ST1201 Statistiske metoder

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen desember 2004

Oppgave 1

a)

Testobservator er

$$T = \frac{\bar{X} - 100}{\sqrt{\frac{S^2}{n}}} \quad \text{der} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{og} \quad n = 10.$$

Når H_0 er riktig er T t-fordelt med $n-1$ frihetsgrader. Beslutningsregelen blir dermed: Forkast H_0 dersom $T < -t_{\alpha,n-1}$.

Innsatt tallverdier får man

$$\bar{x} = \frac{954.2}{10} = 95.42 \quad , \quad s^2 = \frac{1}{n}\left[\sum_{i=1} x_i^2 - n\bar{x}^2\right] = \frac{1}{9}\left[91511.58 - 10 \cdot 95.42^2\right] = 51.31,$$

$$t = \frac{95.42 - 100}{\sqrt{\frac{51.31}{10}}} = -2.02 \quad \text{og} \quad -t_{\alpha,n-1} = -t_{0.05,9} = -1.833.$$

Siden $t = -2.02 < -1.833$ blir konklusjonen: Forkast H_0.

b)

Testobservator for tegntesten blir $Y_+ = \#\{X_i < 100\}$. Generelt vil man ha at

$$Y_+ \sim \text{binomisk}(n, p) \quad \text{der} \quad n = 10 \quad \text{og} \quad p = \mathrm{P}(X_i < 100).$$

Når H_0 er riktig er $p = 0.5$.
Forkaster H_0 dersom $Y_+ \geq k$ der k må velges størst mulig slik at $\mathrm{P}(Y_+ \geq k | H_0) \leq \alpha = 0.05$.
Finner at

$$\mathrm{P}(Y_+ \geq 9 | H_0) = 0.011 \quad \text{og} \quad \mathrm{P}(Y_+ \geq 8 | H_0) = 0.055.$$

Dermed blir $k = 9$ og beslutningsregelen blir: Forkast H_0 dersom $Y_+ \geq 9$.
Lises observasjon gir $y_+ = 7$ slik at man skal ikke forkaste H_0.

c)

Når $X_i \sim \mathrm{N}(95, 7^2)$ får man at $p = \mathrm{P}(X_i < 100) = \Phi\left(\frac{100-95}{7}\right) = \Phi(0.71) = 0.7611$. Teststyrken blir dermed

$$\mathrm{P}(\text{Forkast} H_0 | p = 0.7611) = \mathrm{P}(Y_+ \geq 9 | p = 0.7611)$$

$$= \binom{10}{9} 0.7611^9 \cdot (1 - 0.7611) + 0.7611^{10} = \underline{\underline{0.27}}.$$

Ser at teststyrken for tegntesten er mye lavere enn teststyrken for t-testen. Dette er som man skulle forvente siden t-testen i tilfellet betraktet her benytter seg av mer informasjon enn hva tegntesten gjør. Tegntesten benytter ikke informasjonen om at X_i-ene er normalfordelt, noe t-testen gjør.

Anbefaler at Lise benytter t-testen. Denne har større teststyrke og siden t-testen er robust vil T være tilnærmet t-fordelt selv om vektene ikke er normalfordelt. Dersom man lager et histogram over de observerte vektene vil man dessuten se at fordelingen ikke later til å være spesielt skjev.

Oppgave 2

a)

Finner først forventning og varians for $\widehat{\beta}$,

$$\mathrm{E}(\widehat{\beta}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n Z_i\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^n Z_i\right) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(Z_i) = \frac{1}{n}\sum_{i=1}^n \beta = \frac{1}{n} n\beta = \beta,$$

$$\mathrm{Var}(\widehat{\beta}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Z_i\right) = \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^n Z_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(Z_i)$$
$$= \frac{1}{n^2}\sum_{i=1}^n \beta^2 = \frac{1}{n^2} n\beta^2 = \frac{\beta^2}{n}.$$

Må så bestemme grensen i Cramer-Raos ulikhet,

$$\ln f_Z(x; \beta) = -\ln\beta - \frac{z}{\beta},$$

$$\frac{\partial \ln f_Z(x;\beta)}{\partial \beta} = -\frac{1}{\beta} + \frac{z}{\beta^2},$$
$$\frac{\partial^2 \ln f_Z(z;\beta)}{\partial \beta^2} = \frac{1}{\beta^2} - \frac{2z}{\beta^3},$$
$$\mathrm{E}\left[\frac{\partial^2 f_Z(Z;\beta)}{\partial \beta^2}\right] = \mathrm{E}\left[\frac{1}{\beta^2} - \frac{2Z}{\beta^3}\right] = \frac{1}{\beta^2} - \frac{2}{\beta^3}\mathrm{E}(Z). = \frac{1}{\beta^2} - \frac{2}{\beta^3}\beta = -\frac{1}{\beta^2},$$
$$\left\{-n\mathrm{E}\left[\frac{\partial^2 f_Z(Z;\beta)}{\partial \beta^2}\right]\right\}^{-1} = \left\{-n\cdot\left(-\frac{1}{\beta^2}\right)\right\}^{-1} = \frac{\beta^2}{n}.$$

Ser at $\widehat{\beta}$ er forventningsrett og at $\mathrm{Var}(\widehat{\beta})$ er lik Cramer-Raos nedre grense. Har dermed vist at $\widehat{\beta}$ er en *beste* estimator.

b)

Har $v = 2z/\beta \Leftrightarrow z = \beta v/2$ og dermed

$$\frac{dz}{dv} = \frac{\beta}{2}.$$

Transformasjonsformelen gir da

$$f_V(v) = f_Z\left(\frac{\beta v}{2}\right)\cdot\left|\frac{\beta}{2}\right| = \frac{1}{\beta}\exp\left\{-\frac{\frac{\beta v}{2}}{\beta}\right\}\frac{\beta}{2} = \frac{1}{2}e^{-\frac{v}{2}}.$$

Setter man $n = 2$ inn i formelen for sannsynlighetstettheten for en χ^2_n-fordeling får man

$$\frac{1}{2^{\frac{2}{2}}}\Gamma\left(\frac{2}{2}\right)z^{\frac{2}{2}-1}e^{\frac{z}{2}} = \frac{1}{2}e^{\frac{z}{2}}.$$

Ser dermed at V er χ^2_2-fordelt. [Alternativt kan man vise dette ved bruk av momentgenererende funksjoner.]
Siden Z_i-ene er uavhengige blir også V_i-ene uavhengige. Dessuten er

$$\frac{2n\widehat{\beta}}{\beta} = \frac{2n}{\beta}\frac{1}{n}\sum_{i=1}^{n} Z_i = \sum_{i=1}^{n}\frac{2Z_i}{\beta} = \sum_{i=1}^{n} V_i.$$

Siden en sum av uavhengige χ^2-fordelte variable er χ^2-fordelt der frihetsgradene summerer seg har vi dermed vist at $2n\widehat{\beta}/\beta$ er χ^2-fordelt med $\sum_{i=1}^{n} 2 = 2n$ frihetsgrader.

c)

Siden $2n\widehat{\beta}/\beta \sim \chi^2_{2n}$ har vi at

$$\mathrm{P}\left(\chi^2_{1-\frac{\alpha}{2},2n} \leq \frac{2n\widehat{\beta}}{\beta} \leq \chi^2_{\frac{\alpha}{2},2n}\right) = 1-\alpha.$$

Dette gir

$$\mathrm{P}\left(\frac{2n\widehat{\beta}}{\chi^2_{\frac{\alpha}{2},2n}} \leq \beta \leq \frac{2n\widehat{\beta}}{\chi^2_{1-\frac{\alpha}{2},2n}}\right) = 1-\alpha.$$

$(1-\alpha)\cdot 100\%$-konfidensintervall for β er dermed

$$\left[\frac{2n\widehat{\beta}}{\chi^2_{\frac{\alpha}{2},2n}}, \frac{2n\widehat{\beta}}{\chi^2_{1-\frac{\alpha}{2},2n}}\right].$$

Innsatt tall får vi $\widehat{\beta} = 89.878$, $\chi^2_{\frac{\alpha}{2},2n} = \chi^2_{0.05,20} = 31.410$ og $\chi^2_{1-\frac{\alpha}{2},2n} = \chi^2_{0.95,20} = 10.851$. Konfidensintervaller blir dermed

$$\left[\frac{2\cdot 10\cdot 89.878}{31.410}, \frac{2\cdot 10\cdot 89.878}{10.851}\right] = [57.23, 165.66].$$

Oppgave 3

a)

Rimelighetsfunksjonen blir

$$\begin{aligned} L(a,b) &= f_{X_A,X_B,X_{AB},X_0}(x_A, x_B, x_{AB}, x_0; \theta_A, \theta_B, \theta_{AB}, \theta_0) \\ &= \binom{n}{x_A x_B x_{AB} x_0} \theta_A^{x_A}\cdot\theta_B^{x_B}\cdot\theta_{AB}^{\theta_{AB}}\cdot\theta_0^{x_0} \\ &= \binom{n}{x_A x_B x_{AB} x_0}(a(1-b))^{x_A}\cdot((1-a)b)^{x_B}\cdot(ab)^{\theta_{AB}}\cdot((1-a)(1-b))^{x_0} \end{aligned}$$

slik at log-rimelighetsfunskjonen blir

$$\begin{aligned} l(a,b) = \ln L(a,b) &= \ln\binom{n}{x_A x_B x_{AB} x_0} + x_A \ln a + x_A\ln(1-b) + x_B\ln(1-a) + x_B \ln b \\ &+ x_{AB}\ln a + x_{AB}\ln b + x_0\ln(1-a) + x_0\ln(1-b). \end{aligned}$$

$$= \ln\binom{n}{x_A x_B x_{AB} x_0} + (x_A+x_{AB})\ln a + (x_B+x_{AB})\ln b + (x_B+x_0)\ln(1-a) + (x_A+x_0)\ln(1-b).$$

Deriverer for å finne maksimimuspunkt,

$$\frac{\partial l}{\partial a} = \frac{x_A + x_{AB}}{a} + \frac{x_B + x_0}{1-a} \cdot (-1) = \frac{x_A + x_{AB}}{a} - \frac{x_B + x_{AB}}{1-a},$$

$$\frac{\partial l}{\partial b} = \frac{x_B + x_{AB}}{b} + \frac{x_A + x_0}{1-b} \cdot (-1) = \frac{x_B + x_{AB}}{b} - \frac{x_A + x_0}{1-b}.$$

Finner SME ved å sette de partiellderiverte lik null,

$$\frac{\partial l}{\partial a} = 0 \Rightarrow \frac{x_A + x_{AB}}{a} = \frac{x_B + x_0}{1-a} \Rightarrow a = \frac{x_A + x_{AB}}{x_A + x_{AB} + x_B + x_0} = \frac{x_A + x_{AB}}{n},$$

$$\frac{\partial l}{\partial b} = 0 \Rightarrow \frac{x_B + x_{AB}}{b} = \frac{x_A + x_0}{1-b} \Rightarrow b = \frac{x_B + x_{AB}}{x_B + x_{AB} + x_A + x_0} = \frac{x_B + x_{AB}}{n}.$$

Dermed

$$\widehat{a} = \frac{X_A + X_{AB}}{n} \quad \text{og} \quad \widehat{b} = \frac{X_B + X_{AB}}{n}.$$

Innsatt tallverdier får vi

$$\widehat{a} = \frac{212 + 39}{502} = 0.500 \quad \text{og} \quad \widehat{b} = \frac{103 + 39}{502} = 0.283.$$

b)

Konservativ og alternativ hypoteser blir

$$H_0 : \text{to-loci-teorien er korrekt} \quad \text{og} \quad H_1 : \text{to-loci-teorien er ikke korrekt.}$$

Goodness-of-fit-testobservatoren blir her

$$C = \frac{(X_A - n\widehat{a}(1-\widehat{b}))^2}{n\widehat{a}(1-\widehat{b})} + \frac{(X_B - n(1-\widehat{a})\widehat{b})^2}{n(1-\widehat{a})\widehat{b}} + \frac{(X_{AB} - n\widehat{a}\widehat{b})^2}{n\widehat{a}\widehat{b}} + \frac{(X_0 - n(1-\widehat{a})(1-\widehat{b}))^2}{n(1-\widehat{a})(1-\widehat{b})}$$

og under H_0 er denne χ^2-fordelt med $4-1-2=1$ frihetsgrad. Forkaster dermed H_0 dersom $C \geq \chi^2_{\alpha,1}$.

Innsatt tallverdier får man

$$c = 5.7017 + 14.3861 + 14.4456 + 5.6782 = 40.21 \quad \text{og} \quad \chi^2_{0.01,1} = 2.706.$$

Dvs, forkaster H_0 og konkluderer med at to-loci-teorien er feil.