

Løsningsforslag ST2301 Øving 8

Kapittel 4

Exercise 1

For tre alleler, finn et sett med genfrekvenser for to populasjoner, som gir flere heterozygoter enn forventet ut fra Hardy-Weinberg-andeler for minst én av de heterozygote genotypene, dersom vi ser på et utvalg fra begge populasjonene.

Svar:

Har et utvalg fra to populasjoner. Sannsynligheten for at et individ kommer fra populasjon 1 er m , og sannsynligheten for at det kommer fra populasjon 2 er $1 - m$. La p_{ij} være frekvens av allel i i populasjon j . De forventede allel-frekvensene i utvalget er gitt ved

$$\begin{aligned}E[p_1] &= mp_{11} + (1 - m)p_{12} \\E[p_2] &= mp_{21} + (1 - m)p_{22} \\E[p_3] &= mp_{31} + (1 - m)p_{32}\end{aligned}$$

Frekvensen heterozygoter av type ij i utvalget er gitt ved

$$\begin{aligned}P_{ij} &= m2p_{i1}p_{j1} + (1 - m)2p_{i2}p_{j2} \\&= E[2p_i p_j] \\&= 2E[p_i]E[p_j] + 2\text{Cov}(p_i, p_j)\end{aligned}$$

Hvis utvalget hadde vært i Hardy-Weinberglikevekt ville $P_{ij} = 2E[p_i]E[p_j]$. De fleste allel-par har negativ kovarians (side 131), slik at for de fleste heterozygotene oppstår et underskudd av heterozygoter i forhold til Hardy-Weinbergandeler. Dersom andelen heterozygoter skal være større, må kovariansen være positiv. Kravet blir derfor at

$$\begin{aligned}\text{Cov}(p_i, p_j) &> 0 \\E[p_i p_j] - E[p_i]E[p_j] &> 0 \\E[p_i p_j] &> E[p_i]E[p_j]\end{aligned}$$

Vi trenger bare å finne ett sett med genotypefrekvenser som tilfredsstiller kravet. For eksempel, la $m = 0.5$ og la genfrekvensene i de to populasjonene være

	p_1	p_2	p_3
Pop.1	0.1	0.1	0.8
Pop.2	0.3	0.3	0.4

Det gir

$$\begin{aligned}
 E[p_1 p_3] &= m p_{11} p_{31} + (1 - m) p_{12} p_{32} \\
 &= 0.5 \cdot 0.1 \cdot 0.8 + 0.5 \cdot 0.3 \cdot 0.4 \\
 &= 0.1 \\
 E[p_1] &= m p_{11} + (1 - m) p_{12} \\
 &= 0.5 \cdot 0.1 + 0.5 \cdot 0.3 \\
 &= 0.2 \\
 E[p_3] &= m p_{31} + (1 - m) p_{32} \\
 &= 0.5 \cdot 0.8 + 0.5 \cdot 0.4 \\
 &= 0.6 \\
 E[p_1] E[p_3] &= 0.2 \cdot 0.6 = 0.12 > 0.1
 \end{aligned}$$

Dvs individene i utvalget kommer fra populasjoner med genfrekvensene over, har iallefall én heterozygot (type 13) større andel en forventet ut fra Hardy-weinbergandeler.

Exercise 2

Dersom koplingsulikevekt D blir skapt ved at to populasjoner blandes, hva er uttrykket for endring av D over tid?

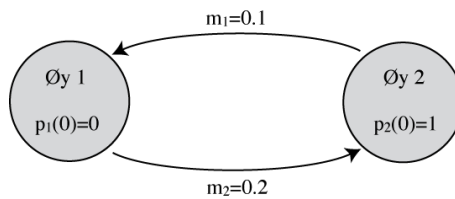
Svar:

Antar to loci A og B , og lar $D = D_{AB}(t)$ være koplingsulikevekten ved generasjon t etter blanding, mens $D_{AB}(0)$ er koplingsulikevekten rett etter blanding. Antar tilfeldig parring etter at populasjonene har blitt blandet og at r er rekombinasjonsraten mellom lociene. Da gjelder likning I-44 side 18:

$$\begin{aligned}
 D_{AB}(t) &= (1 - r)^t D_{AB}(0) \\
 &= (1 - r)^t \text{Cov}(p, q)
 \end{aligned}$$

Exercise 3

Har en modell med to øyer, med immigrasjonsrater $m_1 = 0.1$ fra øy 2 til øy 1, og $m_2 = 0.2$ fra øy 1 til øy 2. Dersom start-genfrekvensene er hhv. $p_1(0) = 0$ og $p_2(0) = 1$, hva er likevektsgenfrekvensen? (Hint: Finn en størrelse a som har egenskapen at dersom $p_1(t)$ og $p_2(t)$ er genfrekvensene på hhv. øy 1 og 2, så vil det vektete gjennomsnittet $ap_1(t) + (1 - a)p_2(t)$ være uendra fra en generasjon til neste.



Svar:

Frekvensen av allel A ved øy 1 er $p_1(t)$ etter t generasjoner, mens frekvensen ved øy 2 er $p_2(t)$. La likevektsfrekvensene på øy 1 og 2 for hhv. p_1 og p_2 . Frekvensen av A ved øy 1 er lik frekvensen blant individene som ble værende fra forrige generasjon, pluss frekvensen blant nye individer som kom fra øy 2. Samme argument gjelder frekvensen ved øy 2. Likningene for endring av genfrekvensen fra en generasjon til neste er derfor

$$\begin{aligned} p_1(t) &= (1 - m_1) p_1(t - 1) + m_1 p_2(t - 1) \\ p_2(t) &= m_2 p_1(t - 1) + (1 - m_2) p_2(t - 1) \end{aligned}$$

Har at

$$\begin{aligned} p_1(t) - p_2(t) &= (1 - m_1 - m_2) p_1(t - 1) - (1 - m_1 - m_2) p_2(t - 1) \\ &= (1 - m_1 - m_2) [p_1(t - 1) - p_2(t - 1)] \end{aligned}$$

Nå kan vi bruke hintet, dvs finne en a som gjør at $ap_1(t) + (1 - a)p_2(t)$ blir uforandra fra en generasjon til neste.

$$\begin{aligned}
ap_1(t) + (1-a)p_2(t) &= ap_1(t-1) + (1-a)p_2(t-1) \\
a[p_1(t) - p_2(t)] + p_2(t) &= a[p_1(t-1) - p_2(t-1)] + p_2(t-1) \\
a(1-m_1-m_2)[p_1(t-1) - p_2(t-1)] + m_2[p_1(t-1) - p_2(t-1)] &= a[p_1(t-1) - p_2(t-1)] \\
a(1-m_1-m_2) + m_2 &= a \\
m_2 - a(m_1 - m_2) &= 0 \\
a &= \frac{m_2}{m_1 + m_2} \\
&= \frac{0.2}{0.1 + 0.2} \\
&= \frac{2}{3}
\end{aligned}$$

Nå kan vi finne likevektsfrekvensen p_2 uttrykt ved p_1 , siden $ap_1(t) + (1-a)p_2(t)$ ikke forandres med t .

$$\begin{aligned}
ap_1(t) + (1-a)p_2(t) &= ap_1(t-1) + (1-a)p_2(t-1) \\
&= ap_1(t-2) + (1-a)p_2(t-2) \\
&\vdots \\
&= ap_1(0) + (1-a)p_2(0) \\
ap_1 + (1-a)p_2 &= ap_1(0) + (1-a)p_2(0) \\
ap_1 + (1-a)p_2 &= 1-a \\
p_2 &= 1 - \frac{a}{1-a} p_1 \\
p_2 &= 1 - 2p_1
\end{aligned}$$

Likevektsfrekvensene må også oppfylle de første likningene, dvs

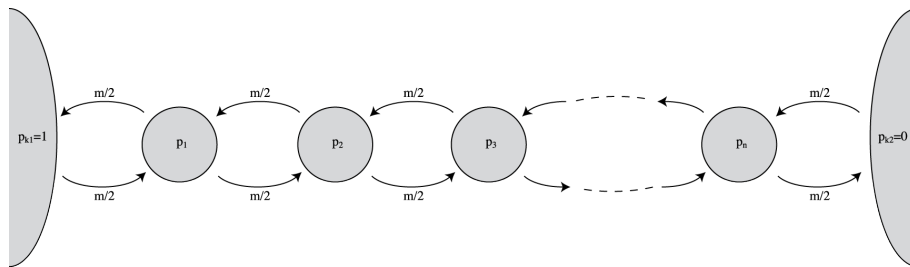
$$\begin{aligned}
p_1 &= (1-m_1)p_1 + m_1 p_2 \\
p_2 &= m_2 p_1 + (1-m_2)p_2
\end{aligned}$$

Setter inn $p_2 = 1 - 2p_1$.

$$\begin{aligned}
p_1 &= p_1 - m_1 p_1 + m_1 - 2m_1 p_1 \\
3m_1 p_1 &= m_1 \\
p_1 &= \frac{1}{3} \\
p_2 &= 1 - 2p_1 = \frac{1}{3}
\end{aligned}$$

Exercise 5

Anta at vi har to kontinenter. Frekvensen av allel A er lik 1 ved det ene, og 0 ved det andre. Mellom kontinentene er en rekke øyer, som lager en perfekt "stepping stone"-modell med migrasjonsrate $m/2$ mellom naboøyer, og migrasjonsrate $m/2$ fra hvert av kontinentene til nærmeste øy. Genfrekvensen på kontinentene antas uforandret fordi populasjonen er mye større her enn på øyene. Det er hverken seleksjon eller mutasjon. Hva er likevekts-rekkefølgen av genfrekvenser på øyene? (Hint: Mønsteret er det samme uavhengig av antall øyer. Det kan være lurt å gjøre en "kvalifisert gjetning" og så vise at denne stemmer.)



Svar:

Siden migrasjonsraten er den samme på alle øyer virker det logisk at likevektsfrekvensen endres like mye fra en øy til den neste. Dersom det bare er én øy, for eksempel, er likevektsfrekvensen antakelig $p_1 = 1/2$. Dersom det er to øyer, tipper vi at likevektsfrekvensene er $p_1 = 2/3$ og $p_2 = 1/3$. For tre øyer får vi:

$$\begin{aligned}
 p_{k1} &= 1 \\
 p_1 &= \frac{3}{4} = \frac{n}{n+1} \\
 p_2 &= \frac{2}{4} = \frac{n-1}{n+1} \\
 p_3 &= \frac{1}{4} = \frac{n-2}{n+1} \\
 p_{k2} &= 0
 \end{aligned}$$

Det ser ut som at likevektsfrekvensen ved øy i er gitt ved

$$p_i = \frac{n+1-i}{n+1} \quad (*)$$

Dette er en gjetning, og det gjenstår å vise at det stemmer. Setter først inn $i = 0$ i (*).

$$p_0 = \frac{n+1-0}{n+1} = 1 = p_{k1}$$

Setter inn $i = n + 1$ i (*).

$$p_{n+1} = \frac{n+1-n-1}{n+1} = 0 = p_{k2}$$

Ser på øy nummer i , $1 \leq i \leq n$. Rekursjonslikningen for endring av genfrekvens fra en generasjon til neste er

$$p_i(t) = (1-m)p_i(t-1) + \frac{m}{2} p_{i-1}(t-1) + \frac{m}{2} p_{i+1}(t-1) \quad (**)$$

Ved likevekt:

$$p_i = (1-m)p_i + \frac{m}{2} p_{i-1} + \frac{m}{2} p_{i+1}$$

Setter inn (*) i rekursjonslikningen (**) for å se om dette er løsningen. Høyre side:

$$\begin{aligned} (1-m)p_i + \frac{m}{2} p_{i-1} + \frac{m}{2} p_{i+1} &= (1-m) \frac{n+1-i}{n+1} + \frac{m}{2} \frac{n+2-i}{n+1} + \frac{m}{2} \frac{n-i}{n+1} \\ &= \frac{1}{n+1} \left((1-m)(n+1-i) + \frac{m}{2} (n+2-i) + \frac{m}{2} (n-i) \right) \\ &= \frac{1}{n+1} \left(n+1-i - mn - m - mi + \frac{mn}{2} + m - \frac{mi}{2} + \frac{mn}{2} - \frac{mi}{2} \right) \\ &= \frac{n+1-i}{n+1} \end{aligned}$$

Venstre side:

$$p_i = \frac{n+1-i}{n+1}$$

Høyre og venstre side er like, så vi har vist at $p_i = \frac{n+1-i}{n+1}$ er løsning på (**).

Complement 2

I en blanding gameter som kommer fra populasjoner som *ikke* er i koplingslikevekt,

1. Hva er uttrykket for D som funksjon av m_i , p_i , q_i , og D_i ?
2. Hva er uttrykket for D som funksjon av kovariansene mellom og innen populasjoner?

Svar:

Dersom populasjonene hadde vært i koplingslikevekt, kunne vi brukt likningene side 132 for å finne D . Siden populasjonene ikke er i koplingslikevekt kan ikke likningene brukes direkte, men vi kan bruke en liknende fremgangsmåte.

La P_{AB} være frekvensen av gamet AB i den blandede populasjonen, og P_{ABi} være frekvensen i populasjon i . La p_i og q_i være frekvensen av hhv. allel A og B i populasjon i , og la m_i være andelen av blandingspopulasjonen som kommer fra populasjon i .

Dersom blandingspopulasjonen hadde vært i koplingslikevekt, ville verdien av gametfrekvensen P_{AB} ha vært

$$\left(\sum_i m_i p_i\right) \left(\sum_i m_i q_i\right)$$

Koplingslikevekten D er den virkelige gametfrekvensen P_{AB} minus verdien over.

$$D = P_{AB} - \left(\sum_i m_i p_i\right) \left(\sum_i m_i q_i\right)$$

Vi trenger altså et uttrykk for P_{AB} .

I populasjon i er koplingslikevekten gitt ved (fra kapittel 1, side 18)

$$\begin{aligned} D_i &= P_{ABi} - p_i q_i \\ P_{ABi} &= D_i + p_i q_i \end{aligned}$$

I blandingspopulasjonen er gametfrekvensen dermed

$$\begin{aligned} P_{AB} &= \sum_i m_i P_{ABi} \\ &= \sum_i m_i (D_i + p_i q_i) \end{aligned}$$

Koplingsulikevekten i blandingspopulasjonen blir

$$D = \sum_i m_i(D_i + p_i q_i) - \left(\sum_i m_i p_i \right) \left(\sum_i m_i q_i \right)$$

For å skrive om uttrykket som funksjon av kovariansene mellom og innen populasjoner, bruk likningene side 132. Det gir

$$E[p] = \sum_i m_i p_i$$

$$E[q] = \sum_i m_i q_i$$

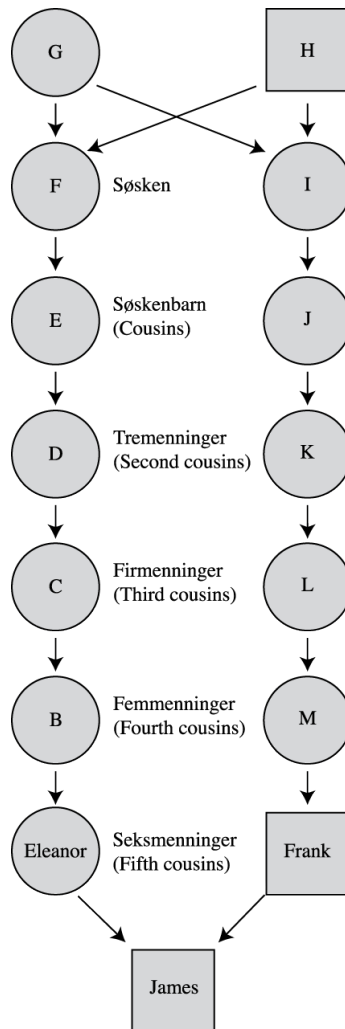
$$E[pq] = \sum_i m_i p_i q_i$$

$$\begin{aligned} D &= \sum_i m_i(D_i + p_i q_i) - \left(\sum_i m_i p_i \right) \left(\sum_i m_i q_i \right) \\ &= \sum_i m_i D_i + \sum_i m_i p_i q_i - \left(\sum_i m_i p_i \right) \left(\sum_i m_i q_i \right) \\ &= \sum_i m_i D_i + E[pq] - E[p]E[q] \\ &= \sum_i m_i D_i + \text{Cov}(p, q) \end{aligned}$$

Kapittel 5

Exercise 2

James Roosevelt var sønn av Eleanor og Franklin Delano Roosevelt, som var seksmenninger (fifth cousins). Hva var innavlskoeffisienten hans? (Merk: Seksmenninger har en forelder hver som er femmenninger til hverandre, osv for firmenninger, tremenninger og søskenbarn. Se figur neste side.)



Svar:

La f_{James} være innavlskoeffisienten til James Roosevelt. Det er to mulige løkker:

1. James Eleanor B C D E F G I J K L M Frank James
2. James Eleanor B C D E F H I J K L M Frank James

Hver løkke går gjennom 13 personer i tillegg til James Roosevelt. Det gir et bidrag på $(\frac{1}{2})^{13}$ fra hver løkke. Innavlskoeffisienten blir

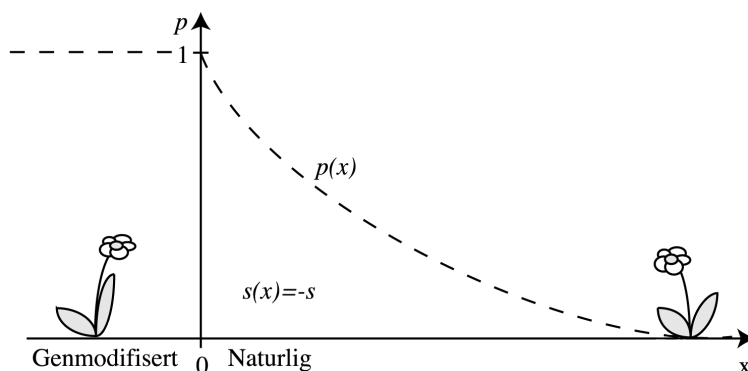
$$f_{\text{James}} = \left(\frac{1}{2}\right)^{13} + \left(\frac{1}{2}\right)^{13} \approx 0.000244$$

Ekstra oppgave

Anta at vi har en populasjon raps som er geografisk fordelt i et lineært habitat. I en del av populasjonen (i geografiske posisjoner $x < 0$) dyrkes det hvert år en genmodifisert rapsvariant. Frekvensen av genet innsatt ved genmodifisering A (dette kalles et transgen) i denne delen av populasjonen er dermed hele tiden $p(x) = 1$. Inntil den genmodifiserte delen av populasjonen befinner det seg (for $x > 0$) en naturlig forekommende rapsvariant som krysser seg med den genmodifiserte varianten. Anta at migrasjon og seleksjon virker som beskrevet ved likning IV-56 og IV-59 og at seleksjonskoeffisienten er $s(x) = -s$ i området $x > 0$ (slik at det er seleksjon mot transgenet i den naturlige delen av populasjonen).

Anta at s og σ^2 er små slik at genfrekvensgradienten som oppstår ved migrasjon-seleksjonsbalanse tilfredsstiller IV-66 (men merk at nå har s skiftet fortegn).

Hva blir nå de nye randbetingelsene og tilhørende løsning av IV-66? Hva blir stigningstallet til gradienten i grensen $x \rightarrow 0$? Bruk dette til å definere et passende karakteristisk mål l på hvor langt inn i den naturlige populasjonen transgenet vil etablere seg når migrasjons-seleksjonsbalanse har inntruffet. Hva blir l dersom $\sigma = 100 \text{ m} \cdot \text{år}^{-1/2}$ og $s = 0.01 \text{ år}^{-1/2}$?



Svar:

Har at

$$y = \frac{dp(x)}{dx}$$

dvs y er stigningstallet til gradienten. Likning IV-66 gir (for $s(x) = -s$)

$$y(x) \frac{dy(x)}{dp(x)} = \frac{2s}{\sigma^2} p(x)(1 - p(x))$$

Tar integralet på begge sider.

$$\frac{y(x)^2}{2} + C = \frac{2s}{\sigma^2} \left(\frac{p(x)^2}{2} - \frac{p(x)^3}{3} \right)$$

Nye randbetingelser: Ser av figuren at ettersom $x \rightarrow \infty$ vil stigningstallet $y(x) = 0$ for $p(x) = 0$. Dette gir $C = 0$, og

$$\frac{y(x)^2}{2} = \frac{2s}{\sigma^2} \left(\frac{p(x)^2}{2} - \frac{p(x)^3}{3} \right)$$

I grensa når $x \rightarrow 0$ er $p(0) = 1$,

$$\begin{aligned} \frac{y(x)^2}{2} &= \frac{2s}{\sigma^2} \left(\frac{1}{2} - \frac{1}{3} \right) \\ y(0) &= \pm \sqrt{\frac{2s}{3\sigma^2}} \end{aligned}$$

Her er det den negative løsningen som er relevant, dvs stigningstallet til $p(x)$ i $x = 0$ er

$$y(0) = -\sqrt{\frac{2s}{3\sigma^2}}$$

Som mål på hvor langt inn i populasjonen de genmodifiserte plantene etablerer seg, kan man bruke $l = |1/y(0)|$, dvs

$$l = \sqrt{\frac{3\sigma^2}{2s}}$$

Dersom $\sigma = 100 \text{ m} \cdot \text{år}^{-1/2}$ og $s = 0.01 \text{ år}^{-1/2}$ får vi

$$l = \sqrt{\frac{3 \cdot 100^2}{2 \cdot 0.01}} = \sqrt{150000} \approx 387$$

Dvs ca 387 m.