Løsningsforslag

Eksamen i Statistikk SIF5060 Aug 2002

Oppgave 1

a)

En god estimator er *forventningsrett* og har liten varians.

Vi tester forventningsretthet:

$$
\begin{aligned}
\mathrm{E}[\hat{\mu}] &= \mathrm{E}[Y] = \mu \\
\mathrm{E}[\tilde{\mu}] &= \mathrm{E}[\frac{1}{2}X + \frac{1}{2}Y] = \frac{1}{2}\,\mathrm{E}[X] + \frac{1}{2}\,\mathrm{E}[Y] = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu \\
\mathrm{E}[\mu^*] &= \mathrm{E}[\frac{1}{5}X + \frac{4}{5}Y] = \frac{1}{5}\,\mathrm{E}[X] + \frac{4}{5}\,\mathrm{E}[Y] = \frac{1}{5}\mu + \frac{4}{5}\mu = \mu.
\end{aligned}
$$

Alle tre estimatorene er altså forventningsrette. Vi sammenligner variansene:

$$
\begin{aligned}
\mathrm{Var}[\hat{\mu}] &= \mathrm{Var}[Y] = \sigma_A^2 = 0.01 \\
\mathrm{Var}[\tilde{\mu}] &= \mathrm{Var}[\frac{1}{2}X + \frac{1}{2}Y] = \frac{1}{2^2}\,\mathrm{Var}[X] + \frac{1}{2^2}\,\mathrm{Var}[Y] = \frac{1}{4}{\sigma_A}^2 + \frac{1}{4}\sigma_B^2 = 0.0125 \\
\mathrm{Var}[\mu^*] &= \mathrm{Var}[\frac{1}{5}X + \frac{4}{5}Y] = \frac{1}{5^2}\,\mathrm{Var}[X] + \frac{4^2}{5^2}\,\mathrm{Var}[Y] = \frac{1}{25}\sigma_A^2 + \frac{16}{25}\sigma_B^2 = 0.008.
\end{aligned}
$$

Her bruker vi at X og Y er uavhengige. Vi ser at μ^* har minst varians, og foretrekker denne som estimator for μ.

b)

μ^* er normalfordelt siden den er en lineærkombinasjon av to uavhengige normalfordelte variabler. Forventningen og variansen er beregnet i oppgave a), og gir at

$$\mu^* \sim N(\mu, 0.008).$$

90%-konfidensintervall for μ finnes da fra at

$$
\begin{aligned}
0.90 &= \mathrm{P}\left(-z_{0.05} < \frac{\mu^* - \mu}{\sqrt{0.008}} < z_{0.05}\right) \\
&= \mathrm{P}\left(\mu^* - z_{0.05}\sqrt{0.008} < \mu < \mu^* + z_{0.05}\sqrt{0.008}\right),
\end{aligned}
$$

der $z_{0.05} = 1.645$ er 5%-kvantilen i standard normalfordelingen, slik at

$$\left(\mu^* - 1.645\sqrt{0.008},\ \mu^* + 1.645\sqrt{0.008}\right) \quad = \quad (\mu^* - 0.15,\ \mu^* + 0.15)$$

er et 90%-konfidensintervall for μ.

Oppgave 2

a)

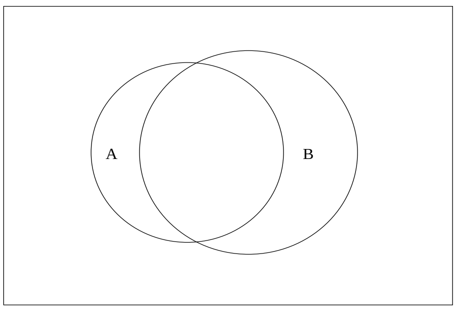
Vi har at

- A = Personen har sykdommen
- B = Prøven indikerer at personen har sykdommen

Sannsynlighetene som er oppgitt kan skrives som

$$\begin{aligned} \mathrm{P}(B|A) &= \frac{9}{10}, & \mathrm{P}(B^c|A) &= \frac{1}{10} \\ \mathrm{P}(B|A^c) &= \frac{1}{20}, & \mathrm{P}(B^c|A^c) &= \frac{19}{20} \\ \mathrm{P}(A) &= \alpha, \end{aligned}$$

og Venn-diagrammet er gitt i figur 1.



Figur 1: Venn-diagram i oppgave 2a).

Sannsynligheten for at en tilfeldig valgt person både har sykdommen og at prøven slår ut:

$$\mathrm{P}(A \cap B) \quad = \quad \mathrm{P}(B|A)\mathrm{P}(A) \quad = \quad \underline{\frac{9}{10}\alpha}.$$

Sannsynligheten for at prøven til en tilfeldig valgt person slår ut:

$$\begin{aligned} P(B) &= P(A\cap B) + P(A^c\cap B) \quad = \quad P(A\cap B) + P(B|A^c)P(A^c) \\ &= \frac{9}{10}\alpha + \frac{1}{20}(1-\alpha) \quad = \quad \underline{\frac{17\alpha+1}{20}}. \end{aligned}$$

b)

$\hat{\alpha}$ er SME for α og er gitt ved den $\alpha \in [0,1]$ som maksimerer rimelighetsfunksjonen til den binomiske fordelingen

$$L(\alpha) = \binom{n}{k}\left(\frac{17\alpha+1}{20}\right)^X\left(\frac{19-17\alpha}{20}\right)^{n-X},$$

og siden ln-funksjonen er monotont stigende, blir dette det samme som å finne maksimum til funksjonen

$$\begin{aligned} l(\alpha) &= \ln(L(\alpha)) \\ &= \text{konst} + X\ln(17\alpha+1) + (n-X)\ln(19-17\alpha). \end{aligned}$$

$l(\alpha)$ har et lokalt maksimum i α hvis

$$\begin{aligned} l'(\alpha) &= 17\left(\frac{X}{17\alpha+1} - \frac{n-X}{19-17\alpha}\right) = 0 \\ &\Downarrow \\ \alpha &= \frac{1}{17}\left(\frac{20X}{n} - 1\right), \end{aligned}$$

og hvis $0 < \alpha < 1$. Dette er tilfellet når $\frac{n}{20} < X < \frac{18n}{20}$. Når X er i dette intervallet har vi dessuten at α er et maksimumspunkt siden $l'(\alpha)$ skifter fortegn fra positiv til negativ i α.

Hvis $l(\alpha)$ ikke har et lokalt maksimum på intervallet $(0,1)$, må maksimum ligge i et av endepunktene.

Når $X < \frac{n}{20}$, har $l(\alpha)$ maksimum i 0 (siden $l'(\alpha)$ er negativ på hele $(0,1)$).

Når $X > \frac{18n}{20}$, har $l(\alpha)$ maksimum i 1 (siden $l'(\alpha)$ er positiv på hele $(0,1)$).

Dermed følger det oppgitte uttrykket for $\hat{\alpha}$.

c)

Forventning til $\tilde{\alpha} = \frac{1}{17}\left(\frac{20X}{n} - 1\right)$:

$$\begin{aligned}
\mathrm{E}[\tilde{\alpha}] &= \mathrm{E}\left[\frac{1}{17}\left(\frac{20X}{n} - 1\right)\right] = \frac{1}{17}\left(\frac{20\,\mathrm{E}[X]}{n} - 1\right) \\
&= \frac{1}{17}\left(\frac{20n(17\alpha+1)/20}{n} - 1\right) = \alpha,
\end{aligned}$$

siden $\mathrm{E}[X] = np = n(17\alpha+1)/20$.

Variansen er gitt ved

$$\begin{aligned}
\mathrm{Var}[\tilde{\alpha}] &= \mathrm{Var}\left[\frac{1}{17}\left(\frac{20X}{n} - 1\right)\right] = \frac{1}{17^2}\,\mathrm{Var}\left[\frac{20X}{n} - 1\right] \\
&= \frac{20^2}{n^2 17^2}\,\mathrm{Var}\left[X\right] = \frac{20^2}{n^2 17^2}\frac{n}{20^2}(17\alpha+1)(19-17\alpha) \\
&= \frac{1}{n}\left(\alpha + \frac{1}{17}\right)\left(\frac{19}{17} - \alpha\right),
\end{aligned}$$

siden $\mathrm{Var}[X] = np(1-p) = n(17\alpha+1)(19-17\alpha)/20^2$.

d) Sentralgrenseteorem sier at hvis $\bar{X}$ er gjennomsnittet til et tilfeldig utvalg av størrelse n fra en populasjon med forventningsverdi μ og standardavvik σ, så vil fordelingen til

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{1}$$

gå mot en standard normalfordeling når n går mot uendelig.

Dette betyr at når n er stor vil Z gitt over være tilnærmet standard normalfordelt. Siden $\bar{X} = \mu + Z\sigma/\sqrt{n}$ er en lineær funksjon av Z vil dermed også $\bar{X}$ være tilnærmet normalfordelt.

I vårt tilfelle er X/n gjennomsnittet av en binomisk forsøksrekke der hvert forsøk resulterer i verdien 0 eller 1. X/n er derfor ifølge sentralgrenseteoremet tilnærmet normalfordelt når n er stor. Dessuten har vi at

$$\tilde{\alpha} = \frac{1}{17}(20n(X/n) - 1),$$

dvs at $\tilde{\alpha}$ er en lineær funksjon av X/n. En linær funksjon av en normalfordelt variabel er også normalfordelt, og dermed er $\tilde{\alpha}$ tilnærmet normalfordelt med den forventning og varians som ble regnet ut i oppgave c).

e)

Undersøkelsene som er gjort i Trøndelag gir en observasjon X som er binomisk fordelt

$$X \quad \sim \quad \text{Bin}(n, (17\alpha + 1)/20).$$

α er imidlertid ukjent, og det er den vi vil forsøke å si noe om.

Når vi formulerer en hypotesetest, lar vi nullhypotesen H_0 være en konservativ oppfatning av verden, mens den alternative hypotesen H_1 er formulert slik at en bekreftelse av den, fører til at vi kan påvise forskjeller av et eller annet slag. I dette tilfellet har vi før forsøket er utført ingen grunn til å anta at det er mer hyppighet av sykdommen i Trøndelag enn andre steder. Derfor må vi anta at som H_0 at $\alpha = 0.053$, i Trøndelag som ellers i landet.

$$H_0: \ \alpha = 0.053$$

$$H_1: \ \alpha > 0.053$$

Vi bruker $\tilde{\alpha}$ som observator og velger å forkaste H_0 hvis $\tilde{\alpha}$ blir mye større enn 0.0053, siden det tyder på at α er større enn 0.0053. Vi vet fra 2c) at $\tilde{\alpha}$ er tilnærmet normalfordelt, med forventning α og varians $\frac{1}{n}\left(\alpha + \frac{1}{17}\right)\left(\frac{19}{17} - \alpha\right)$, slik at en tilnærmet standard normalfordelt størrelse er

$$\begin{aligned} Z &= \frac{\tilde{\alpha} - \alpha}{\sqrt{\frac{1}{n}\left(\alpha + \frac{1}{17}\right)\left(\frac{19}{17} - \alpha\right)}} \\ &= \frac{\tilde{\alpha} - 0.053}{\frac{0.345}{\sqrt{n}}}, \qquad \text{når } \alpha = 0.053 \text{ under } H_0. \end{aligned}$$

Under H_0 er dermed

$$Z \quad = \quad \frac{\tilde{\alpha} - 0.053}{0.345/\sqrt{n}} \quad \sim \quad N(0,\ 1)$$

og vi forkaster H_0 hvis Z får verdi større enn $z_{0.05} = 1.645$ siden det tilsvarer at $\tilde{\alpha}$ er uforholdsmessig stor.

Hvis vi setter inn de oppgitte verdiene, får vi at

$$\begin{aligned} \tilde{\alpha} &= \frac{1}{17}\left(\frac{20 \cdot 101}{1000} - 1\right) \;=\; 0.060 \\ &\Downarrow \\ Z &= \frac{0.060 - 0.053}{0.345/\sqrt{1000}} \;=\; 0.64, \end{aligned}$$

og siden $Z < z_{0.05} = 1.645$, forkaster vi *ikke* H_0.

Oppgave 3

a)

Fra formelsamlingen ser vi at både X og Y er eksponensialfordelte variable:

$$\begin{aligned} X &\sim \text{Exp}(\beta_1) \quad \Rightarrow \quad \text{E}[X] = \beta_1 \\ Y &\sim \text{Exp}(\beta_2) \quad \Rightarrow \quad \text{E}[Y] = \beta_2, \end{aligned}$$

og vi får at

$$\frac{\text{E}[X]}{\text{E}[Y]} = \frac{\beta_1}{\beta_2}.$$

b)

Først gjør vi et variabelskifte der $X = UV$ og $Y = V$ slik at $U = X/Y$ er den variabelen vi er ute etter. Deretter finner vi simultantettheten til U og V, før vi til slutt bruker denne til å finne marginaltettheten til U. Vi vet at simultantettheten til U og V er gitt ved

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v))|J|,$$

der Jacobideterminanten $|J|$ er gitt ved

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = |v|.$$

Dermed blir

$$\begin{aligned} f_{U,V}(u,v) &= \begin{cases} \frac{1}{\beta_1} e^{-uv/\beta_1} \cdot \frac{1}{\beta_2} e^{-v/\beta_2} \cdot |v| & u, v > 0 \\ 0 & \text{ellers} \end{cases} \\ &= \begin{cases} \frac{v}{\beta_1 \beta_2} e^{-\left(\frac{u}{\beta_1} + \frac{1}{\beta_2}\right)v} & u, v > 0 \\ 0 & \text{ellers.} \end{cases} \end{aligned}$$

Marginaltettheten til U finner vi ved

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} f_{U,V}(u,v) dv \\ &= \begin{cases} \int_0^{\infty} \frac{v}{\beta_1 \beta_2} e^{-\left(\frac{u}{\beta_1} + \frac{1}{\beta_2}\right)v} dv = \frac{\beta_1 \beta_2}{(\beta_1 + \beta_2 u)^2} & u > 0 \\ 0 & \text{ellers.} \end{cases} \end{aligned}$$

c)

$$\begin{aligned}
\mathrm{E}[U] &= \int_{-\infty}^{\infty} u \; f_U(u) du \\
&= \int_0^{\infty} \frac{\beta_1\beta_2 \; u}{(\beta_1 \; + \; \beta_2 u)^2} du \\
&= \frac{\beta_1}{\beta_2}\int_1^{\infty} \left(\frac{1}{s} - \frac{1}{s^2}\right) ds \quad = \quad \infty,
\end{aligned}$$

hvor vi bruker substitusjonen $\beta_1 s = \beta_1 + \beta_2 u$ i siste linje.