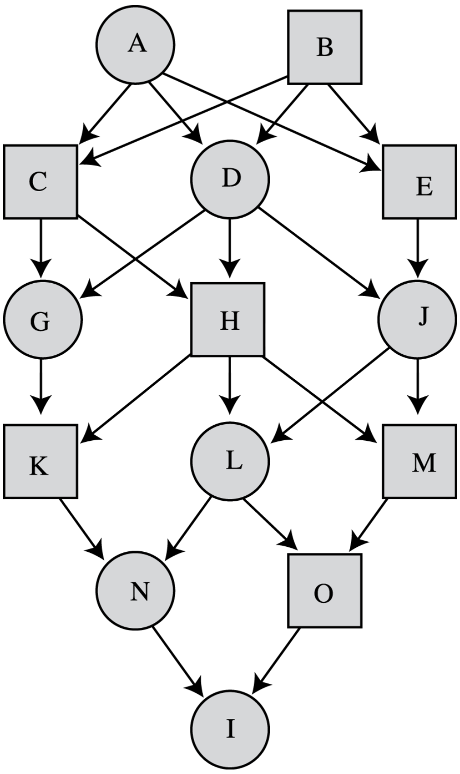
Løsningsforslag ST2301 Øving 10

Kapittel 5

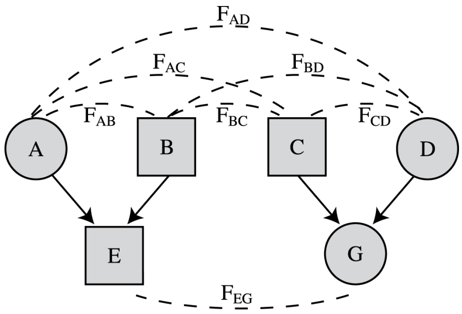
Exercise 6

Hva er innavlskoeffisienten for individ I i følgende stamtre?



Svar:

Her er det best å bruke en annen metode enn løkkemetoden. Slektskapskoeffisient-metoden beskrevet i kapittel V.6 er mer systematisk. Slektskapskoeffisienten mellom to individer er sannsynligheten for at et tilfeldig valgt allel hos det ene individet er "identical by descent" (IBD) med et tilfeldig valgt allel hos det andre individet.



Dersom A og B er foreldre til E, og C og D er foreldre til G, slik som på figuren over, har vi følgende regneregler:

1. Ved å bruke loven om total sannsynlighet blir

$$F_{EG} = \frac{1}{4}F_{AD} + \frac{1}{4}F_{AC} + \frac{1}{4}F_{BC} + \frac{1}{4}F_{BD}$$

   Denne setningen gjelder også dersom A eller B er samme individ som C eller D.

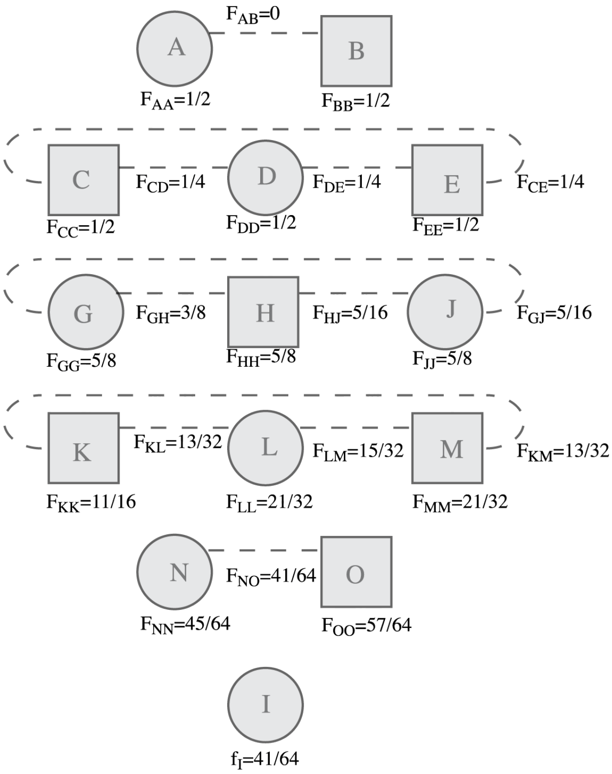
2. Slektskapskoeffisienten til individ E er

$$\begin{aligned} F_{EE} &= \frac{1}{2} + \frac{1}{2}f_E \\ &= \frac{1}{2}(1 + f_E) \\ &= \frac{1}{2}(1 + F_{AB}) \end{aligned}$$

Disse reglene kan brukes på stamtreet i oppgava, ved å starte øverst og så regne ut koeffisienter nedover treet. Det gir:

$$\begin{aligned} F_{CC} &= \frac{1}{2}(1 + F_{AB}) = \frac{1}{2} \\ F_{CD} &= \frac{1}{4}F_{AA} + \frac{1}{4}F_{BB} + \frac{1}{4}F_{AB} + \frac{1}{4}F_{AB} = \frac{1}{4} \\ F_{GG} &= \frac{1}{2}(1 + F_{CD}) = \frac{5}{8} \\ F_{GJ} &= \frac{1}{4}F_{CD} + \frac{1}{4}F_{CE} + \frac{1}{4}F_{DD} + \frac{1}{4}F_{DE} \\ &= \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{4} = \frac{5}{16} \end{aligned}$$

Fortsetter man å regne nedover i treet, finner man tilslutt at $f_i = \frac{41}{64} \approx 0.64$. Figuren under viser slektskapskoeffisientene.



Exercise 8

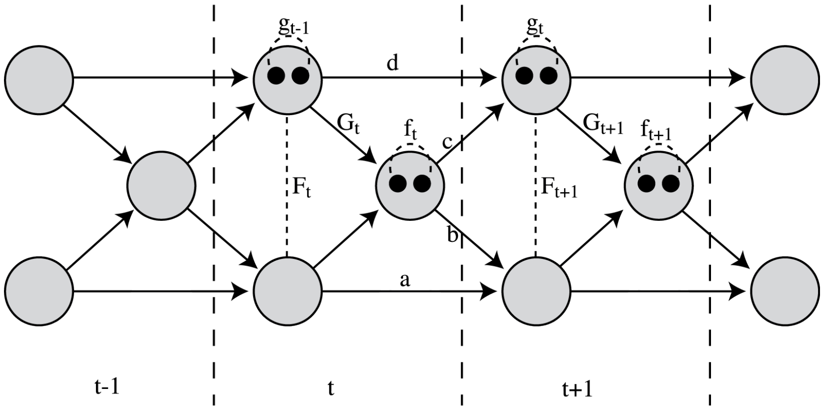
For dette regulære innavls-systemet, finn rekursjonslikningene som trengs for å analysere det.

Svar:

Figuren over viser slektskapskoeffisienter og innavlskoeffisienter for ulike generasjoner. Har at

1. $F_t = f_t$
2. $G_t = g_t$
3. Loven om total sannsynlighet gir

$$\begin{aligned}
F_{t+1} &= Pr(\text{IBD}|a,d)\,Pr(a,d) + Pr(\text{IBD}|c,d)\,Pr(c,d) \\
&\quad + Pr(\text{IBD}|(a\cap c)\cup(b\cap d))\,Pr((a\cap c)\cup(b\cap d)) \\
&= \frac{1}{4}\,F_t + \frac{1}{4}\cdot\frac{1}{2}\,(1+F_t) + (\frac{1}{4}\,G_t + \frac{1}{4}\,G_t) \\
&= \frac{3}{8}\,F_t + \frac{1}{2}\,G_t + \frac{1}{8}
\end{aligned}$$



4. Loven om total sannsynlighet gir

$$
\begin{aligned}
G_{t+1} &= \frac{1}{2} \cdot \frac{1}{2}\,(1 + G_t) + \frac{1}{2}\,F_{t+1} \\
&= \frac{1}{2}\,F_{t+1} + \frac{1}{4}\,G_t + \frac{1}{4}
\end{aligned}
$$

Setter 3 inn i 4, og får at

$$
\begin{aligned}
G_{t+1} &= \frac{1}{2}\,\frac{3}{8}\,F_t + \frac{1}{2}\,\frac{1}{2}\,G_t + +\frac{1}{2}\,\frac{1}{8} + \frac{1}{4}\,G_t + \frac{1}{4} \\
&= \frac{3}{16}\,F_{t+1} + \frac{1}{2}\,G_t + \frac{5}{16}
\end{aligned}
$$

På matriseform blir likningssystemet

$$
\begin{bmatrix} F_{t+1} \\ G_{t+1} \end{bmatrix} = \begin{bmatrix} \frac{3}{8} & \frac{1}{2} \\ \frac{3}{16} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} F_t \\ G_t \end{bmatrix} + \begin{bmatrix} \frac{1}{8} \\ \frac{5}{16} \end{bmatrix}
$$

En partikulærløsning av dette systemet er

$$
x_p = \begin{bmatrix} F_t \\ G_t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}
$$

Den generelle løsningen av systemet er på formen

$$x(t) = c_1\lambda_1 y_1 + c_2\lambda_2 y_2 + x_p$$

der λ_1 og λ_2 er gitt ved

$$\begin{aligned}
\begin{vmatrix} \frac{3}{8} - \lambda & \frac{1}{2} \\ \frac{3}{16} & \frac{1}{2} - \lambda \end{vmatrix} &= 0 \\
\left(\frac{3}{8} - \lambda\right)\left(\frac{1}{2} - \lambda\right) - \frac{3}{32} &= 0 \\
\lambda^2 - \frac{7}{8}\lambda + \frac{3}{32} &= 0 \\
\lambda_1 &= \frac{1}{2}\left(\frac{7}{8} + \sqrt{\frac{49}{64} - \frac{12}{32}}\right) = \frac{3}{4} \\
\lambda_2 &= \frac{1}{2}\left(\frac{7}{8} - \sqrt{\frac{49}{64} - \frac{12}{32}}\right) = \frac{1}{8}
\end{aligned}$$

Den første (største) egenverdien λ_1 bestemmer hvor raskt $x(t)$ nærmer seg x_p asymptotisk. Sannsynligheten for at to allel ikke er IBD, dvs $(1-F_t)$ og $(1-G_t)$, vil altså reduseres med 25% hver generasjon.

Exercise 9

Ser på et locus med et recessivt allel a, med frekvens 0.01. Anta at vi deler populasjonen i to og deretter innavler hver subpopulasjon inntil innavlskoeffisienten er f. Som funksjon av f,

1. Hva er den forventa frekvensen av aa i en subpopulasjon?
2. Hva er den forventa frekvensen av aa i F_1-krysningen (individer med én forelder fra hver subpopulasjon) mellom de to subpopulasjonene?
3. Hva er den forventa frekvensen av aa i F_2-krysningen (individer der foreldrene er to ulike F_1-individer) mellom de to subpopulasjonene?

Svar:

1. I hver subpopulasjon er startfrekvensen av a $p_0 = 0.01$. For hver subpopulasjon har vi at

$$\begin{aligned}
P_{aa,i} &= p_0^2(1-f) + p_0 f \\
&= 0.01^2(1-f) + 0.01f \\
&= 0.0001 + 0.0099f
\end{aligned}$$

2. Sannsynligheten for at to tilfeldig valgte gener fra hver subpopulasjon er IBD, er null (ingen felles opphav). Derfor må

$$\begin{aligned} P_{aa,F_1} &= Pr(\text{Trekke } a \text{ fra pop } 1)Pr(\text{Trekke } a \text{ fra pop } 2) \\ &= p_0^2 = 0.0001 \end{aligned}$$

3. I F_1-krysningen er innavlskoeffisienten lik 0. Sannsynligheten for at to tilfeldig valgte gener stammer fra samme subpopulasjon er $\frac{1}{2}$. Slektskapskoeffisienten til F_1-krysningen er dermed

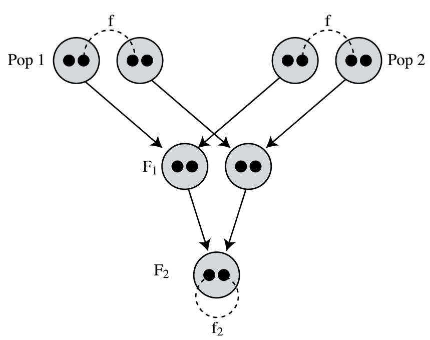
$$F_1 = \frac{1}{2}\,f + \frac{1}{2}\,0$$

Innavlskoeffisienten til et avkom i F_2-krysningen er lik slektskapskoeffisienten til foreldrene i F_1-krysningen.

$$f_2 = F_1 = \frac{1}{2}\,f$$

Dermed blir forventa frekvens av aa i F_2-krysningen

$$\begin{aligned} P_{aa,F_2} &= p_0^2(1 - f_2) + p_0 f_2 \\ &= 0.01^2(1 - \frac{f}{2}) + 0.01\frac{f}{2} \\ &= 0.0001 + 0.00495f \end{aligned}$$



Kapittel 1

Complement 14

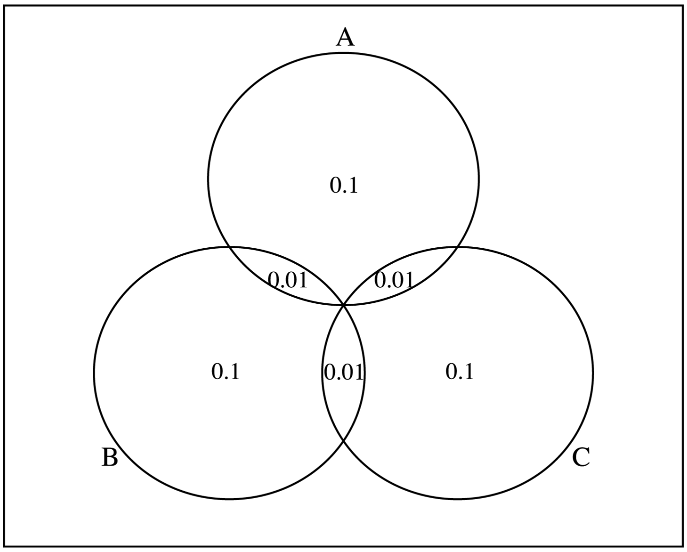
Med tre loci, hvert med to allel, se om du kan finne et sett gametfrekvenser (tre av åtte mulige gameter) som gir $D_{AB} = 0$, $D_{BC} = 0$ og $D_{AC} = 0$, men der gametene ikke er i koplingslikevekt, slik at for eksempel $P_{ABC} \neq p_A p_B p_C$. Kan vi gi gametfrekvensene ved å spesifisere p_A, p_B, p_C, D_{AB}, D_{BC} og D_{AC}? Vis ved eksempler hvorfor dette er/ ikke er mulig.

Svar:

Anta at $p_A = p_B = p_C = 0.1$ og at $P_{AB} = P_{BC} = P_{AC} = 0.01$ slik at $D_{AB} = 0$, $D_{BC} = 0$ og $D_{AC} = 0$. Vi påstår at det er mulig at

$$P_{ABC} = 0 \neq p_A p_B p_C \quad (*)$$

Figuren under viser Venn-diagrammet for utfallsrommet.



Hvis $(*)$ er oppfylt blir sannsynlighetene for hvert av de åtte områdene i Venn-diagrammet

$$
\begin{aligned}
Pr(A \cap B \cap C) &= 0 \\
Pr(A \cap \overline{B} \cap \overline{C}) &= 0.1 - 0.01 - 0.01 = 0.08 \\
Pr(\overline{A} \cap B \cap \overline{C}) &= 0.08 \\
Pr(\overline{A} \cap \overline{B} \cap C) &= 0.08 \\
Pr(A \cap B) = Pr(A \cap B \cap \overline{C}) &- Pr(A \cap B \cap C) = 0.01 \\
Pr(A \cap \overline{B} \cap C) &= 0.01 \\
Pr(\overline{A} \cap B \cap C) &= 0.01 \\
Pr(\overline{A} \cap \overline{B} \cap \overline{C}) &= 1 - 3 \cdot 0.08 - 3 \cdot 0.01 = 0.61
\end{aligned}
$$

Siden alle gametfrekvenser ligger mellom 0 og 1 så er $(*)$ mulig.

Vi har i utgangspunktet åtte frie variable ($8 = 2^3$ gametfrekvenser). Spesifiserer vi p_A, p_B, p_C, D_{AB}, D_{BC}og D_{AC} med tilleggskravet at $\sum_{ijk} P_{ijk} = 1$, får vi sju likninger med åtte ukjente, slik at ikke alle gametfrekvenser P_{ijk} er bestemt.