Løsningsforslag ST2301 Øving 4

Kapittel 1

Exercise 11

Et utvalg på 100 individer trekkes fra en populasjon med tilfeldig parring. Det ble observert $n_{AA} = 63$ individer av genotype AA, $n_{Aa} = 27$, og $n_{aa} = 10$. Lag et 95 % konfidensintervall for frekvensen til A. Hva må du anta?

Svar:

Finner først frekvensen av A i utvalget.

$$\hat{p}_A = \frac{2n_{AA} + n_{Aa}}{2n} = \frac{2 \cdot 63 + 27}{200} = 0.765$$

Variansen til $\hat{p}_A$ må estimeres (bruker $\hat{p}_A$ for å tilnærme variansen til $\hat{p}_A$...).

$$\begin{aligned}\hat{\sigma}^2 &\approx \frac{\hat{p}_A(1-\hat{p}_A)}{2n} = \frac{0.765(1-0.765}{200} \approx 8.99 \cdot 10^{-4} \\ \hat{\sigma} &\approx 0.0300\end{aligned}$$

Et tilnærmet 95 % konfidensintervall for p_a blir dermed

$$\begin{aligned}\hat{p}_A - 1.96\hat{\sigma} &\leq p_A \leq \hat{p}_A + 1.96\hat{\sigma} \\ 0.765 - 1.96 \cdot 0.0300 &\leq p_A \leq 0.765 + 1.96 \cdot 0.0300 \\ 0.7062 &\leq p_A \leq 0.765 + 1.96 \cdot 0.8238\end{aligned}$$

Her måtte vi anta tilnærming av binomisk fordeling til normalfordeling. For å konstruere konfidensintervallet ble det antatt at $\hat{p}_A$ er tilnærmet normalfordelt med forventning p_A og varians σ^2, der

$$\hat{\sigma}^2 \approx \frac{p_A(1-p_A)}{2n}$$

Siden p_A er ukjent (det er den det skulle lages konfidensintervall for), måtte vi bruke et estimat, der $\hat{p}_A$ inngår. Tilnærmingen av binomisk fordeling til normalfordeling fungerer bra så lenge p_A ikke er for nær 0 eller 1, og utvalget er stort nok (et vanlig krav er at $np_A > 5$ og $n(1-p_A) > 5$, som er oppfylt her).

Complement 2

Anta at det er n allel med samme frekvens. Som funksjon av n, hva er andelen homozygoter og heterozygoter i populasjonen?

Svar:

Har at

$$
\begin{aligned}
p_1 = p_2 = ... = p_n = p \\
\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} p = np = 1 \\
p = \frac{1}{n}
\end{aligned}
$$

Andelen homozygoter P_{ii} er

$$
P_{ii} = \sum_{i=1}^{n} p^2 = np^2 = n\,\frac{1}{n^2} = \frac{1}{n}
$$

Andelen heterozygoter er $P_{ij} = 1 - P_{ii} = 1 - \frac{1}{n}$. Eventuelt kan man finne andelen heterozygoter som

$$
P_{ij} = \sum_{i \neq j}^{n} p^2 = (n-1)p\sum_{i=1}^{n} p = (n-1)\frac{1}{n}n\frac{1}{n} = 1 - \frac{1}{n}
$$

Complement 10

Har et recessivt, kjønnskoplet gen A med frekvens p, i et vanlig $XX - XY$-system. Hvilken andel av individene med genet (dvs AA homozygote hunner eller AY hemizygote hanner) er hunner? Anta tilfeldig parring, at populasjonen har kjønnsratio 1:1 og at likevektsfrekvensene av allelene er nådd.

Svar:

Siden kjønnsratioen er 1:1, er sannsynligheten for at et individ er en hann lik sannsynligheten for at det er en hunn, dvs

$$
P(\text{Hann}) = P(\text{Hunn}) = \frac{1}{2}
$$

Når likevektsfrekvensene er nådd er $p_m = p_f = p$. Vi ønsker å finne sannsynligheten for at et individ er en hunn, gitt at det har A-allelet, dvs

$$P(\text{Hunn}|AA \cup AY)$$

Har at

$$\begin{aligned}
P(AA|\text{Hunn}) &= p^2 \\
P(AY|\text{Hann}) &= p \\
P(AA|\text{Hann}) &= P(AY|\text{Hunn}) = 0
\end{aligned}$$

Loven om total sannsynlighet gir de ubetinga genotypefrekvensene.

$$\begin{aligned}
P(AA) &= P(AA|\text{Hunn})P(\text{Hunn}) + P(AA|\text{Hann})P(\text{Hann}) \\
&= p^2 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} \\
&= \frac{p^2}{2} \\
P(AY) &= P(AY|\text{Hunn})P(\text{Hunn}) + P(AY|\text{Hann})P(\text{Hann}) \\
&= 0 \cdot \frac{1}{2} + p \cdot \frac{1}{2} \\
&= \frac{p}{2}
\end{aligned}$$

Siden et individ enten er hann eller hunn, er hendelsene AA og AY disjunkte. Dette gir at

$$P(AA \cup AY) = P(AA) + P(AY) = \frac{p\,(1+p)}{2}$$

Nå kan Bayes regel brukes til å finne den ønska sannsynligheten.

$$\begin{aligned}
P(\text{Hunn}|AA \cup AY) &= \frac{P(\text{Hunn} \cap AA \cup AY)}{P(AA \cup AY)} \\
&= \frac{P(\text{Hunn}capAA) + P(\text{Hann}capAA)}{P(AA \cup AY)} \\
&= \frac{P(AA|\text{Hunn})P(\text{Hunn}) + 0}{P(AA \cup AY)} \\
&= \frac{p^2/2}{p\,(1+p)/2} \\
&= \frac{p}{1+p}
\end{aligned}$$

Complement 11

Et autosomalt locus har et allel a med frekvens p. I en populasjon med tilfeldig parring, hvilken andel av alle eksisterende kopier av a-allelet finnes i aa homozygoter?

Svar:

Andelen a-alleler i populasjonen er p, og andelen aa-homozygoter i populasjonen er p^2 (ved Hardy-Weinberg-andeler). Andelen av a-allel som er i homozygoter er derfor $\frac{p^2}{p} = p$.

En alternativ løsning: Hvis populasjonsstørrelsen er N er det $2NP_{aa} = 2Np^2$ allel av type a som befinner seg i homozygoter aa, og $1NP_{Aa} = N2p(1-p)$ som befinner seg i heterozygoter Aa. Andel allel i homozygoter blir da

$$\frac{2Np^2}{2Np^2 + N2p(1-p)} = p$$

En annen alternativ løsning: Det det er spurt om er det samme som den betingede sannsynligheten for at det andre allelet i et gitt individ er av type a gitt at det første er av type a. Under Hardy-Weinberg-likevekt er disse hendelsene uavhengige, dermed er

$$P(\text{Andre allel type}\, a | \text{Første allel type } \ a) = P(\text{Andre allel type A}) = p$$

Kapittel 2

Exercise 3

Har en haploid populasjon med to alleler A og a, med fitnesser $1 + s : 1$.

1. Hvor lang tid vil det ta å endre frekvensen av A fra 0.1 til 0.2, dersom $s = 0.01$?

2. Hvor lang tid vil det ta å endre frekvensen av A fra 0.9 til 0.8 dersom $s = -0.01$?

3. Forklar hvorfor disse tidene er eller ikke er de samme.

Svar:

1. Bruker likning II-67 side 49, for å finne tida t det tar å endre frekvensen fra $p_A(0) = 0.1$ til $p_A(t) = 0.2$, når $s = 0.01$.

$$
\begin{aligned}
t &= \frac{\ln\left(\frac{p_A(t)}{1-p_A(t)}\right) - \ln\left(\frac{p_A(0)}{1-p_A(0)}\right)}{\ln(1+s)} \\
&= \frac{\ln\left(\frac{0.2}{0.8}\right) - \ln\left(\frac{0.1}{0.9}\right)}{\ln(1.01)} \\
&\approx 81.5
\end{aligned}
$$

2. Bruker igjen likning II-67 side 49, nå er $p_A(0) = 0.9$, $p_A(t) = 0.8$, og $s = -0.01$.

$$
\begin{aligned}
t &= \frac{\ln\left(\frac{p_A(t)}{1-p_A(t)}\right) - \ln\left(\frac{p_A(0)}{1-p_A(0)}\right)}{\ln(1+s)} \\
&= \frac{\ln\left(\frac{0.8}{0.2}\right) - \ln\left(\frac{0.9}{0.1}\right)}{\ln(0.99)} \\
&\approx 80.7
\end{aligned}
$$

3. Årsaken til at de to tidene ikke er like, er at fitnessene ikke er like i de to tilfellene. I det første tilfellet er det frekvensen av A-allelet som endres fra 0.1 til 0.2, mens det i det andre tilfellet er frekvensen av a-allelet som endres fra 0.1 til 0.2. Fitnessene var $1 + s : 1$ i det første tilfellet. For at de to tilfellene skulle vært like, måtte man hatt at fitnessene var $1 : 1 + s$ i det andre tilfellet (standardiserer mot A her, fordi A og a har "byttet plass" sammenliknet med det første tilfellet). Fitnessen i det andre tilfellet var $1 - s : 1$ (når s holdes lik som i første tilfellet). Standardiseres disse fitnessene mot A, får man

$$
\begin{aligned}
w_A &= \frac{1-s}{1-s} = 1 \\
w_a &= \frac{1}{1-s} \neq 1 + s
\end{aligned}
$$

Man kan også se direkte av likning II-67 at de to tilfellene blir ulike. Har at for $s \neq 0$, er

$$
\ln(1+s) \neq \ln(1-s)
$$

Telleren i likning II-67 blir den samme i de to tilfellene, men med motsatt fortegn.

$$\ln\left(\frac{0.2}{0.8}\right) - \ln\left(\frac{0.1}{0.9}\right) = \ln 0.2 - \ln 0.8 - \ln 0.1 + \ln 0.9$$
$$= -\left[\ln\left(\frac{0.8}{0.2}\right) - \ln\left(\frac{0.9}{0.1}\right)\right]$$

Nevneren, som var $\ln(1+s)$, blir altså forskjellig for $s \neq 0$.

Exercise 4

Hvor stor må seleksjonskoeffisienten mot et dominant allel være for å kunne endre allelfrekvensen fra 0.5 til 0.51 i løpet av én generasjon? Regn ut dette eksakt, og sammenlikn med en passende tilnærming.

Svar:

Kaller det dominante allelet A, og det recessive a. Da er fitnessene for hver genotype gitt ved (side 48)

AA	Aa	aa
1+s	1+s	1

Bruker likning II-59 side 48, setter inn $p = 0.5$ og $p' = 0.51$, og løser ut s.

$$\frac{p'}{1-p'} = \frac{p}{1-p}\frac{1+s}{1-sp}$$
$$\frac{0.51}{0.49} = \frac{1+s}{1-0.5s}$$
$$\frac{0.51}{0.49} - \frac{0.51 \cdot 0.5s}{0.49} = 1+s$$
$$\frac{0.51}{0.49} - 1 = s\left(1 + \frac{0.51 \cdot 0.5}{0.49}\right)$$
$$s = \frac{0.51 - 0.49}{0.49 + 0.51 \cdot 0.5}$$
$$\approx 0.08511$$

En passende tilnærming er gitt ved likning II-97 side 57. Setter $t = 1$, $p_t = p_1 = 0.51$ og $p_0 = 0.5$, og løser ut s.

$$t \approx \frac{1}{s}\left[\frac{1}{1-p_t} + \ln\left(\frac{p_t}{1-p_t}\right) - \frac{1}{1-p_0} - \ln\left(\frac{p_0}{1-p_0}\right)\right]$$
$$s \approx \frac{1}{0.49} + \ln\left(\frac{0.51}{0.49}\right) - \frac{1}{0.5} - \ln 1$$
$$\approx 0.08082$$