Problem Set 3, Problem 2: Population Genetics

Vasiliki Kostara, Karim El-Nahass

February 2023

The probability distribution of time T_n to the first coalescent event backwards in time with n alleles can be expressed as:

$$P(T_n = T) = \lambda_n \, e^{-\lambda_n T} \quad \text{with coalescence rate } \lambda_n = \frac{1}{N}\binom{n}{2}$$

The probability distribution of the number of mutations j on a genealogy is:

$$P(S_n = j) = \frac{(\mu T_c)^j}{j!} \, e^{-\mu T_c} \quad \text{with rate } \mu T_c \text{ , where } T_c = \sum_{j=2}^{n} j \, T_j$$

a)

The probability of n alleles to be identical with non-correlated coalescent events, F_n, is the expected value of $P(S_n = 0)$, $n = 2,\ 3,\ ...$, at time $T = T_n$, where:

$$P(S_n = 0) = e^{-\mu T_c} = e^{-\mu \sum_{j=2}^{n} j \, T_j} = \prod_{n \geq 2} e^{-\mu n T_n}$$

Therefore, F_n will be expressed as:

$$F_n = \prod_{n \geq 2} \int_0^{+\infty} e^{-\mu n T_n} \cdot P(Tn) \, dT_n = \prod_{n \geq 2} \int_0^{+\infty} e^{-\mu n T_n} \binom{n}{2} \, e^{-\binom{n}{2}\frac{T_n}{N}} \, \frac{dT_n}{N}$$

Let the transformation $T_n = Nt \longrightarrow dT_n = N\,dt$. Then, F_n becomes:

$$\begin{aligned}
F_n &= \prod_{n\geq 2} \binom{n}{2} \int_0^{+\infty} e^{-\left[\mu N n + \binom{n}{2}\right]\, t}\, dt = \\
&= \prod_{n\geq 2} \left[-\frac{\binom{n}{2}}{\mu N n + \binom{n}{2}} \left[e^{-\left[\mu N n + \binom{n}{2}\right]\, t} \right]_0^{+\infty} \right] = \\
&= \prod_{n\geq 2} \frac{\binom{n}{2}}{\mu N n + \binom{n}{2}} = \prod_{n\geq 2} \frac{\frac{n!}{2!\,(n-2)!}}{\frac{n\theta}{2} + \frac{n!}{2!\,(n-2)!}} = \\
&= \prod_{n\geq 2} \frac{n\,(n-1)\,(n-2)!}{n\theta\,(n-2)! + n\,(n-1)\,(n-2)!} = \\
&= \prod_{n\geq 2} \frac{n-1}{\theta + n - 1} \Longleftrightarrow
\end{aligned}$$

$$\Longleftrightarrow F_n = \frac{(n-1)!}{(1+\theta)\,(2+\theta)\,...\,(n-2+\theta)\,(n-1+\theta)}\,, \qquad n = 2,\, 3,\, ...$$

b)

The probability of two alleles to have j number of mutations with non-correlated coalescent events, F_2, is the expected value of $P(S_2 = j)$ at time $T = T_2$, where:

$$T_c = 2T_2 \quad \text{and} \quad P(T_2) = \frac{e^{-T_2}}{N}$$

Therefore, F_2 will be expressed as:

$$F_2 = \int_0^{+\infty} P(S_2 = j)\,\cdot\,P(T_2)\,dT_2 \;=\; \int_0^{+\infty} \frac{(2\mu T_2)^j}{j!}\, e^{-\frac{(2\mu N+1)T_2}{N}}\, \frac{dT_2}{N}$$

Let the transformation $T_2 = N\,t \longrightarrow dT_2 = N\,dt$. Then, F_2 becomes:

$$F_2 = \int_0^{+\infty} \frac{(\theta t)^j}{j!}\, e^{-(\theta+1)t}\, dt = \frac{\theta^j}{j!} \int_0^{+\infty} t^j\, e^{-(\theta+1)t}\, dt =$$

$$= \frac{\theta^j}{j!} \left[\left[-\left(\frac{e^{-(\theta+1)t}\, t^j}{\theta+1} \right) \right]_0^{+\infty} + \frac{j}{\theta+1} \int_0^{+\infty} t^{j-1}\, e^{-(\theta+1)t}\, dt \right] =$$

$$= \frac{\theta^j}{j!} \left[\left[-\left(\frac{e^{-(\theta+1)t}\, t^j}{\theta+1} \right) \right]_0^{+\infty} + \left[-\left(\frac{e^{-(\theta+1)t}\, t^{j-1}}{\theta+1} \right) \right]_0^{+\infty} + \right.$$
$$\left. + \frac{j\,(j-1)}{(\theta+1)^2} \int_0^{+\infty} t^{j-2}\, e^{-(\theta+1)t}\, dt \right] =$$

$$= \frac{\theta^j}{j!} \left[\left[-\left(\frac{e^{-(\theta+1)t}\, t^j}{\theta+1} \right) \right]_0^{+\infty} + \left[-\left(\frac{e^{-(\theta+1)t}\, t^{j-1}}{\theta+1} \right) \right]_0^{+\infty} + \ldots + \left[-\left(\frac{e^{-(\theta+1)t}\, t}{\theta+1} \right) \right]_0^{+\infty} \right] +$$
$$+ \frac{\theta^j}{j!} \frac{j!}{(\theta+1)^j} \int_0^{+\infty} e^{-(\theta+1)t}\, dt =$$

$$= -\frac{\theta^j}{j!(\theta+1)} \left[\sum_{i=1}^{j} e^{-(\theta+1)\, t}\, t^i \right]_0^{+\infty} + \left(\frac{\theta}{\theta+1} \right)^j \int_0^{+\infty} e^{-(\theta+1)t}\, dt$$

We will now show how the first term in the above expression becomes zero:

$$
\begin{aligned}
\left[\sum_{i=1}^{j} e^{-(\theta+1)\,t}\,t^i\right]_0^{+\infty} &= \lim_{t\to+\infty}\sum_{i=1}^{j} e^{-(\theta+1)\,t}\,t^i - \lim_{t\to 0}\sum_{i=1}^{j} e^{-(\theta+1)\,t}\,t^i = \\
&= \sum_{i=1}^{j}\lim_{t\to+\infty} e^{-(\theta+1)\,t}\,t^i - \sum_{i=1}^{j}\lim_{t\to 0} e^{-(\theta+1)\,t}\,t^i = \\
&= \sum_{i=1}^{j}\lim_{t\to+\infty}\frac{t^i}{e^{(\theta+1)\,t}} - 0 = \\
&= \sum_{i=1}^{j}\lim_{t\to+\infty}\frac{i\,t^{i-1}}{(\theta+1)\,e^{(\theta+1)\,t}} = \ldots = \\
&= \sum_{i=1}^{j}\lim_{t\to+\infty}\frac{i!}{(\theta+1)^i\,e^{(\theta+1)\,t}} = 0
\end{aligned}
$$

Then, the value of the integral in the second term will be:

$$
\int_0^{+\infty} e^{-(\theta+1)t}\,dt = -\frac{1}{\theta+1}\left[e^{-(\theta+1)t}\right]_0^{+\infty} = \frac{1}{\theta+1}
$$

After inserting the above derivations into the last expression of F_2, it becomes:

$$
F_2 = \left(\frac{\theta}{\theta+1}\right)^j \frac{1}{\theta+1}
$$