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 $P = 2 \to 1 \to 3$? **(1 point)**

(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. **(1 point)**

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

Problem 1

independant mutations $\{1,2,3\}$

waiting time: $T_i \sim e^{\lambda i}$, $i=1,2,3$

a) what is the probability for the path $P=2\to1\to3$?

The path undergoes 3 stages:

$Exit_1 = \{1,2,3\}$

$Exit_2 = \{1,2\}$

$Exit_3 = \{2\}$

Probability given by:

$$p(path) = \prod_{i=1}^{k} \frac{\lambda_{i,j}}{\sum \lambda_j}$$

$$= \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} \cdot \frac{\lambda_1}{\lambda_1+\lambda_2} \cdot \frac{\lambda_2}{\lambda_2}$$

b) How many possible genotypes are there?
Expected waiting time any 2 of the 3 genotypes are mutated.

- Number of genotypes is binomal distributed: $\binom{d}{k} = \binom{3}{2} = 3$
- Number of posible pathways: $k! = 3\cdot2\cdot1 = 6$

Computation of expected waiting time:

ce,

$E[\tau_k] = \frac{1}{\lambda}\sum_{j=1}^{k}\frac{1}{d-j+1}$

$$E[\tau_k] = \frac{1}{\lambda}\sum_{j=1}^{k}\frac{1}{d-j+1}$$

$$= \frac{1}{\lambda}\left(\frac{1}{3-1+1}+\frac{1}{3-2+1}+\frac{1}{3-3+1}\right) = \frac{1}{\lambda}\left(\frac{1}{3}+\frac{1}{2}+\frac{1}{1}\right) = \frac{11}{6\lambda}$$ hmm...

c) d independant mutations

# of paths to genotype $\{1, \dots, d\}$: $P_{tot} = d!$

# of paths to genotype with k out of d mutations : $P_k = \binom{d}{k} k!$

paths are there? (1 point)

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

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

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

problem 2

- neutra wright-Fisher process
- System of N cells
- Types = $\{A, B\}$
- $X(t) = n_{At}$

Transition probability:

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

b) compute conditional variance $Var[X(t) \mid X(0) = i]$

$V_t := Var[X(t) \mid X(0) = i]$; $(X(1) \mid X(0) = i) \sim Bin(N, \frac{i}{N})$

$\Rightarrow V_1 = Var[X(1) \mid X(0) = i] = N \cdot \frac{i}{N}\left(1 - \frac{i}{N}\right) = i\left(1 - \frac{i}{N}\right)$

Since binomial distribution all time steps are identically distributed.

$$\Rightarrow Var[X(t) | X(t-1) = i] = V_1 \quad \text{given} \quad E[X(t) | X(t-1)=i] = i$$

use law of total variance: $A = X(t),\ B = (X(t-1) = i)$

$$Var(A) = E_B[Var_A(A|B)] + Var_B[E_A(A|B)]$$

$$\Leftrightarrow Var(A) = E[Var[X(t) | X(t-1) = i]] + Var[E[X(t) | X(t-1) = i]]$$

$$= E[V_1] + Var(i)$$

$$= E[B(1-\tfrac{B}{N})] + Var(B)$$

$$= E[B] - E[\tfrac{B^2}{N}] + E[B^2] - E[B]^2$$

$$= E[B] + (1-\tfrac{1}{N})E[B^2] - E[B]^2 = (1-\tfrac{1}{N})E[B]^2 + \tfrac{1}{N}E[B]^2$$

$$= E[B] + (1-\tfrac{1}{N})E[B^2] - (1-\tfrac{1}{N})E[B]^2 - \tfrac{1}{N}E[B]^2$$

$$= E[B](1-\tfrac{1}{N}E(B)) + (1-\tfrac{1}{N})Var(B)$$

$X(0) = i,$

$$\Rightarrow V_t = i(1-\tfrac{i}{N}) + (1-\tfrac{1}{N})V_{t-1} = V_1 + (1-\tfrac{1}{N})V_{t-1}$$

Solve for V_t:

$$V_t - NV_1 = (1-\tfrac{1}{N})^{t-1}(V_1 - NV_1$$

$$V_t = NV_1(1-(1-\tfrac{1}{N})^t)$$

c) approximation for $Var[X(t)|X(0)=i]$ for large population size N

- compare variance to moran process

Depends on if V_1 is constant or not

(i) Constant V_1

$1-\frac{1}{N}=\frac{N-1}{N}$

$$\lim_{N\to\infty} V_t = V_1 \lim_{N\to\infty} \frac{1-(1-\frac{1}{N})^t}{\frac{1}{N}} = \{\text{l'hôpital}\} = V_1 \lim_{N\to\infty} \frac{① \frac{d}{dN} 1-(1-\frac{1}{N})^t}{② \frac{d}{dN} \frac{1}{N}} = \begin{cases} ① \; 0 - \frac{t\left(\frac{N-1}{N}\right)^{t-1}}{N^2} \\ ② \; -\frac{1}{N^2} \end{cases}$$

$$= V_1 \lim_{N\to\infty} t\left(\frac{N-1}{N}\right)^{t-1} = V_1 \cdot t$$

Same approximation for moran process, but V_1 differs

$V_{1,m} = \frac{2i}{N}(1-\frac{i}{N})$; $V_{1,WF} = i(1-\frac{i}{N})$

$\frac{V_{1WF}}{V_{1m}} = \frac{N}{2}$

(ii) V_1 not constant

$$\lim_{N\to\infty} V_t = \lim_{N\to\infty} V_1 \frac{1-(1-\frac{1}{N})^t}{\frac{1}{N}} = \lim_{N\to\infty} i(1-\tfrac{i}{N})\frac{1-(1-\frac{1}{N})^t}{\frac{1}{N}} = \lim_{N\to\infty} \frac{\frac{d}{dN}(1-\frac{i}{N})(1-(1-\frac{1}{N})^t) \; ①}{\frac{d}{dN} \frac{1}{N} \; ②}$$

$$\left\{ \begin{array}{l} ① = \frac{d}{dN} 1-(1-\frac{1}{N})^t - \frac{i}{N} + \frac{i}{N}(1-\frac{1}{N})^t = \frac{t\left(\frac{N-1}{N}\right)^{t-1}}{N^2} + \frac{\frac{i}{N^2} + i(t-N+1)\left(\frac{N-1}{N}\right)^{t-1}}{N^3} \\ ② -\frac{1}{N^2} \end{array} \right\}$$

$$= \lim_{N\to\infty} \frac{\left(\frac{t\left(\frac{N-1}{N}\right)^{t-1}}{N^2} + \frac{\frac{i}{N^2} + i(t-N+1)\left(\frac{N-1}{N}\right)^{t-1}}{N^3}\right)}{N^2} = \quad t + it$$

for Moran: $\lim_{N\to\infty} V_t = V_1 t$; $V_1 = \frac{2i}{N(1-\frac{i}{N})}$