



Matematisk evolusjonær genetik, ST2301
Fredag 14. desember 2007
Løsningsforslag

Oppgave 1

- a) Hvert observert individ er en av to mulige fenotyper. Fenotypene til ulike individ er uavhengige. Dermed er antall av fenotype aa , N_{aa} , binomisk fordelt med parameter lik genotypefrekvensen P_{aa} og utvalgsstørrelsen n .
- b) I følge vanlig teori for binomisk fordeling blir sannsynlighetsmaksimeringsestimatet av P_{aa} , $\hat{P}_{aa} = N_{aa}/n = 9/100 = 0.09$. Parameteren p_a (allelfrekvensen) er funksjonen $p_a = \sqrt{P_{aa}}$ av parameteren P_{aa} (genotypefrekvensen). Følgelig er SMEen av p_a samme funksjon av SMEen av P_{aa} , d.v.s. $\hat{p}_a = \sqrt{\hat{P}_{aa}} = 0.3$.
- c) Standard metode for konfidensintervall med utgangspunkt i at $N_{aa} \sim \text{bin}(P_{aa}, n)$ gir 95%-konfidensintervallet $(\hat{P}_{aa} - z_{\alpha/2} \sqrt{\hat{P}_{aa}(1 - \hat{P}_{aa})/n}, \hat{P}_{aa} + z_{\alpha/2} \sqrt{\hat{P}_{aa}(1 - \hat{P}_{aa})/n}) = (0.034, 0.146)$ for P_{aa} .

Dette betyr at

$$P(0.034 \leq P_{aa} \leq 0.146) = 0.05 \quad (1)$$

som er ekvivalent med at

$$P(\sqrt{0.034} \leq \sqrt{P_{aa}} \leq \sqrt{0.146}) = 0.05 \quad (2)$$

som betyr at $(0.184, .382)$ er et 95%-konfidensintervall for $p_a = \sqrt{P_{aa}}$.

- d) Sålenge vi antar Hardy-Weinberg-likevekt er nullhypotesen $p_a = 0.5$ samme utsagn som som nullhypotesen $P_{aa} = 0.25$. Standard testmetode basert på normaltilnærming av binomisk modell gir da testobservatoren

$$Z^* = \frac{0.09 - 0.25}{\sqrt{.25(1 - .25)/100}} = -3.69. \quad (3)$$

P-verdien for testen blir dermed

$$p = P(Z < Z^* \cup Z > -Z^*) = 2P(Z < -3.69) = 0.00022, \quad (4)$$

altså forkasning av nullhypotesen $P_{aa} = 0.25$ (og dermed også $p_a = 0.5$).

Oppgave 2

- a) Transformasjonen $P' = \log_2(P)$ gir følgende fenotypiske verdier for ulike genotyper:

	BB	Bb	bb
AA	2	3	1
Aa	3	4	2
aa	1	2	0

Velger vi å jobbe med disse transformert verdiene kan vi skrive P' som en summen

$$P' = \begin{cases} 1 & \text{for } AA \\ 2 & \text{for } Aa \\ 0 & \text{for } aa \end{cases} + \begin{cases} 1 & \text{for } BB \\ 2 & \text{for } Bb \\ 0 & \text{for } bb \end{cases}. \quad (5)$$

- b) Siden P_1 og P_2 er fiksert for ulike allel vil F_1 generasjonen bare bestå av doble heterozygoter med transformert fenotype $P' = 4$.
- c) Koblingsulikevekten i F_1 generasjonen blir $D_{AB} = P_{AB} - p_A p_B = 1/2 - (1/2)^2 = 1/4$ siden F_1 generasjonen er dannet av like store andeler AB og ab gameter.
- d) Siden vi har additivitet mellom loci kan vi bruk standard teori for forventingsverdier til tilbakekrysninger som gir at

$$BC_2 = \frac{1}{2}P_2 + \frac{1}{2}F_1 = \frac{1}{2}0 + \frac{1}{2}4 = 2. \quad (6)$$

- e) I F_2 generasjonen dannet ved tilfeldig parring innad i F_1 får vi H-W-likevekt. Dekomposisjonen for begge loci blir identiske med

$$\mu = \frac{1}{4}1 + \frac{1}{2}2 + \frac{1}{4}0 = \frac{5}{4}, \quad (7)$$

$$\alpha_1 = p_1 a_{11} + p_2 a_{12} - \mu \quad (8)$$

$$= \frac{1}{2}1 + \frac{1}{2}2 - \mu = \frac{1}{4}, \quad (9)$$

$$\alpha_2 = -\frac{1}{4}, \quad (10)$$

$$(11)$$

$$\delta_{11} = a_{11} - \alpha_1 - \alpha_2 - \mu \quad (12)$$

$$= 1 - 2\frac{1}{4} - \frac{5}{4} = -\frac{3}{4}, \quad (13)$$

$$\delta_{12} = \frac{3}{4}, \quad (14)$$

$$\delta_{22} = -\frac{3}{4}. \quad (15)$$

Dette gir additiv varians for hvert loci

$$V_A^{(i)} = 2 \text{Var}(\alpha_I) = 2(p_1(\alpha_1)^2 + p_2(\alpha_2)^2) = \frac{1}{8}, \quad (16)$$

og dominansvarianser lik

$$V_D^{(i)} = \text{Var}(\delta_{IJ}) = p_1^2(\delta_{11})^2 + 2p_1p_2(\delta_{12})^2 + p_2^2(\delta_{22})^2 = \frac{9}{16}. \quad (17)$$

Summerer vi bidragen fra hvert loci får vi da total additiv og dominansvariens $V_A = \frac{1}{4}$ og $V_D = \frac{9}{8}$. Fenotypisk varians blir dermed $V_P = V_A + V_D + V_E = \frac{1}{4} + \frac{9}{8} + 0 = \frac{11}{8}$. Forventet fenotype blir $E(P') = 2\mu = \frac{5}{2}$.

- f) Velger vi ut individ med $P \geq 8$ svarer dette til $P' \geq 3$. Om vi antar at fenotypiske verdier er normalfordelt i populasjonen velger vi dermed ut en andel

$$S = P(P' \geq 3) = P(Z \geq \frac{3 - 5/2}{\sqrt{11/8}}) = 1 - P(Z < 0.42) = 0.33 \quad (18)$$

Dette gir en seleksjonsintensitet

$$i = \frac{1}{0.33} \frac{1}{\sqrt{2\pi}} e^{-0.5^2/2} = 1.06, \quad (19)$$

og et seleksjonsdifferensial

$$S = \bar{X} - \mu = i\sqrt{V_P} = 1.06\sqrt{(11/8)} = 1.24. \quad (20)$$

Arvegraden blir

$$h^2 = \frac{V_A}{V_P} = \frac{1/4}{11/8} = \frac{2}{11} = 0.182. \quad (21)$$

Breeders equation gir da respons til seleksjon lik

$$R = h^2 S = 0.182 \cdot 1.24 = 0.22 \quad (22)$$

- g) Den faktiske fordelingen av fenotypiske verdier i populasjonen er ikke kontinuerlig og normal men diskret og består av verdiene 0,1,2,3 og 4. Genotypefrekvensene er gitt ved

	BB	Bb	bb
AA	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
Aa	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
aa	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

og frekvensen av individer med fenotype $P' = 3$ og $P' = 4$ blir dermed henholdsvis $1/4$ og $1/4$. Dermed blir gjennomsnittlig fenotype blant selekterte foreldre ($P' \geq 3$) lik $\bar{X} = 7/2$ og seleksjonsdifferensialet $S = \bar{X} - \mu = 7/2 - 5/2 = 1$. Dette gir respons til seleksjon $R = h^2 S = 0.18$.

- h) De tre genotypene selekterte foreldre produserer fire forskjellige gamettyper med følgende frekvenser

Genotype	Frekvens	AB	Ab	aB	aa
AABb	$1/4$	$1/2$	$1/2$	0	0
AaBB	$1/4$	$1/2$	0	$1/2$	0
AaBb	$1/2$	$1/4$	$1/4$	$1/4$	$1/4$
Total gametfrekvens		$3/8$	$1/4$	$1/4$	$1/8$

Forenes disse fire gamettyper ved tilfeldig parring får vi dermed følgende genotypefrekvenser

	AB	Ab	aB	ab
AB	$9/64$	$3/32$	$3/32$	$3/64$
Ab	$3/32$	$1/16$	$1/16$	$1/32$
aB	$3/32$	$1/32$	$1/16$	$1/32$
ab	$3/64$	$1/32$	$1/32$	$1/64$

Tilsvarende tabell for fenotypiske verdier blir

	AB	Ab	aB	ab
AB	2	3	3	4
Ab	3	1	4	2
aB	3	4	1	2
ab	4	2	2	0

Dette gir gjennomsnittlig fenotypisk verdi etter seleksjon lik $E(Y) = 85/32 = 2.66$ og respons til seleksjon lik $2.66 - 5/2 = 0.16$.

Forskjellen mellom resultatet i punkt f) og g) skyldes at seleksjonsdifferensialet i f) er feil siden fenotypiske verdier ikke er normalfordelt.

Breeders equation bygger også på at fenotypene til foreldre avkom er multinormalfordelt slik at sammenhengen mellom forventet fenotype til avkom og midparental value blir lineær. Her har vi ikke multinormalfordeling fordi fenotypen er bestemt av bare 2 loci.