Evolutionary Dynamics Homework 6

Minghang Li

November 07, 2024 08:11 (Europe/Berlin)

Problem 1: One-dimensional Fokker-Planck equation

Consider the one-dimensional Fokker-Planck equation with constant coefficients,

$$\partial_t \psi(p,t) = -m\partial_p \psi(p,t) + \frac{v}{2}\partial_p^2 \psi(p,t), \tag{1}$$

with $p \in \mathbb{R}, v > 0$.

(a) See below.

Show that for vanishing selection, $m = 0$,

$$\psi(p,t) = \frac{1}{\sqrt{2\pi v t}} \exp\left(-\frac{p^2}{2vt}\right) \tag{2}$$

solves the Fokker-Planck equation. To which initial condition does this solution correspond?

For vanishing selection $m = 0$, the Fokker-Planck equation can be written as simply

$$\partial_t \psi(p,t) = \frac{v}{2}\partial_p^2 \psi(p,t) \tag{3}$$

Let's compute the left-hand side and the right-hand side to show that the solution is correct.

LHS of (3)

$$\begin{aligned}
\partial_t \psi(p,t) &= \partial_t \left(\frac{1}{\sqrt{2\pi v t}} \exp\left(-\frac{p^2}{2vt}\right)\right) \\
&= -\frac{1}{\sqrt{2\pi v t}\cdot 2t} \cdot \exp\left(-\frac{p^2}{2vt}\right) + \frac{1}{\sqrt{2\pi v t}} \cdot \frac{p^2}{2vt^2} \cdot \exp\left(-\frac{p^2}{2vt}\right) \\
&= \frac{p^2 - vt}{2vt^2} \cdot \frac{1}{\sqrt{2\pi v t}} \cdot \exp\left(-\frac{p^2}{2vt}\right)
\end{aligned} \tag{4}$$

RHS of (3)

$$
\begin{aligned}
\frac{v}{2}\partial_p^2\psi(p,t) &= \frac{v}{2}\partial_p\left(\partial_p\left(\frac{1}{\sqrt{2\pi vt}}\exp\left(-\frac{p^2}{2vt}\right)\right)\right)\\
&= \frac{v}{2\sqrt{2\pi vt}}\partial_p\left(\partial_p\left(\exp\left(-\frac{p^2}{2vt}\right)\right)\right)\\
&= \frac{v}{2\sqrt{2\pi vt}}\partial_p\left(\left(-\frac{p}{vt}\right)\cdot\exp\left(-\frac{p^2}{2vt}\right)\right)\\
&= \frac{v}{2\sqrt{2\pi vt}}\cdot\left(-\frac{1}{vt}\exp\left(-\frac{p^2}{2vt}\right)+\frac{p^2}{v^2t^2}\cdot\exp\left(-\frac{p^2}{2vt}\right)\right)\\
&= -\frac{1}{\sqrt{2\pi vt}\cdot 2t}\cdot\exp\left(-\frac{p^2}{2vt}\right)+\frac{1}{\sqrt{2\pi vt}}\cdot\frac{p^2}{2vt^2}\cdot\exp\left(-\frac{p^2}{2vt}\right)\\
&= \frac{p^2-vt}{2vt^2}\cdot\frac{1}{\sqrt{2\pi vt}}\cdot\exp\left(-\frac{p^2}{2vt}\right)
\end{aligned}
\tag{5}
$$

It's clear that (4) equals (5), so (2) is indeed the solution.

To investigate the initial condition, we need to examine the behavior of $\psi(p,t)$ as $t\to 0^+$. As $t\to 0^+$, we have:

- The exponential term $\exp(-p^2/(2vt))$ approaches:
  - $0, \forall p\neq 0$ becuase $-p^2/(2vt)\to-\infty$
  - 1 for $p=0$.
- The prefactor $\frac{1}{\sqrt{2\pi vt}}$ approaches ∞ for $t\to 0^+$.

This behavior is consistent with the Dirac delta function $\delta(p)$. Hence, the initial condition is

$$\psi(p,0)=\delta(p),$$

where $\delta(p)$ is the Dirac delta function defined as

$$\delta(p)=\begin{cases}0, & \text{if } p\neq 0,\\ \infty, & \text{if } p=0.\end{cases}$$

This means the initial distribution is concentrated at $p=0$, which represents the fact that the starting population is initially at a single allele frequency.

(b) Construct a solution for constant selection, $m\neq 0$, by substituting $z=p-mt$ for p in (1) . What is the mean and variance?

Let $z(p,t)=p-mt$, then we have $\partial_p z(p,t)=1$, and $\partial_t z(p,t)=-m$. Substitute $z(p,t)$ into $\psi(p,t)$ to obtain $\psi(z,t)$, using the chain rule, we have

$$
\begin{aligned}
\frac{\partial\psi(p,t)}{\partial t} &= \frac{\partial\psi(z,t)}{\partial z}\frac{\partial z}{\partial t}+\frac{\partial\psi(z,t)}{\partial t}\frac{\cancel{\partial t}}{\cancel{\partial t}}\\
&= -m\frac{\partial\psi(z,t)}{\partial z}+\frac{\partial\psi(z,t)}{\partial t}
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
\frac{\partial\psi(p,t)}{\partial p} &= \frac{\partial\psi(z,t)}{\partial z}\frac{\partial z}{\partial p}\\
&= \frac{\partial\psi(z,t)}{\partial z}
\end{aligned}
\tag{7}
$$

Using (6) and (7), we can rewrite the Fokker-Planck equation:

$$\begin{aligned}
\frac{\partial \psi(p,t)}{\partial t} &= -m\frac{\partial \psi(p,t)}{\partial p} + \frac{v}{2}\frac{\partial^2 \psi(p,t)}{\partial p^2} \\
\cancel{-m\frac{\partial \psi(z,t)}{\partial z}} + \frac{\partial \psi(z,t)}{\partial t} &= \cancel{-m\frac{\partial \psi(z,t)}{\partial z}} + \frac{v}{2}\frac{\partial^2 \psi(z,t)}{\partial z^2} \\
\frac{\partial \psi(z,t)}{\partial t} &= \frac{v}{2}\frac{\partial^2 \psi(z,t)}{\partial z^2}
\end{aligned} \tag{8}$$

Voila! This equation (8) has exactly the same form as (3), so the solution is of the form of (2) also. We use the same form and subsitute $z = p - mt$ into it:

$$\begin{aligned}
\psi(z,t) &= \frac{1}{\sqrt{2\pi vt}} \exp\left(-\frac{z^2}{2vt}\right) \\
\psi(p,t) &= \frac{1}{\sqrt{2\pi vt}} \exp\left(-\frac{(p-mt)^2}{2vt}\right) \sim \mathcal{N}(mt, vt)
\end{aligned}$$

Hence, the mean is mt and the variance is vt.

Problem 2: Diffusion approximation of the Moran process

Derive a diffusion approximation for the Moran process of two species. Assume the first species has a small selective advantages.

(a) See below.

The general definition for the drift coefficient is

$$M(p) = E[X(t) - X(t-1) | X(t-1) = i]/N, \tag{9}$$

where $p = i/N$ and $X(t)$ denotes the abundance of the first allele. Evaluate this expression for the Moran process with selection. Show that this yields the result for the Wright-Fisher process from the lecture, divided by N.

From lecture 3 we know that for a *neutral* Moran process we have transition probabilities:

- $P_{i,i+1} = p(1-p)$
- $P_{i,i-1} = (1-p)p$
- $P_{i,i} = p^2 + (1-p)^2$

However, since the first species has a small selective advantage (denoted as s hereafter), we need to modify the transition probabilities. We have:

- $P_{i,i+1} = (p(1+s)/(p(1+s) + (1-p))) \cdot (1-p) = (p(1+s)/(1+sp)) \cdot (1-p)$
- $P_{i,i-1} = (1-p)/(1+sp) \cdot p$
- $P_{i,i} = 1 - P_{i,i+1} - P_{i,i-1}$

And the expectation derived from the transition probabilities is:

$$
\begin{aligned}
E[X(t)|X(t-1)=i] &= i \cdot P_{i,i} + (i+1) \cdot P_{i,i+1} + (i-1) \cdot P_{i,i-1} \\
&= i \cdot (1 - P_{i,i+1} - P_{i,i-1}) + (i+1) \cdot P_{i,i+1} + (i-1) \cdot P_{i,i-1} \\
&= i + P_{i,i+1} - P_{i,i-1} \\
&= i + \frac{p(1+s)}{1+sp} \cdot (1-p) - \frac{1-p}{1+sp} \cdot p \\
&= i + \frac{sp(1-p)}{1+sp}
\end{aligned} \tag{10}
$$

Using (10) and (9), we can calculate the drift coefficient $M(p)$:

$$
\begin{aligned}
M(p) &= \frac{E[X(t) - X(t-1)|X(t-1)=i]}{N} \\
&= \frac{E[X(t)|X(t-1)=i] - E[X(t-1)|X(t-1)=i]}{N} \\
&= \frac{E[X(t)|X(t-1)=i] - i}{N} \\
&= \left(i + \frac{sp(1-p)}{1+sp} - i \right) / N \\
&= \frac{sp(1-p)}{N(1+sp)}
\end{aligned} \tag{11}
$$

From slide 33, we know that the drift coefficient for the Wright-Fisher process is:

$$
M_{WF}(p) = \frac{sp(1-p)}{1+sp},
$$

and (11) is exactly the result for the Wright-Fisher process divided by N.

(b) By a similar argument calculate the diffusion coefficient $V(p)$.

From slides, we have

$$
\begin{aligned}
V(p) &= E\left[Var[p(t+1)]|p(t)\right] \\
&= \frac{1}{N^2} E\left[Var[X(t+1)|X(t)]\right]
\end{aligned} \tag{12}
$$

To derive $Var[X(t+1)|X(t)]$ again, we first derive $Var[X(t+1)|X(t)=i]$:

$$
\begin{aligned}
Var[X(t+1)|X(t)=i] &= Var[X(t)] + Var[X(t+1)-X(t)|X(t)=i] \\
&= 0 + E[(X(t+1)-X(t))^2|X(t)=i] - E[X(t+1)-X(t)|X(t)=i]^2 \\
&= (i-1-i)^2 \cdot P_{i,i-1} + (i+1-i)^2 \cdot P_{i,i+1} + (i-i)^2 \cdot P_{i,i} - E[X(t+1)-X(t)|X(t)=i]^2 \\
&= P_{i,i-1} + P_{i,i+1} - E[X(t+1)|X(t)=i]^2 \\
&= \frac{p(1-p)}{1+sp} + \frac{(1+s)p(1-p)}{1+sp} - \left(\frac{sp(1-p)}{1+sp}\right)^2 \\
&= \frac{p(1-p)(s+2)}{1+sp} - \left(\frac{sp(1-p)}{1+sp}\right)^2 \\
&= p(1-p)\left(1 + \frac{1+s}{(1+sp)^2}\right)
\end{aligned}
\tag{13}
$$

As (13) is not related to i at all, we can say that the expression of $Var[X(t+1)|X(t)]$ is the same as (13). Plug (13) back to (12), we obtain the diffusion coefficient:

$$
V(p) = \frac{1}{N^2}\left(p(1-p)\left(1+\frac{1+s}{(1+sp)^2}\right)\right) \tag{14}
$$

(c) Use your results from (a) and(b) in the forward Kolmogorov equation to present a diffusion equation for the Moran model.

Plug in (11) and (14) into the forward Kolmogorov equation:

$$
\begin{aligned}
\frac{\partial \psi(p,t)}{\partial t} &= -\frac{\partial}{\partial p}\left[\psi(p,t)M(p)\right] + \frac{1}{2}\frac{\partial^2}{\partial p^2}\left[\psi(p,t)V(p)\right] \\
&= -\frac{\partial}{\partial p}\left[\psi(p,t)\frac{sp(1-p)}{N(1+sp)}\right] + \frac{1}{2}\frac{\partial^2}{\partial p^2}\left[\psi(p,t)\frac{1}{N^2}p(1-p)\left(1+\frac{1+s}{(1+sp)^2}\right)\right]
\end{aligned}
$$

(d) See below.

Now assume that $s \leq 1$. Approximate your results from (a) and (b) and use the general expression for the fixation probability $\rho(p_0)$ to show that the fixation probability is given by:

$$
\rho(p_0 = 1/N) = \frac{1-e^{-s}}{1-e^{-Ns}} \tag{15}
$$

The general expression for fixation probability $\rho(p_0)$ is (given in the slides) as:

$$
\rho(p_0) - \frac{\int_0^{p_0} \exp\left(-\int_0^p \frac{2M(q)}{V(q)}dq\right)dp}{\int_0^1 \exp\left(-\int_0^p \frac{2M(q)}{V(q)}dq\right)dp} \tag{16}
$$

Now we want to simplify the expression of $2M(q)/V(q)$. Since we have $s \leq 1$, we can write:

$$
\begin{aligned}
\frac{2M(q)}{V(q)} &= \frac{2sp(1-p)}{N(1+sp)} \Big/ \left(\frac{p(1-p)}{N^2} \left(1 + \frac{1+s}{(1+sp)^2} \right) \right) \\
&= \frac{2Ns}{1+sp} \Big/ \frac{(1+sp)^2 + (1+s)}{(1+sp)^2} \\
&= \frac{2Ns(1+sp)}{(1+sp)^2 + (1+s)} \\
&\approx \frac{2Ns \cdot 1}{1^2 + 1} && (p \text{ is also small}) \\
&= Ns
\end{aligned} \tag{17}
$$

Given (17), we have:

$$
\int_0^p \frac{2M(q)}{V(q)} = \int_0^p Nsdq = Nsp \tag{18}
$$

Plug (18) back to (16), we have:

$$
\begin{aligned}
\rho(p_0 = 1/N) &= \frac{\int_0^{p_0} \exp\left(-Ns\right) dp}{\int_0^1 \exp\left(-Ns\right) dp} \\
&= \frac{1 - \exp(-Nsp_0)}{1 - \exp(-Ns)} \ , \\
&= \frac{1 - e^{-s}}{1 - e^{-Ns}}
\end{aligned}
$$

which is exactly the form of (15).

(e) Take the limit to derive a result for the fixation probability of a neutral allele, $s = 0$. Evaluate (15) for $N = 10$ and $N = 1000$ for both positive and negative selection, $s = 2\%$, and $s = -2\%$, respectively. Compare your result with ρ_1 of the exact Moran process.

Fixation probability for neutral allele

$$
\lim_{s \to 0} \frac{1 - e^{-s}}{1 - e^{-Ns}} = \frac{1 - e^0}{1 - e^0} = 1
$$

Approximation v.s. exact Moran process

The author preferred to write code block (although sometimes get the parentheses wrong) to do numeric calcualtions.

```
library(tidyr)
library(dplyr)
library(kableExtra)
```

```
moran_fixation_prob <- function(N, s) {
  return((1 - 1 / (1+s)) / (1 - 1 / (1+s)^N))
}
```

```
diffusion_fixation_prob <- function(N, s) {
  return((1 - exp(-s)) / (1 - exp(-N * s)))
}
```

```r
rho_exact_N10_spos <- moran_fixation_prob(10, 0.02)
rho_exact_N10_sneg <- moran_fixation_prob(10, -0.02)
rho_exact_N1000_spos <- moran_fixation_prob(1000, 0.02)
rho_exact_N1000_sneg <- moran_fixation_prob(1000, -0.02)

rho_approx_N10_spos <- diffusion_fixation_prob(10, 0.02)
rho_approx_N10_sneg <- diffusion_fixation_prob(10, -0.02)
rho_approx_N1000_spos <- diffusion_fixation_prob(1000, 0.02)
rho_approx_N1000_sneg <- diffusion_fixation_prob(1000, -0.02)
```

```r
fixation_probs <- data.frame(
  Population_Size = c(10, 10, 1000, 1000),
  Selection_Coefficient = c(0.02, -0.02, 0.02, -0.02),
  Exact_Probability = c(rho_exact_N10_spos, rho_exact_N10_sneg,
                        rho_exact_N1000_spos, rho_exact_N1000_sneg),
  Diffusion_Approximation = c(rho_approx_N10_spos, rho_approx_N10_sneg,
                              rho_approx_N1000_spos, rho_approx_N1000_sneg)
)

fixation_probs %>%
  kable(format = "latex",
        col.names = c("N", "s", "Exact ( )", "Diffusion Approx. ( )"),
        digits = 6,
        caption = "Fixation Probabilities: Exact vs Diffusion Approximation") %>%
  kable_styling(latex_options = c("striped", "hold_position"),
                full_width = FALSE) %>%
  row_spec(0, bold = TRUE)
```

Table 1: Fixation Probabilities: Exact vs Diffusion Approximation

N	**s**	**Exact (ρ)**	**Diffusion Approx. (ρ)**
10	0.02	0.109144	0.109237
10	-0.02	0.091156	0.091242
1000	0.02	0.019608	0.019801
1000	-0.02	0.000000	0.000000

Problem 3: Absorption time in the diffusion approximation

In the diffusion approximation of a process with absorbing states 0 and 1 the expected fixation time, conditioned on absorption in state 1,reads:

$$\tau_1(p_0) = 2(S(1) - S(0)) \left(\int_{p_0}^{1} \frac{\rho(p)(1-\rho(p))}{e^{-A(p)}V(p)} dp + \frac{1-\rho(p_0)}{\rho(p_0)} \int_0^{p_0} \frac{\rho(p)^2}{e^{-A(p)}V(p)} dp \right)$$

where $\rho(p)$ denotes the fixation probability,

$$A(p) = \int_0^p \frac{2M(p)}{V(p)} dp, \quad \text{and} \quad S(p) = \exp(-A(p))dp.$$

(a) Calculate the *conditional expected waiting time for fixation*, $\tau_1(p_0)$, of an allele of frequency p_0 in the neutral Wright-Fisher process. Approximate the result for small p_0.

The original formula seems intimidating, so there must be some way to make it simpler. Fortunately, we have the following information:

- $S(0) = 0$, by definition.
- For neutral Wright-Fisher process, $M(p) = 0$, hence $A(p) = 0, \forall p$ (From slides, $M(p) = p(1-p)s/(1+ps)$, and $s \to 0$ for neutral Wright-Fisher process).
- $S(1) = 1$ follows the previous point.
- $V(p) = p(1-p)/N$ as from slide 19.
- $\rho(p) = p$ for neutral Wright-Fisher process.

With the above information, we can simplify the formula for $\tau_1(p_0)$:

$$\begin{aligned}
\tau_1(p_0) &= 2\left(\int_{p_0}^{1} \frac{p(1-p)}{e^0 p(1-p)/N} dp + \frac{1-p_0}{p_0}\int_0^{p_0} \frac{p^2}{e^0 p(1-p)/N} dp\right) \\
&= 2\left(\int_{p_0}^{1} N dp + \frac{1-p_0}{p_0}\int_0^{p_0} \frac{Np}{1-p} dp\right) \\
&= 2\left(N(1-p_0) + \frac{N(1-p_0)}{p_0}\left(-p_0 - \log(1-p_0)\right)\right) \\
&= 2\left(N(1-p_0) - N(1-p_0) - \frac{N(1-p_0)\log(1-p_0)}{p_0}\right) \\
&= -\frac{2N(1-p_0)\log(1-p_0)}{p_0}
\end{aligned}$$

For small p_0, we can use the Taylor expansion of $\log(1-p_0)$:

$$\begin{aligned}
\tau_1(p_0) &= -\frac{2N(1-p_0)(-p_0 - p_o^2/2 - O(p_0^3))}{p_0} \\
&= 2N(1-p_0)(1 + p_0/2 + O(p_0^2)) \\
&= 2N(1 + p_0/2 + O(p_0^2)) - 2N(p_0 + O(p_0^2)) \\
&\approx 2N(1 - p_0/2) \\
&\approx 2N \qquad\qquad \text{(for small } p_0 \ll 1\text{)}
\end{aligned}$$

Therefore, for small p_0, the conditional expected waiting time for fixation is approximately $2N$ generations.

(b) Compute τ_0, the conditional expected waiting time until *extinction* (absorption in state 0) in the neutral Wright-Fisher process. Also derive the unconditioned expected waiting time $\bar{\tau}$ until either fixation or extinction.

If the current allele of interest (say A) goes to extinction, it means the other allele a goes to fixation. If A has initial abundance p_0, then a has initial abundance $1 - p_0$.

$$\begin{aligned}
\tau_0(p_0) &= \tau_1(1-p_0) \\
&= -\frac{2N(1-(1-p_0))\log(1-(1-p_0))}{1-p_0} \\
&= -\frac{2Np_0\log(p_0)}{1-p_0}
\end{aligned}$$

For the unconditioned expected waiting time $\bar{\tau}(p_0)$

- The probability of fixation is p_0
- The probability of extinction is $1 - p_0$

Therefore:

$$\begin{aligned}
\bar{\tau}(p_0) &= p_0\tau_1(p_0) + (1-p_0)\tau_0(p_0) \\
&= p_0\left(-\frac{2N(1-p_0)\log(1-p_0)}{p_0}\right) + (1-p_0)\left(-2N\left(\frac{p_0\log(p_0)}{1-p_0}\right)\right) \\
&= -2N(1-p_0)\log(1-p_0) - 2Np_0\log(p_0) \\
&= -2Np_0(1-p_0)[\log(1-p_0)/p_0 + \log(p_0)/(1-p_0)]
\end{aligned}$$

This is the general formula for the unconditional expected waiting time until absorption in either state 0 or 1.