Chapter 3
One Dimensional Diffusions

3.1 Diffusions

In this chapter we are going to remind ourselves of some useful facts about one-dimensional diffusions. It is not an exhaustive study. Excellent references for this material are Karlin and Taylor (1981) and Knight (1981). We start in a fairly general setting.

Definition 3.1 (One-dimensional diffusion). A one-dimensional diffusion process $\{X_t\}_{t\geq 0}$ is a strong Markov process on $\mathbb{R}$ which traces out a continuous path as time evolves.

At any instant in time, X_t is a continuous random variable but also any realisation of $\{X_t\}_{t\geq 0}$ is a continuous function of time. Its range need not be the whole of $\mathbb{R}$ and indeed for the most part we'll be interested in diffusions on $(0,1)$ (possibly union either or both of the endpoints $\{0,1\}$). For the time being let us take the state space to be an interval (a,b) (possibly infinite), again possibly union one or both of the endpoints $\{a,b\}$.[1]

Remark 3.2. In our discussion of diffusions, we shall generally use x to denote a generic point in (a,b), but for consistency with our previous notation we shall use p for points in $(0,1)$ when discussing models for allele frequencies in genetics.

The generator of the diffusion takes the form

$$\mathscr{L}f(x) = \frac{1}{2}\sigma^2(x)\frac{d^2 f}{dx^2}(x) + \mu(x)\frac{df}{dx}(x). \tag{3.1}$$

Evidently for this to be defined f must be twice differentiable on (a,b). Depending on the behaviour of the diffusion close to the boundaries of its domain, f may also have to satisfy boundary conditions at a and b. We'll specify these precisely in Theorem 3.17, but for now assume that if we apply the generator to a function, f,

---

[1] We wish to include all *accessible* endpoints, defined in Definition 3.16, in the state space.

A. Etheridge, *Some Mathematical Models from Population Genetics*, Lecture Notes in Mathematics 2012, DOI 10.1007/978-3-642-16632-7_3,

then f is in the domain $\mathscr{D}(\mathscr{L})$ of the generator $\mathscr{L}$. To avoid pathologies, we make the following assumptions:

1. For any compact interval $I \subset (a,b)$, there exists $\varepsilon > 0$ such that $\sigma^2(x) > \varepsilon$ for all $x \in I$.
2. The coefficients $\mu(x)$ and $\sigma^2(x)$ are continuous functions of $x \in (a,b)$.

Note that (crucially for applications in genetics) we *do* allow $\sigma^2(x)$ to vanish at the boundary points $\{a,b\}$.

Let us write $\Delta_h X(t) = X_{t+h} - X_t$, then taking $f_1(x) = x$ in the generator (and using the Markov property) we see that

$$\mathscr{L} f_1(X_t) = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}[\Delta_h X(t)|X_t] = \mu(X_t)$$

and so

$$\mathbb{E}[\Delta_h X(t)|X_t] = h\mu(X_t) + o(h) \qquad \text{as } h \downarrow 0. \tag{3.2}$$

Now observe that we can write $(X_{t+h} - X_t)^2 = X_{t+h}^2 - X_t^2 - 2X_t(X_{t+h} - X_t)$ and so, taking $f_2(x) = x^2$,

$$\mathscr{L} f_2(X_t) - 2X_t \mathscr{L} f_1(X_t) = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left[(\Delta_h X(t))^2 \middle| X_t\right] = \sigma^2(X_t),$$

which yields

$$\mathbb{E}\left[(\Delta_h X(t))^2 \middle| X_t\right] = h\sigma^2(X_t) + o(h) \qquad \text{as } h \downarrow 0. \tag{3.3}$$

This motivates the following terminology.

Definition 3.3 (Infinitesimal drift and variance). The coefficients $\mu(x)$ and $\sigma^2(x)$ are called the (infinitesimal) *drift* and *variance* of the diffusion $\{X_t\}_{t\geq 0}$.

In fact, if a strong Markov process $\{X_t\}_{t\geq 0}$ is càdlàg (that is its paths are right continuous with left limits) and satisfies (3.2), (3.3) and the additional condition

$$\lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left[|\Delta_h X(t)|^p | X_t = x\right] = 0 \qquad \text{for some } p > 2,$$

where the convergence is uniform in (x,t) on compact subsets of $(a,b) \times \mathbb{R}_+$, then $\{X_t\}_{t\geq 0}$ is necessarily a diffusion (see Karlin and Taylor (1981), Sect. 15.1, Lemma 1.1).

The canonical example of a one-dimensional diffusion is one-dimensional Brownian motion which has generator

$$\mathscr{L}_B f(x) = \frac{1}{2}\frac{d^2 f}{dx^2}(x).$$

It has transition density function

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right).$$

In other words, if $\{W_t\}_{t\geq 0}$ denotes Brownian motion then

$$\mathbb{P}[W_t \in A | X_0 = x] \equiv \mathbb{P}_x[W_t \in A] = \int_A \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right) dy$$

for any subset $A \subseteq \mathbb{R}$.

Brownian motion can be thought of as a building block from which other one-dimensional diffusions are constructed. One approach is to observe that the diffusion corresponding to the generator $\mathscr{L}$ of (3.1) can be expressed as the solution of a stochastic differential equation driven by Brownian motion (with appropriate boundary conditions)

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t. \tag{3.4}$$

Remark 3.4 (Mathematical drift versus genetic drift). We have already encountered the Wright–Fisher diffusion several times, corresponding to the solution of the stochastic differential equation

$$dp_t = (\nu_2(1-p_t) - \nu_1 p_t)dt + \sqrt{p_t(1-p_t)}dW_t.$$

It is an unfortunate accident of history that the standard terminology for the stochastic term (driven by Brownian motion) is *genetic drift*, whereas to a mathematician it is the *deterministic* mutation term that corresponds to drift.

We can see from (3.4) and Itô's Lemma that

$$f(X_t) - \int_0^t \mathscr{L}f(X_s)ds \tag{3.5}$$

is a martingale for all $f \in \mathscr{D}(\mathscr{L})$. Stroock and Varadhan (1979) use this martingale property as a way of characterising the Markov process associated with a given generator. Solving the corresponding martingale problem requires all the expressions in (3.5) to be martingales.

Definition 3.5. We shall say that the *martingale problem* for μ, σ is well-posed if for each $x \in (a,b)$ there is a unique probability measure $\mathbb{P}_x$ on the continuous functions from $[0,1]$ to $\mathbb{R}$ (with the σ-field generated by the coordinate maps) such that $\mathbb{P}[X_0 = x] = 1$ and the quantities in (3.5) are martingales.

We refer to Stroock and Varadhan (1979) for a thorough introduction to martingale problems. In particular, the martingale problem for a diffusion process on $\mathbb{R}$ is certainly well-posed if μ and σ^2 are bounded measurable functions with σ^2 uniformly strictly positive. (This last condition is violated by the Wright–Fisher diffusion, but it turns out that the martingale problem is nonetheless well-posed, see Sect. 3.2 for references.)

Our approach to constructing one-dimensional diffusions from Brownian motion will not be via stochastic differential equations or martingale problems, but rather through the theory of speed and scale. Before introducing that, let's fill a gap that we left at the end of Chap. 2.

3.2 Convergence to Diffusions

In the case where the martingale problem of Definition 3.5 is well-posed, Stroock and Varadhan (1979) provide elementary criteria for convergence of discrete or continuous time Markov chains to a diffusion, which we record in Theorem 3.6 below. We follow Sect. 8.7 of Durrett (1996) which treats discrete and continuous time together. We need some notation. Suppose that we have a series of discrete time Markov chains, $\{Y^h_{nh}\}_{n\in\mathbb{N}}$ say, indexed by h and taking values in $S_h \subseteq \mathbb{R}$, with the chain with index h jumping at time intervals of length h. Write

$$\mathbb{P}[Y^h_{(n+1)h} \in A | Y^h_{nh} = x] = \Pi_h(x,A), \qquad \text{for } x \in S_h, A \subset \mathbb{R}.$$

(When we write $A \subseteq \mathbb{R}$ we implicitly assume that A is a Borel subset of $\mathbb{R}$.) We define $X^h_t = Y^h_{h[t/h]}$, where $[u]$ denotes the integer part of $u \in \mathbb{R}$. In other words we extend Y^h to all times $t \geq 0$ by setting it to be constant on time intervals $[nh,(n+1)h)$.

Now suppose that we have continuous time chains $\{X^h_t\}_{t\geq 0}$ taking values in $S_h \subseteq \mathbb{R}$. In place of the sequence of transition probabilities Π_h for the discrete time chain, we have a sequence of transition rates:

$$\frac{d}{dt}\mathbb{P}[X^h_t \in A | X^h_0 = x]\bigg|_{t=0} = Q_h(x,A), \qquad \text{for } x \in S_h, A \subset \mathbb{R}, x \notin A.$$

We assume that for any compact set K,

$$\sup_{x\in K} Q_h(x,\mathbb{R}) < \infty. \tag{3.6}$$

Let us write

$$K_h(x,dy) = \begin{cases} \frac{1}{h}\Pi_h(x,dy) & \text{in discrete time} \\ Q_h(x,dy) & \text{in continuous time} \end{cases}$$

and define

$$(\sigma^2)^h(x) = \int_{|y-x|\leq 1} (y-x)^2 K_h(x,dy),$$

$$\begin{aligned}\mu^h(x) &= \int_{|y-x|\le 1} (y-x)K_h(x,dy),\\ \Delta_\varepsilon^h(x) &= K_h(x,B(x,\varepsilon)^c),\end{aligned}$$

where $B(x,\varepsilon) = (x-\varepsilon, x+\varepsilon)$.

Theorem 3.6. *Suppose that μ and σ are continuous coefficients for which the martingale problem for $\mathscr{L}$ of (3.1) is well-posed. In continuous time we assume (3.6). Suppose further that for each $R < \infty$ and $\varepsilon > 0$*

1.

$$\lim_{h\downarrow 0} \sup_{x\in S_h, |x|\le R} |\mu^h(x) - \mu(x)| = 0,$$

2.

$$\lim_{h\downarrow 0} \sup_{x\in S_h, |x|\le R} |(\sigma^2)^h(x) - \sigma^2(x)| = 0,$$

3.

$$\lim_{h\downarrow 0} \sup_{x\in S_h, |x|\le R} \Delta_\varepsilon^h(x) = 0.$$

If $X_0^h = x_h \to x$ then we have $\{X_t^h\}_{t\ge 0} \Rightarrow \{X_t\}_{t\ge 0}$, the solution of the martingale problem with $X_0 = x$. (Here $\Rightarrow$ denotes convergence in the sense of finite-dimensional distributions.)

This is a special case of Theorem 8.7.1 of Durrett (1996) which in turn is based upon Chap. 11 of Stroock and Varadhan (1979). The first two conditions of the Theorem ensure that infinitesimal drift and variance of the sequence of Markov chains converge (uniformly on compact sets) to the right thing, while the third rules out jumps in the limit.

Of course it remains to check that the martingale problem *is* well-posed for our Wright–Fisher diffusion. That result is really due to Feller (1951) (although he didn't use this language). It can be found in Ethier and Kurtz (1986) who consider convergence of a Wright–Fisher model (with possibly more than two alleles) to the Wright–Fisher diffusion in their Chap. 10. They invoke much more powerful weak convergence results that are beyond our scope here.

Remark 3.7. This sort of convergence is enough to justify using our limiting Wright–Fisher diffusion to approximate things like time to fixation and fixation probabilities. However, if we are really interested in the genealogies of populations, then we need more. For our Moran models, the Donnelly–Kurtz lookdown construction gave us a much stronger result, namely the joint convergence of the forwards in time model for the evolution of the population and the (backwards in time) genealogical trees relating individuals in that population. In general we must be careful. It is possible to arrive at the same diffusion for allele frequencies from many different individual based models for our population, and it is *not* always the case that the genealogies converge to the same limit (see Taylor (2009) for some examples).

3.3 Speed and Scale

A nice feature of one dimensional diffusions is that many quantities can be calculated explicitly. This is because (except at certain singular points which will only ever be at a or b under our conditions) all one-dimensional diffusions can be transformed into Brownian motion first by a change of space variable (through the so-called scale function) and then a timechange (through what is known as the speed measure).

To see how this works, we first investigate what happens to a diffusion when we change the timescale. Suppose that a diffusion $\{Z_t\}_{t\geq 0}$ has generator $\mathscr{L}_Z$, with infinitesimal drift $\mu_Z(x)$ and infinitesimal variance $\sigma_Z^2(x)$. We define a new process $\{Y_t\}_{t\geq 0}$ by $Y_t = Z_{\tau(t)}$ where

$$\tau(t) = \int_0^t \beta(Y_s)ds,$$

for some function $\beta(x)$ which we assume to be bounded, continuous and strictly positive. So if $Y_0 = Z_0$, then the increment of Y_t over an infinitesimal time interval $(0,dt)$ is that of Z_t over the interval $(0,d\tau(t)) = (0,\beta(Y_0)dt)$. In our previous notation,

$$\mathbb{E}[\Delta_h Y(0)|Y_0 = y] = \beta(Y_0)h\mu_Z(Z_0) + o(h) = \beta(y)\mu_Z(y)h + o(h),$$

and

$$\mathbb{E}[(\Delta_h Y(0))^2|Y_0 = y] = \beta(Y_0)h\sigma_Z^2(Z_0) + o(h) = \beta(y)\sigma_Z^2(y)h + o(h).$$

In other words,

$$\mathscr{L}_Y f(x) = \beta(x)\mathscr{L}_Z f(x).$$

In the simplest example, β is a constant and we are simply changing our time units in a spatially homogeneous way. In general, the rate of our ‘clock’ depends upon where we are in space. We are now in a position to understand speed and scale. Let $\{X_t\}_{t\geq 0}$ be governed by the generator (3.1). Suppose now that $S(x)$ is a strictly increasing function on (a,b) and consider the new process $Z_t = S(X_t)$. Then the generator $\mathscr{L}_Z$ of Z can be calculated as

$$\begin{aligned}
\mathscr{L}_Z f(x) &= \frac{d}{dt}\mathbb{E}\left[f(Z_t)|Z_0 = x\right]\Big|_{t=0} \\
&= \frac{d}{dt}\mathbb{E}\left[f(S(X_t))|S(X_0) = x\right]\Big|_{t=0} \\
&= \mathscr{L}_X(f\circ S)(S^{-1}(x)) \\
&= \frac{1}{2}\sigma^2(S^{-1}(x))\frac{d^2}{dx^2}(f\circ S)(S^{-1}(x)) + \mu(S^{-1}(x))\frac{d}{dx}(f\circ S)(S^{-1}(x))
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2}\sigma^2(S^{-1}(x))\left\{(S'(S^{-1}(x)))^2\frac{d^2f}{dx^2}(x) + S''(S^{-1}(x))\frac{df}{dx}(x)\right\}\\
&\quad + \mu(S^{-1}(x))S'(S^{-1}(x))\frac{df}{dx}(x)\\
&= \frac{1}{2}\sigma^2(S^{-1}(x))S'(S^{-1}(x))^2\frac{d^2f}{dx^2}(x) + \mathscr{L}S(S^{-1}(x))\frac{df}{dx}(x). \qquad (3.7)
\end{aligned}$$

Now if we can find a strictly increasing function S that satisfies $\mathscr{L}S \equiv 0$, then the drift term (in the mathematical sense) in (3.7) will vanish and so Z_t will just be a time change of Brownian motion on the interval $(S(a), S(b))$. Such an S is provided by the scale function of the diffusion.

Definition 3.8 (Scale function). For a diffusion X_t on (a,b) with drift μ and variance σ^2, the *scale function* is defined by

$$S(x) = \int_{x_0}^{x} \exp\left(-\int_{\eta}^{y}\frac{2\mu(z)}{\sigma^2(z)}dz\right)dy,$$

where x_0, η are points fixed (arbitrarily) in (a,b).

Definition 3.9 (Natural scale). We shall say that a diffusion is in *natural scale* if $S(x)$ can be taken to be linear.

The scale change $X_t \mapsto S(X_t)$ resulted in a timechanged Brownian motion on $(S(a), S(b))$. The change of time required to transform this into standard Brownian motion is dictated by the speed measure.

Definition 3.10 (Speed measure). The function $m(\xi) = \frac{1}{\sigma^2(\xi)S'(\xi)}$ is the *density of the speed measure* or just the *speed density* of the process X_t. We write

$$M(x) = \int_{x_0}^{x} m(\xi)d\xi.$$

Remark 3.11. The function m plays the rôle of β before. Naively, looking at (3.7), we might expect to timechange via $\beta(\xi) = 1/(\sigma^2(\xi)S'(\xi)^2)$. However, notice that

$$\int_{x_0}^{x} m(\xi)d\xi = \int_{S(x_0)}^{S(x)} m(S^{-1}(y))\frac{1}{S'(S^{-1}(y))}dy = \int_{S(x_0)}^{S(x)} \frac{1}{\sigma^2(S^{-1}(y))\big(S'(S^{-1}(y))\big)^2}dy.$$

The additional $S'(y)$ in the generator (3.7) has been absorbed in the change of coordinates since our time change is applied to $S(X_t)$ on $(S(a), S(b))$, not to X_t itself.

In summary, we have the following.

Lemma 3.12. *Denoting the scale function and the speed measure by S and M respectively we have*

$$\mathscr{L}f = \frac{1}{2}\frac{1}{dM/dS}\frac{d^2f}{dS^2} = \frac{1}{2}\frac{d}{dM}\left(\frac{df}{dS}\right).$$

Proof.

$$\begin{aligned}\frac{1}{2}\frac{d}{dM}\left(\frac{df}{dS}\right) &= \frac{1}{2}\frac{1}{dM/dx}\frac{d}{dx}\left(\frac{1}{dS/dx}\frac{df}{dx}\right)\\ &= \frac{1}{2}\sigma^2(x)S'(x)\frac{d}{dx}\left(\frac{1}{S'(x)}\frac{df}{dx}\right)\\ &= \frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2} - \frac{1}{2}\sigma^2(x)S'(x)\frac{S''(x)}{(S'(x))^2}\frac{df}{dx}\\ &= \frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2} + \mu(x)\frac{df}{dx}\end{aligned}$$

(since S solves $\mathscr{L}S = 0$) as required. □

3.4 Hitting Probabilities and Feller's Boundary Classification

Before going further, let's see how we might apply this. Suppose that a diffusion process on $(0,1)$ represents the frequency of an allele, a say, in a population and that zero and one are traps for the process. One question that we should like to answer is "What is the probability that the a-allele is eventually lost from the population?" In other words, what is the probability that the diffusion hits zero before one? To prove a general result we need first to be able to answer this question for Brownian motion.

Lemma 3.13. *Let $\{W_t\}_{t\geq 0}$ be standard Brownian motion on the line. For each $y \in \mathbb{R}$, let T_y denote the random time at which it hits y for the* first *time. Then for $a < x < b$,*

$$\mathbb{P}[T_a < T_b | W_0 = x] = \frac{b-x}{b-a}.$$

Sketch of Proof. Let $u(x) = \mathbb{P}[T_a < T_b | W_0 = x]$ and assume that $\mathbb{P}[T_a \wedge T_b < h | W_0 = x] = o(h)$ as $h \to 0$. If we suppose that u is sufficiently smooth, then, using the Markov property,

$$\begin{aligned}u(x) &= \mathbb{E}\left[u(W_h) | W_0 = x\right] + o(h)\\ &= \mathbb{E}\left[u(x) + (W_h - x)u'(x) + \frac{1}{2}(W_h - x)^2 u''(x)\right] + o(h)\\ &= u(x) + \frac{1}{2}hu''(x) + o(h).\end{aligned}$$

Subtracting $u(x)$ from each side, dividing by h and letting h tend to zero, we obtain $u''(x) = 0$. We also have the boundary conditions $u(a) = 1$ and $u(b) = 0$. This is easily solved to give

$$u(x) = \frac{b-x}{b-a}$$

as required. □

Of course this reflects the corresponding result for simple random walk that we used in the proof of Lemma 2.31. In general we can reduce the corresponding question for $\{X_t\}_{t\geq 0}$ to solution of the equation $\mathscr{L}u(x) = 0$ with $u(a) = 1$ and $u(b) = 0$, but in fact we have already done all the work we need. We have the following result.

Lemma 3.14 (Hitting probabilities). *Let $\{X_t\}_{t\geq 0}$ be a one-dimensional diffusion on (a,b) with infinitesimal drift $\mu(x)$ and variance $\sigma^2(x)$ satisfying the conditions above. If $a < a_0 < x < b_0 < b$ then writing T_y for the first time at which $X_t = y$,*

$$\mathbb{P}[T_{a_0} < T_{b_0} | X_0 = x] = \frac{S(b_0) - S(x)}{S(b_0) - S(a_0)}, \tag{3.8}$$

where S is the scale function for the diffusion.

Remark 3.15. Our definition of the scale function, S, depended upon arbitrary choices of η and x_0, but η cancels in the ratio and x_0 in the difference, so that the expression on the right hand side of (3.8) is well-defined.

Proof. Evidently it is enough to consider the corresponding hitting probabilities for the process $Z_t = S(X_t)$, where S is the scale function. The process $\{Z_t\}_{t\geq 0}$ is a time changed Brownian motion, but since we only care about *where* not *when* the process exits the interval $(S(a_0), S(b_0))$, then we need only determine the hitting probabilities for Brownian motion and the result follows immediately from Lemma 3.13. □

Before continuing to calculate quantities of interest, we fill in a gap left earlier when we failed to completely specify the domain of the generators of our one-dimensional diffusions. Whether or not functions in the domain must satisfy boundary conditions at a and b is determined by the nature of those boundaries from the perspective of the diffusion. More precisely, we have the following classification.

Definition 3.16 (Feller's boundary classification). For a one-dimensional diffusion on the interval with endpoints a, b (with $a < b$), define

$$u(x) = \int_{x_0}^{x} M dS, \qquad v(x) = \int_{x_0}^{x} S dM,$$

where S is the scale function of Definition 3.8 and M the speed measure of Definition 3.10. The boundary b is said to be

$$\begin{array}{ll} \textit{regular} & \text{if } u(b) < \infty \text{ and } v(b) < \infty \\ \textit{exit} & \text{if } u(b) < \infty \text{ and } v(b) = \infty \\ \textit{entrance} & \text{if } u(b) = \infty \text{ and } v(b) < \infty \\ \textit{natural} & \text{if } u(b) = \infty \text{ and } v(b) = \infty \end{array}$$

with symmetric definitions at a.

Regular and exit boundaries are said to be *accessible* while entrance and natural boundaries are called *inaccessible*.

Theorem 3.17. *If neither a nor b is regular, the domain of the generator (3.1) is the continuous functions f on $[a,b]$ which are twice continuously differentiable on the interior and for which*

1. *if a and b are inaccessible there are no further conditions,*
2. *if b (resp. a) is an exit boundary, then*

$$\lim_{x \to b} \mathscr{L} f(x) = 0$$

$$\left(\textit{resp. } \lim_{x \to a} \mathscr{L} f(x) = 0\right).$$

If b (resp. a) is a regular boundary, then for each fixed $q \in [0,1]$ there is a Feller semigroup corresponding to the generator (3.1) with domain as above plus the additional condition

$$q \lim_{x \to b} \mathscr{L} f(x) = (1-q) \lim_{x \to b} \frac{1}{S'(x)} f'(x) \tag{3.9}$$

$$\left(\textit{resp. } q \lim_{x \to a} \mathscr{L} f(x) = -(1-q) \lim_{x \to a} \frac{1}{S'(x)} f'(x)\right).$$

For a more careful discussion see Ethier and Kurtz (1986), Chap. 8.

Remark 3.18. For each fixed $q \in [0,1]$, condition (3.9) is enough to specify the boundary behaviour of the diffusion at a regular boundary uniquely. It is easy to check that the Wright–Fisher diffusion with mutation with generator (2.11) has a regular boundary at 0 (resp. 1) if $\nu_2 \in (0, \frac{1}{2})$ (resp. if $\nu_1 \in (0, \frac{1}{2})$), but the condition (3.9) is in fact *the same* for all $q > 0$.

3.5 Green's Functions

Lemma 3.14 tells us the probability that we exit (a,b) for the first time through a, but can we glean some information about how long we must wait for $\{X_t\}_{t \geq 0}$ to exit the interval (a,b) (either through a or b) or, more generally, writing T^* for the first exit time of (a,b), can we say anything about $\mathbb{E}[\int_0^{T^*} g(X_s) ds | X_0 = x]$? (Putting $g = 1$ gives the mean exit time.) Let us write

$$w(x) = \mathbb{E}\left[\int_0^{T^*} g(X_s)ds|X_0 = x\right]$$

and we'll derive the differential equation satisfied by w.

Suppose for simplicity that g is Lipschitz continuous on (a,b) with Lipschitz constant K. First note that $w(a) = w(b) = 0$. Now consider a small interval of time of length h. We're going to split the integral into the contribution up to time h and after time h. Because $\{X_t\}_{t\geq 0}$ is a Markov process,

$$\mathbb{E}\left[\int_h^{T^*} g(X_s)ds|X_h = z\right] = \mathbb{E}\left[\int_0^{T^*} g(X_s)ds|X_0 = z\right] = w(z)$$

and so for $a < x < b$

$$w(x) \approx \mathbb{E}\left[\int_0^{h} g(X_s)ds|X_0 = x\right] + \mathbb{E}\left[w(X_h)|X_0 = x\right]. \tag{3.10}$$

The '$\approx$' here is because we have ignored the possibility that $h > T^*$. Since g is Lipschitz continuous, we have the approximation

$$\begin{aligned}&\left|\mathbb{E}\left[\int_0^{h} g(X_s)ds|X_0 = x\right] - hg(x)\right| = \mathbb{E}\left[\left|\int_0^{h} g(X_s)ds - hg(x)\right| |X_0 = x\right]\\ &\leq \mathbb{E}\left[\int_0^{h} K|X_s - x|ds|X_0 = x\right] \leq K\int_0^h \sqrt{\mathbb{E}\left[\,|X_s - x|^2|X_0 = x\right]} = \mathscr{O}(h^{3/2}).\end{aligned}$$

Now substitute this estimate in (3.10), subtract $w(x)$ from both sides, divide by h and let $h \downarrow 0$ to obtain

$$\mu(x)w'(x) + \frac{1}{2}\sigma^2(x)w''(x) = -g(x), \quad w(a) = 0 = w(b). \tag{3.11}$$

Let us now turn to solving this equation. Using Lemma 3.12 with $w = f$,

$$\mathscr{L}w(x) = \frac{1}{2}\frac{1}{m(x)}\frac{d}{dx}\left(\frac{1}{S'(x)}w'(x)\right)$$

and so we have

$$\frac{d}{dx}\left(\frac{1}{S'(x)}w'(x)\right) = -2g(x)m(x),$$

whence

$$\frac{1}{S'(x)}w'(x) = -2\int_a^x g(\xi)m(\xi)d\xi + \beta$$

where β is a constant. Multiplying by $S'(x)$ and integrating gives

$$w(x) = -2\int_a^x S'(\xi)\int_a^{\xi} g(\eta)m(\eta)d\eta d\xi + \beta(S(x) - S(a)) + \alpha$$

for constants α, β. Since $w(a) = 0$, we immediately have that $\alpha = 0$. Reversing the order of integration,

$$\begin{aligned} w(x) &= -2\int_a^x \int_\eta^x S'(\xi)d\xi g(\eta)m(\eta)d\eta + \beta(S(x) - S(a)) \\ &= -2\int_a^x (S(x) - S(\eta))g(\eta)m(\eta)d\eta + \beta(S(x) - S(a)) \end{aligned}$$

and $w(b) = 0$ now gives

$$\beta = \frac{2}{S(b) - S(a)}\int_a^b (S(b) - S(\eta))g(\eta)m(\eta)d\eta.$$

Finally then

$$\begin{aligned} w(x) &= \frac{2}{S(b) - S(a)}\Big\{(S(x) - S(a))\int_a^b (S(b) - S(\eta))g(\eta)m(\eta)d\eta \\ &\qquad -(S(b) - S(a))\int_a^x (S(x) - S(\eta))g(\eta)m(\eta)d\eta\Big\} \\ &= \frac{2}{S(b) - S(a)}\Big\{(S(x) - S(a))\int_x^b (S(b) - S(\eta))g(\eta)m(\eta)d\eta \\ &\qquad +(S(b) - S(x))\int_a^x (S(\eta) - S(a))g(\eta)m(\eta)d\eta\Big\} \end{aligned}$$

where the last line is obtained by splitting the first integral into $\int_a^b = \int_x^b + \int_a^x$.

Theorem 3.19. *For a continuous function g,*

$$\mathbb{E}\left[\int_0^{T*} g(X_s)ds|X_0 = x\right] = \int_a^b G(x,\xi)g(\xi)d\xi,$$

where for $a < x < b$ we have

$$G(x,\xi) = \begin{cases} 2\dfrac{(S(x) - S(a))}{(S(b) - S(a))}(S(b) - S(\xi))m(\xi), \textit{for } x < \xi < b \\ \\ 2\dfrac{(S(b) - S(x))}{(S(b) - S(a))}(S(\xi) - S(a))m(\xi), \textit{for } a < \xi < x, \end{cases}$$

with S the scale function given in Definition 3.8 and $m(\xi) = \frac{1}{\sigma^2(\xi)S'(\xi)}$, the density of the speed measure.

Definition 3.20. The function $G(x,\xi)$ is called the *Green's function* of the process $\{X_t\}_{t\geq 0}$.

By taking g to approximate $\mathbf{1}_{(x_1,x_2)}$ we see that $\int_{x_1}^{x_2} G(x,\xi)d\xi$ is the mean time spent by the process in (x_1,x_2) before exiting (a,b) if initially $X_0 = x$. Sometimes, the Green's function is called the *sojourn density*.

Example 3.21. Consider the Wright–Fisher diffusion with generator

$$\mathscr{L}f(p) = \frac{1}{2}p(1-p)f''(p).$$

Notice that since it has no drift term ($\mu = 0$) it is already in natural scale, $S(p) = p$ (up to an arbitrary additive constant). What about $\mathbb{E}[T^*]$?

Using Theorem 3.19 with $g = 1$ we have

$$\begin{aligned}
\mathbb{E}_p[T^*] &= \mathbb{E}\left[\int_0^{T*} 1ds\middle| X_0 = p\right] = \int_0^1 G(p,\xi)d\xi \\
&= 2\int_p^1 p(1-\xi)\frac{1}{\xi(1-\xi)}d\xi + 2\int_0^p (1-p)\xi\frac{1}{\xi(1-\xi)}d\xi \\
&= 2p\int_p^1 \frac{1}{\xi}d\xi + 2(1-p)\int_0^p \frac{1}{1-\xi}d\xi \\
&= -2\{p\log p + (1-p)\log(1-p)\}.
\end{aligned}$$

□

This suggests that in our Moran model, at least if the population is large, if the current proportion of a-alleles is p, the time until either the a-allele or the A-allele is fixed in the population should have mean approximately

$$-2\{p\log p + (1-p)\log(1-p)\}. \tag{3.12}$$

In fact by conditioning on whether the proportion of a-alleles increases or decreases at the first reproduction event, one obtains a recurrence relation for the *number of jumps* until the Moran process first hits either zero or one. This recurrence relation can be solved explicitly and since jumps occur at independent exponentially distributed times with mean $1/\binom{N}{2}$, it is easy to verify that (3.12) is indeed a good approximation. For the Wright–Fisher model, in its original timescale, there is no explicit expression for the expected time to fixation, $t(p)$. However, since changes in p over a single generation are typically small, one can expand $t(p)$ in a Taylor series, in just the same way as we did to derive equation (2.1), and thus verify that for a large population,

$$p(1-p)t''(p) = -2N, \quad t(0) = 0 = t(1).$$

This is readily solved to give

$$t(p) = -2N\{p\log p + (1-p)\log(1-p)\},$$

as predicted by our diffusion approximation. (Recall that our Moran model is already in the diffusive timescale, whereas the Wright–Fisher model is not, accounting for the extra factor of N.)

3.6 Stationary Distributions and Reversibility

Before moving on to models in which a gene is allowed to have more than two alleles, we consider one last quantity for our one-dimensional diffusions. First a general definition.

Definition 3.22 (Stationary distribution). Let $\{X_t\}_{t\geq 0}$ be a Markov process on the space E. A *stationary distribution* for $\{X_t\}_{t\geq 0}$ is a probability distribution ψ on E such that if X_0 has distribution ψ, then X_t has distribution ψ for all $t \geq 0$.

In particular this definition tells us that if ψ is a stationary distribution for $\{X_t\}_{t\geq 0}$, then

$$\frac{d}{dt}\mathbb{E}\left[f(X_t)|X_0 \sim \psi\right] = 0,$$

where we have used $X_0 \sim \psi$ to indicate that X_0 is distributed according to ψ. In other words

$$\frac{d}{dt}\int_E \mathbb{E}\left[f(X_t)|X_0 = x\right]\psi(dx) = 0.$$

Evaluating the time derivative at $t = 0$ gives

$$\int_E \mathscr{L}f(x)\psi(dx) = 0.$$

Sometimes this allows us to find an explicit expression for $\psi(dx)$. Let $\{X_t\}_{t\geq 0}$ be a one-dimensional diffusion on (a,b) with generator given by (3.1). We're going to suppose that there is a stationary distribution which is absolutely continuous with respect to Lebesgue measure. Let us abuse notation a little by using $\psi(x)$ to denote the density of $\psi(dx)$ on (a,b). Then, integrating by parts, we have that for all $f \in \mathscr{D}(\mathscr{L})$,

$$\begin{aligned} 0 &= \int_a^b \left\{\frac{1}{2}\sigma^2(x)\frac{d^2 f}{dx^2}(x) + \mu(x)\frac{df}{dx}(x)\right\}\psi(x)dx \\ &= \int_a^b f(x)\left\{\frac{1}{2}\frac{d^2}{dx^2}\left(\sigma^2(x)\psi(x)\right) - \frac{d}{dx}\left(\mu(x)\psi(x)\right)\right\}dx + \text{boundary terms.} \end{aligned}$$

This equation must hold for all f in the domain of $\mathscr{L}$ and so, in particular, choosing f and f' to vanish on the boundary,

$$\frac{1}{2}\frac{d^2}{dx^2}\left(\sigma^2(x)\psi(x)\right) - \frac{d}{dx}\left(\mu(x)\psi(x)\right) = 0 \quad \text{for } x \in (a,b). \tag{3.13}$$

Integrating once gives

$$\frac{1}{2}\frac{d}{dx}\left(\sigma^2(x)\psi(x)\right) - \mu(x)\psi(x) = C_1,$$

for some constant C_1 and then using $S'(x)$ as an integrating factor we obtain

$$\frac{d}{dy}\left(S'(y)\sigma^2(y)\psi(y)\right) = C_1 S'(y),$$

from which

$$\psi(x) = C_1\frac{S(x)}{S'(x)\sigma^2(x)} + C_2\frac{1}{S'(x)\sigma^2(x)} = m(x)\left[C_1 S(x) + C_2\right].$$

If we can arrange constants so that $\psi \geq 0$ and

$$\int_a^b \psi(\xi)d\xi = 1$$

then the stationary distribution exists and has density ψ. In particular, if $\int_a^b m(y)\, dy < \infty$, then taking $C_1 = 0$,

$$\psi(x) = \frac{m(x)}{\int_a^b m(y)dy} \tag{3.14}$$

is the density of a stationary distribution for the diffusion.

We know from the theory of Markov chains that uniqueness of the stationary measure of a chain requires irreducibility. The corresponding condition here is regularity.

Definition 3.23. For a one dimensional diffusion process on the interval I, let us write

$$H_y = \inf\{t > 0 : X_t = y\}.$$

The diffusion is said to be *regular* if for all $x \in I^0$ (the interior of I) and all $y \in I$ (including finite endpoints) $\mathbb{P}_x[H_y < \infty] > 0$.

Theorem 3.24 (Watanabe and Motoo 1958). *A regular diffusion in natural scale with no absorbing boundary points has a stationary distribution if and only if the speed measure is finite and then it is given by (3.14).*

Under these conditions there is also an ergodic theorem.

Example 3.25. Recall the generator of the Wright–Fisher diffusion with mutation,

$$\mathscr{L}f(p) = \frac{1}{2}p(1-p)\frac{d^2 f}{dp^2} + \Big(\nu_2(1-p) - \nu_1 p\Big)\frac{df}{dp}.$$

What is the stationary distribution?

For this diffusion

$$\begin{aligned} S'(p) &= \exp\left(-\int_{p_0}^{p} \frac{2\mu(z)}{\sigma^2(z)}dz\right) \\ &= \exp\left(-\int_{p_0}^{p} \frac{2\nu_2(1-z) - 2\nu_1 z}{z(1-z)}dz\right) \\ &= C\exp\left(-2\nu_2 \log p - 2\nu_1 \log(1-p)\right) \\ &= Cp^{-2\nu_2}(1-p)^{-2\nu_1}, \end{aligned}$$

where the value of the constant C depends on p_0. In this case we have

$$m(p) = \frac{1}{\sigma^2(p)S'(p)} = Cp^{2\nu_2-1}(1-p)^{2\nu_1-1}.$$

Now

$$\int_0^1 m(p)dp = \int_0^1 Cp^{2\nu_2-1}(1-p)^{2\nu_1-1}dp = C\frac{\Gamma(2\nu_1)\Gamma(2\nu_2)}{\Gamma(2(\nu_1+\nu_2)}$$

(where Γ is Euler's Gamma function) and so the stationary distribution is just

$$\psi(p) = \frac{\Gamma(2(\nu_1+\nu_2))}{\Gamma(2\nu_1)\Gamma(2\nu_2)}p^{2\nu_2-1}(1-p)^{2\nu_1-1}. \tag{3.15}$$

Ethier and Kurtz (1986), Chap. 10, Lemma 2.1 gives a direct proof of uniqueness of this stationary distribution. □

The stationary distribution gives us some understanding of the longterm balance between the competing forces of mutation (which maintains genetic diversity) and genetic drift (which removes variation from the population). Figure 3.1 shows the density of the stationary distribution of the Wright–Fisher diffusion with mutation for a variety of parameters. When $2\nu_1$ and $2\nu_2$ are both bigger than 1, the stationary distribution is peaked around its mean, but when they are both less than one it has singularities at $\{0,1\}$. Of course, if there is no mutation, then the process eventually becomes trapped in 0 and 1.

One can also calculate simple summary statistics.

Definition 3.26. The *gene diversity* or *heterozygosity*, H, is the probability that two randomly chosen genes are of different allelic types.

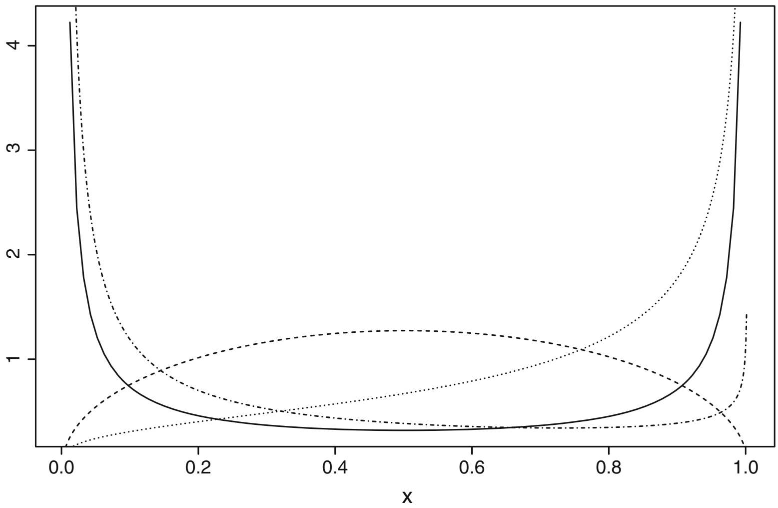

Fig. 3.1 Stationary distribution of the Wright–Fisher diffusion. The graphs plot the density ψ, given by (3.15) for: $2\nu_1 = 2\nu_2 = 0.2$ (*solid line*), $2\nu_1 = 2\nu_2 = 1.5$ (*dashed line*), $2\nu_1 = 0.5, 2\nu_2 = 1.3$ (*dotted line*) and $2\nu_1 = 0.7, 2\nu_2 = 0.2$ (alternating *dashes* and *dots*)

If the allele frequency, P say, is at stationarity under the Wright–Fisher diffusion with mutation, then

$$\begin{aligned} H = \mathbb{E}[2P(1-P)] &= \int_0^1 2p(1-p)\psi(p)dp \\ &= 2\int_0^1 p^{2\nu_2}(1-p)^{2\nu_1}dp\frac{\Gamma(2(\nu_1+\nu_2))}{\Gamma(2\nu_1)\Gamma(2\nu_2)} \\ &= 2\frac{\Gamma(2\nu_1+1)\Gamma(2\nu_2+1)}{\Gamma(2\nu_1+2\nu_2+2)}\frac{\Gamma(2(\nu_1+\nu_2))}{\Gamma(2\nu_1)\Gamma(2\nu_2)} \\ &= \frac{2\cdot 2\nu_1\cdot 2\nu_2}{(2\nu_1+2\nu_2)(2\nu_1+2\nu_2+1)} \\ &= \frac{4\nu_1\nu_2}{(\nu_1+\nu_2)(2(\nu_1+\nu_2)+1)}. \end{aligned}$$

Now in ‘real’ units, $\nu_1 = N\mu_1$ and $\nu_2 = N\mu_2$ (or $2N_e\mu_1$ and $2N_e\mu_2$ for a diploid population) and so

$$H = \frac{2\mu_1\mu_2}{(\mu_1+\mu_2)(\mu_1+\mu_2+\frac{1}{2N})}.$$

Notice in particular that gene diversity increases with population size. For a larger population the force of genetic drift is weaker.

Finally let us demonstrate one very powerful technique that is often applied in settings where the speed measure is a stationary distribution. The idea is familiar from the study of discrete time and space Markov chains.

Definition 3.27. A discrete time and space Markov chain with transition probabilities $p(i,j)$ is said to be *reversible* with respect to the stationary measure π if it satisfies the *detailed balance* equation:

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

for all i and j in the state space.

For such chains we can say things about events backwards in time by considering the forwards in time transition probabilities. The analogue of the detailed balance equation for a one-dimensional diffusion is

$$\psi(x)p(t,x,y) = \psi(y)p(t,y,x) \quad \text{for all } x,y,t.$$

Now multiplying by arbitrary functions $f(x)$ and $g(y)$ in the domain of the generator of the diffusion we obtain

$$\int \psi(x)f(x)\left(\int p(t,x,y)g(y)dy\right)dx = \int \psi(y)g(y)\left(\int p(t,y,x)f(x)dx\right)dy.$$

Now observe that the inner integrals are

$$\mathbb{E}[g(X_t)|X_0 = x] \quad \text{and} \quad \mathbb{E}[f(X_t)|X_0 = y]$$

and differentiate with respect to t at $t = 0$ to obtain

$$\int f(x)\mathscr{L}g(x)\psi(x)dx = \int \mathscr{L}f(y)g(y)\psi(y)dy. \tag{3.16}$$

Definition 3.28. If the identity (3.16) is satisfied for all f and g, then ψ is called a *reversible stationary distribution* and we say that the diffusion is *reversible* with respect to ψ.

Now suppose that the stationary distribution of the diffusion is given by $\psi(x) = m(x)/\int m(y)dy$. Then choosing f and g to vanish at the boundary of the domain to force the boundary terms to vanish when we integrate by parts (twice), we obtain

$$\begin{aligned}\int_a^b f(x)\mathscr{L}g(x)m(x)dx &= \frac{1}{2}\int_a^b f(x)\frac{1}{m(x)}\frac{d}{dx}\left(\frac{1}{S'(x)}\frac{dg}{dx}\right)m(x)dx \\ &= \frac{1}{2}\int_a^b \frac{d}{dx}\left(\frac{1}{S'(x)}\frac{df}{dx}\right)g(x)dx\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\int_a^b \frac{1}{m(x)}\frac{d}{dx}\left(\frac{1}{S'(x)}\frac{df}{dx}\right) g(x)m(x)dx \\
&= \int_a^b \mathscr{L}f(x)g(x)m(x)dx,
\end{aligned}
$$

so this is indeed a *reversible* stationary distribution.

Example 3.29 (Which allele is the oldest?). Suppose that a population consists of two allelic types, a and A. We assume that one of the two alleles arose through mutation onto a background consisting entirely of the other type, since when there have been no further mutations. If the a-allele is currently at frequency p, what is the probability that it is the older allele?

The usual way to handle questions like this is to think of the model as one arising in the limit of very low mutation rates. If mutation rates are low then the process of allele frequencies consists of a sequence of excursions away from the boundaries. The a-allele is the oldest if as we trace backwards in time the allele frequency hits the boundary point 1 before it hits 0. Reversing with respect to the speed measure we see that this probability is the same as the probability that we hit 1 before 0 forwards in time. And (using Lemma 3.14) this in turn, as the mutation rates tend to zero, converges to the current frequency of the a-allele, that is p. For more details see, for example, Watterson (1977). □