



Matematisk evolusjonær genetikk, ST2301 – Løsningsforslag
Torsdag 8. desember 2005

Oppgave 1

- a) Sannsynlighetsmaksimeringsestimaterne av allelfrekvensene blir lik frekvensen av hvert allel i utvalget, altså $\hat{p}_1 = \frac{2 \cdot 2 + 3 + 14}{2 \cdot 100} = 0.105$, $\hat{p}_2 = 0.185$, og $\hat{p}_3 = 0.71$.
- b) Dersom populasjonen er i Hardy-Weinberg likevekt er genotypfrekvensene funksjoner av allelfrekvensene og SMEene av genotypfrekvensene er samme funksjoner av SMEene av allelfrekvensene. Dette gir at

$$\hat{P}_{11} = \hat{p}_1^2 = 0.011 \quad (1)$$

$$\hat{P}_{12} = 2\hat{p}_1\hat{p}_2 = 0.0389 \quad (2)$$

$$\hat{P}_{13} = 0.1491 \quad (3)$$

$$\hat{P}_{22} = 0.0342 \quad (4)$$

$$\hat{P}_{23} = 0.2627 \quad (5)$$

$$\hat{P}_{33} = 0.5041 \quad (6)$$

- c) Under H_0 er testobservatoren

$$\chi^2 = \sum_{\text{alle genotyper } ij} \frac{(n_{ij} - n\hat{P}_{ij})^2}{n\hat{P}_{ij}} \quad (7)$$

kji-kvadrat fordelt med $k - 1 - m = 3$ frihetsgrader når $k = 6$ (antall kategorier) og $m = 2$ (sannsynlighetene for observasjoner i de 6 ulike kategoriene kan skrives som funksjoner av 2 estimerte parametere (2 av 3 allelfrekvenser)).

Observert verdi av testobservatoren blir (innsetting fra forrige punkt)

$$\chi^2 = \frac{(2 - 1.1)^2}{1.1} + \frac{(3 - 3.89)^2}{3.89} + \dots = 1.09 \quad (8)$$

Dersom vi velger signifikansnivå lik 0.05 blir kritisk verdi for testen øvre 0.05-kvantil i en kji-kvadratfordeling, $\chi_{3,0.05}^2 = 7.81$. Siden observert verdi er mindre enn kritisk verdi kan vi ikke forkaste nullhypotesen at populasjonen er i Hardy-Weinberg likevekt.

- d) Vi tar utgangspunkt i at antall kopier av allel A_2 i utvalget, $N_2 = N_{12} + 2N_{22} + N_{23}$ vil være binomisk fordelt med parametere p_2 og $2n$ hvis populasjonen er i Hardy-Weinberg-likevekt fordi hendelsene at hver genkopi i et gitt individ er av type A_2 da er uavhengige slik at vi får en binomisk forsøksrekke når vi teller opp totalt antall kopier av A_2 i utvalget (av totalt $2n$ mulige). Da har vi videre at størrelsen

$$\frac{\hat{p}_2 - p_2}{\sqrt{\hat{p}_2(1 - \hat{p}_2)/(2n)}} \quad (9)$$

er tilnærmet standardnormalfordelt slik at

$$P(-z_{\alpha/2} < \frac{\hat{p}_2 - p_2}{\sqrt{\hat{p}_2(1 - \hat{p}_2)/(2n)}} < z_{\alpha/2}) = 1 - \alpha. \quad (10)$$

Isolerer vi p_2 midt i dobbeltulikheten får vi at

$$(\hat{p}_2 - z_{\alpha/2}\sqrt{\hat{p}_2(1 - \hat{p}_2)/(2n)}, \hat{p}_2 + z_{\alpha/2}\sqrt{\hat{p}_2(1 - \hat{p}_2)/(2n)}) \quad (11)$$

blir et $(1 - \alpha)$ -konfidensintervall for p_2 . For $\alpha = 0.05$ blir intervallet $(0.131, 0.238)$.

Oppgave 2

- a) Tilnærmingen basert på liten s og kontinuerlig tid gir

$$t = \frac{1}{s} \left(\ln\left(\frac{p_t}{1 - p_t}\right) - \ln\left(\frac{p_0}{1 - p_0}\right) \right). \quad (12)$$

Setter vi inn $p_t = 0.0001$, $p_0 = 0.5$, og $s = -1/2$ får vi $t = 18.4$. Dette betyr at p_t passerer 0.0001 etter 19 generasjoner.

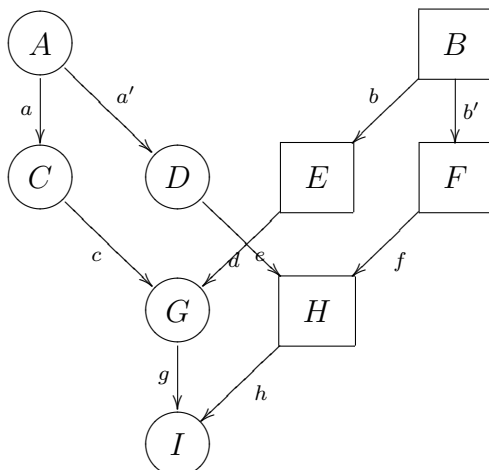
- b) For $s = -1/2$ vil rekursjonsligningen som beskriver endring i genfrekvens være $p_{t+1} = p_t/2$. Dette gir

$$t = (\ln p_0 - \ln p_t) / \ln 2 = 12.28, \quad (13)$$

altså passering etter 13 generasjoner. Tilnærmingen brukt i punkt a) fungerer m.a.o. ikke spesielt bra når s er så stor som $-1/2$.

Oppgave 3

- a) Vi setter først navn på gametene som nedarves.



Løkkemetoden gir at $f_I = \frac{1}{25} + \frac{1}{25} = 1/16$ for autosomale loci.

For kjønnskoblede loci vil (bl.a.) $P(e \leftarrow b) = 0$ slik at løkken gjennom B blir brutt. Dessuten blir $P(h \leftarrow d) = 1$ fordi det X-kromosomet som nedarves fra H til I må komme fra H 's mor siden H har fått et Y-kromosom fra F . Dermed blir for $f_I = \frac{1}{25-1} = 1/16$ også ved kjønnskobling.

- b) Slektskapskoeffisienten mellom G og H blir $F_{GH} = f_I = 1/16$. Forøvrig er G og H doble halvsøskenbarn.

- c) Innsetting i f.eks. (IX-42) og (IX-45) gir at den totale additive variansen blir

$$V_A = 2p(1-p)[pa_{11} + (1-p)a_{12} - pa_{12} - (1-p)a_{22}]^2 = 0.375, \quad (14)$$

og dominansvariansen

$$V_D = p^2(1-p)^2(a_{11} + a_{22} - 2a_{12})^2 = 0.5625. \quad (15)$$

- d) At de maternale og paternale genkopiene i G og H er IBD er hendelsen $\{c \equiv d \cap e \equiv f\}$. Dette kan bare inntreffe ved at $\{c \leftarrow a \equiv a' \rightarrow d\} \cap \{e \leftarrow b \equiv b' \rightarrow f\}$. Sannsynligheten for denne hendelsen (ett snitt av 6 uavhengige enkelthendelser) blir $\frac{1}{2^6} = 1/64$.

- e) Sannsynligheten beregnet i forrige punkt kalles P_2 i boka (også kalt "coefficient of fraternity"). Slektskapskoeffisienten F er kjent fra punkt b) og varianskomponentene V_A og V_D fra punkt c). Dermed får vi at kovariansen mellom G og H 's fenotyper blir

$$\text{Cov}(X, Y) = 2FV_A + P_2V_D = 2\frac{1}{16}0.375 + \frac{1}{64}0.563 = 0.0557 \quad (16)$$

f) Setter vi (14) lik 0 får vi at

$$2p + 4 - 4p - 4p - 2 + 2p = 0 \quad (17)$$

som gir $p = 1/2$. Ved denne allelfrekvensen blir arvegraden (heritability) $h^2 = V_A/V_P$ også lik null og dermed også respons til seleksjon $R = h^2 S$ uansett hvordan vi selekterer. Dette stemmer forøvrig bra men single-locus modell for endring i genfrekvens ved over eller underdominans. Om vi f.eks. selekterer alle individ større enn f.eks. 4 vil heterozygoten få størst fitness mens homozygotene for mindre og like store relative fitnesser. Da vet vi at $\Delta p = 0$ ved $p = 1/2$. Prediksjonene basert på kvantitativ genetikk er m.a.o. i overenstemmelse med prediksjonene basert på single-locus teori i dette tilfelle.