FFR110 - Homework Problem 3 - Part 2

Computational Biology

Felix Waldschock
000420-T398

Chalmers University of Technology
February 2024

Contents

1 Population genetics

1.1 a) Derive an expression for P(Sn = 0), i.e. the probability to not have any SNPs in a sample of size n.

Problem 2 a

F_2 describes 2 alleles to

in prev. generation

$$F_2^{(t+1)} = \underbrace{(1-\mu)^2}_{\text{no mutation}} \cdot \left[\frac{1}{N} + \underbrace{\left(1 - \frac{1}{N}\right)}_{\text{diffrent parents}} F_2^{(t)} \right]$$

prob 2 child are same

same parent

parents are same

for $\mu = 0$ and $N \to \infty$

this leads to:

$$F_2 = \frac{1}{1+\theta}$$

What we want to find is $P(S_n = 0)$ which describes the probability of n alleles not being the same.

from the lecture notes:

$$P(S=j) = \frac{(\mu T_c)^j}{j!} e^{-\mu T_c} \qquad \text{where } T_c = \sum_{j=2}^{n} j T_j \qquad (1)$$

number of possible pairs

(2)

and $P(T_j) = \lambda_j e^{-\lambda_j T_j}$ with $\lambda_j = \dfrac{\binom{j}{2}}{N}$

Time to coalescent event backward in time when starting with j lines.

now $P(S_n = 0)$:

$$P(S_n = 0) = e^{\mu T_C} = e^{\mu \sum_{j=2}^{n} j \cdot T_j}$$

rewrite as product:

$$\prod_{n \geq 2} e^{-\mu T_n n}$$

now integrate this probability from $0 \to \infty$ to get F_n.

$$F_n = \prod_{n \geq 2} \int_0^\infty e^{-\mu \cdot T_n \cdot n} \cdot P(T_n) \cdot dT_n$$

now insert (2)

$$= \prod_{n \geq 2} \int_0^\infty e^{-\mu T_n n} \cdot \frac{1}{N} \binom{n}{2} e^{-\binom{n}{2} \frac{T_n}{N}} dT_n$$

$$= \prod_{n \geq 2} \binom{n}{2} \int_0^\infty e^{-\mu T_n n} \cdot e^{-\binom{n}{2} \frac{T_n}{N}} \frac{1}{N} dT_n$$

simplify the exp expression with $T_n = N \cdot t$; $e^{-\mu N t n}$ & $e^{-\binom{n}{2} t}$; $dT_n \to dt$

$$= \prod_{n \geq 2} \binom{n}{2} \int_0^\infty e^{-t \left(\mu N n + \binom{n}{2}\right)} dt .$$

now:

$$\binom{n}{2} = \frac{n(n-1)}{2} \; ; \; \prod_{n \geq 2} \binom{n}{2} = n-1 \qquad (3)$$

solve the integral: $\int_0^\infty e^{-nx} = -\frac{e^{-nx}}{n}$ gives:

$$\prod_{n\geq 2} \left(-\frac{\binom{n}{2}}{\mu N n + \binom{n}{2}} \left[e^{-(\mu N n + \binom{n}{2})t} \right]_0^\infty \right)$$

with (3) this gives:

$$= \prod_{n\geq 2} \frac{\cancel{n}(n-1)\cancel{(n-2)!}}{\cancel{n}\,\theta\,\cancel{(n-2)!} + \cancel{n}(n-1)\cancel{(n-2)!}}$$

$$= \prod_{n\geq 2} \frac{(n-1)}{\theta + n - 1}$$

$$\prod \frac{n-1}{\theta+n-1} = \underline{\underline{\frac{(n-1)!}{(1+\theta)(2+\theta)\dots(n-1+\theta)}}}$$

1.2 b) Derive the distribution of the number of SNPs in a sample of size n = 2.

$$P(S_2 = j) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^j \tag{1}$$

Problem 2b

Derive distribution of the number of SNPs in a sample of size n=2.

should find:

$$\boxed{P(S_2 = j) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^j} \qquad (2b)$$

from lecture notes:

$$T_c = \sum_{j=2}^{n} j\, T_j \quad ; \quad \text{with } S_2 = j \Rightarrow T_c = 2T_2$$

$$P(T_j) = \lambda_j \cdot e^{-\lambda_j T_j} \quad \text{with } \lambda_j = \frac{\binom{j}{2}}{N} \quad ; \quad j=2 \Rightarrow P(T_j) = \frac{1}{N} \cdot e^{-T_2}$$

F_2 is found when integrating the product of the 2 above:

$$F_2 = \int_0^\infty \underbrace{P(S_2 = j)}_{\substack{\text{expression}\\ \text{from a)}}} \cdot \underbrace{P(T_2)}_{\substack{\text{also}\\ \text{see a)}}} dT_2 = \int_0^\infty \frac{(2\mu T_2)^j}{j!} e^{-\frac{(2\mu N+1)T_2}{N}} \cdot \frac{1}{N} \cdot dT_2$$

transform $T_2 = N \cdot t$ leads to:

$$= \int_0^\infty \frac{(2\mu N \cdot t)^j}{j!} \cdot e^{-(2\mu N+1)\cdot t} \cdot dt$$

now we rewrite $2\mu T_2$ $\qquad \theta = 2N \cdot \mu$ (lecture notes p. 7)

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

$\parallel$ take t independent variables out of $\int$

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

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

this will give new integrals until $j=0$ and $t^j = 1$

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

as we want to find (2b) there are some terms that seem familiar. Therefore compute $(*)$.

$$\int_0^\infty e^{-(\theta+1)t}\,dt = -\frac{1}{\theta+1}\underbrace{\left[e^{-(\theta+1)t}\right]_0^\infty}_{1/0} = \frac{1}{1+\theta}$$

this looks again familar.

Therefore lets see if $(\#)$ cancels out.

$$\underbrace{-\frac{\theta^j}{j!\,(\theta+1)}}_{!0}\underbrace{\left[\sum_{k=1}^{j} e^{-(\theta+1)t}\,t^k\right]_0^\infty}_{?0}$$

$$\left[\sum_{k=1}^{j} e^{-(\theta+1)t}\,t^k\right]_0^\infty = \lim_{t\to\infty}\sum_{k=1}^{j} e^{-(\theta+1)t}\,t^j - \underbrace{\lim_{t\to 0}\sum_{k=1}^{j} e^{-(\theta+1)t}\,t^k}_{=0}$$

$$\lim_{t\to\infty}\sum_{k=1}^{j} e^{-(\theta+1)t}\,t^k = \sum_{k=1}^{j}\lim_{t\to\infty} e^{-(\theta+1)t}\,t^k$$

new look with l'Hospital

$$t^k \xrightarrow{'} k \cdot t^{k-1}$$

$$e^{-(\theta+1)t} = \frac{1}{e^{-(\theta+1)t}} \xrightarrow{'} \left((\theta+1)\left(e^{(\theta+1)t}\right)\right)^{-1}$$

$$\sum_{k=1}^{j} \lim_{t\to\infty} \frac{k \cdot t^{k-1}}{(\theta+1)\left(e^{(\theta+1)t}\right)}$$

and again l'Hospital ... leads to

$$\sum_{k=1}^{j} \lim_{t\to\infty} \frac{i!}{(\theta+1)^i\left(e^{(\theta+1)t}\right)} = \underline{0}$$

So this then shows:

$$= -\frac{\theta^j}{j!\,(\theta+1)} \underbrace{\cancel{\left[\sum_{k=1}^{j} e^{-(\theta+1)t}\, t^k\right]_0^\infty}}_{=0} + \left(\frac{\theta}{\theta+1}\right)^j \int_0^\infty e^{-(\theta+1)t}\, dt$$

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