

Evolutionary Dynamics Homework 5

Minghang Li

October 30, 2024 20:10 (America/New_York)

Problem 1: Pathways of carcinogenesis

Consider three independent mutations $\{1, 2, 3\}$. Each mutation occurs after an exponentially distributed waiting time $T_i \sim \exp(\lambda_i), i = 1, 2, 3$.

(a) What is the probability for the path $P = 3 \rightarrow 1 \rightarrow 2$?

Since the mutations are independent, let Exit_j denote the set of all possible mutations in step j , then we have:

- $\text{Exit}_1 = \{1, 2, 3\}$.
- $\text{Exit}_2 = \{1, 2\}$ (as 3 has already occurred).
- $\text{Exit}_3 = \{2\}$.

$$\begin{aligned}\text{Prob}(P) &= \prod_{j=1}^k \frac{\lambda_{i_j}}{\sum_{i \in \text{Exit}_j} \lambda_i} \\ &= \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot \frac{\lambda_2}{\lambda_2} \\ &= \frac{\lambda_1 \cdot \lambda_3}{(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_2 + \lambda_3)}\end{aligned}$$

(b) Assume cancer arises if any two of the three genes are mutated. How many possible genotypes are there? How many pathways? Compute the expected waiting time until any two out of three genes are mutated.

If cancer arises if any two of the three genes are mutated, then the possible genotypes are

$$\text{number of genotypes} = \binom{3}{2} = 3,$$

which are $\{1, 2\}$, $\{1, 3\}$, and $\{2, 3\}$.

Since the order of mutation does not matter, the number of pathways is:

$$\text{number of pathways} = \text{number of genotypes} \cdot 2! = 3 \cdot 2 = 6.$$

The expected waiting time of 2 out of 3 independent mutations is:

$$E(\tau_2) = \sum_P \text{Prob}(P) E(\tau_P),$$

where we have

$$\text{Prob}(P) = \prod_{j=1}^k \frac{\lambda_{i_j}}{\sum_{i \in \text{Exit}_i} \lambda_i}, \quad E(\tau_P) = \sum_{j=1}^k \frac{1}{\sum_{i \in \text{Exit}_i} \lambda_i}.$$

For the ease of typing, λ_1 , λ_2 , and λ_3 will be denoted as A , B , C respectively. Then we have:

$$\begin{aligned} E(\tau_2) &= \sum_P \text{Prob}(P) E(\tau_P) \\ &= \frac{AB}{(A+B+C)(B+C)} \left(\frac{1}{A+B+C} + \frac{1}{B+C} \right) \\ &\quad + \frac{AC}{(A+B+C)(B+C)} \left(\frac{1}{A+B+C} + \frac{1}{B+C} \right) \\ &\quad + \frac{BA}{(A+B+C)(A+C)} \left(\frac{1}{A+B+C} + \frac{1}{A+C} \right) \\ &\quad + \frac{BC}{(A+B+C)(A+C)} \left(\frac{1}{A+B+C} + \frac{1}{A+C} \right) \\ &\quad + \frac{CA}{(A+B+C)(A+B)} \left(\frac{1}{A+B+C} + \frac{1}{A+B} \right) \\ &\quad + \frac{CB}{(A+B+C)(A+B)} \left(\frac{1}{A+B+C} + \frac{1}{A+B} \right) \\ &= \frac{A(B+C)}{(A+B+C)(B+C)} \left(\frac{1}{A+B+C} + \frac{1}{B+C} \right) \\ &\quad + \frac{B(A+C)}{(A+B+C)(A+C)} \left(\frac{1}{A+B+C} + \frac{1}{A+C} \right) \\ &\quad + \frac{C(A+B)}{(A+B+C)(A+B)} \left(\frac{1}{A+B+C} + \frac{1}{A+B} \right) \\ &\quad \text{(Rewrite A, B, C back to } \lambda_1, \lambda_2, \lambda_3 \dots) \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \left(\frac{1}{\lambda_1 + \lambda_2 + \lambda_3} + \frac{1}{\lambda_2 + \lambda_3} \right) \\ &\quad + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \left(\frac{1}{\lambda_1 + \lambda_2 + \lambda_3} + \frac{1}{\lambda_1 + \lambda_3} \right) \\ &\quad + \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \left(\frac{1}{\lambda_1 + \lambda_2 + \lambda_3} + \frac{1}{\lambda_1 + \lambda_2} \right) \end{aligned}$$

(c) Now consider d independent mutations. How many paths exist leading to the genotype $\{1, \dots, d\}$ with all mutations present? If cancer already arises after any k mutations, how many different paths are there?

Since all mutations are independent, the number of paths lead to the genotype $\{1, \dots, d\}$ is $d!$.

If cancer already arises after k mutations, then we need to:

1. Choose k mutation out of d
2. Permute the k mutations

to calculate the number of different paths. Therefore, the number of different paths is:

$$\binom{d}{k} \cdot k! = \frac{d!}{k!(d-k)!} \cdot k! = \frac{d!}{(d-k)!}.$$

Problem 2: Neutral Wright-Fisher process

Consider the neutral Wright-Fisher process for a system of N cells of two different types $\{A, B\}$. Let $X(t)$ denote the number of A -cells at time t . The process has the transition matrix

$$P_{i,j} = \text{Prob}[X(t) = j | X(t-1) = i] = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$$

that is, $X(t) | X(t-1) = i$ is binomially distributed with parameter $p = i/N$.

(a) Compute the conditional expectation $E[X(t) | X(0) = i]$.

(b) Compute the conditional variance $\text{Var}[X(t) | X(0) = i]$.

Hint: Show that

$$\text{Var}[X(t) | X(0) = i] = V_1 + \left(1 - \frac{1}{N}\right) \text{Var}[X(t-1) | X(0) = i],$$

where $V_1 = \text{Var}[X(1) | X(0) = i]$. You can then use the expression above to derive the final result (no explicit calculation needed for this last step).

(c) Derive an approximation for $\text{Var}[X(t) | X(0) = i]$ for large population size N . Compare the variance of the Wright Fisher process to the variance of the Moran process, explain the difference(s).

(d) See below.

Show that in the Wright-Fisher process, the heterozygosity H_t at time t satisfies

$$E[H_t | X_0 = i] = H_0(i) \left(1 - \frac{1}{N}\right)^t,$$

and hence decreases exponentially at rate $1/N$. Compare this behaviour with the Moran model.

Note: Heterozygosity in this context is defined as the probability that two individuals chosen at random from the population are of different types.

Problem 3: Wave approximation

Consider the wave approximation of the Wright-Fisher model for cancer progression. Here, the growth of a clone with j mutations is given by

$$\dot{x}_j = s x_j (j - \langle j \rangle).$$

For small times, the average fitness $s\langle j \rangle = s \sum_j x_j$ can be considered constant. Use this throughout your calculations.

- (a) Find the analytic solution for the initial condition $x_j(0) = 1/N$.
- (b) The rate at which an additional mutation in a clone with j mutations occurs is given by $udx_j(t)$. Find the time τ when the cumulative probability exceeds $1/N$.
- (c) Compute the waiting time until the next mutation for a mutation rate $u = 10^{-7}$ per cell generation, $d = 80$ genes and a fitness advantage of $s = 1.15\%$ per mutation. Use that $j - \langle j \rangle \approx \sqrt{\ln N}$ with $N = 10^7$ cells and assume a cell generation time of 1 day.