Solutions 4: Finite number fluctuations and random walks, models with discrete time

BIO-369

Prof. Anne-Florence Bitbol
EPFL

1 Simulation of the Moran process

a) Because total population size is fixed at N individuals, the state of the population is completely described by the number i of A individuals. Indeed, the number of B individuals is given by $N-i$. The number i can take values from 0 to N. Upon each step, the number i can either:

- increase by 1, if an individual of type A divides while a B dies,
- decrease by 1, if an individual of type B divides while an A dies,
- stay constant, if the individual that divides and the one that dies are of the same type.

b) Upon a given step, the number i increases by 1 if an individual of type A divides while a B dies, which occurs with probability

$$\alpha_i = \frac{ri}{ri+N-i} \times \frac{N-i}{N}\,, \tag{1}$$

where the first term is the probability that an A individual is chosen to divide, and the second one is the probability that a B individual is chosen to die.

Upon a given step, the number i decreases by 1 if an individual of type B divides while an A dies, which occurs with probability

$$\beta_i = \frac{N-i}{ri+N-i} \times \frac{i}{N}\,, \tag{2}$$

where the first term is the probability that a B individual is chosen to divide, and the second one is the probability that an A individual is chosen to die.

c) Upon a given step, the number i stays constant if the individual that divides and the one that dies are of the same type. This occurs with probability

$$\frac{ri}{ri+N-i} \times \frac{i}{N} + \frac{N-i}{ri+N-i} \times \frac{N-i}{N}\,, \tag{3}$$

where the first term of the sum is the probability that an A individual is chosen to divide, and that an A individual is chosen to die, and the second term of the sum is the same for B individuals. One can check that

$$\frac{ri}{ri+N-i} \times \frac{i}{N} + \frac{N-i}{ri+N-i} \times \frac{N-i}{N} = 1-\alpha_i-\beta_i\,, \tag{4}$$

This makes sense: as there is no fourth possibility, by normalization, the probability that i stays constant is $1-\alpha_i-\beta_i$.

d) If $r=1$ the above expressions show that $\alpha_i=\beta_i$: thus, at each time step, it is as likely for i to increase by one or to decrease by one. This correspond to the case of a neutral mutant, i.e. $f_A=f_B$. The random walk has no bias in this case, i.e. there is no natural selection. In this case, any variation in the proportion of mutant organisms will just arise from finite size fluctuations associated to birth and death events (genetic drift), while in the presence of fitness differences, both natural selection and genetic drift will come into play.

e) To choose whether an A or a B individual is going to die, we draw a uniformly distributed random number between 0 and 1 and then we compare it to i/N: if it is smaller than i/N then an A dies, while if it is larger, then a B dies. This procedure is consistent with the Moran model, because then, the probability that an A dies is indeed i/N. The same reasoning holds for division.

f) The final values of i that are obtained are 0 and N. Intermediate outcomes are not possible. Indeed, eventually, after a sufficient number of generations, because N is finite, all individuals will be descended from just one single ancestor, and thus one of the two types will take over (we say that it fixes) and the other one will disappear.

g) Trajectories of i versus t observed until type A disappears or takes over are very noisy. This corresponds to a random walk.

h) See Jupyter notebook.

2 Fixation probability and fixation time

a) See Jupyter notebook.

b) See Jupyter notebook.

c) We find that p_A is very close to $1/N$. In the neutral case, each individual in the initial population is as likely to take over (i.e. fix), and thus each of them has a probability $1/N$ to fix, meaning that in the end all individuals in the population are descended from it. So in fact, $p_A = 1/N$. Hence, the ability of a neutral mutant to take over in a population strongly depends on population size, and is larger for smaller populations.

We observe that t_f, t_{fA} and t_{fB} all seem to increase with N, and t_{fA} is much larger than others and seems to increase faster with N than others. This makes sense as fixation of A corresponds to a trajectory where the number of A goes from 1 to N, and thus it takes more time than when B fixes since B just needs to go from 1 to 0 (but can do so in very meandering ways). Whether A or B takes over, fixation takes longer in a larger population because there are more intermediate compositions that can be explored between the two extremes.

d) We observe that p_A is very small for deleterious mutants ($r < 1$), and then substantially increases as r is increased above 1. This demonstrates the effect of natural selection.

We observe that t_{fB} seems to feature a maximum around $r = 1$. Qualitatively, in that case, trajectories are not biased and tend to meander for a long time. The behavior of t_{fA} is more complicated to interpret because data is very scarce or lacking for $r < 1$ due to the fact that p_A is very small in this case. But in fact it also has a similar maximum. Finally, the behavior of t_f is a combination of those described until now, since $t_f = p_A t_{fA} + (1 - p_A) t_{fB}$.

e) A good match with the analytical prediction for p_A is obtained.

3 Additional problem: Transcription factor moving on DNA

a) Using the Boltzmann distribution, we obtain:

$$\frac{P(E_2)}{P(E_1)} = \exp\left(-\frac{E_2 - E_1}{k_\mathrm{B}T}\right) = \exp\left(-\frac{\Delta E}{k_\mathrm{B}T}\right) . \tag{5}$$

b) Using the binding energy between a transcription factor and a binding site on DNA $\Delta E = 10 k_\mathrm{B}T$, and the formula above, we obtain:

$$\frac{P(E_2)}{P(E_1)} = \exp(-10) = 4.5 \times 10^{-5} , \tag{6}$$

The bound state (state 1) where the transcription factor is bound to the binding site is much more likely than the unbound state (state 2). Thus it is much more difficult for the transcription factor to unbind from the binding site than to bind to it.

c) The non-specific binding energy should be substantially smaller than the specific one, $\Delta E = 10 k_{\rm B} T$. (This means that the transcription factor can unbind from non-specific sites rather easily – but we ignore this point here and just consider the fact that it can jump to other non-specific sites.)

d) The motion of the transcription factor on DNA is a random walk or diffusion, more precisely a one-dimension random walk or diffusion. This problem is analogous to the Moran model. The Moran model is also a random walk in one dimension, but the quantity studied is very different (number of mutants vs. position of the transcription factor).

e) If we start from one transcription factor at site i with $1 \leq i < N$, after a sufficiently long time, either it exits the DNA fragment on the left, or it binds to the specific transcription factor binding site on the right. Once these edges are reached, the transcription factor stays there for ever. Due to finite size fluctuations, one of these edges has to be reached in finite time.

f) Consider the first jump of a transcription factor starting at site i with $1 \leq i < N$: it can be to the site immediately on the right, site $i+1$, with probability α, or to the site immediately on the left, site $i-1$, with probability $1-\alpha$.

g) Denoting by ρ_i the probability that the transcription factor finally binds to the specific binding site when it starts at site i with $1 \leq i < N$, and discriminating over the possibilities at the first jump seen in the previous question, we can write the following equation relating ρ_i to ρ_{i-1} and ρ_{i+1}:

$$\rho_i = \alpha \rho_{i+1} + (1-\alpha) \rho_{i-1} \, . \tag{7}$$

h) Because the transcription factor falls off the DNA segment if it reaches 0, we have $\rho_0 = 0$. Because the transcription factor has to move to the binding site if it reaches site N, we have $\rho_N = 1$.

i) Introducing $y_i = \rho_i - \rho_{i-1}$, Eq. 7 becomes

$$y_{i+1} = \frac{y_i}{r} \, , \tag{8}$$

with $r = \alpha/(1-\alpha)$. Thus,

$$y_i = \frac{\rho_1}{r^{i-1}} \, , \tag{9}$$

for $1 \leq i \leq N$.

j) We have

$$\sum_{i=1}^{N} y_i = \rho_N - \rho_0 = 1 \tag{10}$$

and using Eq. 9, we have

$$\sum_{i=1}^{N} y_i = \rho_1 \frac{1-r^{-N}}{1-r^{-1}} \, , \tag{11}$$

for $\alpha \neq 1/2$, and

$$\sum_{i=1}^{N} y_i = \rho_1 N \, , \tag{12}$$

for $\alpha = 1/2$. Therefore, we obtain

$$\rho_1 = \frac{1-r^{-1}}{1-r^{-N}} \tag{13}$$

for $\alpha \neq 1/2$, and

$$\rho_1 = 1/N \, , \tag{14}$$

for $\alpha = 1/2$. The impact of α is to bias the random walk: if $\alpha > 1/2$, it is more likely for the transcription factor to go right than to go left, and therefore this makes ρ_1 larger.