



Matematisk evolusjonær genetikk (ST2301)

Tirsdag 19. mai 2009

Løsningsforslag

(For flere av oppgavene finnes det også alternative fremgangsmåter)

Oppgave 1

a) Først gjør vi om til relative fitnesser,

Genotype	Relativ fitness
A	1.2
a	1.

For dette tilfellet har vi formelen

$$\frac{p_A(t)}{p_a(t)} = (1 + s)^t \frac{p_A(0)}{p_a(0)},$$

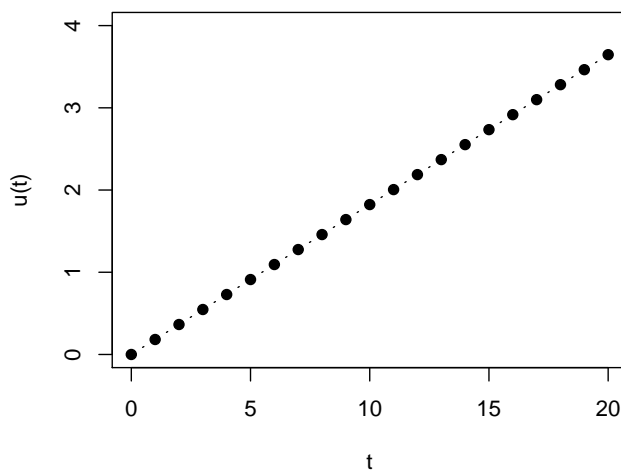
der $s = 0.2$. Setter inn verdiene fra oppgaveteksten, og finner

$$\begin{aligned} \frac{2/3}{1/3} &= (1 + 0.2)^t \frac{0.5}{0.5} \\ 1.2^t &= 2 \\ t &= \frac{\ln 2}{\ln 1.2} = 3.801784. \end{aligned}$$

b) Vi har at $p_A(t) = \frac{w_A}{w} p_A(t-1)$, og tilsvarende for frekvensen av a . Det gir

$$\begin{aligned}u(t) &= \ln\left(\frac{p_A(t)}{p_a(t)}\right) \\&= \ln\left(\frac{w_A p_A(t-1)}{w_a p_a(t-1)}\right) \\&= \ln\left(\frac{w_A}{w_a}\right) + \ln\left(\frac{p_A(t-1)}{p_a(t-1)}\right) \\&= \ln\left(\frac{w_A}{w_a}\right) + u(t-1) \\&= 2 \ln\left(\frac{w_A}{w_a}\right) + u(t-2) \\&= \dots \\&= t \ln\left(\frac{w_A}{w_a}\right) + u(0).\end{aligned}$$

Variabelen $u(t)$ vokser altså lineært med t , som vist i figuren under (for parameterverdier fra oppgave 1 a).



- c) Sannsynligheten for å trekke en hulepartner med genotype A eller a er hhv $p_A = p$ og $p_a = (1 - p)$. Vi kan definere en midlere relativ fitness \bar{w}_A for A -individene, ved å bruke loven om total sannsynlighet og betinge på om hulepartneren har genotype A eller a . Dermed får vi for A -individene at

$$\bar{w}_A = pw_{AA} + (1 - p)w_{Aa}.$$

Tilsvarende får vi for individene med genotype a at

$$\bar{w}_a = pw_{aA} + (1 - p)w_{aa}.$$

Gjennomsnittlig fitness for hele populasjonen blir da

$$\begin{aligned}\bar{w} &= p\bar{w}_A + (1 - p)\bar{w}_a \\ &= p^2w_{AA} + p(1 - p)(w_{Aa} + w_{aA}) + (1 - p)^2w_{aa}.\end{aligned}$$

For å finne neste års frekvens av A kan vi bruke samme type argumentasjon som for det diploide tilfellet (likning II-31 til II-33), men vi må huske på at w_{Aa} generelt ikke er lik w_{aA} i denne modellen. Neste års frekvens av A blir

$$\begin{aligned}p' &= \frac{\bar{w}_A}{\bar{w}}p \\ &= \frac{p^2w_{AA} + p(1 - p)w_{Aa}}{p^2w_{AA} + p(1 - p)(w_{Aa} + w_{aA}) + (1 - p)^2w_{aa}}\end{aligned}$$

- d) For å finne likevektene kan vi (for eksempel) undersøke når $\Delta p = 0$. Her er endringen i genfrekvens Δp gitt ved

$$\begin{aligned}
\Delta p &= p' - p \\
&= p \frac{\bar{w}_A - \bar{w}}{\bar{w}} \\
&= p \frac{pw_{AA} + (1-p)w_{Aa} - p^2w_{AA} + p(1-p)(w_{Aa} + w_{aA}) + (1-p)^2w_{aa}}{\bar{w}} \\
&= p(1-p) \frac{pw_{AA} + (1-p)w_{Aa} - [pw_{aA} + (1-p)w_{aa}]}{\bar{w}} \\
&= p(1-p) \frac{\bar{w}_A - \bar{w}_a}{\bar{w}}
\end{aligned}$$

Vi får altså likevekt for $p = 0$ (bare a -individer), $p = 1$ (bare A -individer), samt for alle p_e som oppfyller $\bar{w}_A = \bar{w}_a$ ($\bar{w} \rightarrow \infty$ er ikke noen reell mulighet). For den siste muligheten har vi

$$\begin{aligned}
\bar{w}_A &= \bar{w}_a \\
pw_{AA} + (1-p)w_{Aa} &= pw_{aA} + (1-p)w_{aa} \\
p_e &= \frac{w_{aa} - w_{Aa}}{w_{AA} + w_{aa} - w_{Aa} - w_{aA}}.
\end{aligned}$$

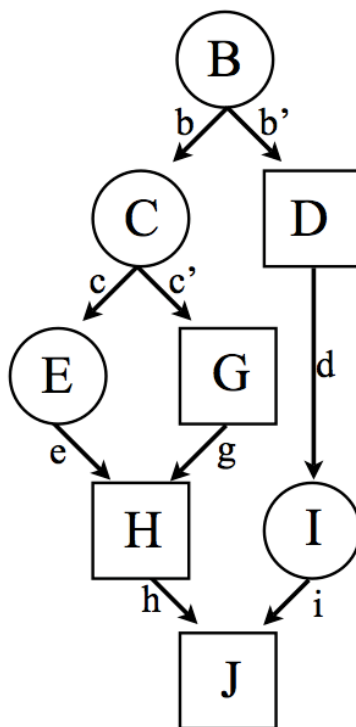
For at denne likevekten skal eksistere må den ligge mellom 0 og 1, dvs vi må ha at $|w_{aa} - w_{Aa}| \leq |w_{AA} + w_{aa} - w_{Aa} - w_{aA}|$ samt at fortegnet til teller og nevner må være det samme.

Dersom $w_{aa} > w_{Aa}$ og $w_{AA} > w_{aA}$, dvs hvis alle individene gir en større fitnessfordel til hulepartnere av samme genotype enn til hulepartnere med motsatt genotype, så vil likevekten eksistere men være ustabil (vi vil ende opp i $p = 1$ eller $p = 0$ avhengig av hvilken som er størst av \bar{w}_A og \bar{w}_a).

Dersom $w_{aa} < w_{Aa}$ samt $w_{AA} < w_{aA}$, dvs hvis alle individer gir større fitnessfordel til en hulepartner av motsatt genotype enn til en hulepartner av samme genotype, så vil likevekten eksistere og være stabil.

- e) Denne modellen er ikke helt ekvivalent med det diploide tilfellet for over- og underdominans. Det kan f.eks. hende A har større fordel av å bo med a enn omvendt, slik at vi generelt har to ulike fitnesser for "heterozygotene" (ulike hulepartnere). I det diploide tilfellet er begge allelene i samme individ, og vi skiller derfor ikke mellom Aa - og aA -individer. For huleboer-organismen ser vi at dersom to individer av ulik genotype får samme fitness når de deler hule, altså hvis $w_{Aa} = w_{aA}$, så får vi en situasjon som tilsvarer over- og underdominans.

Oppgave 2 Setter navn på allelene som nedarves:



- a) For det autosomale tilfellet kan vi bruke løkkemetoden for å finne innavlskoeffisienten til J , f_J . Det er to løkker i dette stamtreet.

$$P(JHEC\bar{B}DIJ) = \left(\frac{1}{2}\right)^6$$

$$P(JHGC\bar{B}DIJ) = \left(\frac{1}{2}\right)^6.$$

Det gir $f_J = \left(\frac{1}{2}\right)^6 + \left(\frac{1}{2}\right)^6 = \frac{1}{32} = 0.03125$. Dersom locuset er kjønnskoplede blir innavlskoeffisienten til J lik 0, fordi dette individet er en hann og derfor har kun én kopi av hvert kjønnskoplede gen.

- b) Slektsskapskoeffisienten F_{ij} mellom to individ i og j er sannsynligheten for at to tilfeldig valgte alleler fra hhv i og j er IBD (“identical by descent”). Vi ser fra stamtreet at individ D er onkelen til individ E . Slektsskapskoeffisienten mellom disse blir

$$\begin{aligned}
 F_{DE} &= P(e \leftarrow c \leftarrow b \equiv b' \rightarrow d) \\
 &= \frac{1}{16} = 0.0625.
 \end{aligned}$$

Individ D og H er i slekt ved at D er onkel til både E og G , som er foreldre til H . Her får vi

$$\begin{aligned}
 F_{DH} &= P(h \leftarrow e \leftarrow c \leftarrow b \equiv b' \rightarrow d) + P(h \leftarrow g \leftarrow c \leftarrow b \equiv b' \rightarrow d) \\
 &= \frac{1}{16} = 0.0625.
 \end{aligned}$$

Slektskapskoeffisientene F_{DE} og F_{DH} blir altså her like, dette skyldes at foreldrene til H er halvsøsken.

c) Ved innsetting i formlene i boka (kap 9) finner vi

$$\begin{aligned}
 \mu &= p^2 a_{11} + 2p(1-p)a_{12} + (1-p)^2 a_{22} \\
 &= 0.5^2 \cdot 2 + 2 \cdot 0.5 \cdot 0.5 \cdot 6 + 0.5^2 \cdot 5 \\
 &= 4.75
 \end{aligned}$$

$$\begin{aligned}
 \alpha_1 &= p a_{11} + (1-p)a_{12} - \mu \\
 &= 0.5 \cdot 2 + 0.5 \cdot 6 - 4.75 \\
 &= -0.75 \\
 \alpha_2 &= \dots = 0.75
 \end{aligned}$$

$$\begin{aligned}
 \delta_{11} &= a_{11} - 2\alpha_1 - \mu \\
 &= 2 - 2 \cdot (-0.75) - 4.75 \\
 &= -1.25 \\
 \delta_{22} &= \dots = -1.25 \\
 \delta_{12} &= \dots = 1.25
 \end{aligned}$$

Det gir additiv varians

$$\begin{aligned}
 V_A &= 2(p\alpha_1^2 + (1-p)\alpha_2^2) \\
 &= 2(0.5 \cdot (-0.75)^2 + 0.5 \cdot 0.75^2) \\
 &= 1.125,
 \end{aligned}$$

og dominansvariansen er gitt ved

$$\begin{aligned}
 V_D &= p^2\delta_{11}^2 + 2p(1-p)\delta_{12}^2 + (1-p)^2\delta_{22}^2 \\
 &= 0.5^2 \cdot (-1.25)^2 + 2 \cdot 0.5 \cdot 0.5 \cdot 1.25^2 + 0.5^2 \cdot (-1.25)^2 \\
 &= 1.5625.
 \end{aligned}$$

- d) Fra punkt b vet vi at slektskapskoeffisienten mellom D og E er $\frac{1}{16}$. I tillegg trenger vi å finne P_2 ("coefficient of fraternity"), sannsynligheten for at begge alleler i D er IBD med alleler i E . Vi ser imidlertid av stamtreet at maksimalt ett allelpar fra disse individene kan være IBD, siden fedrene til E og D (som ikke er tegnet i stamtreet) ikke er i slekt (allelene som nedarves fra disse er helt sikkert ikke IBD). Dermed blir $P_2 = 0$. Lar vi X og Y være fenotypene til E og D , får vi dermed

$$\begin{aligned}
 \text{Cov}(X, Y) &= 2FV_A + P_2V_D \\
 &= 2 \cdot \frac{1}{16} \cdot 1.125 \\
 &\approx 0.14
 \end{aligned}$$

Vi kan ikke bruke samme metode til å finne F_{DH} , fordi vi da måtte tatt hensyn til innavl (foreldrene til D og H er halvsøsken). For systemer uten innavl trenger vi kun å vurdere P_0 , P_1 og P_2 (sannsynlighetene for at hhv 0, 1, eller 2 av allelene i det ene individet er IBD med et allel hos slektningen). Men når vi har innavl får vi flere muligheter for nedarving av allelene. I dette eksempelet kan individ H f.eks motta samme genkopi fra mor og far. Det er også mulig å regne på slike tilfeller med innavl, men det gir komplikasjoner som ikke tas hensyn til i boka (kap 9).

Oppgave 3

- a) Vi antar at populasjonsstørrelsen er konstant hver generasjon. For å finne effektiv populasjonsstørrelse trenger vi variansen i antall gameter som et individ viderefører til neste generasjon. Dette tilsvarer her variansen i antall avkom, men siden selvertilisering er mulig må vi anta at avkom produsert ved selvertilisering teller som to avkom.

Ser på et vilkårlig individ og lar X være antall avkom det får i løpet av livet. Siden det bare er to muligheter (0 eller 10 avkom) tilsvarer det et Bernoulliforsøk. La $Y = 1$ dersom individet overlever vinteren (sannsynlighet for dette er 0.2) og $Y = 0$ ellers. Bruker vi loven om total forventning får vi

$$\begin{aligned} E[X] &= 10 \cdot 0.2 + 0 \cdot 0.8 = 2 \\ E[X^2] &= 100 \cdot 0.2 + 0 \cdot 0.8 = 20 \\ \text{Var}(X) &= E[X^2] - E[X]^2 = 16. \end{aligned}$$

Dermed blir effektiv populasjonsstørrelse

$$\begin{aligned} N_e &= \frac{4N - 2}{2 + \text{Var}(X)} \\ &= \frac{400 - 2}{2 + 16} \\ &\approx 22.1. \end{aligned}$$

- b) Siden vi har drift uten seleksjon blir forventet allelfrekvens lik startfrekvensen, altså $E[p_t] = p_0 = 0.6$. Variansen er gitt ved

$$\begin{aligned} \text{Var}(p_t) &= p_0(1 - p_0) \left[1 - \left(1 - \frac{1}{2N_e} \right)^t \right] \\ &= 0.6(1 - 0.6) \left[1 - \left(1 - \frac{1}{2 \cdot 22.1} \right)^t \right]. \end{aligned}$$

Etter 20 generasjoner får vi

$$\text{Var}(p_{20}) = 0.6(1 - 0.6) \left[1 - \left(1 - \frac{1}{2 \cdot 22.1} \right)^{20} \right] \approx 0.09.$$

- c) Vi starter med en andel på 0.48 heterozygoter (fra HW). Etter 20 generasjoner er denne andelen blitt

$$\left(1 - \frac{1}{2 \cdot 22.1}\right)^{20} \cdot 0.48 \approx 0.30.$$