Evolutionary Dynamics: Homework 06

Guido Putignano, Lorenzo Tarricone, Gavriel Hannuna,
Athanasia Sapountzi

February 22, 2024

Problem 1: One-dimensional Fokker-Planck equation

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

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

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

(a) Vanishing Selection, $m = 0$

Show that for vanishing selection, $m = 0$, the following solution satisfies the Fokker-Planck equation:

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

Solution
To show it we have to calculate the derivative of the function respect to t and p. Before doing that it's important to remember the product rule equation $\frac{d}{dx}(u \cdot v) = \frac{du}{dx} \cdot v + u \cdot \frac{dv}{dx}$.

Derivative respect to t

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

$$\partial_t \left(\frac{1}{\sqrt{2\pi vt}} \exp\left(-\frac{p^2}{2vt}\right)\right) = \left(\frac{p^2}{2vt^2\sqrt{2\pi vt}} \exp\left(-\frac{p^2}{2vt}\right)\right) - \left(\frac{v}{2vt\sqrt{2\pi vt}} \exp\left(-\frac{p^2}{2vt}\right)\right)$$
$$= \left(\frac{-vt + p^2}{(2vt)^2\sqrt{2\pi vt}}\right) \exp\left(-\frac{p^2}{2vt}\right)\Bigg)$$

Derivative respect to p

$$\partial_p\left(\frac{1}{\sqrt{2\pi vt}}\exp\left(-\frac{p^2}{2vt}\right)\right) = -\left(\frac{p}{2vt\sqrt{2\pi vt}}\exp\left(-\frac{p^2}{2vt}\right)\right).$$

$$\partial_p^2\left(\frac{1}{\sqrt{2\pi vt}}\exp\left(-\frac{p^2}{2vt}\right)\right) =$$

$$-\left(\frac{1}{vt\sqrt{2\pi vt}}\exp\left(-\frac{p^2}{2vt}\right)\right) - \left(\frac{p}{vt\sqrt{2\pi vt}}\cdot\frac{-p}{vt}\exp\left(-\frac{p^2}{2vt}\right)\right).$$

$$\frac{v}{2}\partial_p^2\psi(p,t) = \left(\frac{-vt+p^2}{(2vt)^2\sqrt{2\pi vt}}\right)\exp\left(-\frac{p^2}{2vt}\right)\Bigg).$$

From here we obtain

$$\left(\frac{-vt+p^2}{(2vt)^2\sqrt{2\pi vt}}\right)\exp\left(-\frac{p^2}{2vt}\right) = \left(\frac{-vt+p^2}{(2vt)^2\sqrt{2\pi vt}}\right)\exp\left(-\frac{p^2}{2vt}\right)$$

$$0 = 0$$

Hence we show that the previous solution satisfies the Fokker-Planck equation

(b) Constant Selection, $m \neq 0$

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

Solution

From here it's possible to solve the exercise by remembering the chain rule equation $\frac{\partial y}{\partial x_i} = \sum_{\ell=1}^{m}\frac{\partial y}{\partial u_\ell}\frac{\partial u_\ell}{\partial x_i}$.

As we are taking into account $m \neq 0$, the text defines $z(p,t) = p - mt$. For this reason $\frac{\partial z}{\partial p} = 1$ and $\frac{\partial z}{\partial t} = -m$. Let $\Psi(p,t) = \psi(z(p,t),t)$.

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

$$\frac{\partial\Psi(p,t)}{\partial p} = \frac{\partial\psi(z,t)}{\partial z}\frac{\partial z}{\partial p} = \frac{\partial\psi(z,t)}{\partial z}$$

$$\Rightarrow \frac{\partial^2\Psi(p,t)}{\partial p^2} = \frac{\partial^2\psi(z,t)}{\partial z^2}$$

$$-m\frac{\partial\psi(z,t)}{\partial z} + \frac{\partial\psi(z,t)}{\partial t} = -m\frac{\partial\psi(z,t)}{\partial z} + \frac{v}{2}\frac{\partial^2\psi(z,t)}{\partial z^2}$$

$$\Rightarrow \frac{\partial\psi(z,t)}{\partial t} = \frac{v}{2}\frac{\partial^2\psi(z,t)}{\partial z^2}$$

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

What is noticeable is that as it's possible to remember from a normal distribution, there is a shift over time. Moreover, it's possible to see that the variance increases over time.

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 advantage s.

(a) Drift Coefficient for the Moran Process with Selection

The general definition for the drift coefficient is

$$M(p) = E[X(t) - X(t-1)|X(t-1) = i]/N, \tag{3}$$

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. (tutorial question)

(b) Diffusion Coefficient for the Moran Process

By a similar argument, calculate the diffusion coefficient $V(p)$. Use your findings to set up the diffusion equation for the Moran model.

Solution

Given the Variance equation described at slide 7 in class we have

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

Using the same reasoning of the tutorial question (a):

$$\begin{aligned} &= \frac{1}{N^2}(\frac{i(N-i)}{N}\frac{r+1}{ri-i+N} - (\frac{i(N-i)}{N}\frac{r-1}{ri-i+N})^2) \\ &= \frac{1}{N^2}(p(1-p)\frac{s+2}{ps+1} - (\frac{sp(1-p)}{ps+1})^2) \\ &= \frac{1}{N^2}(p(1-p)\frac{s+1}{(ps+1)^2+1}) \end{aligned}$$

To set up the diffusion equation for the Moran model, we compute the Fokker-Planck equation (Kolmogorov forward equation):

$\frac{\partial \psi(p,t)}{\partial t} = -\frac{\partial}{\partial p}\left[\psi(p,t)M(p)\right] + \frac{1}{2}\frac{\partial^2}{\partial p^2}\left[\psi(p,t)V(p)\right]$

We replace $M(p)$ with the solution from the tutorial question (a) and $V(p)$ with the solution above

Thus, we calculate:

$\frac{\partial \psi(p,t)}{\partial t} = -\frac{\partial}{\partial p}\left[\psi(p,t)\frac{ps(1-p)}{(1+sp)N}\right] + \frac{1}{2}\frac{\partial^2}{\partial p^2}\left[\psi(p,t)\frac{1}{N^2}(p(1-p)\frac{s+1}{(ps+1)^2+1})\right]$

(c) Fixation Probability for Small s

Now assume that $s \ll 1$. 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:

$$\rho(p_0 = 1/N) = (1 - e^{-s})/(1 - e^{-Ns}). \tag{4}$$

Evaluate Eq. (4) for $N = 10$ and $N = 1000$ for both positive ($s = 1.5\%$) and negative selection ($s = -1.5\%$), respectively. Compare your results with ρ_1 of the exact Moran process.

Solution

The general expression for the fixation probability $\rho(p_0)$ is:

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

Since we assume that $s \ll 1$ we compute:

$\frac{2M(q)}{V(q)} = \frac{2sp(1-p)}{(1+ps)N} * \frac{N^2(1+ps)^2}{p(1-p)((s+1)+(1+ps)^2)} = \frac{2Ns(1+ps)}{(s+1)+(1+ps)^2} \sim \frac{2Ns*1}{1+1} = Ns$

Thus $\int_0^{p_0} \frac{2M(q)}{V(q)} dq = Nsp_0$

Therefore,

$\rho(p_0 = 1/N) = \frac{\int_0^{p_0} \exp(-Nsp) dp}{\int_0^1 \exp(-Nsp) dp} = \frac{1-\exp(-Nsp_0)dp}{1-\exp(-Ns)dp} = \frac{1-e^{-s}}{1-e^{-Ns}}$

We calculate the ρ_1 of the exact Moran process:

$\rho_1 = \frac{1-\frac{1}{r}}{1-\frac{1}{r^N}} = \frac{1-\frac{1}{1+s}}{1-\frac{1}{(1+s)^N}}$

For $N = 10$ and for positive selection ($s = 1.5\%$) we calculate:

$\rho(p_0 = 1/10) = \frac{1-e^{-0.015}}{1-e^{-10*0.015}} = 0.107$

$\rho_1 = \frac{1-\frac{1}{1+0.015}}{1-\frac{1}{(1+0.015)^{10}}} = 0.107$

For $N = 10$ and for negative selection ($s = -1.5\%$) we calculate:

$\rho(p_0 = 1/10) = \frac{1-e^{0.015}}{1-e^{10*0.015}} = 0.093$

$\rho_1 = \frac{1-\frac{1}{1-0.015}}{1-\frac{1}{(1-0.015)^{10}}} = 0.093$

For $N = 1000$ and for positive selection ($s = 1.5\%$) we calculate:

$\rho(p_0 = 1/1000) = \frac{1-e^{-0.015}}{1-e^{-1000*0.015}} = 0.015$

$\rho_1 = \frac{1-\frac{1}{1+0.015}}{1-\frac{1}{(1+0.015)^{1000}}} = 0.015$

For $N = 1000$ and for negative selection ($s = -1.5\%$) we calculate:

$\rho(p_0 = 1/1000) = \frac{1-e^{0.015}}{1-e^{1000*0.015}} = 4 * 10^{-9}$

$\rho_1 = \frac{1-\frac{1}{1-0.015}}{1-\frac{1}{(1-0.015)^{1000}}} = 4 * 10^{-9}$

We observe that the numerical results of the fixation probability and the Moran process are the same.

(d) Fixation Probability for Neutral Allele

Take the limit to derive a result for the fixation probability of a neutral allele with $s = 0$.

Solution

To calculate the fixation probability of a neutral allele with $s = 0$, we take the following limit and we solve it using L'Hôpital's Rule:
$\lim_{s\to 0} \frac{1-e^{-s}}{1-e^{-Ns}} = \frac{1}{N}$

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:

$$\tau_1(p_0) = 2(S(1){-}S(0))\left(\int_{p_0}^{1} \frac{\rho(p)(1-\rho(p))}{e^{-A(p)}V(p)}dp + \frac{1-\rho(p_0)}{\rho(p_0)}\int_0^{p_0} \rho\frac{(p)^2}{e^{-A(p)}V(p)}dp\right), \tag{5}$$

where $\rho(p)$ denotes the fixation probability, $A(p) = \int_0^p 2M(p')/V(p')dp'$, and $S(p) = \int_0^p e^{-A(p')}dp'$.

(a) Conditional Expected Fixation Time in Neutral Wright-Fisher Process

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

In order to explicitly find the expression for $\tau_1(p_0)$ we want first to use some results obtained in class regarding the Wright-Fisher process and in general natural processes.

- $S(0) = 0$ by definition of s
- For the **neutral** Wright-Fisher process $M(p) = 0 \forall p \in [0,1]$ which implies $A(p) = 0 \forall p \in [0,1]$ [1]
- $S(1) = 1$ [2]
- $V(p) = \frac{p(1-p)}{N}$ as derived in class
- $\rho(p) = p$ again thanks to neutrality as seen during the lecture

This allows us to rewrite the expression as:

$$\tau_1(p_0) = 2\left(\int_{p_0}^{1} N dp + \frac{1-p_0}{p_0}\int_0^{p_0} \frac{pN}{1-p}dp\right)$$

[1]We know that in general $M(p) = \frac{p(1-p)s}{1+ps}$ which limit goes to zero as $s \to 0$ (neutral case)

[2]Using the fact that $A(p) = 0$ by definition $S(1) = \int_0^1 1dp = 1$

After collection the factor N we can focus on solving the second interval, which leads to:

$$\int_0^{p_0} \frac{pN}{1-p} dp = \int_0^{p_0} -1 - \frac{1}{p-1} dp = -p_0 - ln(|p_0 - 1|) = -p_0 - ln(1-p_0)$$

Where in the last passage we use the fact that $p_0 \in [0, 1]$.

Now the equation becomes:

$$2N \left[1 - p_0 + \frac{1-p_0}{p_0} [-p_0 - ln(1-p_0)] \right] = -2N \frac{1-p_0}{p_0} ln(1-p_0)$$

Notably, this coincides with the expression given during the lecture

(b) Conditional Expected Waiting Time in Neutral Wright-Fisher Process

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 I want my first allele to go extinct we need the second allele to reach fixation. This will follow the same dynamics derived above, but starting from an initial abundance of $1 - p_0$ instead of p_0. The expression becomes therefore:

$$\tau_0(p_0) = \tau_1(1-p_0) = -2N \frac{p_0}{1-p_0} ln(p_0)$$

For the unconditioned waiting time, the question itself seems vague. If we define the time in which at least one of the two may happen is very the minimum between $\tau_0(p_0)$ and $\tau_1(p_0)$ This can be seen graphically in the plot below

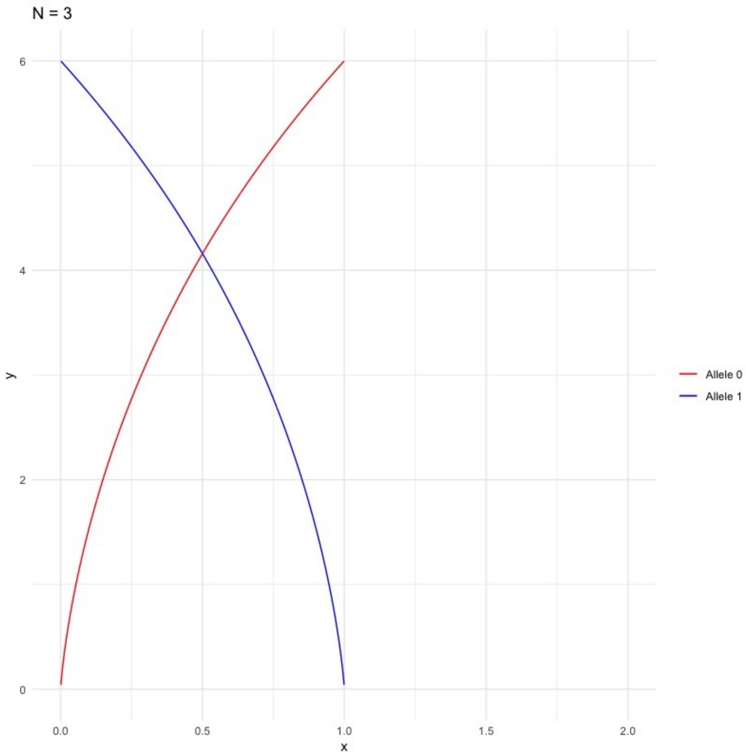

Figure 1: Graphical solution for $\overline{\tau}$

We notice that independently from the value of N chosen, the intersection will always be for $p_0 = 0.5$ therefore $\overline{\tau}$ will be:

$$\begin{cases} \tau_0 & if\ p_0 \in [0, 0.5) \\ \tau_1 & if\ p_0 \in [0.5, 1] \end{cases} . \tag{6}$$

If we define unconditioned expected waiting time $\overline{\tau}$ as the average waiting time until either fixation or extinction, without specifying which one will occur, considering both outcomes and providing the average time it takes for the process to reach either state, we have:

$$\bar{\tau} = \tau_1 \rho_1 + \tau_0 \rho_0 = -2N \left((1 - p_0) \log (1 - p_0) + p_0 \log (p_0) \right)$$

Plotting the result

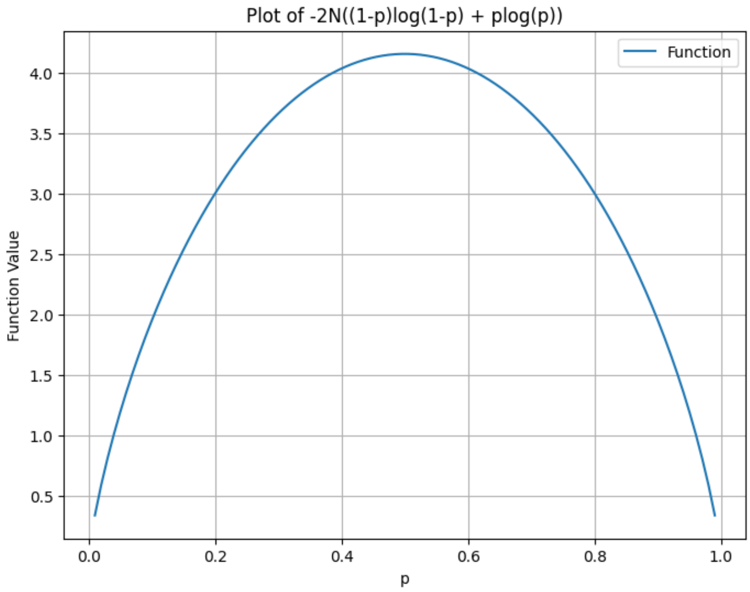

Figure 2: Graphical solution for $\bar{\tau}$ given N=3

(c) Comparison with Numerical Simulations

Compare your analytical results for the absorption times τ_1, τ_0, and $\bar{\tau}$ with those from numerical simulations of the neutral Wright-Fisher process. Use $N = 100$ individuals and initial frequencies of $p_0 = 0.5$ and $p_0 = 1/N$. Perform 1,000 simulations for each case (or more), and ensure that the simulation time is sufficiently long.

Solution

When doing our numerical simulation, we can use the formula indicated previously for τ_0 to find the appropriate time (number of generations) to perform our simulation. With that time, we can be sure that in most of the trajectories we will reach one of the 2 absorption states (0 or N).

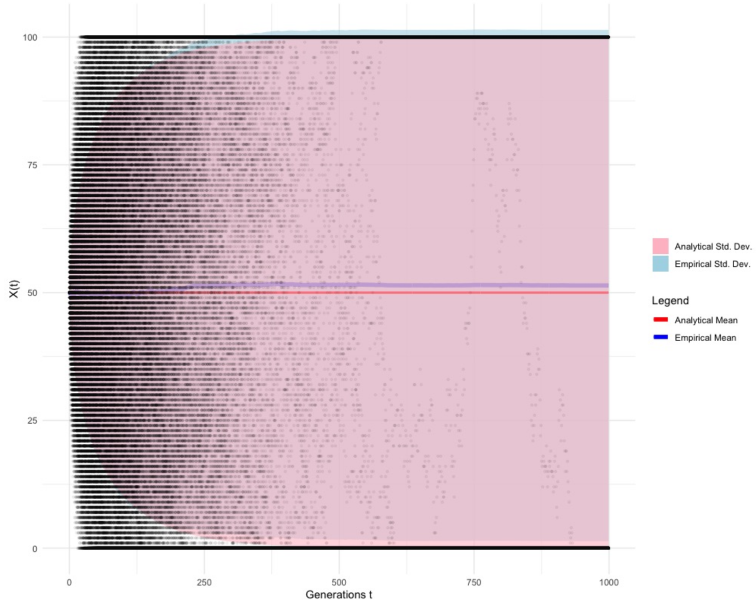

Figure 3: WF process simulation of 1000 trajectories, $p_0 = 0.5$

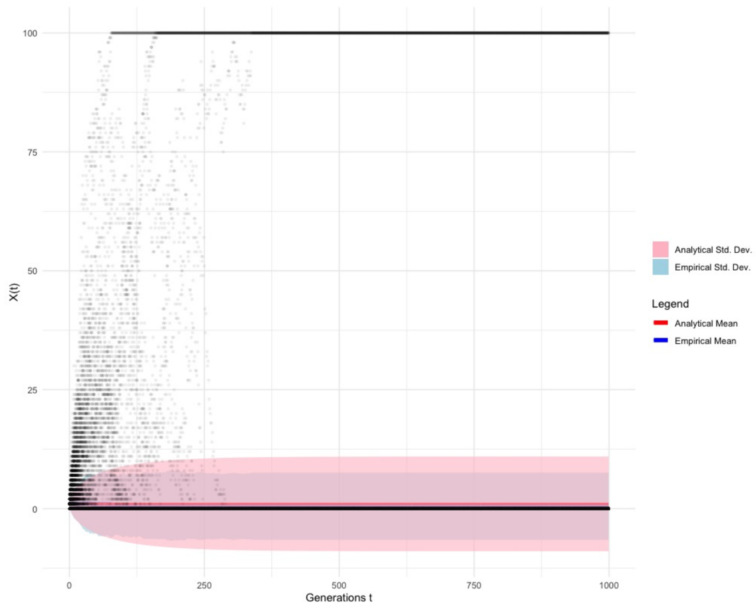

Figure 4: WF process simulation of 1000 trajectories, $p_0 = 0.01$

When simulating, we can calculate the average fixation time of alleles 0 and 1 over the 1000 trajectories. Here are examples of values that we can reach:

$$\tau_0(p_0 = 0.5) = 137.0226$$

$$\tau_1(p_0 = 0.5) = 137.7802$$

$$\tau_0(p_0 = 0.01) = 9.953629$$

$$\tau_1(p_0 = 0.01) = 201.00$$

As we can see, the fixation times for p_0 are quite similar between themselves, and fluctuate enough so that sometimes τ_0 is bigger than τ_1, or vice-versa. The probability of one being bigger than the other is the same, this indicates an equal probability of reaching each of the 2 fixation states, a situation very different from when we start the simulation at $p_0 = 0.01$, which has a much higher tendency to reach the fixation state at 0 instead of N.
Those τs are calculated by averaging the position of the first encounter of one of the fixation states in all of the trajectories, while the value that we get using the formula shown in section b ensures us that in all of the simulation (trajectories), we will reach one of the fixation states. The following τ_0s are the ones calculated analytically, which are an approximation of the time needed for fixation.

$$\tau_0(p_0 = 0.5) = \tau_1(p_0 = 0.5) = 138.63$$

$$\tau_0(p_0 = 0.01) = 9.3$$

$$\tau_1(p_0 = 0.01) = 198.996$$

Although τ is calculated through various approximations, the resulting empirical values are extremely similar to the analytical ones.
Given the explanation described here, $\overline{\tau}$ can be readily derived from the consequential calculations of τ_1 and τ_0 outlined in point (b). In this context, our primary focus lies in the comparison of the latter, as $\overline{\tau}$ emerges as a result of these calculations.