{2N} < T_0)$$

since $h(0) = 0$ and $h(2N) = 1$. □

The next result is the continuous-time analogue of Theorem 7.4.

Theorem 7.5. *Suppose that it is possible to reach 0 and $2N$ from each $0 < i < 2N$. $h(i) = P_i(T_{2N} < T_0)$ is the unique solution of $Qh = 0$ for $0 < i < 2N$ with $h(0) = 0$ and $h(2N) = 1$.*

Proof. Theorem 7.3 implies $\sup_{0 < i < 2N} E_i \tau < \infty$. To check that the equation is satisfied, we note that

$$\left. \frac{d}{ds} E_i h(X_s) \right|_{s=t} = E_i Qh(X_t) = 0$$

so $E_i h(X_{t \wedge \tau})$ is constant in time, and we can repeat the argument from discrete time to conclude $h(i) = E_i h(X_\tau) = P_i(T_{2N} < T_0)$. \square

Turning to the case of a diffusion process, let $T_a = \inf\{t : X_t = a\}$. Again, we begin with a technical result.

Theorem 7.6. *Let $y < z$ and $\tau_{y,z} = T_y \wedge T_z$. Suppose that it is possible to reach y and z from each $y < x < z$. Then $\sup_{x \in (y,z)} E_x \tau_{y,z} < \infty$.*

Proof. Pick $w \in (y, z)$. Pick M large enough so that $P_w(T_y \leq M) \geq \epsilon > 0$ and $P_w(T_z \leq M) \geq \epsilon > 0$. By the argument in discrete time, it is enough to show

$$\sup_{x \in (y,z)} P_x(\tau_{y,z} > M) \leq 1 - \epsilon$$

By considering the first time the process starting from w hits z and using the Markov property, it follows that if $w < x < z$ then

$$\begin{aligned} P_w(T_z \leq t) &= E_w(P_x(T_z \leq t - T_x); T_x \leq t) \\ &\leq P_x(T_z \leq t)P_w(T_z \leq t) \leq P_x(T_z \leq t) \end{aligned}$$

A similar argument shows that for $y < x < w$, $P_x(T_y \leq t) \geq P_w(T_y \leq t)$, and the desired result follows. \square

Theorem 7.7. *Let $y < z$. Suppose that it is possible to reach y and z from each $y < x < z$. $h(x) = P_x(T_z < T_y)$ is the unique solution of $Lh = 0$ for $y < x < z$ with $h(y) = 0$ and $h(z) = 1$.*

Proof. Theorem 7.6 implies that $\sup_{x \in (y,z)} E_x \tau_{y,z} < \infty$, so h is well defined. To check that the equation is satisfied, we note that the Markov property implies that

$$\left. \frac{d}{ds} E_x h(X_s) \right|_{s=t} = E_x Lh(X_t) \tag{7.18}$$

$Lh = 0$ implies $E_x h(X_{t \wedge \tau_{x,y}})$ is constant in time, so we can argue as before that $h(x) = E_x h(X_\tau) = P_x(T_1 < T_0)$. \square

Comparing the last three theorems shows that hitting probabilities satisfy

$$\begin{aligned} (P - I)h &= 0 && \text{discrete time Markov chain} \\ Qh &= 0 && \text{continuous time Markov chain} \\ Lh &= 0 && \text{diffusion process} \end{aligned}$$

so the three operators $P - I$, Q , and L are analogous.

Diffusion hitting probabilities

Based on Theorem 7.7, we want to solve

$$L\phi = \frac{1}{2}a(x)\frac{d^2}{dx^2}\phi + b(x)\frac{d}{dx}\phi = 0$$

ϕ is called the *natural scale* for the diffusion process because $\phi(X_t)$ is a martingale. To solve this equation, we let $\psi = \phi'$ and note that

$$\frac{1}{2}a(x)\psi' + b(x)\psi = 0 \quad \text{or} \quad \psi' = \frac{-2b(x)}{a(x)}\psi$$

As one can check by differentiating, this equation is solved by

$$\psi(y) = \exp\left(\int^y \frac{-2b(z)}{a(z)} dz\right)$$

where the lack of a lower limit indicates that we can choose any convenient value, or what is the same, use any antiderivative of $-2b(z)/a(z)$. ϕ can be obtained by ψ by integrating:

$$\phi(x) = \int^x \psi(y) dy$$

To have the boundary conditions $h(y) = 0$, $h(z) = 1$ satisfied

$$\begin{aligned} P_x(T_z < T_y) &= \frac{\phi(x) - \phi(y)}{\phi(z) - \phi(y)} \\ P_x(T_z > T_y) &= \frac{\phi(z) - \phi(x)}{\phi(z) - \phi(y)} \end{aligned} \tag{7.19}$$

The second equation follows from $P_x(T_z > T_y) = 1 - P_x(T_z < T_y)$.

Turning to special cases:

Example 7.11. Martingale diffusions. Suppose $b(x) = 0$, $\psi'(x) = 0$ and hence $\phi(x) = x$. X_t is a martingale and we have

$$P_x(T_z < T_y) = \frac{x - y}{z - y} \quad P_x(T_z > T_y) = \frac{z - x}{z - y}$$

In the next three examples, the mutation rates are zero and we are considering a special case of the general selection model so $a(x) = x(1 - x)$, $b(x) = x(1 - x)(\delta + \eta x)$ and hence

$$-\frac{2b(x)}{a(x)} = -2(\delta + \eta x)$$

Example 7.12. Additive selection. In this case, by (7.6) $-2b(x)/a(x) = -2\gamma$ so

$$\psi(y) = e^{-2\gamma y} \quad \text{and} \quad \phi(y) = [1 - e^{-2\gamma y}]/2\gamma \tag{7.20}$$

and the hitting probabilities are

$$P_x(T_1 < T_0) = \frac{1 - e^{-2\gamma x}}{1 - e^{-2\gamma}} \tag{7.21}$$

which agrees with (6.3). When $x = 1/2N$, using $\gamma/2N = s$ and $1 - e^{-2s} \approx 2s$ we have

$$P_{1/2N}(T_1 < T_0) \approx \frac{2s}{1 - e^{-2\gamma}} \approx 2s$$

when γ is large. The next figure shows the hitting probabilities when $\gamma = 0, 2, 5, 10$.

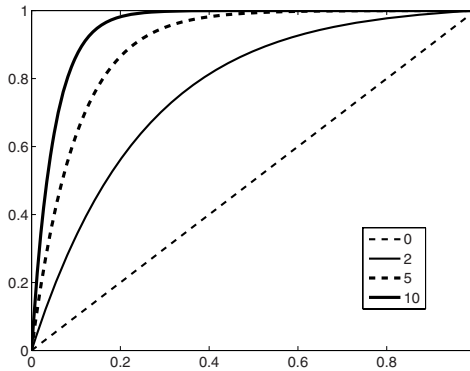


Fig. 7.2. Hitting probabilities for additive selection.

Example 7.13. Dominant advantageous allele. In this case (7.9) implies that $-2b(x)/a(x) = -2\gamma(1 - x)$, so $\psi(x) = e^{-\gamma(2x-x^2)}$, and

$$\phi(x) = \int_0^x e^{-\gamma(2y-y^2)} dy$$

Example 7.14. Recessive advantageous allele. In this case (7.10) implies that $-2b(x)/a(x) = -2\gamma x$, so $\psi(x) = e^{-\gamma x^2}$, and

$$\phi(x) = \int_0^x e^{-\gamma y^2} dy$$

Figure 7.3 shows $P_x(T_1 < T_0) = \phi(x)/\phi(1)$ when $\gamma = 0, 2, 5, 10$.

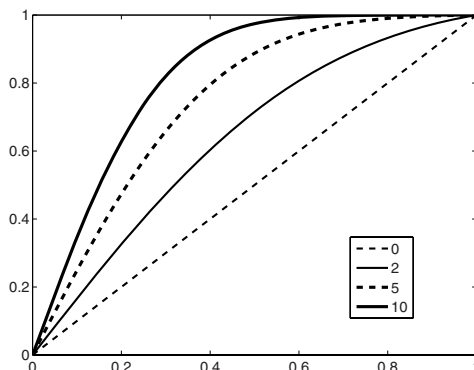


Fig. 7.3. Recessive advantageous allele hitting probabilities

The integrand is a constant multiple of the normal density with mean 0 and variance $1/2\gamma$, so if γ is large

$$\phi(1) \approx \frac{1}{2} \sqrt{\frac{\pi}{\gamma}}$$

If x is small $\phi(x) \approx x$ so if $x = 1/2N$

$$\frac{\phi(1/2N)}{\phi(1)} \approx \frac{1}{N} \sqrt{\frac{\gamma}{\pi}} = \sqrt{\frac{2s}{\pi N}}$$

which is (15) of Kimura (1962). This is larger than the neutral fixation probability $1/2N$, but smaller than the $2s/(1 - e^{-\gamma})$ for additive selection.

Example 7.15. Symmetric balancing selection. By (7.8) $-2b(x)/a(x) = -2\gamma(1 - 2x)$ so

$$\psi(y) = e^{-2\gamma y(1-y)} \quad \text{and} \quad \phi(x) = \int_0^x \psi(y) dy$$

Figure 7.4 shows $P_x(T_1 < T_0) = \phi(x)/\phi(1)$ when $\gamma = 0, 2, 5, 10$.

If γ is large then most of the contribution to $\phi(1)$ comes from values within $O(1/2\gamma)$ of the boundary so

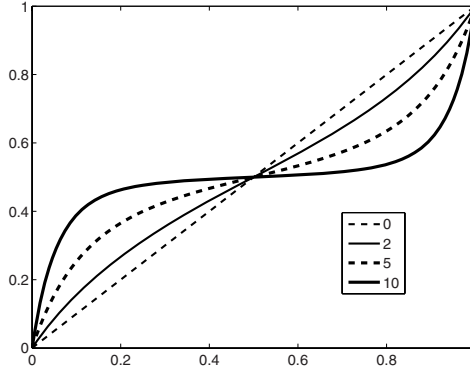


Fig. 7.4. Balancing selection hitting probabilities.

$$\phi(1) \approx 2 \int_0^\infty e^{-2\gamma y} dy = \frac{1}{\gamma}$$

If $x = c/(2\gamma)$ then

$$\phi(x) \approx \int_0^x e^{-2\gamma y} dy = \frac{1 - e^{-2\gamma x}}{2\gamma} = \frac{1 - e^{-c}}{2\gamma}$$

and we have

$$P_{c/(2\gamma)}(T_1 < T_0) \approx \frac{1}{2}(1 - e^{-c})$$

Using $P_{1-x}(T_0 < T_1) = P_x(T_1 < T_0)$, we see that if $x \gg 1/\gamma$ and $(1 - x) \gg 1/\gamma$ then $P_x(T_1 < T_0) \approx 1/2$.

Up to this point we have ignored the hypothesis “Suppose that it is possible to reach y and z from each $y < x < z$.” It follows from Theorem 7.2 that this holds for $0 < y < z < 1$ whenever the coefficients a and b are Hölder continuous on $[0, 1]$, and $a(x) > 0$ on $(0, 1)$, which is true in all of our examples. In all of the results above we can make the derivation rigorous by first computing $P_x(T_y < T_z)$ and then letting $y \rightarrow 0$ and $z \rightarrow 1$. To show that problems can occur even in natural examples, we consider

Example 7.16. Wright-Fisher model with mutation. For simplicity we assume that there is no selection, so

$$Lf = \frac{1}{2}x(1 - x)\frac{d^2}{dx^2}f + (\beta_1(1 - x) - \beta_2x)\frac{d}{dx}f$$

In this case, we have

$$\begin{aligned}\psi(x) &= \exp\left(\int^x -2b(y)/a(y) dy\right) \\ &= \exp\left(\int^x -\frac{2\beta_2}{1-y} - \frac{2\beta_1}{y} dy\right) \\ &= x^{-2\beta_1}(1-x)^{-2\beta_2}\end{aligned}$$

and $\phi(x) = \int_{1/2}^x y^{-2\beta_1}(1-y)^{-2\beta_2} dy$. $\phi(0) = -\infty$ if $\beta_1 \geq 1/2$ so taking $0 < y < x < z < 1$,

$$P_x(T_y < T_z) = \frac{\phi(z) - \phi(x)}{\phi(z) - \phi(y)} \rightarrow 0 \quad \text{as } y \rightarrow 0$$

and the process cannot get to 0. In words the mutation rate is so strong that the allele frequency cannot reach 0.

Likewise $\phi(1) = \infty$ if $\beta_2 \geq 1/2$ and $0 < y < x < z < 1$,

$$P_x(T_z < T_y) = \frac{\phi(x) - \phi(y)}{\phi(z) - \phi(y)} \rightarrow 0 \quad \text{as } z \rightarrow 1$$

so the process cannot get to 1. Of course if β_1 and β_2 are both $\geq 1/2$, $P_x(T_0 < T_1)$ is meaningless. We will return to this issue in Section 7.9, when we consider the boundary behavior of diffusion processes.

7.5 Stationary measures

Stationary distributions for Markov processes are important because they represent equilibrium states and are (under mild regularity conditions) the limiting distribution as time $t \rightarrow \infty$. In discrete time a nonnegative solution of

$$\sum_i \pi(i)p(i, j) = \pi(j) \tag{7.22}$$

is called a *stationary measure*. A solution with $\sum_i \pi(i) = 1$ is called a *stationary distribution*. If (7.22) holds then

$$\sum_i \pi(i)p^n(i, j) = \sum_{i,k} \pi(i)p(i, k)p^{n-1}(k, j) = \sum_k \pi(k)p^{n-1}(k, j)$$

and it follows by induction that $P_\pi(X_n = i) = \pi(i)$. Results from Markov chain theory imply that if there is a stationary distribution π , and $p(i, j)$ is irreducible and aperiodic (terms defined in Section 4.5) then $p^n(i, j) \rightarrow \pi(j)$ as $n \rightarrow \infty$.

To see what the condition for stationarity should be in continuous time, we note that if E_π is the expected value starting at π then

$$\left. \frac{d}{dt} E_\pi(f(X_t)) \right|_{t=0} = \sum_i \pi(i) \sum_j q(i, j) f(j) = \sum_j \left(\sum_i \pi(i) q(i, j) \right) f(j)$$

In order for this to be 0 for all f we must have

$$(\pi Q)(j) = \sum_i \pi(i)q(i, j) = 0 \quad \text{for all } j \tag{7.23}$$

To see that this is sufficient, note that the forward equation implies

$$\frac{d}{dt} \sum_i \pi(i)p_t(i, j) = \sum_{i,k} \pi(i)Q(i, k)p_t(k, j) = 0$$

Results from Markov chain theory imply that if there is a stationary distribution π and $p_t(i, j)$ is irreducible then $p_t(i, j) \rightarrow \pi(j)$ as $n \rightarrow \infty$.

For a diffusion process we want

$$0 = \left. \frac{d}{dt} E_\pi f(X_t) \right|_{t=0} = \int \pi(x)Lf(x) dx$$

If f is 0 outside $[\delta, 1 - \delta]$ for some $\delta > 0$ then integrating by parts twice converts this into

$$\int L^* \pi(x)f(x) dx = 0$$

where L^* is the adjoint operator

$$L^* \pi = \frac{1}{2} \frac{d^2}{dx^2} (a(x)\pi(x)) - \frac{d}{dx} (b(x)\pi(x))$$

If this holds for all f then we must have

$$L^* \pi = 0 \tag{7.24}$$

To see that this is sufficient, note that the backward equation and integration by parts imply

$$\frac{d}{dt} \int \pi(x)p_t(x, y) dx = \int \pi(x)L_x p_t(x, y) dx = \int L^* \pi(x)p_t(x, y) dx = 0$$

From Theorem 7.2 and the theory of Harris chains, it follows that if there is a stationary distribution π , and $a(x) > 0$ for $x \in (0, 1)$ then $p_t(x, y) \rightarrow \pi(y)$ as $t \rightarrow \infty$.

Comparing the last three equations shows that the stationary measures satisfy

$$\begin{aligned} \pi(P - I) = 0 & \quad \text{discrete time Markov chain} \\ \pi Q = 0 & \quad \text{continuous time Markov chain} \\ L^* \pi = 0 & \quad \text{diffusion process} \end{aligned}$$

Again the three operators $P - I$, Q , and L are analogous, but as in Section 7.3, multiplying the matrices on the left by π corresponds to using the adjoint of the diffusion's generator.

Diffusion stationary measures

Theorem 7.8. *If $\psi(x)$ is the derivative of the natural scale then $m(x) = 1/a(x)\psi(x)$ is a stationary measure.*

m is sometimes called the *speed measure*, although as we will see in Section 7.7, this term is misleading. If $\int_0^1 m(x) dx < \infty$ then we can convert $m(x)$ into a stationary distribution by multiplying by a constant to make the integral equal to 1.

Proof. To solve $L^*\pi = 0$, it is convenient to note that since $\psi'(x)/\psi(x) = -2b(x)/a(x)$

$$\begin{aligned} & \frac{1}{2}a(x)\psi(x)\frac{d}{dx}\left(\frac{1}{\psi(x)}\frac{d}{dx}f(x)\right) \\ &= \frac{1}{2}a(x)\psi(x)\frac{1}{\psi(x)}\frac{d^2}{dx^2}f(x) + \frac{1}{2}a(x)\psi(x)\frac{-\psi'(x)}{\psi(x)^2}\frac{d}{dx}f(x) = Lf \end{aligned}$$

Thus, if we let $m(x) = 1/a(x)\psi(x)$ then

$$Lf = \frac{1}{2m(x)}\frac{d}{dx}\left(\frac{1}{\psi(x)}\frac{d}{dx}f(x)\right) \quad (7.25)$$

Writing L in this form before we integrate by parts, it follows that

$$L^*m = \frac{d}{dx}\left(\frac{1}{\psi(x)}\left[\frac{d}{dx}\frac{1}{2m(x)}m(x)\right]\right) = 0 \quad \square$$

Example 7.17. General diploid selection and mutation. In this case, the generator is

$$\frac{1}{2}x(1-x)\frac{d^2}{dx^2}f + \{x(1-x)[\delta + \eta x] + \beta_1(1-x) - \beta_2x\}\frac{d}{dx}f$$

so we have

$$\begin{aligned} \psi(x) &= \exp\left(\int^x -2b(y)/a(y) dy\right) \\ &= \exp\left(\int^x -2[\delta + \eta y] + \frac{2\beta_2}{1-y} - \frac{2\beta_1}{y} dy\right) \\ &= x^{-2\beta_1}(1-x)^{-2\beta_2}e^{-2\delta x - \eta x^2} \end{aligned} \quad (7.26)$$

Since $a(x) = x(1-x)$ the stationary measure is

$$m(x) = \frac{1}{a(x)\psi(x)} = x^{2\beta_1-1}(1-x)^{2\beta_2-1}e^{2\delta x + \eta x^2} \quad (7.27)$$

In the case of additive selection, $\delta = \gamma$ and $\eta = 0$ so

$$m(x) = x^{2\beta_1-1}(1-x)^{2\beta_2-1}e^{2\gamma x} \quad (7.28)$$

If either $\beta_i = 0$ this is not integrable. If $\beta_1 = 0$ then there are no mutations from a to A , so 0 is an absorbing state. Likewise if $\beta_2 = 0$ then there are no mutations from A to a , so 1 is an absorbing state. If both $\beta_i > 0$ then there is a stationary distribution. In the case of no selection this is the beta distribution

$$\pi(x) = \frac{\Gamma(2\beta_1 + 2\beta_2)}{\Gamma(2\beta_1)\Gamma(2\beta_2)} x^{2\beta_1-1}(1-x)^{2\beta_2-1}$$

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t} dt$ is the usual gamma function.

This formula can be found on page 123 of Wright's (1931) seminal paper on evolutionary theory. According to Will Provine's annotation of a collection of Wright's papers, Wright (1986), "this paper resulted in his admission to the National Academy of Science at a young age." It is interesting to note that the computation of the stationary distribution precedes the equation $L^*\pi = 0$, which first appears in the biology literature in Wright's (1945) work. In that paper, Wright says "Dr. A. Kolmogorov has recently been kind enough to send me a reprint of an important paper on this subject which was published in 1935, but which had not previously come to my attention." As one can see from the dates, the 10 year delay was likely due to World War II.

Suppose X has distribution π . Using this recursion and the fact that the constant makes $\int \pi(x) dx = 1$, we can compute

$$\begin{aligned} EX &= \frac{\Gamma(2\beta_1 + 2\beta_2)}{\Gamma(2\beta_1)} \frac{\Gamma(2\beta_1 + 1)}{\Gamma(2\beta_1 + 2\beta_2 + 1)} = \frac{2\beta_1}{2\beta_1 + 2\beta_2} \\ EX^2 &= \frac{\Gamma(2\beta_1 + 2\beta_2)}{\Gamma(2\beta_1)} \frac{\Gamma(2\beta_1 + 2)}{\Gamma(2\beta_1 + 2\beta_2 + 2)} = \frac{2\beta_1(2\beta_1 + 1)}{(2\beta_1 + 2\beta_2)(2\beta_1 + 2\beta_2 + 1)} \\ \text{var}(X) &= EX^2 - (EX)^2 = \frac{2\beta_1(2\beta_2)}{(2\beta_1 + 2\beta_2)^2(2\beta_1 + 2\beta_2 + 1)} \end{aligned}$$

Using the first two formulas, we can compute the mean of the heterozygosity, i.e., the probability in equilibrium that two randomly chosen individuals are different:

$$\begin{aligned} E(2X(1-X)) &= \frac{2\beta_1}{2\beta_1 + 2\beta_2} \left(1 - \frac{2\beta_1 + 1}{2\beta_1 + 2\beta_2 + 1} \right) \\ &= 2 \cdot \frac{2\beta_1(2\beta_2)}{(2\beta_1 + 2\beta_2)(2\beta_1 + 2\beta_2 + 1)} \end{aligned}$$

Reversibility

As we will now explain, the stationary measures of a one dimensional diffusion process have a very special property. Again, we begin by considering Markov chains. In discrete time the *detailed balance condition*:

$$\pi(i)p(i, j) = \pi(j)p(j, i) \quad (7.29)$$

implies that

$$\sum_i \pi(i)p(i, j) = \pi(j) \sum_i p(j, i) = \pi(j)$$

so π is a stationary distribution. In most cases, there is no π satisfying (7.29). However, if there is and we start the process from π , then the time reversal $Y_m = X_{n-m}$, $0 \leq m \leq n$ is a Markov chain with transition probability

$$p^*(i, j) = P_\pi(X_0 = j | X_1 = i) = \frac{P_\pi(X_0 = j, X_1 = i)}{P_\pi(X_1 = i)} = \frac{\pi(j)p(j, i)}{\pi(i)} = p(i, j)$$

Informally, a movie of a reversible process looks the same running forward or backwards in time.

For a continuous-time Markov chain, the detailed balance condition is

$$\pi(i)q(i, j) = \pi(j)q(j, i)$$

or, in equilibrium, the rate of jumps from i to j is the same as the rate of jumps from j to i . To extend the definition to diffusions, we define an inner product by

$$\langle f, g \rangle_\pi = \sum_i f(i)\pi(i)g(i)$$

Given a linear operator R , we define the adjoint operator R^* with respect to π by

$$\langle f, Rg \rangle_\pi = \langle R^*f, g \rangle_\pi$$

Theorem 7.9. *If R is a matrix $r(i, j)$ then R^* is the matrix*

$$r^*(i, j) = \pi(j)r(j, i)/\pi(i).$$

In words, when R is a transition probability with stationary distribution π , R^* is the transition probability for the chain running backwards in time.

Proof. To check our proposed formula, we note that

$$\begin{aligned} \langle f, Rg \rangle_\pi &= \sum_j f(j)\pi(j) \sum_i r(j, i)g(i) \\ &= \sum_i \left(\sum_j \frac{\pi(j)r(j, i)}{\pi(i)} f(j) \right) \pi(i)g(i) \\ &= \sum_i \left(\sum_j r^*(i, j) f(j) \right) \pi(i)g(i) = \langle R^*f, g \rangle_\pi \quad \square \end{aligned}$$

For a diffusion process, if we write

$$Lf = \frac{1}{2}a(x)\psi(x) \frac{d}{dx} \left(\frac{1}{\psi(x)} \frac{d}{dx} f(x) \right)$$

then using the speed measure $m(x) = 1/a(x)\psi(x)$ to define the inner product we have

$$\langle g, Lf \rangle_m = \frac{1}{2} \int g(x) \frac{d}{dx} \left(\frac{1}{\psi(x)} \frac{d}{dx} f(x) \right) dx$$

If we assume that f and g vanish outside $[\delta, 1 - \delta]$, then integrating by parts twice shows that the above is

$$\begin{aligned} &= \frac{1}{2} \int \frac{d}{dx} g(x) \cdot \frac{1}{\psi(x)} \frac{d}{dx} f(x) dx \\ &= \frac{1}{2} \int \frac{d}{dx} \left(\frac{1}{\psi(x)} \frac{d}{dx} g(x) \right) f(x) dx = \langle Lg, f \rangle_m \end{aligned}$$

so L is self-adjoint with respect to m . As in the case of discrete state space, this implies that the transition probability has a symmetry property that resembles the detailed balance condition

$$p_t(x, y) = \frac{m(y)p_t(y, x)}{m(x)} \quad \text{or} \quad m(x)p_t(x, y) = m(y)p_t(y, x) \quad (7.30)$$

7.6 Occupation times

Let τ be the amount of time it takes for fixation or loss of an allele to occur. In addition to computing the probabilities of the two outcomes, we would like to determine the average time that this will take. For discrete models $\tau = T_0 \wedge T_{2N}$; for diffusions $\tau = T_0 \wedge T_1$. To compute $g(i) = E_i \tau$, it is convenient to consider a more general problem:

$$g(i) = E_i \sum_{0 \leq m < \tau} f(X_m)$$

which reduces to the original question when $f \equiv 1$. If $0 < i < 2N$ then breaking things down according to what happens on the first step

$$g(i) = f(i) + \sum_j p(i, j)g(j) \quad \text{or} \quad E_i g(X_1) = g(i) - f(i) \quad (7.31)$$

The next result shows that this equation together with the boundary conditions $g(0) = 0$ and $g(2N) = 0$ are enough to identify $E_i \sum_{0 \leq m < \tau} f(X_m)$.

Theorem 7.10. *Suppose that it is possible to reach 0 and $2N$ from each $0 < i < 2N$. $g(i) = E_i \sum_{0 \leq m < \tau} f(X_m)$ is the unique solution of $(P - I)g = -f$ with $g(0) = 0$ and $g(2N) = 0$.*

Proof. By Theorem 7.3, our assumption implies $\sup_{0 < i < 2N} E_i \tau < \infty$, so g is well defined. To prove that the equation holds, we note that (7.31) implies that on $\{\tau > n\}$

$$\begin{aligned} & E_i \left(g(X_{(n+1)\wedge\tau}) + \sum_{0 \leq m < (n+1)\wedge\tau} f(X_m) \middle| \mathcal{F}_n \right) \\ &= E_i(g(X_{(n+1)\wedge\tau}) | \mathcal{F}_n) + \sum_{0 \leq m < (n+1)\wedge\tau} f(X_m) \\ &= g(X_{n\wedge\tau}) + \sum_{0 \leq m < n\wedge\tau} f(X_m) \end{aligned}$$

by (7.31). The last equality is trivial on $\{\tau \leq n\}$. Thus the expected value

$$E_i \left(g(X_{(n+1)\wedge\tau}) + \sum_{0 \leq m < (n+1)\wedge\tau} f(X_m) \right)$$

is constant in time, so letting $n \rightarrow \infty$ and using $g(X_\tau) = 0$.

$$g(i) = E_i \left(g(X_\tau) + \sum_{0 \leq m < \tau} f(X_m) \right) = E_i \sum_{0 \leq m < \tau} f(X_m) \quad \square$$

For a continuous-time Markov chain:

Theorem 7.11. *Suppose that it is possible to reach 0 and $2N$ from each $0 < i < 2N$. $g(i) = E_i \int_0^\tau f(X_s) ds$ is the unique solution of $Qg = -f$ for $0 < i < 2N$ with $g(0) = 0$ and $g(2N) = 0$.*

Proof. By Theorem 7.3, our assumption implies $\sup_{0 < i < 2N} E_i \tau < \infty$, so g is well defined. To prove that the equation holds, note that $Qg = -f$ for $0 < i < 2N$ implies

$$\frac{d}{dt} E_i \left(g(X_t) + \int_0^t f(X_s) ds \right) = E_i [Qg(X_t) + f(X_t)] = 0 \quad \text{when } t < \tau$$

so $E_i [g(X_{t\wedge\tau}) + \int_0^{t\wedge\tau} f(X_s) ds]$ is constant in time. If $g(0) = g(2N) = 0$ then letting $t \rightarrow \infty$ we have $g(i) = E_i \int_0^\tau f(X_s) ds$. □

For a diffusion process:

Theorem 7.12. *Suppose that it is possible to reach y and z from each $y < x < z$. $g(x) = E_x \int_0^\tau f(X_s) ds$ is the unique solution of $Lg = -f$ for $0 < x < 1$ with $g(0) = 0$ and $g(1) = 0$.*

Proof. By Theorem 7.6, our assumption implies $\sup_{x \in (y,z)} E_x \tau < \infty$, so g is well defined. To prove that the equation holds, note that $Lg = -f$ for $y < x < z$ implies

$$\frac{d}{dt} E_x \left(g(X_t) + \int_0^t f(X_s) ds \right) = E_x [Lg(X_t) + f(X_t)] = 0 \quad \text{when } t < \tau$$

so $E_x [g(X_{t\wedge\tau}) + \int_0^{t\wedge\tau} f(X_s) ds]$ is constant in time. If $g(y) = g(z) = 0$ then letting $t \rightarrow \infty$ we have $g(x) = E_x \int_0^\tau f(X_s) ds$. □

By now, the reader has probably learned that the three operators $P - I$, Q , and L are analogous, but, again, comparing the last three theorems shows that the occupation times satisfy

$$\begin{aligned} (P - I)g &= -f && \text{discrete time Markov chain} \\ Qg &= -f && \text{continuous time Markov chain} \\ Lg &= -f && \text{diffusion process} \end{aligned}$$

with $g = 0$ at the endpoints.

Exit times

We begin with two examples where $f \equiv 1$.

Example 7.18. Symmetric simple random walk. In this case we guess

$$g(i) = i(2N - i).$$

This obviously satisfies the boundary conditions $g(0) = 0$, $g(2N) = 1$. To check (7.31), we note that

$$\begin{aligned} \sum_j p(i, j)g(j) &= (i + 1)\frac{2N - i - 1}{2} + (i - 1)\frac{2N - i + 1}{2} \\ &= i(2N - i) + \frac{2N - i}{2} - \frac{i + 1}{2} - \frac{2N - i}{2} + \frac{i - 1}{2} \\ &= g(x) - 1 \end{aligned}$$

Example 7.19. Wright-Fisher model. In the case of no selection or mutation, inspired by a result of Kimura we guess

$$\begin{aligned} g(x) &= -2[x \log x + (1 - x) \log(1 - x)] \\ g'(x) &= -2 \left[\log x + 1 - \log(1 - x) + (1 - x) \frac{1}{1 - x} \cdot (-1) \right] \\ g''(x) &= -2 \left[\frac{1}{x} - \frac{1}{1 - x} \cdot (-1) \right] = \frac{-2}{x(1 - x)} \end{aligned}$$

so $(1/2)x(1 - x)g''(x) = -1$.

Example 7.20. Higher moments. As the proof will show, the next result is valid for continuous time Markov chains if we replace L by Q . This is Theorem 13.17 in Dynkin (1965).

Theorem 7.13. *If $f(x) = E_x \tau^k / k!$ then the solution of $Lg = -f$ is $g(x) = E \tau^{k+1} / (k + 1)!$*

Proof. To begin, we recall that $E_x \tau^k / k! = \int_0^\infty \frac{u^{k-1}}{(k-1)!} P_x(\tau > u) du$ and write

$$g(x) = E_x \int_0^\tau E_{X_s}(\tau^k / k!) ds = E_x \int_0^\infty \int_0^\infty 1_{(\tau > s)} \frac{u^{k-1}}{(k-1)!} P_{X_s}(\tau > u) du ds$$

The Markov property implies that if \mathcal{F}_s is the σ -field generated by the process up to time s then

$$P_x(\tau > t > s | \mathcal{F}_s) = 1_{(\tau > s)} P_{X_s}(\tau > t - s)$$

Taking expected values and changing variables $u = t - s$, we can write

$$g(x) = \int_0^\infty \int_s^\infty \frac{(t-s)^{k-1}}{(k-1)!} P_x(\tau > t > s) dt ds$$

Interchanging the order of integration:

$$\begin{aligned} &= \int_0^\infty \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} P_x(\tau > t) ds dt \\ &= \int_0^\infty \frac{t^k}{k!} P_x(\tau > t) dt = E_x \tau^{k+1} / (k+1)! \end{aligned}$$

which completes the proof. □

7.7 Green's functions

For a discrete-time Markov chain on $\{0, 1, \dots, 2N\}$, we define the *Green's function* $G(i, j)$ to be the solution of $(P - I)g = -1_j$ with $g(0) = g(2N) = 0$, where 1_j is the function that is 1 at j and 0 otherwise. It follows from Theorem 7.10 that $G(i, j)$ is the expected number of visits to j starting from i before $\tau = T_0 \wedge T_{2N}$ and

$$E_i \sum_{0 \leq m < \tau} f(X_m) = \sum_j G(i, j) f(j).$$

Theorem 7.14. *Suppose that it is possible to reach 0 and $2N$ from each $0 < i < 2N$. If we let $T_j^+ = \min\{n \geq 1 : X_n = j\}$ then*

$$G(i, j) = \frac{P_i(T_j < \tau)}{P_j(T_j^+ > \tau)} \tag{7.32}$$

Proof. The first factor is the probability we visit j at least once. If this occurs, then the number of visits to j has a geometric distribution with mean $1/P_j(T_j^+ > \tau)$. □

In continuous time, we define the Green's function, $G(i, j)$, to be the solution of $Qg = -1_j$. It follows from Theorem 7.11 that $G(i, j)$ is the expected occupation time of j starting from i and

$$E_i \int_0^\tau f(X_s) ds = \sum_j G(i, j) f(j).$$

Theorem 7.15. *Suppose that it is possible to reach 0 and 2N from each 0 < i < 2N. Let $R_j = \min\{t : X_t = j \text{ and } X_s \neq j \text{ for some } s < t\}$ be the first time the process returns to j.*

$$G(i, j) = \frac{P_i(T_j < \tau)}{q_j P_j(R_j > \tau)} \tag{7.33}$$

where $q_j = -Q(j, j)$ is the rate at which the process jumps out of j.

Proof. Again the first factor is the probability we visit j at least once. If this occurs, the number of visits to j has a geometric distribution with mean $1/P_j(R_j > \tau)$, and each visit lasts for an average amount of time $1/q_j$. \square

When space is continuous, we could, by analogy with the two previous cases, define the *Green's function* $G(x, y)$ to be the solution of $Lg = -\delta_y$, where δ_y is a point mass at y. However, as we explain in (7.44), solving this equation requires the use of calculus for "generalized functions." To keep things simple, we will instead define the Green's function $G(x, y)$ for the interval $[u, v]$ by the property that

$$g(x) = \int G(x, y)f(y) dy \quad \text{satisfies} \quad Lg = -f$$

for $u < x < v$ with $g(u) = g(v) = 0$.

Theorem 7.16. *Suppose that it is possible to reach u and v from each $u < x < v$. The Green's function $G(x, y)$ for the interval $[u, v]$ is*

$$\begin{aligned} &2 \frac{(\phi(v) - \phi(x))(\phi(y) - \phi(u))}{\phi(v) - \phi(u)} \cdot m(y) && y \leq x \\ &2 \frac{(\phi(x) - \phi(u))(\phi(v) - \phi(y))}{\phi(v) - \phi(u)} \cdot m(y) && x \leq y \end{aligned} \tag{7.34}$$

where $\phi(x)$ is the natural scale and $m(x) = 1/\phi'(x)a(x)$ is the speed measure.

Proof. To solve equation $Lg = -f$ now, we use (7.25) to write

$$\frac{d}{dx} \left(\frac{1}{\psi(x)} \frac{dg}{dx} \right) = -2m(x)f(x)$$

and integrate to conclude that for some constant C

$$\frac{1}{\psi(y)} \frac{dg}{dy} = C - 2 \int_u^y dz m(z)f(z)$$

Multiplying by $\psi(y)$ on each side, integrating y from u to x, and recalling that $g(u) = 0$ and $\psi = \phi'$ we have

$$g(x) = C(\phi(x) - \phi(u)) - 2 \int_u^x dy \psi(y) \int_u^y dz m(z)f(z) \tag{7.35}$$

In order to have $g(v) = 0$, we must have

$$C = \frac{2}{\phi(v) - \phi(u)} \int_u^v dy \psi(y) \int_u^y dz m(z) f(z)$$

Plugging the formula for C into (7.35) and writing

$$h_1(x) = \frac{\phi(x) - \phi(u)}{\phi(v) - \phi(u)} = P_x(T_v < T_u)$$

we have

$$\begin{aligned} f(x) &= 2h_1(x) \int_u^v dy \psi(y) \int_u^y dz m(z) f(z) \\ &\quad - 2 \int_u^x dy \psi(y) \int_u^y dz m(z) f(z) \end{aligned}$$

Interchanging the order of integration gives

$$\begin{aligned} f(x) &= 2h_1(x) \int_u^v dz m(z) f(z) (\phi(v) - \phi(z)) \\ &\quad - 2 \int_u^x dz m(z) f(z) (\phi(x) - \phi(z)) \end{aligned}$$

The integral over $[x, v]$ in the first term is

$$2h_1(x) \int_x^v dz m(z) f(z) (\phi(v) - \phi(z)) \tag{7.36}$$

Adding the integral over $[u, x]$ from the first term to the second gives

$$2 \int_u^x dz m(z) f(z) \left[\frac{\phi(x) - \phi(u)}{\phi(v) - \phi(u)} (\phi(v) - \phi(z)) - (\phi(x) - \phi(z)) \right]$$

A little algebra shows

$$\begin{aligned} &(\phi(x) - \phi(u)) \cdot (\phi(v) - \phi(z)) - (\phi(x) - \phi(z)) \cdot (\phi(v) - \phi(u)) \\ &= -\phi(u)\phi(v) - \phi(x)\phi(z) + \phi(z)\phi(v) + \phi(x)\phi(u) \\ &= (\phi(v) - \phi(x)) \cdot (\phi(z) - \phi(u)) \end{aligned}$$

so the second part of our formula becomes

$$2 \frac{\phi(v) - \phi(x)}{\phi(v) - \phi(u)} \int_u^x dz m(z) f(z) [\phi(z) - \phi(u)]$$

Adding this to (7.36) gives the desired result. \square

An important consequence of (7.34) is:

Corollary. If $\tau_{u,v} = T_u \wedge T_v$ is the exit time from (u, v) then

$$E_x \tau_{u,v} = \int_u^v G(x, y) dy \tag{7.37}$$

Speed (?) measure

Suppose first that the diffusion is on its natural scale, i.e., $\phi(x) = x$. In this case, if we take $\alpha = x - h$ and $\beta = x + h$ then the Green's function becomes

$$\begin{aligned} (x + h - y)m(y) & \quad x \leq y \leq x + h \\ (y - x + h)m(y) & \quad x - h \leq y \leq x \end{aligned}$$

so (7.37) implies

$$E_x \tau_{x-h, x+h} = \int_x^{x+h} (x + h - y)m(y) dy + \int_{x-h}^x (y - x + h)m(y) dy \tag{7.38}$$

When h is small, $m(y) \approx m(x)$ for $y \in [x - h, x + h]$ so the above is

$$\approx m(x) \left(\int_x^{x+h} (x + h - y) dy + \int_{x-h}^x (y - x + h) dy \right) = m(x)h^2$$

Thus, $m(x)$ gives the *time* that X_t takes to exit a small interval centered at x , or to be precise, the ratio of the time for X_t to the time for a standard Brownian motion, which is h^2 . Since speed is inversely proportional to the exit time, the term speed measure is a misnomer, but it is too late to change its name.

To treat a general diffusion, we have to transform it to its natural scale. Writing $\psi = \phi'$ and noting $L\phi = 0$, Theorem 7.1 implies that $Y_t = \phi(X_t)$ is a diffusion with coefficients

$$\bar{a}(y) = (a\psi^2)(\phi^{-1}(y)) \quad \bar{b}(y) = 0 \tag{7.39}$$

Using the previous calculation for Y , if $\bar{m}(y) = 1/(a\psi^2)(\phi^{-1}(y))$ is the speed measure for Y then $E_{\phi(x)} \tau_{\phi(x-h), \phi(x+h)}$ is

$$\approx \bar{m}(\phi(x)) \left(\int_{\phi(x)}^{\phi(x+h)} (\phi(x+h) - z) dz + \int_{\phi(x-h)}^{\phi(x)} (z - \phi(x-h)) dz \right)$$

Changing variables $z = \phi(w)$, $dz = \psi(w)dw$, we see that this is

$$\begin{aligned} &= \bar{m}(\phi(x))\psi(x) \left(\int_x^{x+h} (x + h - w) dw + \int_{x-h}^x (w - x + h) dw \right) \\ &= \frac{1}{a(x)\psi(x)} h^2 = m(x)h^2 \end{aligned}$$

so the interpretation of $m(x)$ given above holds in general. As a final check that the speed measure is indeed the opposite of what the name suggests, recall that the stationary measure is $m(x)$, and the long run occupation time of a region is inversely proportional to the speed at which the process leaves it.

7.8 Examples

In all of the genetics examples in this section, there is no mutation. The calculations of the Green's functions require a fair amount of algebra, but they inform us about where we can expect the process to spend its time before reaching a boundary point at time τ and they allow us to compute $E\tau$. We begin by considering what happens when there is

No selection

Example 7.21. Symmetric simple random walk. Suppose that up jumps and down jumps each occur with probability $1/2$. X_n is a martingale, and it follows that if $a < x < b$

$$P_x(T_b < T_a) = \frac{x - a}{b - a} \quad P_x(T_a < T_b) = \frac{b - x}{b - a}$$

Using the second formula and then the first, the numerator in (7.32) is

$$P_i(T_j < T_{2N}) = \frac{2N - i}{2N - j} \quad j \leq i$$

$$P_i(T_j < T_0) = \frac{i}{j} \quad i \leq j$$

To compute the denominator of (7.32), we note that if $\tau = T_0 \wedge T_{2N}$ then

$$P_j(T_j^+ > \tau) = \frac{1}{2}P_{j+1}(T_{2N} < T_j) + \frac{1}{2}P_{j-1}(T_0 < T_j)$$

$$= \frac{1}{2} \cdot \frac{1}{2N - j} + \frac{1}{2} \cdot \frac{1}{j} = \frac{1}{2} \cdot \frac{2N}{j(2N - j)}$$

Combining the results, we can write $G(i, j)$ as

$$2 \frac{(2N - i)j}{2N} \quad j \leq i$$

$$2 \frac{i(2N - j)}{2N} \quad i \leq j \tag{7.40}$$

Summing over j and letting $k = 2N - j$, we have

$$\begin{aligned}
 E_{i\tau} &= 2 \frac{(2N-i)}{2N} \sum_{j=1}^i j + 2 \cdot \frac{i}{2N} \sum_{k=1}^{2N-i-1} k \\
 &= 2 \left[\frac{(2N-i)}{2N} \cdot \frac{i(i+1)}{2} + \frac{i}{2N} \cdot \frac{(2N-i)(2N-i-1)}{2} \right] \\
 &= \frac{i(2N-i)}{2N} [(i+1) + (2N-i-1)] = i(2N-i)
 \end{aligned}$$

in agreement with the result in Example 7.18.

Example 7.22. Moran model with no selection. In this case $P_i(T_j < \tau)$ and $P_j(R_j > \tau) = P_j(T_j^+ > \tau)$ are the same as for the symmetric simple random walk considered above, while $q_j = 2j(2N-j)/2N$. It follows from the calculation in the previous example that $q_j P_j(R_j > \tau)$ in the denominator of (7.33) is 1 and we can write $G(i, j)$ as

$$\begin{aligned}
 \frac{2N-i}{2N-j} & \quad j \leq i \\
 i/j & \quad i \leq j
 \end{aligned} \tag{7.41}$$

Summing over j we have

$$E_{i\tau} = (2N-i) \sum_{j=1}^i \frac{1}{2N-j} + i \sum_{j=i+1}^{2N} \frac{1}{j}$$

If $i = 2Nx$ with $0 < x < 1$ then

$$\sum_{j=i+1}^{2N} \frac{1}{j} = \sum_{j=i+1}^{2N} \frac{1}{j/2N} \cdot \frac{1}{2N} = \int_x^1 \frac{dx}{x} = -\log x$$

so we have

$$\frac{1}{N} E_{2Nx} - 2x \log x - 2(1-x) \log(1-x) \tag{7.42}$$

Example 7.23. Wright-Fisher diffusion with no selection. $\phi(x) = x$, $\psi(x) = \phi'(x) = 1$, and $a(x) = x(1-x)$ so using (7.34), $G(x, y) =$

$$\begin{aligned}
 \frac{2(1-x)y}{y(1-y)} &= \frac{2(1-x)}{1-y} & y \leq x \\
 \frac{2x(1-y)}{y(1-y)} &= \frac{2x}{y} & x \leq y
 \end{aligned} \tag{7.43}$$

If we set $i = 2Nx$ and $j = 2Ny$ in the Moran model formula, we get $(1-x)/(1-y)$ and x/y . The missing factor of 2 comes from the fact that i corresponds to $[x, x + 1/2N]$, so the occupation time density is multiplied by $2N$, but time is run at rate N , so it is divided by N .

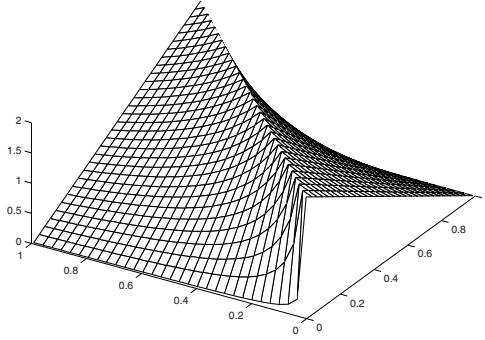


Fig. 7.5. Green's function for Wright-Fisher diffusion with no selection.

Integrating $G(x, y)$ we have

$$\begin{aligned} E_x \tau &= \int_x^1 \frac{2x}{y} dy + \int_0^x \frac{2(1-x)}{1-y} dy \\ &= -2x \log x - 2(1-x) \log(1-x) \end{aligned}$$

which agrees with (7.42) and our computation in Example 7.19.

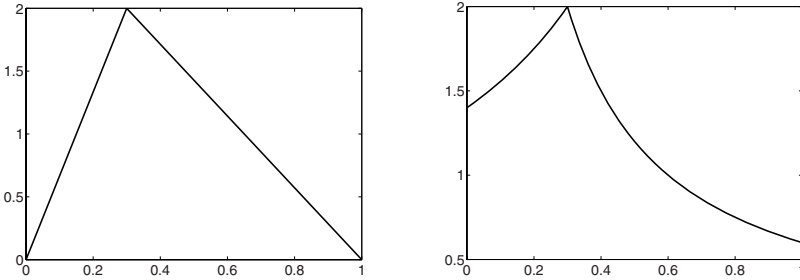


Fig. 7.6. Slices $x \rightarrow G(x, 0.3)$ and $y \rightarrow G(0.3, y)$ of the previous graph.

For fixed y , $x \rightarrow G(x, y)$ is linear on $[0, y]$ and $[y, 1]$ and vanishes at 0 and 1. To explain the form of the answer we return to our remark that for fixed y , $g(x) = G(x, y)$ is a solution of

$$\frac{1}{2}x(1-x)\frac{d^2}{dx^2}g = -\delta_y \tag{7.44}$$

When $x \neq y$, $g''(x) = 0$ so $g(x)$ is linear on $[0, y]$ and $[y, 1]$. The integral of $-\delta_y$ is 0 for $x < y$ and -1 for $x > y$, so with a little thought we realize that (7.44) can be written as $g'(y+) - g'(y-) = 2/y(1-y)$, which is correct since

$$g'(y+) - g'(y-) = \frac{2}{1-y} - \frac{2}{y} = \frac{2}{y(1-y)}$$

Selection

Example 7.24. Asymmetric simple random walk. Suppose that up jumps occur with probability $p = 1/(2-s)$ and down jumps with probability $1-p = (1-s)/(2-s)$. We have chosen these values of p so that this is the embedded discrete-time jump chain for the Moran model with selection. The ratio $(1-p)/p = 1-s$, so if we let $h(x) = 1 - (1-s)^x$ then by calculations in Section 6.1, $h(X_n)$ is a martingale and it follows that if $a < x < b$

$$P_x(T_b < T_a) = \frac{h(a) - h(x)}{h(a) - h(b)} \quad P_x(T_a < T_b) = \frac{h(x) - h(b)}{h(a) - h(b)}$$

We have reversed the usual order of the numerator and denominator to make the next few calculations easier to see. Using the second formula and then the first, the numerator in (7.32) is

$$P_i(T_j < T_{2N}) = \frac{(1-s)^i - (1-s)^{2N}}{(1-s)^j - (1-s)^{2N}} \quad j \leq i$$

$$P_i(T_j < T_0) = \frac{1 - (1-s)^i}{1 - (1-s)^j} \quad i \leq j$$

To compute the denominator of (7.32), we note that

$$P_j(T_j^+ > \tau) = \frac{1}{2-s} P_{j+1}(T_{2N} < T_j) + \frac{1-s}{2-s} P_{j-1}(T_0 < T_j)$$

$$= \frac{1}{2-s} \frac{(1-s)^j - (1-s)^{j+1}}{(1-s)^j - (1-s)^{2N}} + \frac{1-s}{2-s} \frac{(1-s)^{j-1} - (1-s)^j}{1 - (1-s)^j}$$

The two numerators are $(1-s)^j - (1-s)^{j+1} = s(1-s)^j$, so the above is

$$= \frac{s(1-s)^j}{2-s} \cdot \frac{[1 - (1-s)^{2N}]}{[1 - (1-s)^j][(1-s)^j - (1-s)^{2N}]} \quad (7.45)$$

Reintroducing $h(x) = 1 - (1-s)^x$, we can write $G(i, j)$ as

$$\frac{(h(2N) - h(i)) \cdot (h(j) - h(0))}{h(2N) - h(0)} \cdot \frac{2-s}{s(1-s)^j} \quad j \leq i$$

$$\frac{(h(i) - h(0)) \cdot (h(2N) - h(j))}{h(2N) - h(0)} \cdot \frac{2-s}{s(1-s)^j} \quad i \leq j \quad (7.46)$$

Example 7.25. Moran model with selection. In this case, $P_i(T_j < \tau)$ and $P_j(R_j > \tau) = P_j(T_j^+ > \tau)$ are the same as for the asymmetric simple random

walk considered above, while $q_j = (2 - s)j(2N - j)/2N$. It follows from (7.45) that the denominator of (7.33) is

$$q_j P_j(R_j > \tau) = \frac{j(2N - j)}{2N} \frac{s(1 - s)^j \cdot [1 - (1 - s)^{2N}]}{[1 - (1 - s)^j][(1 - s)^j - (1 - s)^{2N}]}$$

and the Green's function becomes

$$\begin{aligned} & \frac{(h(2N) - h(i)) \cdot (h(j) - h(0))}{h(2N) - h(0)} \cdot \frac{2N}{s(1 - s)^j j(2N - j)} \quad 0 < j < i \\ & \frac{(h(i) - h(0)) \cdot (h(2N) - h(j))}{h(2N) - h(0)} \cdot \frac{2N}{s(1 - s)^j j(2N - j)} \quad i < j < 2N \end{aligned} \quad (7.47)$$

Example 7.26. Wright-Fisher diffusion with additive selection. To make it easier to relate the results for this case to the Moran model, we will define the natural scale to be $\phi(y) = 1 - \exp(-2\gamma y)$ which makes

$$\begin{aligned} \psi(y) &= 2\gamma \exp(-2\gamma y) \\ m(y) &= \frac{1}{a(y)\psi(y)} = \frac{1}{2\gamma \exp(-2\gamma)y(1 - y)} \end{aligned}$$

Recalling the formula for $G(x, y)$

$$\begin{aligned} & 2 \frac{(\phi(1) - \phi(x))(\phi(y) - \phi(0))}{\phi(1) - \phi(0)} \cdot m(y) \quad y \leq x \\ & 2 \frac{(\phi(x) - \phi(0))(\phi(1) - \phi(y))}{\phi(1) - \phi(0)} \cdot m(y) \quad x \leq y \end{aligned}$$

we see that $G(x, y)$ is given by

$$\begin{aligned} & 2(e^{-2\gamma x} - e^{-2\gamma}) \frac{1 - e^{-2\gamma y}}{1 - e^{-2\gamma}} \cdot \frac{1}{2\gamma e^{-2\gamma y} y(1 - y)} \quad y \leq x \\ & 2(1 - e^{-2\gamma x}) \frac{e^{-2\gamma y} - e^{-2\gamma}}{1 - e^{-2\gamma}} \cdot \frac{1}{2\gamma e^{-2\gamma y} y(1 - y)} \quad x \leq y \end{aligned} \quad (7.48)$$

To connect with the Moran model, note that if $x = i/2N$, $y = j/2N$, and $s = 2\gamma/2N$ then

$$\begin{aligned} h(i) &= 1 - (1 - \gamma/2N)^{2Nx} \rightarrow 1 - e^{-2\gamma x} = \phi(x) \\ \frac{2N \cdot 2N}{2Ns(1 - s)^j j(2N - j)} &\rightarrow \frac{1}{2\gamma e^{-2\gamma y} y(1 - y)} = m(y) \end{aligned}$$

As in the case of no selection, the missing factor of 2 comes from the fact that i corresponds to $[x, x + 1/2N]$, so the density is multiplied by $2N$, but time is run at rate N to get the diffusion limit, so it is divided by N .

To help understand the Green's function, it is useful to look at slices through the graph. If y is fixed then for $x < y$ we have $G(x, y) = A(y)(1 -$

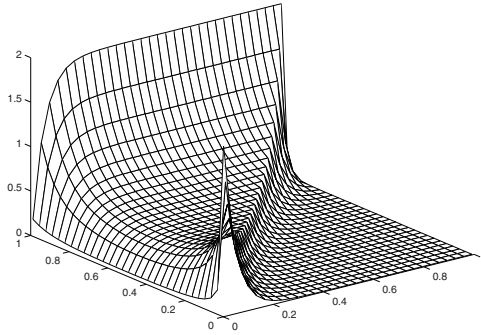


Fig. 7.7. Green's function for Wright-Fisher diffusion with additive selection $\gamma = 10$

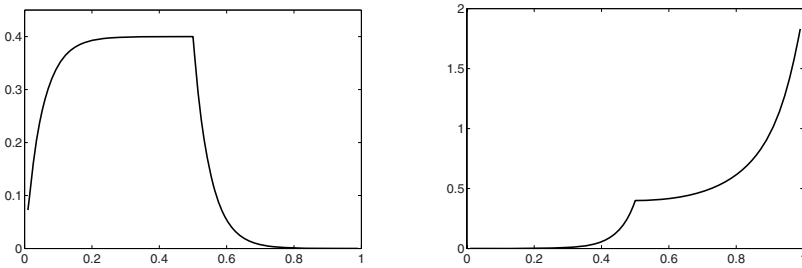


Fig. 7.8. Slices $x \rightarrow G(x, 1/2)$ and $y \rightarrow G(1/2, y)$ of the previous graph.

$e^{-2\gamma x}$ so if $x \gg 1/2\gamma$, then $G(x, y)$ is roughly constant because x will hit y with probability close to 1. For $x > y$, $G(x, y) = B(y)(e^{-2\gamma x} - e^{-2\gamma})$ since the probability of hitting y decays exponentially fast. Let $g(x) = G(x, y)$. For $x \neq y$ we have $Lg = 0$ where

$$Lg = \frac{1}{2}x(1-x) \left[\frac{d^2g}{dx^2} + 2\gamma \frac{dg}{dx} \right]$$

A little calculus shows that again we have

$$g'(y+) - g'(y-) = -2\gamma e^{-2\gamma y} \frac{2}{m(y)} = \frac{2}{y(1-y)}$$

To understand the behavior for x fixed, it is useful to multiply top and bottom of (7.48) by $e^{2\gamma y}$ to rewrite $G(x, y)$ as

$$\begin{aligned} & \frac{e^{-2\gamma x} - e^{-2\gamma}}{1 - e^{-2\gamma}} \cdot \frac{e^{2\gamma y} - 1}{\gamma y(1-y)} & y \leq x \\ & \frac{1 - e^{-2\gamma x}}{1 - e^{-2\gamma}} \cdot \frac{1 - e^{-2\gamma(1-y)}}{2\gamma y(1-y)} & x \leq y \end{aligned} \tag{7.49}$$

If x is fixed and γ is large, then $G(x, y)$ is approximately

$$\begin{aligned} \frac{e^{-2\gamma(x-y)}}{\gamma y(1-y)} & \quad 0 \leq y \leq x \\ \frac{1}{\gamma y(1-y)} & \quad x \leq y \quad \text{and} \quad \gamma(1-y) \gg 1 \\ \frac{1 - e^{-2c}}{c} & \quad x \leq y = 1 - c/\gamma \end{aligned}$$

This shows that the process spends a negligible amount of time $< x$ and moves through values $y < 1$ at the rate predicted by the logistic differential equation until $1 - y = O(\gamma^{-1})$. Note that as $y \rightarrow 1$, $c \rightarrow 0$ and $(1 - e^{-2c})/c \rightarrow 2$.

Example 7.27. Symmetric balancing selection. In this case $\psi(x) = e^{-2\gamma x(1-x)}$, $m(x) = e^{2\gamma x(1-x)}/x(1-x)$, and $\phi(x) = \int_0^x \psi(y) dy$, so $G(x, y)$ is

$$\begin{aligned} 2 \frac{(\phi(1) - \phi(x))\phi(y)}{\phi(1)} \cdot \frac{e^{2\gamma y(1-y)}}{y(1-y)} & \quad y \leq x \\ 2 \frac{(\phi(1) - \phi(y))\phi(x)}{\phi(1)} \cdot \frac{e^{2\gamma y(1-y)}}{y(1-y)} & \quad x \leq y \end{aligned} \tag{7.50}$$

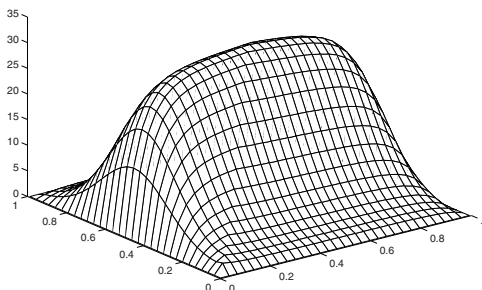


Fig. 7.9. Green’s function for Wright-Fisher diffusion with symmetric balancing selection $\gamma = 10$

To help understand the Green’s function, it is useful to look at slices through the graph. As we computed in Example 7.15 at the end of Section 7.4, when γ is large and x is away from the boundaries at 0 and 1,

$$\begin{aligned} \phi(1) & \approx 2 \int_0^\infty e^{-2\gamma y} dy = \frac{1}{\gamma} \\ \phi(x), 1 - \phi(x) & \approx \int_0^\infty e^{-2\gamma y} dy = \frac{1}{2\gamma} \end{aligned}$$

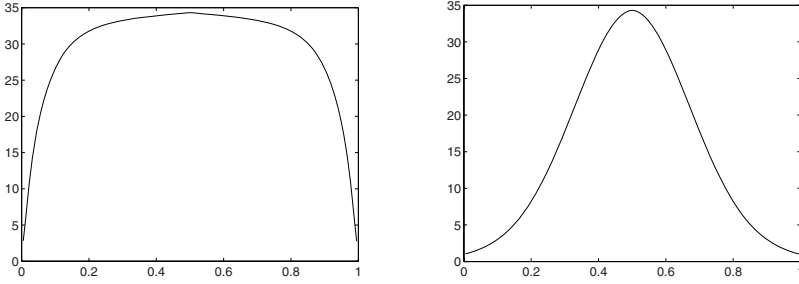


Fig. 7.10. Slices $x \rightarrow G(x, 1/2)$ and $y \rightarrow G(1/2, y)$ of the previous graph.

so the first factor in (7.50) is $\approx 1/4\gamma$ and the two cases collapse to

$$\frac{e^{2\gamma y(1-y)}}{2\gamma y(1-y)}$$

The approximation does not depend on the starting point because no matter where the diffusion starts, the frequency quickly moves to $1/2$.

Changing variables $y = 1/2 + z$, which makes $y(1 - y) = 1/4 - z^2$, the above becomes

$$\frac{2e^{\gamma/2}}{\gamma} \frac{e^{-2\gamma z^2}}{1 - 4z^2}$$

Realizing that most of the contribution will come from values of z of order $O(1/\sqrt{\gamma})$,

$$E_x \tau = \int_0^1 G(x, y) dy \approx \frac{2e^{\gamma/2}}{\gamma} \int e^{-2\gamma z^2} dz$$

The integrand resembles the normal density with mean 0 and variance $1/4\gamma$, so its value is $\sqrt{\pi/2\gamma}$ and we have

$$E_x \tau \approx e^{\gamma/2} \gamma^{-3/2} \sqrt{2\pi}$$

7.9 Conditioned processes

In many situations we are interested in conditioning that the current mutation fixes or dies out. If $h(x) = P_x(T_1 < T_0)$ and we condition on fixation then the new transition probability

$$\bar{p}_t(x, y) = p_t(x, y)h(y)/h(x).$$

The same result holds for conditioning on loss, with $h(x) = P_x(T_0 < T_1)$. Integrating, we have that the conditioned Green's function

$$\bar{G}(x, y) = G(x, y)h(y)/h(x). \tag{7.51}$$

To compute the generator of the conditioned process, we note that

$$\begin{aligned} \bar{L}f(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \left(\int \frac{p_t(x, y)h(y)}{h(x)} f(y) dy - f(x) \right) \\ &= \frac{1}{h(x)} \lim_{t \rightarrow 0} \frac{1}{t} \left(\int p_t(x, y)h(y)f(y) dy - h(x)f(x) \right) = \frac{1}{h(x)}L(hf) \end{aligned}$$

Working out the derivatives

$$\frac{1}{h(x)}L(hf) = \frac{1}{h(x)} \left(b(x)(h'f + hf') + \frac{1}{2}a(x)(h''f + 2h'f' + hf'') \right)$$

Using $Lh = 0$ and simplifying

$$\bar{L}f = \frac{1}{2}a(x)f'' + \left(b(x) + a(x)\frac{h'(x)}{h(x)} \right) f'(x) \tag{7.52}$$

in agreement with (32) in Ewens (1973).

Example 7.28. Wright-Fisher diffusion with no selection. $h(x) = x$ is the probability of fixation, so using (7.43) and (7.51), $\bar{G}(x, y) =$

$$\begin{aligned} \frac{2(1-x)}{1-y} \cdot \frac{y}{x} & \quad y \leq x \\ \frac{2x}{y} \cdot \frac{y}{x} = 2 & \quad x \leq y \end{aligned} \tag{7.53}$$

and using (7.52) we have

$$\bar{L}v_1 = \frac{1}{2}x(1-x)f''(x) + (1-x)f'(x) \tag{7.54}$$

Theorem 7.17. *For the Wright-Fisher model conditioned on fixation, $\tau = T_0 \wedge T_1$ has*

$$E_x(\tau|T_1 < T_0) = -2\frac{(1-x)}{x} \log(1-x) \tag{7.55}$$

$$E_x(\tau^2|T_1 < T_0) = 8 \left(\frac{(1-x) \log(1-x)}{x} - \int_x^1 \frac{\log(1-y)}{y} dy \right) \tag{7.56}$$

The first formula is (14) in Kimura and Ohta (1969a). This second can be obtained from (A7) of Kimura and Ohta (1969b), which is for conditioning on extinction. Both are on page 29 of Ewens (1973), but in the second case he has an erroneous minus sign. In the two references cited, formulas are given on the original time scale, so the first formula is multiplied by $2N$ and the second by $(2N)^2$.

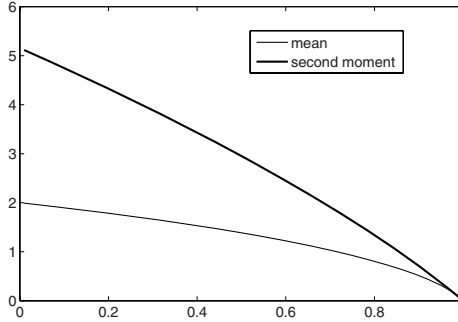


Fig. 7.11. First and second moments of τ conditioned on $T_1 < T_0$, as a function of the starting point.

Proof. We can derive the first formula from (7.53). Integrating using $y/(1 - y) = 1/(1 - y) - 1$, we have

$$\begin{aligned} E_x(\tau|T_1 < T_0) &= \int_x^1 2 \, dy + \frac{2(1-x)}{x} \int_0^x \frac{y}{1-y} \, dy \\ &= 2 \left((1-x) + \frac{(1-x)}{x} (-\log(1-x) - x) \right) \end{aligned}$$

and after a little arithmetic, we have (7.55).

A second approach is to use Theorem 7.12. Let $v_1(x)$ be our formula for $E_x(\tau|T_1 < T_0)$. The first step is to check that $Lv_1 = -1$. To do this, we note that writing $-(1-x)/x = -1/x + 1$

$$\begin{aligned} v_1'(x) &= \frac{2}{x^2} \log(1-x) - 2 \frac{1-x}{x} \cdot \frac{-1}{1-x} = \frac{2}{x^2} \log(1-x) + \frac{2}{x} \\ v_1''(x) &= -\frac{4}{x^3} \log(1-x) + \frac{2}{x^2} \cdot \frac{-1}{1-x} - \frac{2}{x^2} \end{aligned}$$

so we have

$$\begin{aligned} \bar{L}v_1 &= \frac{1}{2}x(1-x)v_1''(x) + (1-x)v_1'(x) \\ &= -\frac{2(1-x)}{x^2} \log(1-x) - \frac{1}{x} - \frac{(1-x)}{x} \\ &\quad + \frac{2(1-x)}{x^2} \log(1-x) + \frac{2(1-x)}{x} \\ &= -\frac{1}{x} + \frac{1-x}{x} = -1 \end{aligned}$$

To examine the boundary conditions, we note that $y \log(y) \rightarrow 0$ as $y \rightarrow 0$ so $v_1(1) = 0$. We do not have $v_1(0) = 0$. The easiest way to see this is to note

that $\log(1-x) \leq -x$, so $v_1'(x) \leq 0$, i.e., $v_1(x)$ is decreasing, which is what one should expect for $v_1(x) = E_x(\tau|T_1 < T_0)$. As $x \rightarrow 0$, $\log(1-x) \sim -x$ so $v_1(0) = 2$. As we will see in Section 7.10, the apparent inconsistency with Theorem 7.12 comes because it is impossible for the conditioned process to get to 0, so one of the assumptions of that theorem does not hold.

Let $v_2(x)$ be our formula for $E_x(\tau^2|T_1 < T_0)$. To verify the second formula using Theorem 7.13, we want to show $Lv_2 = -2v_1$. To check this guess, we note that

$$\begin{aligned} \frac{d}{dx} \frac{v_2}{8} &= -\frac{1}{x^2} \log(1-x) + \frac{(1-x)}{x} \cdot \frac{-1}{1-x} + \frac{\log(1-x)}{x} \\ &= \left(-\frac{1}{x^2} + \frac{1}{x}\right) \log(1-x) - \frac{1}{x} \\ \frac{d^2}{dx^2} \frac{v_2}{8} &= \left(\frac{2}{x^3} - \frac{1}{x^2}\right) \log(1-x) - \frac{1-x}{x^2} \cdot \frac{-1}{1-x} + \frac{1}{x^2} \\ &= \frac{2-x}{x^3} \log(1-x) + \frac{2}{x^2} \end{aligned}$$

Combining the last two results, we see that

$$\begin{aligned} \bar{L}v_2 &= \frac{1}{2} \cdot \frac{2-x}{x^2} (1-x) \log(1-x) + \frac{1-x}{x} \\ &\quad + \frac{(x-1)(1-x)}{x^2} \log(1-x) - \frac{(1-x)}{x} \\ &= \frac{(1-x)}{x} \log(1-x) \left[\frac{1}{x} - \frac{1}{2} + 1 - \frac{1}{x} \right] \end{aligned}$$

and it follows that

$$\bar{L}v_2 = 4 \frac{(1-x)}{x} \log(1-x) = -2v_1$$

Clearly $v_2(1) = 0$, so the only relevant boundary condition is satisfied. □

Example 7.29. Age of alleles. By symmetry

$$E_x(\tau|T_0 < T_1) = E_{1-x}(\tau|T_1 < T_0) = -2 \frac{x}{1-x} \log(x)$$

As we will now show, this gives the average age of an allele A observed to be at frequency x , a classic result of Kimura and Ohta (1973). To argue this, we note that the density of the age of A given that it has frequency x is

$$f_x(t) = \lim_{\epsilon \rightarrow 0} \frac{p_t(\epsilon, x)}{\int_0^\infty p_s(\epsilon, x) ds}$$

Multiplying top and bottom by the speed measure $m(\epsilon)$, then using reversibility, (7.30), and noting the factors of m cancel, the above

$$\begin{aligned} &= \lim_{\epsilon \rightarrow 0} \frac{m(\epsilon)p_t(\epsilon, x)}{\int_0^\infty m(\epsilon)p_s(\epsilon, x) ds} \\ &= \lim_{\epsilon \rightarrow 0} \frac{p_t(x, \epsilon)}{\int_0^\infty p_s(x, \epsilon) ds} = g_x(t) \end{aligned}$$

the density for the hitting time of 0 starting from x . Therefore,

$$\int t f_x(t) dt = \int t g_x(t) dt = E_x(\tau | T_0 < T_1)$$

I learned this argument from Griffiths (2003). See his paper for an account of the history and results for the expected values of the ages of alleles observed to occur k times in a sample of size n .

Example 7.30. Wright-Fisher diffusion with additive selection. In this case the probability of fixation is $h(x) = (1 - e^{-2\gamma x}) / (1 - e^{-2\gamma})$ when the initial frequency is x , so

$$\bar{G}(x, y) = G(x, y) \frac{1 - e^{-2\gamma y}}{1 - e^{-2\gamma x}}$$

Using (7.49) now, we have that $\bar{G}(x, y)$ is

$$\begin{aligned} &\frac{1 - e^{-2\gamma y}}{1 - e^{-2\gamma x}} \cdot \frac{e^{-2\gamma x} - e^{-2\gamma}}{1 - e^{-2\gamma}} \cdot \frac{e^{2\gamma y} - 1}{\gamma y(1 - y)} && y \leq x \\ &\frac{1 - e^{-2\gamma y}}{1 - e^{-2\gamma}} \cdot \frac{1 - e^{-2\gamma(1-y)}}{\gamma y(1 - y)} && x \leq y \end{aligned}$$

Note that the second formula does not depend on x (except through the condition $x \leq y$). As Figure 7.12 shows, the conditioning does not change the picture very much except near $x = 0$, where we no longer have $G(x, y) \rightarrow 0$ as $x \rightarrow 0$.

If we integrate this with respect to y , then we get a result first derived by Kimura and Ohta (1969a), see their (17).

$$\begin{aligned} \bar{E}_x \tau &\approx \int_x^1 \frac{[1 - e^{-2\gamma y}] \cdot [1 - e^{-2\gamma(1-y)}]}{[1 - e^{-2\gamma}] \cdot \gamma y(1 - y)} dy \\ &+ \frac{e^{-2\gamma x} - e^{-2\gamma}}{1 - e^{-2\gamma x}} \int_0^x \frac{[1 - e^{-2\gamma y}] \cdot [e^{\gamma y} - 1]}{[1 - e^{-2\gamma}] \cdot 2\gamma y(1 - y)} dy \end{aligned}$$

Since $1 - e^{-2\gamma a} \leq 2\gamma a$, the two integrals are finite. However, they must be evaluated numerically. As a check on the last formula, we note $(e^{a\gamma} - 1) / \gamma \rightarrow a$ as $\gamma \rightarrow 0$ so

$$\frac{1 - e^{-2\gamma y}}{\gamma y} \cdot \frac{1 - e^{-2\gamma(1-y)}}{\gamma(1 - y)} \cdot \frac{\gamma}{1 - e^{-2\gamma}} \rightarrow 2$$

Using similar reasoning on the other terms,

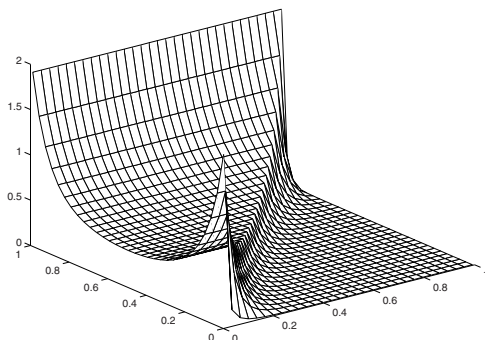


Fig. 7.12. Green’s function for conditioned Wright-Fisher with selection $\gamma = 10$

$$\begin{aligned} \bar{E}_x \tau &\rightarrow \int_x^1 2 \, dy + 2N \left(\frac{1-x}{x} \right) \int_0^x \frac{2y}{1-y} \, dy \\ &= -2 \cdot \frac{(1-x)}{x} \log(1-x) \end{aligned}$$

by the calculation in Example 7.28.

Example 7.31. General diffusion. Maruyama and Kimura (1974) observed that for Wright-Fisher diffusions with general diploid selection

$$\lim_{x \rightarrow 0} E_x(\tau|T_1 < T_0) = \lim_{x \rightarrow 1} E_x(\tau|T_0 < T_1) \tag{7.57}$$

As we will now show, and presumably the authors also realized, this is a general property of one dimensional diffusions.

Proof. We begin by recalling the formula for the Green’s function given in (7.34), which we simplify by supposing $\phi(0) = 0$ and $\phi(1) = 1$.

$$\begin{aligned} 2(1 - \phi(x))\phi(y)m(y) & \quad y \leq x \\ 2\phi(x)(1 - \phi(y))m(y) & \quad x \leq y \end{aligned}$$

The Green’s function $G_1(x, y)$ for the process starting from x and conditioned on $T_1 < T_0$ is

$$\begin{aligned} 2(1 - \phi(x))\phi(y)m(y) \cdot \frac{\phi(y)}{\phi(x)} & \quad y \leq x \\ 2\phi(x)(1 - \phi(y))m(y) \cdot \frac{\phi(y)}{\phi(x)} & \quad x \leq y \end{aligned}$$

in agreement with (12) of Maruyama and Kimura (1971). The Green’s function $G_0(x, y)$ for the process starting from x and conditioned on $T_0 < T_1$ is

$$2(1 - \phi(x))\phi(y)m(y) \cdot \frac{1 - \phi(y)}{1 - \phi(x)} \quad y \leq x$$

$$2\phi(x)(1 - \phi(y))m(y) \cdot \frac{1 - \phi(y)}{1 - \phi(x)} \quad x \leq y$$

In each case, after cancellation we have for $x \leq y$

$$G_0(x, y) = G_1(x, y) = 2(\phi(1) - \phi(y))\phi(y) \cdot m(y)$$

and integrating gives the desired result. \square

Taylor, Iwasa, and Nowak (2006) have shown that this result holds for reversible Markov chains on $\{0, 1, \dots, N\}$ in which 0 and N are absorbing and these two states can only be reached directly from 1 and $N - 1$ respectively.

7.10 Boundary behavior

The consideration of diffusion processes leads to two questions that have no analogues for discrete models: “Can the process get to the boundary?” and “Once it gets to the boundary can it return to the interior of the state space?” To build some suspense, we invite the reader to guess what happens for the Wright-Fisher diffusion with mutation

$$Lf = \frac{1}{2} \frac{d^2}{dx^2} f + (\beta_1(1 - x) - \beta_2 x) \frac{d}{dx} f$$

It should not be surprising that if $\beta_1 = \beta_2 = 0$ then the diffusion stops the first time it hits 0 or 1, but what if one or both of the $\beta_i > 0$?

It is enough to consider the boundary at 0. Consider a diffusion on $(0, r)$ where $r \leq \infty$, let $q \in (0, r)$, and let

$$I = \int_0^q (\phi(z) - \phi(0)) m(z) dz$$

$$J = \int_0^q (M(z) - M(0)) \psi(z) dz$$

where M is an antiderivative of m . Writing iff as short for “if and only if,” we have the following results for a diffusion process X_t .

Theorem 7.18. X_t can get IN to the boundary point 0 iff $I < \infty$.
 X_t can get OUT from the boundary point 0 iff $J < \infty$.

Note that $\phi(0) = -\infty$ implies $I = \infty$ and $M(0) = -\infty$ implies $J = \infty$.

Proof. To start to prove the first result, we will show

Theorem 7.19. *Let $1/2 < b < 1$. The following are equivalent:*

- (i) $\phi(0) > -\infty$ and $\int_0^{1/2} (\phi(z) - \phi(0)) m(z) dz < \infty$
- (ii) $\inf_{0 < a < 1/2} P_{1/2}(T_a < T_b) > 0$ and $\sup_{0 < a < 1/2} E_{1/2}(T_a \wedge T_b) < \infty$
- (iii) $P_{1/2}(T_0 < T_b) > 0$

Proof. We first show that (i) and (ii) are equivalent.

$$P_{1/2}(T_a < T_b) = \frac{\phi(b) - \phi(1/2)}{\phi(b) - \phi(a)}$$

so $\inf_{\alpha < a < 0} P_0(T_a < T_b) > 0$ if and only if $\phi(0) > -\infty$. Using (7.37) and (7.34)

$$\begin{aligned} E_x \tau_{a,b} &= 2 \frac{\phi(x) - \phi(a)}{\phi(b) - \phi(a)} \int_x^b (\phi(b) - \phi(z)) m(z) dz \\ &\quad + 2 \frac{\phi(b) - \phi(x)}{\phi(b) - \phi(a)} \int_a^x (\phi(z) - \phi(a)) m(z) dz \end{aligned}$$

The first integral always stays bounded as $a \downarrow 0$. So $E_0 \tau_{(a,b)}$ stays bounded as $a \rightarrow 0$ if and only if $\phi(0) > -\infty$ and

$$\int_0^{1/2} (\phi(z) - \phi(0)) m(z) dz < \infty$$

which completes the proof of the equivalence of (i) and (ii).

It is easy to see that (ii) implies (iii). For the other direction, note that Theorem 7.6 implies that if $P_{1/2}(T_0 < T_b) > 0$ then $E_{1/2} \tau_{0,b} < \infty$. \square

To try to start the process X_t from 0, let ϕ be its natural scale. As (7.39) shows, $Y_t = \phi(X_t)$ has coefficients $\bar{b}(y) = 0$ and

$$\bar{a}(y) = (a\psi^2)(\phi^{-1}(y))$$

To see if we can start the process Y_t at 0, we extend \bar{a} to the negative half-line by setting $\bar{a}(-y) = \bar{a}(y)$ and let Z_t be the associated diffusion. If we let $\bar{m}(y) = 1/\bar{a}(|y|)$ be the speed measure for Z_t , which is on its natural scale, we can use (7.39) and the symmetry $\bar{m}(-y) = \bar{m}(y)$ to conclude

$$\frac{1}{2} E_0 \tau_{-\epsilon, \epsilon} = \int_0^\epsilon (\epsilon - y) \bar{m}(y) dy$$

Changing variables $y = \phi(x)$, $dy = \psi(x) dx$, $\epsilon = \phi(\delta)$, the above

$$\begin{aligned} &= \int_0^\delta (\phi(\delta) - \phi(x)) \frac{1}{\psi(x)a(x)} dx \\ &= \int_0^\delta \left(\int_x^\delta \psi(z) dz \right) m(x) dx \end{aligned}$$

Interchanging the order of integration, the above

$$= \int_0^\delta \int_0^z m(x) dx \psi(z) dz$$

At this point we have shown

$$E_0 \tau_{-\epsilon, \epsilon} = 2 \int_0^\delta (M(z) - M(0)) \psi(z) dz \quad (7.58)$$

To see that $E_0 \tau_{-\epsilon, \epsilon} = \infty$ means that the process cannot escape from 0, we note that Theorem 7.6 implies that if $P_0(\tau_{-\epsilon, \epsilon} < \infty) > 0$ then $E_0 \tau_{-\epsilon, \epsilon} < \infty$. This completes the proof of the second result and of the theorem. \square

There are four possible combinations of I and J being finite or infinite, which were named by Feller as follows

$$\begin{array}{lll} I < \infty & J < \infty & \text{regular} \\ I < \infty & J = \infty & \text{absorbing} \\ I = \infty & J < \infty & \text{entrance} \\ I = \infty & J = \infty & \text{natural} \end{array}$$

The second case is called absorbing because we can get in to 0 but cannot get out. The third is called an entrance boundary because we cannot get to 0 but we can start the process there. Finally, in the fourth case, the process can neither get to nor start at 0, so it is reasonable to exclude 0 from the state space. We will now give examples of the various possibilities.

Example 7.32. Reflecting Brownian motion. Suppose $X_t = |B_t|$. In this case $\phi(x) = x$ and $m(x) = 1$ so

$$\begin{aligned} I &= \int_0^{1/2} (\phi(z) - \phi(0))m(z) dz = \int_0^{1/2} z dz < \infty \\ J &= \int_0^{1/2} (M(z) - M(0))\psi(z) dz = \int_0^{1/2} z dz < \infty \end{aligned}$$

and 0 is a regular boundary point. $\phi(\infty) = \infty$ and $M(\infty) = \infty$, so ∞ is a natural boundary.

Example 7.33. Wright-Fisher diffusion. We begin with the case of no selection. From (7.26), we have

$$\begin{aligned} \psi(x) &= x^{-2\beta_1} (1-x)^{-2\beta_2} \\ m(x) &= x^{2\beta_1-1} (1-x)^{2\beta_2-1} \end{aligned}$$

As $x \rightarrow 0$, $\psi(x) \sim x^{-2\beta_1}$, so if $\beta_1 \geq 1/2$, $\phi(0) = -\infty$ and the boundary cannot be reached. If $\beta_1 < 1/2$ then $\phi(z) - \phi(0) \sim Cx^{-2\beta_1+1}$ so $I < \infty$. If $\beta_1 = 0$ then $M(0) = -\infty$. If $\beta_1 > 0$ then $M(z) - M(0) \sim Cz^{2\beta_1}$, so $J < \infty$. Combining our calculations we see that

if	I	J	0 is
$\beta_1 = 0$	$< \infty$	$= \infty$	absorbing
$\beta_1 \in (0, 1/2)$	$< \infty$	$< \infty$	regular
$\beta_1 \geq 1/2$	$= \infty$	$< \infty$	entrance

Personally, I find it a little surprising that the accessibility of the boundary depends on size of the mutation rate.

If we consider a general selection scheme, then the function ψ is multiplied by $e^{-(2\delta x + \eta x^2)}$, and m by $e^{2\delta x + \eta x^2}$, which are bounded on $[0, 1]$, so the results of the tests do not change.

Example 7.34. Conditioned processes. If $p_t(x, y)$ is the transition probability of one of our diffusions X_t and $h(x) = P_x(T_1 < T_0)$, then, as we observed in Section 7.9, the process conditioned to hit 1 before 0, \bar{X}_t has transition probability $\bar{p}_t(x, y) = p_t(x, y)h(y)/h(x)$, and generator

$$\bar{L}f = Lf + a(x)\frac{h'(x)}{h(x)}f'(x)$$

In the absence of mutation and selection, $h(x) = x$ and $h'(x) = 1$, so

$$\bar{L}f = \frac{1}{2}x(1-x)\frac{d^2f}{dx^2} + (1-x)\frac{df}{dx}$$

Dropping the bars for the rest of the computation, $-2b(x)/a(x) = -2/x$, so

$$\psi(x) = e^{-2\log x} = x^{-2} \quad \text{and} \quad \phi(x) = -x^{-1}$$

$\phi(0) = -\infty$ so $I = \infty$. The speed measure

$$m(x) = \frac{1}{x(1-x)x^{-2}} = \frac{x}{1-x}$$

so $M(z) - M(0) \sim z^2$ as $z \rightarrow 0$. Since $\psi(z) = z^{-2}$, $J < \infty$. Thus, as we should have expected from the beginning, the conditioning makes 0 an entrance boundary. The process started at 0 will immediately become positive and never to return to 0.

In the next two examples, we will examine the influence of the drift and diffusion coefficients on the boundary behavior.

Example 7.35. Bessel processes. Suppose that $a(x) = 1$ and $b(x) = \gamma/2x$ for a diffusion on $[0, \infty)$. The natural scale is

$$\begin{aligned} \phi(x) &= \int_1^x \exp\left(-\int_1^y \gamma/z \, dz\right) dy \\ &= \int_1^x y^{-\gamma} dy = \begin{cases} \ln x & \text{if } \gamma = 1 \\ (x^{1-\gamma} - 1)/(1 - \gamma) & \text{if } \gamma \neq 1 \end{cases} \end{aligned}$$

From the last computation, we see that if $\gamma \geq 1$ then $\phi(0) = -\infty$ and $I = \infty$.

To handle $\gamma < 1$, we observe that the speed measure

$$m(z) = \frac{1}{\phi'(z)a(z)} = z^\gamma$$

So taking $q = 1$ in the definition of I ,

$$I = \int_0^q (\phi(z) - \phi(0)) m(z) dz = \int_0^1 \frac{z^{1-\gamma}}{1-\gamma} z^\gamma dz < \infty$$

To compute J , we observe that for $\gamma \leq -1$, $M(0) = -\infty$ while for $\gamma > -1$, $M(z) = z^{\gamma+1}/(\gamma+1)$ and

$$J = \int_0^q (M(z) - M(0)) \psi(z) dz = \int_0^1 \frac{z^{\gamma+1}}{\gamma+1} z^{-\gamma} dz < \infty$$

Combining the two conclusions about I and J , we see that

if	I	J	0 is
$\gamma \geq 1$	$= \infty$	$< \infty$	entrance
$\gamma \in (-1, 1)$	$< \infty$	$< \infty$	regular
$\gamma \leq -1$	$< \infty$	$= \infty$	absorbing

which makes sense because as γ gets larger, the push away from 0 increases.

Example 7.36. Power law fluctuations. Suppose $a(x) = x^\delta$ and $b(x) = 0$. The natural scale is $\phi(x) = x$ and the speed measure is $m(x) = 1/(\phi'(x)a(x)) = x^{-\delta}$, so

$$I = \int_0^1 x^{1-\delta} dx = \begin{cases} < \infty & \text{if } \delta < 2 \\ = \infty & \text{if } \delta \geq 2 \end{cases}$$

When $\delta \geq 1$, $M(0) = -\infty$ and hence $J = \infty$. When $\delta < 1$

$$J = \int_0^1 \frac{z^{1-\delta}}{1-\delta} dz < \infty$$

Combining the last two conclusions, we see that

if	I	J	0 is
$\delta \in [2, \infty)$	$= \infty$	$= \infty$	natural
$\delta \in [1, 2)$	$< \infty$	$= \infty$	absorbing
$\delta < 1$	$< \infty$	$< \infty$	regular

which makes sense, because as δ gets larger the fluctuations near the boundary are smaller.

7.11 Site frequency spectrum

In this section, we will calculate the site frequency spectrum for our diffusion processes, extending the result for the Moran model in Section 1.5. Special cases of the formula, as well as pictures similar to ones given in this section can be found in Wright's (1942) paper based on his Gibbs lecture to the American Mathematical Society. The general result can be found in formula (9.27) of Kimura's (1964) paper, which appeared in the first volume of the Journal of Applied Probability.

Theorem 7.20. *Under the infinite sites model if mutations occur at rate μ and $\theta = 4N\mu$ then the site frequency spectrum in the diffusion process is*

$$\theta f(y) \quad \text{where} \quad f(y) = \frac{\psi(0)}{m(y)} \cdot \frac{\phi(1) - \phi(y)}{\phi(1) - \phi(0)} \quad (7.59)$$

Proof. We begin by recalling the result for the Moran model. Suppose a mutation occurs at time $-t$ introducing a new allele and no further mutation occurs at that locus, which is the case in the infinite sites model. The probability that there are k copies at time 0 is given by the transition probability $p_t(1, k)$. If mutations occur at times of a Poisson process with rate λ and each mutation occurs at a different site then the number of mutants with k copies at time 0 is Poisson with mean

$$\lambda \int_0^\infty p_t(1, k) dt = \lambda G(1, k)$$

Turning to the diffusion process, suppose without loss of generality that $\phi(0) = 0$. We cannot introduce mutants at frequency 0, so we introduce them at frequency δ at rate $(\theta/2)(\psi(0)/\phi(\delta))$. Here $\theta/2 = 2N\mu$ is the rate at which mutations occur in the population, and the factor $\psi(0)/\phi(\delta)$ is chosen so that if $\delta < \epsilon$ then the mutations that reach frequency ϵ is a Poisson process with rate

$$\frac{\theta\psi(0)}{2\phi(\delta)} \cdot \frac{\phi(\delta)}{\phi(\epsilon)} = \frac{\theta\psi(0)}{2\phi(\epsilon)}$$

The factor $\psi(0)$ is included because the natural scale with $\phi(0) = 0$ is only specified up to a constant multiple.

Using the Green's function formula (7.34) now the number of mutants with frequency in $(y, y + dy)$ with $y > \delta$ is Poisson with mean

$$\frac{\theta\psi(0)}{2\phi(\delta)} \cdot 2 \frac{\phi(\delta)}{\phi(1)} \cdot (\phi(1) - \phi(y))m(y) dy \quad (7.60)$$

Letting $\delta \rightarrow 0$ gives the desired formula.

Remark. Most derivations of this result introduce mutations at frequency $1/2N$. In this approach, which as Sawyer and Hartl (1992) observe at the

top of page 1165, the rigorous justification that they give on pages 1172–1174 is somewhat painful. In the example of additive selection which they were considering, this can be done using our remark in Section 7.8 that the Green's function for the Moran model converges to that of the Wright-Fisher diffusion as the population size $N \rightarrow \infty$. It would not be much fun to do this for every model, so we have taken the approach of introducing mutations at frequency δ and then letting $\delta \rightarrow 0$, which easily gives the result for any diffusion.

Examples

In all of the diffusions we will consider, $a(y) = y(1 - y)$.

Example 7.37. No selection. In the neutral case, $\phi(x) = x$ and (7.59) becomes

$$f(y) = \frac{(1 - y)}{y(1 - y)} = \frac{1}{y}$$

which agrees with the result derived in Section 1.5.

Example 7.38. Additive selection. In this case $\psi(x) = e^{-2\gamma x}$, $\phi(x) = [1 - \exp(-2\gamma x)]/2\gamma$, and $m(x) = e^{2\gamma x}/x(1 - x)$ so (7.59) becomes

$$f(y) = \frac{e^{2\gamma y}}{y(1 - y)} \frac{e^{-2\gamma y} - e^{-2\gamma}}{1 - e^{-2\gamma}} = \frac{1}{y(1 - y)} \frac{1 - e^{-2\gamma(1-y)}}{1 - e^{-2\gamma}} \quad (7.61)$$

This formula can be found in slightly different notation on page 92 of Fisher (1930) and as formula (39) in Wright (1938). The next figure shows the site frequency spectrum for four values of γ . When $y \rightarrow 0$, $f(y) \sim 1/y$, while for $y \rightarrow 1$, we have $(1 - e^{-2\gamma(1-y)})/(1 - y) \rightarrow 2\gamma$ and hence $f(y) \rightarrow 2\gamma/(1 - e^{-2\gamma}) > 1$, so there is an excess of high frequency mutations.

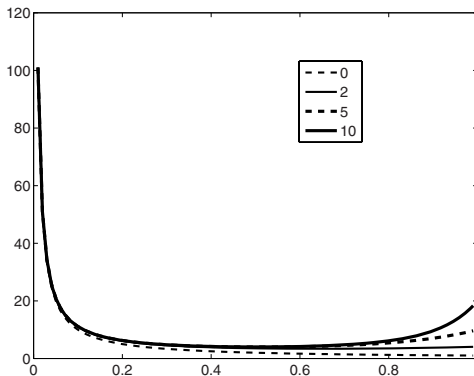


Fig. 7.13. Site frequency spectrum under directional selection.

Example 7.39. Symmetric balancing selection. In this case $\psi(x) = e^{-2\gamma x(1-x)}$ and $m(x) = e^{2\gamma x(1-x)}/x(1-x)$ so (7.59) becomes

$$f(y) = \frac{e^{2\gamma x(1-x)} \int_x^1 e^{-2\gamma y(1-y)} dy}{x(1-x) \int_0^1 e^{-2\gamma y(1-y)} dy} \quad (7.62)$$

As $x \rightarrow 1$, $\int_x^1 e^{-2\gamma y(1-y)} dy \sim 1-x$ so the density does not blow up there. If γ is large and x is away from the boundary then the ratio of the two integrals is close to $1/2$, and the curve is $\approx e^{2\gamma x(1-x)}/x(1-x)$, which reaches a maximum at $x = 1/2$.

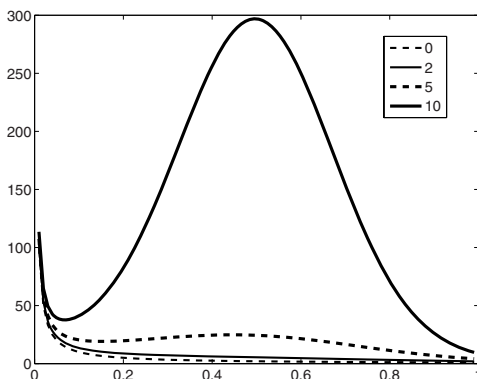


Fig. 7.14. Site frequency spectrum under balancing selection.

Fixation rate

By reasoning similar to that for (7.60), we see that if ϕ is chosen with $\phi(0) = 0$ the rate at which new mutations become fixed (when time is scaled by $2N$ generations) is

$$\frac{\theta\psi(0)}{2\phi(\delta)} \frac{\phi(\delta)}{\phi(1)} = \frac{\theta\psi(0)}{2\phi(1)}$$

In the case of additive selection $\psi(x) = e^{-2\gamma x}$ and $\phi(x) = [1 - \exp(-2\gamma x)]/2\gamma$ so this is

$$\frac{\theta}{2} \cdot \frac{2\gamma}{1 - e^{-2\gamma}} \quad (7.63)$$

7.11.1 Poisson random field model

To set up the problem, we quote from page 1166 of Sawyer and Hartl (1992): “Suppose that two species diverged $t_{div}N_e$ generations ago, and that both have

the same haploid effective population size N_e . Assume that the mutation rate for silent sites in the coding region of a particular gene is μ_s per gene per generation, and that the mutation rate for nonlethal replacement mutations is μ_r per gene per generation. Assume further that (i) all new replacement mutations bestow equal fitness $w = 1 + \gamma/N_e$, (ii) each new mutation since the divergence of species occurred at a different site (in particular, the gene has not been saturated by mutations), and (iii) different sites remain in linkage equilibrium.”

In this case (7.63) and (7.61) give us Table 1 of Sawyer and Hartl (1992)

	Fixation rate	Mutant frequency spectrum
Neutral	μ_s	$2\mu_s \frac{dx}{x}$
$\gamma \neq 0$	$\mu_r \frac{2\gamma}{1 - e^{-2\gamma}}$	$\frac{2\mu_r}{y(1-y)} \frac{1 - e^{-2\gamma(1-y)}}{1 - e^{-2\gamma}}$

To make the connection note that their μ_r and μ_s are our $2N\mu$. To make it easier to compare with their paper we will keep their notation. The expected number of fixed differences between the two species are

$$2\mu_s t_{div} \quad \text{and} \quad 2\mu_r \frac{2\gamma}{1 - e^{-2\gamma}}$$

for silent and replacement sites, respectively.

McDonald-Kreitman tables

Now suppose we have aligned DNA sequences from m chromosomes from the first species and n chromosomes from the second species. An allele with frequency x will be polymorphic in a sample of size m with probability $1 - x^m - (1 - x)^m$, so the expected number of silent polymorphic sites in a sample of size m is

$$2\mu_s \int_0^1 \frac{1 - x^m - (1 - x)^m}{x} dx = 2\mu_s \sum_{k=1}^{m-1} \frac{1}{k}$$

Writing $L(m) = \sum_{k=1}^{m-1} 1/k \approx \log m$, the number of silent polymorphic sites in both samples together is then

$$2\mu_s(L(m) + L(n)) \tag{7.64}$$

A silent site will look like a fixed difference in species 1 in the comparison of the two samples if it is fixed in the population or if it by chance occurs in all m sampled individuals, so the expected value is

$$\mu_s t_{div} + \int_0^1 2\mu_s x^m \frac{dx}{x} = \mu_s \left(t_{div} + \frac{2}{m} \right)$$

Thus the expected number of silent fixed differences is

$$2\mu_s \left(t_{div} + \frac{1}{m} + \frac{1}{n} \right) \tag{7.65}$$

By the same reasoning the number of polymorphic replacement sites in a sample is

$$2\mu_r(H(m) + H(n)) \tag{7.66}$$

where

$$H(m) = \int_0^1 \frac{1 - x^m - (1 - x)^m}{x(1 - x)} \frac{1 - e^{-2\gamma(1-x)}}{1 - e^{-2\gamma}} dx$$

and the number of replacement fixed differences has mean

$$2\mu_r \frac{2\gamma}{1 - e^{-2\gamma}} (t_{div} + G(m) + G(n)) \tag{7.67}$$

where

$$G(m) = \int_0^1 x^{m-1} \frac{1 - e^{-2\gamma(1-x)}}{2\gamma(1 - x)} dx$$

Since $e^{-y} \geq 1 - y$ and hence $(1 - e^{-y})/y \leq 1$ for $y > 0$, $G(m) \leq 1/m$ for $\gamma > 0$.

Our formulas give the expected value of the four entries in the McDonald-Krietman table.

	Divergence	Polymorphism
Silent	(7.65)	(7.64)
Replacement	(7.67)	(7.66)

Bayesian estimation

The number of mutations in one locus typically does not give us enough information to get good estimates of the parameters, so it is natural to combine the information from many loci. To do this we will follow the approach of Bustamante et al. (2002) and Sawyer et al. (2003). Changing to their notation we let $\theta_s = 4N\mu_s$ and $\theta_a = 4N\mu_r$ and denote the entries in the DPRS table by

	Divergence	Polymorphism
Silent	K_s	S_s
Replacement	K_a	S_a

The theoretical expectations for any single DPRS table include four parameters, θ_s , θ_a , γ , and the divergence time t , and contain four observations: K_s , S_s , K_a , and S_a hence there is no meaningful opportunity for model fitting. However, the divergence time is a shared parameter among all sequences. The basic idea of Bayesian analysis is to treat the parameters in a model as random variables with some underlying prior distribution. In Bustamante et al. (2002), it was assumed that for each coding sequence γ was a fixed constant but that across loci the distribution of γ was given by a normal with mean μ and standard deviation σ . The other prior distributions are $q(t)$ is uniform, $p(\theta)$ is gamma, $h(\sigma)$ is such that $1/\sigma^2$ is gamma, and $g(\mu|\sigma)$ is normal.

The posterior distribution $\pi(\gamma, t, \theta, \mu, \sigma)$ is analytically intractable but can be computed by Markov chain Monte Carlo. That is, by simulation of a Markov chain defined in such a way that the stationary distribution is precisely π . One simple method for doing this is the Metropolis algorithm in which a trial value for the new parameter is used to replace the old if the ratio of the posterior probabilities for the trial and the present values is greater than a uniform random number in $[0, 1]$. For more details about the MCMC method see page 533 of Bustamante et al. (2002).

Sawyer et al. (2003) modified the “fixed-effects” model described above to be a random-effects model so that for the i th coding sequence, the selection coefficient for a new mutation is normal with mean γ_i and standard deviation σ_w . Here σ_w is a global parameter that applies to all loci and has a uniform prior distribution.

Sawyer et al. (2003) studied a set of 72 *D. simulans* coding sequences from GenBank, which had sample sizes ranging from 4 to 70 with an average of 10.5. Nucleotide divergence between *D. simulans* and *D. melanogaster* was inferred from the reference sequence for *D. melanogaster*, see Adams (2000). In applying the random-effects model to the DPRS data, they initially found that the Markov chain did not converge, or did so excessively slowly. The output for various runs suggested that the main reason for poor convergence was that values of θ_r could be balanced off by γ . That is, an excess of replacement mutations can be caused either by a stronger intensity of positive selection or a higher mutation rate.

From runs of the fixed-effects model, they noticed that about 80% of the coding sequences had values of $\theta_r/2\theta_s$ near 1/4, or more precisely about 0.28, so they modified the model to include a new parameter $Q = \theta_r/2\theta_s$ with a gamma prior distribution. Among the 72 genes, 14 were excluded because $\theta_r/2\theta_s > 0.28$ and two additional genes were excluded because they appeared to be spurious for other reasons. The list of genes omitted include eight male accessory gland proteins.

For the random effects model they found that the distribution of the γ_i had mean -7.3 and standard deviation $\sigma_b = 5.69$, while the within locus standard deviation was $\sigma_w = 6.79$. Most of the mean selection intensities for the 56 genes were negative but many had 95% credible intervals that overlapped 0. The fraction of beneficial new mutations ranged from 1% for *Pgm* to 62% for *Rel* with an outlier at 90% for *mei-218*. The average for all loci was 19.4%. Among the replacement polymorphisms in the data, an average of 46.9% were estimated to be beneficial. For the Y-linked gene *kl-5* the estimated average selection intensity was -0.38 . All others were positive and ranged from 2.1 for *vermillion* to 9.4 for *Rel*, with an overall mean, excluding *kl-5* of 5.1.

For an application of these methods to a large number of genes in the human genome, see Bustamante et al. (2005).

7.12 Fluctuating selection

Two mechanisms by which evolution can occur are the adaptive processes of natural selection and the neutral processes of genetic drift. Which of these is the principal force in the evolution of a population has been one of the central issues in evolutionary biology. An early exchange in this debate was over the changes in the frequencies of a color polymorphism in a population of the scarlet tiger moth *Callimorpha (Panaxia) dominula* near Oxford, England. Fisher and Ford (1947) argued that the population size was too large for the changes in frequencies to be due to random drift, and were caused by fluctuating selection. Wright (1948) replied by arguing that multiple factors could affect a population, and that the effective population might be much smaller than the census population size. A publicized debate ensued, see Fisher and Ford (1950), and Wright (1951).

Kimura (1954, 1962) and Ohta (1972) studied the question mathematically, but did not find the correct diffusion approximation. A little later Gillespie (1973) and Jensen (1973) did. For more on the early history see Felsenstein (1976). We will follow Karlin and Levikson (1974) and consider a model in which the fitness of A in generation n is $1 + \sigma_n$ and the fitness of a is $1 + \tau_n$ where σ_n, τ_n are independent and identically distributed. Dropping the subscripts to simplify the formulas, we let

$$\begin{aligned}\alpha &= 2N[E(\sigma - \tau) - E(\sigma^2 - \tau^2)/2 + E(\sigma - \tau)^2/2] \\ \beta &= 2NE(\sigma - \tau)^2\end{aligned}$$

Theorem 7.21. *The diffusion approximation for the Karlin-Levikson model has coefficients:*

$$b(x) = x(1-x)(\alpha - \beta x) \quad a(x) = x(1-x)[1 + \beta x(1-x)] \quad (7.68)$$

The drift looks like balancing selection, but the variance has an additional term.

Proof. To derive the diffusion approximation, note that reasoning as in Section 6.2, the change in frequency in one generation is

$$\begin{aligned}\frac{x(1+\sigma)}{x(1+\sigma) + (1-x)(1+\tau)} - x &= \frac{x(1+\sigma) - x - \sigma x^2 - \tau x(1-x)}{1 + \sigma x + \tau(1-x)} \\ &= \frac{(\sigma - \tau)x(1-x)}{1 + \sigma x + \tau(1-x)} \approx (\sigma - \tau)x(1-x)[1 - \sigma x - \tau(1-x)]\end{aligned}$$

Writing $x = 1/2 - (1/2 - x)$ and $1 - x = 1/2 + 1/2 - x$, the above is

$$\begin{aligned}&= (\sigma - \tau)x(1-x)[1 - (\sigma + \tau)/2 + (\sigma - \tau)(1/2 - x)] \\ &= x(1-x)[(\sigma - \tau) - (\sigma^2 - \tau^2)/2 + (\sigma - \tau)^2(1/2 - x)]\end{aligned}$$

Taking expected value and speeding up time by a factor of $2N$ the drift coefficient is

$$b(x) = x(1-x)(2N)[E(\sigma - \tau) - E(\sigma^2 - \tau^2)/2 + E(\sigma - \tau)^2(1/2 - x)]$$

To compute the variance, let ΔX be the change in frequency and Y be the environment.

$$\text{var}(\Delta X) = E \text{var}(\Delta X|Y) + \text{var}(E(\Delta X|Y))$$

To evaluate the first term, recall that the variance of $\text{Binomial}(2N, p)$ is $2Np(1-p)$ and the allele frequencies by $O(1/N)$ in one generation so

$$\text{var}(\Delta X|Y) \approx \frac{x(1-x)}{2N}$$

As we computed above

$$E(\Delta X|Y) = \frac{(\sigma - \tau)x(1-x)}{1 + \sigma x + \tau(1-x)}$$

Since $\sigma x, \tau(1-x) \ll 1$, we can drop these terms from the denominator:

$$\text{var}(E(\Delta X|Y)) = x^2(1-x)^2 E(\sigma - \tau)^2$$

Adding the two results and speeding up time by a factor of $2N$ gives

$$a(x) = x(1-x) + x^2(1-x)^2 2NE(\sigma - \tau)^2$$

and completes the proof. \square

Remark. Takahata, Ishii, Matsuda (1975) considered a Wright-Fisher diffusion with varying selection

$$\frac{1}{4N}x(1-x)\frac{d^2}{dx^2} + s(t)x(1-x)\frac{d}{dx}$$

They let $\bar{s} = Es(t)$ and $V = \int_0^\infty E([s(t) - \bar{s}][s(0) - \bar{s}]) dt$, and found that in the diffusion approximation

$$a(x) = \frac{1}{2N}x(1-x) + 2Vx^2(1-x)^2$$

$$b(x) = \bar{s}x(1-x) + Vx(1-x)(1-2x)$$

To connect with the Karlin-Levikson result, suppose $E(\sigma^2 - \tau^2) = 0$, let $\bar{s} = E(\sigma - \tau)$, and note that in discrete time $V = E(\sigma - \tau)^2/2$. This suggests that if $\bar{s} = E(\sigma - \tau) = 0$ and we have (σ_n, τ_n) that are correlated in time all we do is replace $E(\sigma - \tau)^2/2$ by

$$\sum_{n=0}^{\infty} E[(\sigma_0 - \tau_0)(\sigma_n - \tau_n)].$$

For the rest of the section we will only consider the special case that is closely related to the model of Takahata, Ishii, and Matsuda (1975).

Theorem 7.22. Consider the Karlin-Levikson model with $E(\sigma - \tau) = 0$ and $E(\sigma^2 - \tau^2) = 0$. The derivative of the natural scale

$$\psi(y) = \frac{1}{y(1-y) + 1/\beta}$$

The speed measure is exactly the same as for the neutral case

$$m(y) = \frac{1}{y(1-y)}$$

Proof. Since $a(y) = y(1-y)(1 + \beta y(1-y))$, the second formula follows from the first. To compute ψ , we begin by noting

$$\frac{-2b(x)}{a(x)} = \frac{-2[\alpha - \beta x]}{1 + x(1-x)\beta}$$

To find the roots of the quadratic in the denominator, we write it as $x^2 - x - 1/\beta = 0$ and solve to find roots $r_1 < 0 < 1 < r_2$ given by

$$r_i = \frac{1 \pm \sqrt{1 + 4/\beta}}{2}$$

Note that the two roots are symmetric about $1/2$. To evaluate the integral we write

$$\frac{-2b(x)}{a(x)} = \frac{-2[\alpha - \beta x]}{1 + x(1-x)\beta} = \frac{-2[\alpha/\beta - x]}{1/\beta + x(1-x)} = \frac{C}{x - r_1} + \frac{D}{r_2 - x}$$

To find the constants we solve $-C + D = 2$ and $Cr_2 - Dr_1 = -2\alpha/\beta$ to find

$$C = \frac{2r_1 - 2\alpha/\beta}{r_2 - r_1} \quad D = \frac{2r_2 - 2\alpha/\beta}{r_2 - r_1}$$

which, as the reader can easily check, satisfies the two equations. Integrating

$$\int^y \frac{C}{x - r_1} + \frac{D}{r_2 - x} dx = C \log(y - r_1) - D \log(r_2 - y)$$

so we have

$$\psi(y) = \exp\left(\int^y \frac{-2b(x)}{a(x)}\right) = (y - r_1)^C (r_2 - y)^{-D}$$

Consider now the special case in which σ and τ have the same distribution so $E(\sigma - \tau) = 0$, $E(\sigma^2 - \tau^2) = 0$, and hence $\alpha = \beta/2$.

$$C = \frac{2r_1 - 2\alpha/\beta}{r_2 - r_1} = \frac{2(r_1 - 1/2)}{r_2 - r_1} = -1$$

$$D = \frac{2r_2 - 2\alpha/\beta}{r_2 - r_1} = \frac{2(r_2 - 1/2)}{r_2 - r_1} = 1$$

and we have the very nice formula

$$\psi(y) = (y - r_1)^{-1}(r_2 - y)^{-1} = \frac{1}{y(1 - y) + 1/\beta} \quad \square$$

Karlin and Levikson (1974) find $\psi(y) = [1 + \beta y(1 - y)]^{-1}$ on their page 402, but this agrees with our computation since the solution of $\psi'(y) = -2b(y)\psi(y)/a(y)$ is only determined up to a constant multiple. To make it easier to compare with their formulas, for the rest of the section we will use

$$\psi(y) = \frac{1}{\beta y(1 - y) + 1} = \beta^{-1}(y - r_1)^{-1}(r_2 - y)^{-1} \quad (7.69)$$

Theorem 7.23. *Let $r_1 < r_2$ be the roots $(1 \pm \sqrt{1 + 4/\beta})/2$. Under the assumptions of Theorem 7.22, the probability of fixation starting from frequency x is*

$$\frac{1}{2} + \frac{\log \left[\frac{x - r_1}{r_2 - x} \right]}{2 \log [r_2 / (-r_1)]}$$

This is (8) in Jensen (1973). As $\beta \rightarrow \infty$, $r_1 \rightarrow 0$ and $r_2 \rightarrow 1$ so $\phi(x) \rightarrow 1/2$. The next graph shows the hitting probabilities for $\beta = 0, 10, 40$.

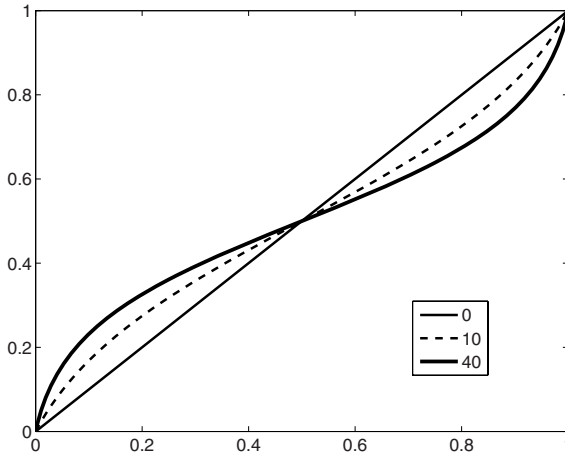


Fig. 7.15. Hitting probabilities for fluctuating selection

Proof. To compute the natural scale ϕ , we integrate to find

$$\begin{aligned} \phi(x) &= \beta^{-1} \int_0^x (y - r_1)^{-1}(r_2 - y)^{-1} dy \\ &= \frac{1}{\beta(r_2 - r_1)} \int_0^x \frac{1}{y - r_1} + \frac{1}{r_2 - y} dy \\ &= \frac{1}{\sqrt{\beta^2 + 4\beta}} [\log(x - r_1) - \log(-r_1) - \log(r_2 - x) + \log(r_2)] \end{aligned} \quad (7.70)$$

This is close to but not exactly the same as Karlin and Levikson (1974). Their roots are $\lambda_2 = r_1$ and $\lambda_1 = r_2$, and they write $w = \beta/2$, so their constant has 2β instead of 4β under the square root.

Since $\phi(0) = 0$, the probability of fixation starting from frequency x is

$$\phi(x)/\phi(1) = \log \left[\frac{x - r_1}{-r_1} \cdot \frac{r_2}{r_2 - x} \right] \bigg/ \log \left[\frac{1 - r_1}{-r_1} \cdot \frac{r_2}{r_2 - 1} \right] \quad (7.71)$$

$r_2 - 1/2 = 1/2 - r_1$ and $r_2 - 1 = -r_1$ so $\phi(1/2)/\phi(1) = 1/2$ and we can write above as

$$\frac{1}{2} + \frac{\phi(x) - \phi(1/2)}{\phi(1)} = \frac{1}{2} + \frac{\log \left[\frac{x - r_1}{r_2 - x} \right]}{2 \log [r_2 / (-r_1)]} \quad \square$$

Theorem 7.24. Let $\tau = T_0 \wedge T_1$ be the time until one allele is lost. Under the assumptions of Theorem 7.22

$$E_x \tau = \begin{cases} \int_0^x \frac{2}{1 + \beta y(1-y)} \log \left(\frac{1-y}{y} \right) dy & \text{when } x \leq 1/2 \\ \int_x^1 \frac{2}{1 + \beta y(1-y)} \log \left(\frac{y}{1-y} \right) dy & \text{when } x \geq 1/2 \end{cases}$$

Note that in each case the log is nonnegative throughout the range of integration, so $E_x \tau$ is a decreasing function of β . This result, which is on page 402 of Karlin and Levikson (1974) is somewhat surprising since (7.68) shows that the diffusion has a drift toward $1/2$, which will encourage it to spend more time at intermediate values. However, this effect is counteracted by the increase in $a(x)$.

Proof. Since $m(y) = 1/y(1-y)$ and $\phi(0) = 0$, the Green's function $G(x, y)$ from (7.34) is

$$\begin{aligned} & 2 \frac{\phi(x)}{\phi(1)} \cdot \frac{\phi(1) - \phi(y)}{y(1-y)} & x \leq y \\ & 2 \frac{\phi(1) - \phi(x)}{\phi(1)} \cdot \frac{\phi(y)}{y(1-y)} & y \leq x \end{aligned}$$

The expected time to fixation is

$$E_x \tau = 2 \frac{\phi(x)}{\phi(1)} \int_x^1 \frac{\phi(1) - \phi(y)}{y(1-y)} dy + 2 \frac{\phi(1) - \phi(x)}{\phi(1)} \int_0^x \frac{\phi(y)}{y(1-y)} dy$$

Since $1/(1-y)y = 1/(1-y) + 1/y$ has antiderivative $-\log(1-y) + \log(y)$, integrating by parts gives

$$\begin{aligned}
 &= 2 \frac{\phi(x)}{\phi(1)} (\phi(1) - \phi(y)) \cdot (-\log(1-y) + \log(y)) \Big|_x^1 \\
 &\quad + 2 \frac{\phi(x)}{\phi(1)} \int_x^1 \psi(y) \log\left(\frac{y}{1-y}\right) dy \\
 &\quad + 2 \frac{\phi(1) - \phi(x)}{\phi(1)} \phi(y) \cdot (-\log(1-y) + \log(y)) \Big|_0^x \\
 &\quad - 2 \frac{\phi(1) - \phi(x)}{\phi(1)} \int_0^x \psi(y) \log\left(\frac{y}{1-y}\right) dy
 \end{aligned}$$

Since $\phi(1) - \phi(y) \sim \phi'(1)(1-y)$ as $y \rightarrow 1$ and $(1-y)\log(1-y) \rightarrow 0$ as $y \rightarrow 1$, evaluating the first term at 1 gives 0. Similarly $\phi(y) \sim \phi'(0)y$ as $y \rightarrow 0$ and $y\log(y) \rightarrow 0$ as $y \rightarrow 0$, so evaluating the third term at 0 gives 0. Evaluating the first term at x cancels with evaluating the third at x , so the above

$$= 2 \frac{\phi(x)}{\phi(1)} \int_x^1 \psi(y) \log\left(\frac{y}{1-y}\right) dy - 2 \frac{\phi(1) - \phi(x)}{\phi(1)} \int_0^x \psi(y) \log\left(\frac{y}{1-y}\right) dy$$

Adding and subtracting $2(\phi(x)/\phi(1)) \int_0^x$, then flipping the fraction inside the log to get rid of the minus sign, the above

$$= 2 \frac{\phi(x)}{\phi(1)} \int_0^1 \psi(y) \log\left(\frac{y}{1-y}\right) dy + \int_0^x 2\psi(y) \log\left(\frac{1-y}{y}\right) dy$$

$\psi(y)$ is symmetric about $1/2$ and $\log(y/(1-y)) = \log(y) - \log(1-y)$ is antisymmetric about $1/2$, so the first integral vanishes, and

$$E_x \tau = \int_0^x \frac{2}{1 + \beta y(1-y)} \log\left(\frac{1-y}{y}\right) dy$$

When $x \geq 1/2$ we can use the fact that the total integral is 0 to write

$$E_x \tau = \int_x^1 \frac{2}{1 + \beta y(1-y)} \log\left(\frac{y}{1-y}\right) dy \quad \square$$

Using Kimura’s formula (7.59), we have

Theorem 7.25. *Under the assumptions of Theorem 7.22, the site frequency spectrum is*

$$\frac{\theta}{y(1-y)} \cdot \log\left(\frac{1-r_1}{y-r_1} \cdot \frac{r_2-y}{r_2-1}\right) \Big/ \log\left(\frac{1-r_1}{-r_1} \cdot \frac{r_2}{r_2-1}\right)$$

Proof. Using either formula for $\psi(y)$

$$\frac{\psi(0)}{\psi(y)a(y)} = \frac{1}{y(1-y)}$$

Using (7.71), shows that $(\phi(1) - \phi(y))/(\phi(1) - \phi(0)) =$ the second factor.

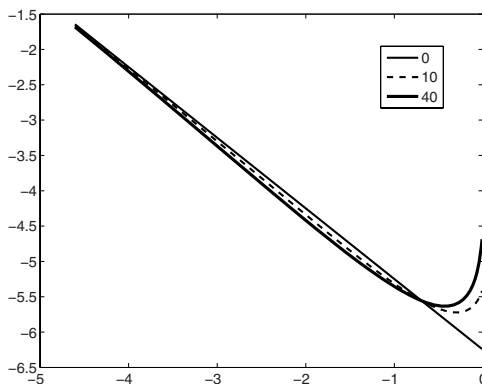


Fig. 7.16. Log-log plot of the site frequency spectrum for fluctuating selection

Figure 7.16 shows the site frequency spectrum for $\beta = 0, 10, 40$. To make the differences more visible we have done a log-log plot. In the presence of fluctuating selection high frequency derived alleles (y near 1) are overrepresented, and intermediate frequency alleles are underrepresented with respect to the neutral case. Somewhat remarkably,

Theorem 7.26. *The integral of the site frequency spectrum does not depend on β .*

Proof. It follows from Theorem 7.23 that

$$g(\beta, y) = \frac{\partial}{\partial \beta} \frac{\phi(1) - \phi(y)}{\phi(1) - \phi(0)}$$

has $g(\beta, y) = -g(\beta, 1 - y)$. From this it follows that

$$\frac{\partial}{\partial \beta} \int_0^1 \frac{\theta}{y(1-y)} \frac{\phi(1) - \phi(y)}{\phi(1) - \phi(0)} dy = \int_0^1 \frac{\theta}{y(1-y)} g(\beta, y) dy = 0 \quad \square$$

Parameter estimation

Most studies of fluctuating selection have based their inferences on time series data for allele frequencies. See Mueller et al (1985), Lynch (1987), Cook and Jones (1996), and O'Hara (2005). Recently, Huerta-Sanchez, Durrett, and Bustamante (2007) have used the Poisson random field framework to develop an alternative approach that uses DNA polymorphism data from a sample of individuals collected at a single point in time. To do this they used methods described in Bustamante et al. (2001), which we will begin by describing in general. Let $f(y, \beta)$ be the site frequency spectrum. Since this represents the

distribution of mutation frequencies at any time, the probability of finding i mutant alleles in a sample of size n

$$F_n(k, \beta) = \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} f(y, \beta) dy$$

Our definition of $f(y, \beta)$ is 1/2 the usual one so there is no factor of 2, as in (2) of Bustamante et al. (2001).

Let y_i be the number of sites at which there are i mutant alleles. In the Poisson random field framework, different sites are independent so the likelihood is given by

$$L(\theta, \beta) = \prod_{i=1}^{n-1} \exp(-\theta F_n(i, \beta)) \frac{(\theta F_n(i, \beta))^{y_i}}{y_i!}$$

Therefore, the log likelihood function (dropping the term $\log(y_i!)$ which is independent of the parameters) is

$$\ell(\theta, \beta) = \sum_{i=1}^{n-1} -\theta F_n(i, \beta) + y_i \log(\theta F_n(i, \beta))$$

Differentiating with respect to θ we see that

$$\frac{\partial}{\partial \beta} \log L(y, \beta) = - \sum_{i=1}^n F_n(i, \beta) + \frac{y_i}{\theta}$$

so for fixed β the maximum likelihood estimate of θ is

$$\hat{\theta}(\beta) = S_n / \sum_{i=1}^n F_n(i, \beta)$$

which is a generalization of Watterson's estimate.

Given the last result we can work with the profile likelihood

$$\ell^*(\beta) = L(\hat{\theta}(\beta), \beta)$$

which can be maximized numerically using standard optimization techniques such as Newton-Raphson iteration. In the current example, that task simplifies because Theorem 7.26 implies

Theorem 7.27. $ES_n = \sum_{i=1}^n F_n(i, \beta)$ does not depend on β .

Proof. Let $h(k, y) = \binom{n}{k} y^k (1-y)^{n-k} + \binom{n}{n-k} y^{n-k} (1-y)^k$. Since $h_n(k, y)$ is symmetric about 1/2,

$$\frac{\partial}{\partial \beta} [F_n(k, \beta) + F_n(n-k, \beta)] = 0$$

Summing from $k = 1$ to $n - 1$ now gives the desired result. \square

To perform the optimization, we have to compute the first and second derivatives of the log likelihood with respect to its parameters. We have nothing insightful to say about these details so we refer the reader to Bustamante et al. (2001) or Huerta-Sanchez, Durrett, and Bustamante (2007) for details, and simulation results which show the performance of the estimators. For another approach to fitting fluctuating selection models to data see Mutsonen and Lässig (2007). Using data for 271 loci in 12 *Drosophila melanogaster* and a *D. simulans* sequence, they find strong support ($p < 10^{-17}$) for time dependent selection.

Before leaving the topic of fluctuating selection, we must mention the work of Gillespie. To quote the preface of his 1991 book *The Causes of Molecular Evolution*: “If we are to propose that molecular evolution is due to the action of natural selection, we need a mathematical theory to demonstrate that the dynamics of selection are compatible with the observations of molecular variation. It is my conviction that the only viable model of selection is one based on temporal and spatial fluctuations in the environment. The mathematics of selection in a random environment have never been systematically developed or brought to a point where they serve as a model of molecular evolution. Both situations will be remedied in Chapter 4. Unfortunately, the mathematics are very difficult. Yet, if molecular evolution is in response to a changing environment, then this is the sort of mathematical challenge we must be willing to face. Chapter 4 is littered with unresolved problems that should prove of interest to those with a mathematical bent.” In addition to the source just cited the reader should consult his more recent papers on the SAS-CFF model (stochastic additive scale-concave fitness function).