

Løsningsforslag ST2301 Øving 11

Kapittel 6

Exercise 2

Hvorfor er ikke prosessen med genetisk drift tilsvarende det å kaste kron/mynt mange ganger med sannsynligheten for kron lik p ? I såfall ville vi i det lange løp forventet å se en andel p kron istedet for å få bare kron eller bare mynt ("fiksering"). Hvor bryter analogien sammen?

Svar:

Genetisk drift innebærer at sannsynligheten for å trekke et allel A fra gamet-poolen en gitt generasjon avhenger av hva som skjedde tidligere generasjoner. Når man kaster kron og mynt avhenger ikke kastene av resultatene fra tidligere kast.

Exercise 5

Ettersom et sjeldent allel blir fiksert ved tilfeldig drift, vil heterozygositeten først øke og deretter minke. Hvordan kan dette forklares samtidig med at f alltid øker over tid?

Svar:

Dersom p_t tilfeldigvis øker kan det virke uforenlig med at F_{Aa} skal avta. p_t er en stokastisk variabel med forventning og varians

$$\begin{aligned} E[p_t] &= p_0 \\ \text{Var}(p_t) &= p_0(1-p_0) \left[1 - \left(1 - \frac{1}{2N} \right)^t \right] \end{aligned}$$

gitt at vi bare kjenner p_0 . Det gir at

$$\begin{aligned}
E[P_{Aa}] &= E[2p_t(1-p_t)] \\
&= 2E[p_t] - 2E[p_t^2] \\
&= 2E[p_t] - 2\text{Var}(p_t) - 2E[p_t]^2 \\
&= 2p_0 - 2p_0(1-p_0) \left[1 - \left(1 - \frac{1}{2N}\right)^t \right] - 2p_0^2 \\
&= 2p_0 - 2p_0 + 2p_0^2 + 2p_0(1-p_0) \left(1 - \frac{1}{2N}\right)^t - 2p_0^2 \\
&= 2p_0(1-p_0) \left(1 - \frac{1}{2N}\right)^t
\end{aligned}$$

Innavlskoeffisienten følger uttrykket

$$f_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

Denne forteller sannsynligheten for IBD. Har at

$$P_{Aa} = 2p_0(1-p_0)(1-f_t)$$

Dermed følger at $E[P_{Aa}] = P_{AB}$.

$$E[P_{Aa}] = 2p_0(1-p_0)(1-f_t)$$

Det er altså ingen motsetning mellom tilfeldig økning i frekvensen av et sjeldent allel og økning i f .

Exercise 7

Hos pelsel vil kanskje 10% av alle hanner reproducere i hver generasjon. Hvor mye påvirker dette den effektive populasjonsstørrelsen?

Svar:

La populasjonsstørrelsen være N , og anta at det er like mange hanner og hunner. Bare 10% av alle hanner reproducerer, men alle hunner reproducerer. Det gir

$$\begin{aligned}
N_m &= \frac{1}{10} \frac{N}{2} \\
N_f &= \frac{N}{2}
\end{aligned}$$

Effektiv populasjonsstørrelse blir

$$\begin{aligned} N_e &\approx \frac{1}{\frac{1}{4N_m} + \frac{1}{4N_f}} \\ &= \frac{1}{\frac{20}{4N} + \frac{2}{4N}} \\ &= \frac{N}{5 + \frac{1}{2}} \\ &= \frac{2N}{11} \\ &\approx 0.18N \end{aligned}$$

Den effektive populasjonsstørrelsen er altså mye mindre enn den faktiske størrelsen N .

Exercise 10

Se på en stor populasjon med størrelse N . Hver generasjon vil halvparten av individene tilfeldigvis finne gode reirplasser og ha gjennomsnittlig fertilitet på 3, mens den andre halvparten må klare seg med dårlige reirplasser og har fertilitet på 1. Hver generasjon dør foreldrene, og det er ingen korrelasjon mellom hvor foreldrene og avkommene finner reirplasser. Hva er den effektive populasjonsstørrelsen som funksjon av N ?

Svar:

Sannsynligheten for fertiliteten (antall videreførte allel n_i) til individ nummer i er gitt ved

$$\begin{aligned} Pr(n_i = 3) &= \frac{1}{2} \\ Pr(n_i = 1) &= \frac{1}{2} \end{aligned}$$

Forventa antall videreførte allel for individ i er dermed

$$E[n_i] = \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 1 = 2$$

Siden hvert individ har to foreldre, betyr dette at populasjonsstørrelsen vil være lik N alle generasjoner. Sannsynligheten for at to tilfeldig valgte allel kommer fra samme individ er

$$Pr(I = J) = \sum_{i=1}^N \frac{n_i(n_i - 1)}{2N(2N - 1)}$$

Den effektive populasjonsstørrelsen er gitt ved (VI-55 side 210)

$$\begin{aligned}\frac{1}{N_e} &= \mathbb{E} \left[\sum_{i=1}^N \frac{n_i(n_i - 1)}{2N(2N - 1)} \right] \\ &= \frac{\sum_{i=1}^N \mathbb{E}[n_i^2] - \mathbb{E}[n_i]}{2N(2N - 1)}\end{aligned}$$

Har at

$$\mathbb{E}[n_i^2] = \frac{1}{2} \cdot 3^2 + \frac{1}{2} \cdot 1 = 5$$

Det gir

$$\begin{aligned}\frac{1}{N_e} &= \frac{\sum_{i=1}^N \mathbb{E}[n_i^2] - \mathbb{E}[n_i]}{2N(2N - 1)} \\ &= \frac{\sum_{i=1}^N 4}{2N(2N - 1)} \\ &= \frac{2}{2N - 1} \\ N_e &= N - \frac{1}{2}\end{aligned}$$

Exercise 11

Blant menn i dette landet opptrer etternavn som om de var Y -koplet. Hva forteller teorien om genetisk drift oss om hvor raskt navnediversiteten vil forsvinne?

Svar:

Et Y -koplet gen vil oppføre seg som et gen i en haploid modell. Sannsynligheten for at et etternavn forsvinner vil være $1 - p_0$.

Exercise 12

Hvis en populasjon starter med størrelse N og vokser med 2% hver generasjon, hvor mye innavl vil akkumuleres? Bruk en tilnærming for store N , og sum av geometrisk rekke.

Svar:

Lar vekstraten være $a = 1.02$. Da er

$$N_t = Na^t$$

Den effektive populasjonsstørrelsen i tidsrommet frem til t er (VI-48 side 208)

$$\begin{aligned} N_e &= \frac{1}{\frac{1}{t} \sum_{j=0}^{t-1} \frac{1}{N_j}} \\ &= \frac{1}{\frac{1}{t} \sum_{j=0}^{t-1} \frac{1}{Na^j}} \\ \frac{1}{2N_e} &= \frac{1}{2Nt} \sum_{j=0}^{t-1} \frac{1}{a^j} \end{aligned}$$

Den gjenværende andelen av det opprinnelige antallet heterozygoter er

$$\begin{aligned} h_t &= \left(1 - \frac{1}{2N_e}\right)^t \\ &= \left(1 - \frac{1}{2Nt} \sum_{j=0}^{t-1} \frac{1}{a^j}\right)^t \\ &= \left(1 - \frac{1 - (1/a)^t}{2Nt(1 - 1/a)}\right)^t \end{aligned}$$

For å finne ut hvor mye innavl som akkumuleres noensinne, la $t \rightarrow \infty$ (Se f.eks. Rottmann side 80 for formel).

$$\begin{aligned} \lim_{t \rightarrow \infty} h_t &= \lim_{t \rightarrow \infty} \left(1 - \frac{1 - (1/a)^t}{2Nt(1 - 1/a)}\right)^t \\ &= \lim_{t \rightarrow \infty} \left(1 - \frac{1}{2Nt(1 - 1/a)}\right)^t \\ &= \exp\left(-\frac{1}{2N(1 - 1/a)}\right) \\ &= \exp\left(-\frac{51}{2N}\right) \end{aligned}$$