Lecture 25 : Fixation in the diffusion limit

MATH285K - Spring 2010 *Lecturer: Sebastien Roch*

References: [Dur08, Chapter 7], [KT81, Chapter 15].

Previous class

Recall, we consider diffusions defined on a closed interval $I = [l, r]$ that satisfy the following properties:

- For every $\varepsilon > 0$,

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}[|X(t+h) - x| > \varepsilon \,|\, X(t) = x] = 0, \tag{1}$$

  for all $x \in I$. (All diffusions satisfy this version of "continuity," unlike jump chains for instance.)

- Let $\Delta_h X(t) = X(t+h) - X(t)$. For all $l < x < r$ and $t \in \mathbb{R}_+$,

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[\Delta_h X(t) \,|\, X(t) = x] = \mu(x), \tag{2}$$

  and

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[(\Delta_h X(t))^2 \,|\, X(t) = x] = \sigma^2(x), \tag{3}$$

  where μ, the *infinitesimal drift* (not to be confused with genetic drift), and σ^2, the *infinitesimal variance*, are continuous functions of x. In particular, the process is *time-homogeneous*. Moreover for $r = 3, 4, \ldots$

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[|\Delta_h X(t)|^r \,|\, X(t) = x] = 0. \tag{4}$$

- The process is *regular*, that is, for all x, y in the interior of I

$$\mathbb{P}[T(y) < \infty \,|\, X(0) = x] > 0, \tag{5}$$

  where $T(y)$ is the hitting time of y, that is, the first time y is reached. (See e.g. [Dur96, (3.1) in Chapter 6])

1 Wright-Fisher diffusion model for diploids

We generalize slightly the model we derived in the previous lecture by considering a diploid population with N individuals. Suppose we have two alleles A and a at a locus where the genotypes have the following relative fitness (one way to think about this is that each genotype survives to maturity with the given probability):

$$\begin{array}{ccc} AA & Aa & aa \\ 1-s_0 & 1-s_1 & 1-s_2. \end{array}$$

Moreover, we have mutations $A \to a$ (respectively $a \to A$) with probability μ_1 (respectively μ_2). Assume the parameters scale with N as

$$\gamma_i = 2Ns_i, \qquad \beta_i = 2N\mu_i.$$

Following [Dur08], we also use the notation $\delta = \gamma_2 - \gamma_2$ and $\eta = 2\gamma_1 - \gamma_0 - \gamma_2$.

The infinitesimal drift and variance are enough to characterize the behaviour of the diffusion limit—except for the boundary conditions which we will discuss in the next lecture. Denoting the rescaled state by x (frequency of A), the infinitesimal drift and variance are given by

$$\mu(x) = [\beta_1(1-x) - \beta_2 x] + x(1-x)[\delta + \eta x], \tag{6}$$

and

$$\sigma^2(x) = x(1-x). \tag{7}$$

The latter comes from the binomial sampling scheme. The first term of (6) reflects the mutation pressure and is straightforwardx to interpret. To understand the second term it is useful to look at examples (where we take $\beta_1 = \beta_2 = 0$).

EX 25.1 (Additive selection) *Here $s_0 = 0$, $s_1 = s$, $s_2 = 2s$ and we let $\gamma = 2Ns$. Then*

$$\mu(x) = \delta x(1-x).$$

EX 25.2 (Balancing selection) *Here $s_1 = 0$. Then*

$$\mu(x) = x(1-x)\left[\gamma_2 - (\gamma_0 + \gamma_2)x\right] = (\gamma_0 + \gamma_2)x(1-x)\left[\frac{\gamma_2}{\gamma_0 + \gamma_2} - x\right].$$

EX 25.3 (Dominant A) *Here $s_0 = s_1 = 0$, $s_2 = s$ and we let $\gamma = 2Ns$. Then*

$$\mu(x) = x(1-x)[\gamma - \gamma x] = \gamma x(1-x)^2.$$

EX 25.4 (Recessive A) *This case is symmetric to the previous one. Here* $s_0 = 0$*,* $s_1 = s_2 = s$ *and* $\gamma = 2Ns$*. Then*

$$\mu(x) = \gamma x^2(1 - x).$$

See [Dur08, Figure 7.1] for an illustration of the mean behavior in each case. However, the variance term cannot in general be ignored as it reflects *genetic drift*, a key factor in shaping the genetic variation of a population. For instance, a positively selected allele can be lost due to chance. We now turn to the analysis of this type of phenomenon.

2 Hitting times

Let $l < a < b < r$ and

$$T^* = \min\{T_a, T_b\},$$

where T_y the first time y is reached.

Problem Statement. We consider the following two problems. For $a < x < b$, we seek to compute

$$u(x) = \mathbb{P}[T_b < T_a \,|\, X(0) = x], \tag{8}$$

and

$$v(x) = \mathbb{E}[T^* \,|\, X(0) = x]. \tag{9}$$

Problem Solution. Under the assumptions above, it can be shown that u and v are bounded and twice continuously differentiable. Moreover, they satisfy

$$0 = Lu, \qquad u(a) = 0, u(b) = 1, \tag{10}$$

and

$$-1 = Lu, \qquad v(a) = 0, v(b) = 0, \tag{11}$$

where the *infinitesimal generator* is

$$Lf(x) = \mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x). \tag{12}$$

Heuristic justification. We give a heuristic argument for the first problem. (The second one is similar.) Fix $a < x < b$. By (1), as $h \downarrow 0$ the probability of reaching a or b in time h is $o(h)$. Hence

$$\begin{aligned}
u(x) &\approx \mathbb{E}_x[u(X(h))] + o(h) \\
&\approx \mathbb{E}_x[u(x + \Delta_h X)] + o(h) \\
&\approx \mathbb{E}_x\left[u(x) + u'(x)\Delta_h X + \frac{1}{2}u''(x)(\Delta_h X)^2 + o(\Delta_h X)^2\right] + o(h) \\
&\approx u(x) + \mu(x)u'(x)h + \frac{1}{2}\sigma^2(x)u''(x)h + o(h).
\end{aligned}$$

For a formal proof using martingales, see [Dur96, (3.2) and (4.2) in Chapter 6].

3 Solving the equations

To solve an equation of the form

$$Lf(x) = C,$$

it useful to let $g = f'$ and multiply both sides by an integrating factor w

$$\begin{aligned}
Cw(x) &= \mu(x)g(x)w(x) + \frac{1}{2}\sigma^2(x)g'(x)w(x) \\
&= \frac{1}{2}\sigma^2(x)\left[\frac{2\mu(x)}{\sigma^2(x)}g(x)w(x) + g'(x)w(x)\right] \\
&= \frac{1}{2}\sigma^2(x)\frac{\mathrm{d}}{\mathrm{d}x}[g(x)w(x)],
\end{aligned}$$

where we need

$$w'(x) = \frac{2\mu(x)}{\sigma^2(x)}w(x),$$

that is,

$$w(x) = \exp\left(\int^x \frac{2\mu(y)}{\sigma^2(y)}\mathrm{d}y\right).$$

Then the solution can be found by integrating twice.

From this argument, it is natural to re-write the equations in the following way. Let

$$s(x) = \exp\left(-\int^x \frac{2\mu(y)}{\sigma^2(y)}\mathrm{d}y\right), \qquad S(x) = \int^x s(y)\mathrm{d}y,$$

and

$$m(x) = \frac{1}{\sigma^2(x)s(x)}, \qquad M(x) = \int^x m(y)\mathrm{d}y.$$

The functions S and m are called the *scale function* and *speed density* respectively. Then we have

$$Lf(x) = \frac{1}{2}\frac{1}{m(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{s(x)}\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right] = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}M}\left[\frac{\mathrm{d}}{\mathrm{d}S}f(x)\right].$$

To solve the first problem, integrate twice to obtain

$$u(x) = C_1 S(x) + C_2.$$

The boundary conditions give

$$u(x) = \frac{S(x) - S(a)}{S(b) - S(a)} = \frac{S[a,x]}{S[a,b]}.$$

Similarly, the solution of the second problem is

$$v(x) = \int_a^b G(x,y)\mathrm{d}y,$$

where

$$G(x,y) = \begin{cases} 2\frac{S[a,x]S[y,b]}{S[a,b]}m(y), & a \leq x \leq y \leq b, \\ 2\frac{S[a,y]S[x,b]}{S[a,b]}m(y), & a \leq y \leq x \leq b, \end{cases}$$

is the *Green function*.

4 Applications to the Wright-Fisher diffusion

EX 25.5 (No mutation/selection.) *Since $\mu(x) = 0$, we have*

$$s(x) = C_1, \qquad S(x) = C_1 x + C_2, \qquad m(x) = \frac{1}{x(1-x)C_1},$$

so that

$$G(x,y) = \begin{cases} \frac{2x}{y}, & 0 < x < y < 1, \\ \frac{2(1-x)}{(1-y)}, & 0 < y < x < 1, \end{cases}$$

and

$$\begin{aligned} v(x) &= 2(1-x)\int_0^x \frac{1}{1-y}\mathrm{d}y + 2x\int_x^1 \frac{1}{y}\mathrm{d}y \\ &= -2[(1-x)\log(1-x) + x\log x]. \end{aligned}$$

EX 25.6 (Additive selection) *Consider the case with additive selection and no mutation. Fix* $0 < a < x < b < 1$. *Then*

$$s(x) = \exp\left(-\int^x \frac{2\delta y(1-y)}{y(1-y)}\mathrm{d}y\right) = C_1 e^{-2\delta x},$$

and

$$S(x) = \int^x C_1 e^{-2\delta y}\mathrm{d}y = C_1 e^{-2\delta x} + C_2.$$

Hence

$$u(x) = \frac{e^{-2\delta x} - e^{2\delta a}}{e^{-2\delta b} - e^{-2\delta a}}.$$

Taking limits $a \downarrow 0$ *and* $b \uparrow 1$, *we have*

$$\mathbb{P}_x[T_1 < T_0] = \frac{1 - e^{-2\delta x}}{1 - e^{-2\delta}} \approx 2s,$$

where the last expression uses $x = \frac{1}{2N}$ *(for a new mutation) and* δ *large enough. In comparison, recall that a neutral mutation fixates with probability* $1/2N$.

EX 25.7 (One-way mutation) *Assume* $\beta_2 = 0$, $\beta_1 = \beta$ *and that there is no selection. Fix* $0 < a < x < b < 1$. *Then*

$$s(x) = \exp\left(\int^x \frac{2\beta y}{y(1-y)}\mathrm{d}y\right) = \frac{C_1}{(1-x)^{2\beta}},$$

and

$$S(x) = \int^x \frac{C_1}{(1-y)^{2\beta}}\mathrm{d}y = C_1\frac{1}{(1-x)^{2\beta-1}} + C_2.$$

Hence

$$u(x) = \frac{(1-x)^{-2\beta+1} - (1-a)^{-2\beta+1}}{(1-b)^{-2\beta+1} - (1-a)^{-2\beta+1}}.$$

Taking limits $a \downarrow 0$, *we have*

$$\mathbb{P}_x[T_b < T_0] = \frac{(1-x)^{-2\beta+1} - 1}{(1-b)^{-2\beta+1} - 1}.$$

So if $2\beta \geq 1$, *this probability goes to* 0 *as* $b \to 1$. *In other words, the mutation pressure is strong enough that the* 1 *boundary cannot be attained. For* $2\beta < 1$, *the limit is positive.*

Further reading

The material in this section was taken from Chapter 15 of [KT81] and Chapter 7 of [Dur08].

References

[Dur96] Richard Durrett. *Stochastic calculus*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996.

[Dur08] Richard Durrett. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer, New York, second edition, 2008.

[KT81] Samuel Karlin and Howard M. Taylor. *A second course in stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.