ST1201 Statistiske metoder

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen desember 2006

Oppgave 1

Regner ut forventningsverdien til hver av de to foreslåtte estimatorene

$$\mathrm{E}[\widehat{\mu}] = \mathrm{E}\left[\frac{X+\frac{1}{2}Y}{1+\frac{1}{2}}\right] = \frac{1}{1+\frac{1}{2}}\mathrm{E}\left[X+\frac{1}{2}Y\right] = \frac{1}{1+\frac{1}{2}}\left[\mathrm{E}[X]+\frac{1}{2}\mathrm{E}[Y]\right]$$

$$= \frac{1}{1+\frac{1}{2}}\left[\mu+\frac{1}{2}\mu\right] = \mu.$$

$$\mathrm{E}[\widehat{\mu}] = \mathrm{E}\left[\frac{X+\frac{1}{\sqrt{2}}Y}{1+\frac{1}{\sqrt{2}}}\right] = \frac{1}{1+\frac{1}{\sqrt{2}}}\mathrm{E}\left[X+\frac{1}{\sqrt{2}}Y\right] = \frac{1}{1+\frac{1}{\sqrt{2}}}\left[\mathrm{E}[X]+\frac{1}{\sqrt{2}}\mathrm{E}[Y]\right]$$

$$= \frac{1}{1+\frac{1}{\sqrt{2}}}\left[\mu+\frac{1}{\sqrt{2}}\mu\right] = \mu.$$

Begge estimatorene er forventningsrette. Må dermed sjekke variansene for å avhjøre hvilken som er å foretrekke (benytter da at X og Y er uavhengige)

$$\mathrm{Var}[\widehat{\mu}] = \mathrm{Var}\left[\frac{X+\frac{1}{2}Y}{1+\frac{1}{2}}\right] = \frac{1}{\left(1+\frac{1}{2}\right)^2}\mathrm{Var}\left[X+\frac{1}{2}Y\right] = \frac{4}{9}\left(\mathrm{Var}[X]+\left(\frac{1}{2}\right)^2\mathrm{Var}[Y]\right)$$

$$= \frac{4}{9}\left(\sigma^2+\frac{1}{4}\cdot 2\sigma^2\right) = \frac{2}{3}\sigma^2.$$

$$\mathrm{Var}[\tilde{\mu}] = \mathrm{Var}\left[\frac{X+\frac{1}{\sqrt{2}}Y}{1+\frac{1}{\sqrt{2}}}\right] = \frac{1}{\left(1+\frac{1}{\sqrt{2}}\right)^2}\mathrm{Var}\left[X+\frac{1}{\sqrt{2}}Y\right]$$

$$= \frac{1}{\left(1+\frac{1}{\sqrt{2}}\right)^2}\left(\mathrm{Var}[X]+\left(\frac{1}{\sqrt{2}}\right)^2\mathrm{Var}[Y]\right) = \frac{1}{\left(1+\frac{1}{\sqrt{2}}\right)^2}\left(\sigma^2+\frac{1}{2}\cdot 2\sigma^2\right)$$

$$= \frac{2}{\left(1+\frac{1}{\sqrt{2}}\right)^2}\sigma^2 = 0.686\sigma^2.$$

Siden begge estimatorene er forventningsrette foretrekker man den med minst varians, dvs. foretrekker $\widehat{\mu}$.

Oppgave 2

a)
Hvis man har r ukjente parametre $\theta_1, \theta_2, \ldots, \theta_r$, må man finne de r første momentene uttrykt ved $\theta_1, \theta_2, \ldots, \theta_r$ og beregne tilsvarende empiriske momenter. Dvs

$$\begin{aligned}
E[X] &= \mu_1(\theta_1, \theta_2, \ldots, \theta_r), \quad \text{som kan estimeres med } \; m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \\
E[X^2] &= \mu_2(\theta_1, \theta_2, \ldots, \theta_r), \quad \text{som kan estimeres med } \; m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
&\ldots \\
E[X^r] &= \mu_r(\theta_1, \theta_2, \ldots, \theta_r), \quad \text{som kan estimeres med } \; m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r
\end{aligned}$$

For å finne momentestimatorene tilnærmer man hver av de teoretiske momentene μ_i med tilhørende empiriske moment m_i, dvs

$$\begin{aligned}
\mu_1(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) &= \frac{1}{n}\sum_{i=1}^{n} X_i; \\
\mu_2(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2; \\
&\ldots \\
\mu_r(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) &= \frac{1}{n}\sum_{i=1}^{n} X_i^r
\end{aligned}$$

og finner så $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r$ ved å løse dette ligningssystemet.
I denne oppgaven har vi at $\mu_1(\beta) = 2\beta$, så

$$2\beta = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Rightarrow \quad \widehat{\beta} = \frac{1}{2n}\sum_{i=1}^{n} X_i.$$

b)
Her kan man benytte transformasjonsformelen, eller man kan observere at fordelingen til X_i er en gammafordeling med parametre $\alpha = 2$ og β. Derfor kjenner man momentgenererende funksjonen til X_i:

$$M_{X_i}(t) = \left(\frac{1}{1-\beta t}\right)^2, \quad \text{for} \quad t < \frac{1}{\beta}.$$

Regneregler for momentgenererende funksjoner gir at

$$M_{Z_i}(t) = M_{X_i}\left(\frac{2}{\beta}\,t\right) = \left(\frac{1}{1-2t}\right)^2$$

som er momentgenererendefunksjon for en χ^2-fordelt stokastisk variabel med fire frihetsgrader. Videre har man at

$$\frac{4n\widehat{\beta}}{\beta} = \frac{4n\sum_{i=1}^{n} X_i}{2n\beta} = \frac{2}{\beta}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n}\frac{2X_i}{\beta}$$

som er en sum av uavhengige χ^2-fordelte stokastiske variabler med fire frihetsgrader hver. Derfor, har vi fra teorem 7.3.1 at $4n\widehat{\beta}/\beta \sim \chi^2_{4n}$.

c)
Vi skal teste

$$H_0 : \beta = \beta_0 \;\; \text{mot} \;\; H_1 : \beta > \beta_0$$

Siden $4n\widehat{\beta}/\beta_0 \sim \chi^2_{4n}$ når H_0 er sann, benytter vi $4n\widehat{\beta}/\beta_0$ som testobservator og beslutningsregelen blir

$$\text{“Forkast } \; H_0 \; \text{ hvis } \; 4n\widehat{\beta}/\beta_0 > \chi^2_{\alpha,4n}\text{”}$$

Med de gitte tallene blir 4n$\widehat{\beta}/\beta_0 = 25.87$ og $\chi^2_{\alpha,4n} = 26.509$ og $\underline{\underline{H_0 \text{ kan ikke forkastes}}}$.

d)

$$\begin{aligned}
1-\boldsymbol{\beta}(\beta) &= P\left\{\frac{4n\widehat{\beta}}{\beta_0} > \chi^2_{\alpha,4n}|\beta\right\} = P\left\{\frac{4n\widehat{\beta}}{\beta} > \frac{\chi^2_{\alpha,4n}\,\beta_0}{\beta}\right|\beta\right\} \\
&= 1-P\left\{\frac{4n\widehat{\beta}}{\beta} \leq \frac{\chi^2_{\alpha,4n}\,\beta_0}{\beta}\right|\beta\right\} = \underline{\underline{1-F_{4n}\left(\frac{\chi^2_{\alpha,4n}\,\beta_0}{\beta}\right)}},
\end{aligned}$$

der $F_{4n}(x)$ er kumulativ fordelingsfunksjon for en χ^2-fordeling med $4n$ frihetsgrader.
Med de gitte tallene blir styrkefunksjonen lik

$$1-\boldsymbol{\beta}(\beta) = \underline{\underline{1-F_{40}\left(\frac{53.018}{\beta}\right)}},$$

og fra tabellene finner vi at teststyrken blir 0.99 hvis $\frac{53.018}{\beta} = \chi^2_{0.01,40} = 22.164$, og $\underline{\underline{\beta \text{ må bli } 5.031}}$.

Hvis $\beta = 5.031$ er sannsynligheten for å konkludere med at H_0 skal forkastes lik 0.99.

Oppgave 3

a)

Rimelighetsfunksjonen blir

$$L(\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2 x_i}} e^{-\frac{(Y_i-\alpha-\beta x_i)^2}{2\sigma_0^2 x_i}} = \frac{1}{(2\pi)^{n/2}\sigma_0^n \prod_{i=1}^n \sqrt{x_i}} e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n \frac{(Y_i-\alpha-\beta x_i)^2}{x_i}}$$

$$l(\alpha,\beta) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma_0 - \frac{1}{2}\sum_{i=1}^n \ln x_i - \frac{1}{2\sigma_0^2}\sum_{i=1}^n \frac{(Y_i-\alpha-\beta x_i)^2}{x_i}.$$

Partiellderiverer med hensyn på hver av α og β og får

$$\begin{aligned}
\frac{\partial l(\alpha,\beta)}{\partial\alpha} &= -\frac{1}{2\sigma_0^2}\sum_{i=1}^n \frac{1}{x_i}\cdot 2(Y_i-\alpha-\beta x_i)(-1) = \frac{1}{\sigma_0^2}\sum_{i=1}^n \frac{Y_i-\alpha-\beta x_i}{x_i} \\
\frac{\partial l(\alpha,\beta)}{\partial\beta} &= -\frac{1}{2\sigma_0^2}\sum_{i=1}^n \frac{1}{x_i}\cdot 2(Y_i-\alpha-\beta x_i)(-x_i) = \frac{1}{\sigma_0^2}\sum_{i=1}^n (Y_i-\alpha-\beta x_i)
\end{aligned}$$

Vi setter de deriverte lik null

$$\sum_{i=1}^n \frac{Y_i-\alpha-\beta x_i}{x_i} = 0 \tag{1}$$

$$\sum_{i=1}^n (Y_i-\alpha-\beta x_i) = 0 \tag{2}$$

Fra (2) følger enkelt at

$$\alpha = \bar{y} - \beta\bar{x}.$$

Setter man dette inn i (1) og løser med hensyn på β får man (etter litt opprydding, som selvfølgelig må være med i en besvarelse) at

$$\beta = \frac{\bar{y}\sum_{i=1}^n \frac{1}{x_i} - \sum_{i=1}^n \frac{y_i}{x_i}}{\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n}.$$

Følgelig er SME som oppgitt i oppgaveteksten.

b)

$$
\begin{aligned}
E[\widehat{\beta}] &= E\left[\frac{\bar{Y}\sum_{i=1}^{n}\frac{1}{x_i}-\sum_{i=1}^{n}\frac{Y_i}{x_i}}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i}-n}\right] = \frac{1}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i}-n}\left(\sum_{i=1}^{n}\frac{1}{x_i}E[\bar{Y}]-\sum_{i=1}^{n}E\left[\frac{Y_i}{x_i}\right]\right) = \\
&= \frac{1}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i}-n}\left(\left(\sum_{i=1}^{n}\frac{1}{x_i}\right)\frac{1}{n}\sum_{j=1}^{n}E[Y_j]-\sum_{i=1}^{n}\frac{E[Y_i]}{x_i}\right) = \\
&= \frac{1}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i}-n}\left(\left(\sum_{i=1}^{n}\frac{1}{x_i}\right)\frac{1}{n}\sum_{j=1}^{n}(\alpha+\beta x_j)-\sum_{i=1}^{n}\frac{\alpha+\beta x_i}{x_i}\right) = \\
&= \frac{1}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i}-n}\left(\left(\sum_{i=1}^{n}\frac{1}{x_i}\right)(\alpha+\beta\bar{x})-\alpha\sum_{i=1}^{n}\frac{1}{x_i}-\beta n\right) = \\
&= \frac{1}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i}-n}\left(\beta\bar{x}\sum_{i=1}^{n}\frac{1}{x_i}-\beta n\right) = \\
&= \beta
\end{aligned}
$$

så $\widehat{\beta}$ er forventnigsrett.

$$
\begin{aligned}
Var[\widehat{\beta}] &= Var\left[\frac{\bar{Y}\sum_{i=1}^n \frac{1}{x_i} - \sum_{i=1}^n \frac{Y_i}{x_i}}{\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n}\right] = \\
&= \frac{1}{(\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n)^2}\left(\left(\sum_{i=1}^n \frac{1}{x_i}\right)^2 Var[\bar{Y}] - \sum_{i=1}^n Var\left[\frac{Y_i}{x_i}\right]\right) = \\
&= \frac{1}{(\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n)^2}\left(\left(\sum_{i=1}^n \frac{1}{x_i}\right)^2 \frac{1}{n^2}\sum_{j=1}^n Var[Y_j] - \sum_{i=1}^n \frac{Var[Y_i]}{x_i^2}\right) = \\
&= \frac{1}{(\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n)^2}\left(\left(\sum_{i=1}^n \frac{1}{x_i}\right)^2 \frac{\sigma_0^2}{n^2}\sum_{j=1}^n x_j - \sigma_0^2\sum_{i=1}^n \frac{x_i}{x_i^2}\right) = \\
&= \frac{\sigma_0^2}{(\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n)^2}\left(\left(\sum_{i=1}^n \frac{1}{x_i}\right)^2 \frac{1}{n^2}\sum_{j=1}^n x_j - \sum_{i=1}^n \frac{1}{x_i}\right) = \\
&= \frac{\sigma_0^2}{(\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n)^2}\left(\sum_{i=1}^n \frac{1}{x_i}\right)\frac{1}{n}\left(\left(\sum_{i=1}^n \frac{1}{x_i}\right)\frac{1}{n}\sum_{j=1}^n x_j - n\right) = \\
&= \frac{\sigma_0^2}{(\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n)^2}\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}\right)\left(\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n\right) = \\
&= \sigma_0^2 \cdot \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}}{\bar{x}\sum_{i=1}^n \frac{1}{x_i} - n} = \\
&= \frac{\sigma_0^2}{n} \cdot \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}}{\bar{x}\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}\right) - 1}
\end{aligned}
$$

c)

$\widehat{\beta}$ er en linærkombinasjon av uavhengige normalfordelte stokastiske variabler, dvs. Y_i-ene. Man kan enkelt vise at $\widehat{\beta}$ kan skrives som $\widehat{\beta} = c_0 + \sum_{i=1}^n c_i Y_i$, der alle c_i-ene er numeriske konstanter (funksjoner av x_i-ene). Fra pensum vet man at en linærkombinasjon av uavhengige normalfordelte stokastiske variabler er normalfordelt, følgelig er $\widehat{\beta}$ normalfordelt med forventning og varians som regnet ut over.

Siden

$$\widehat{\beta} \sim N(\beta, Var[\widehat{\beta}])$$

har man at

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{Var[\widehat{\beta}]}} \sim N(0,1)$$

og

$$P\{-z_{a/2} \leq Z \leq z_{a/2}\} = 1 - a$$

som er det samme som

$$P\left\{-z_{a/2} \leq \frac{\widehat{\beta} - \beta}{\sqrt{Var[\widehat{\beta}]}} \leq z_{a/2}\right\} = 1 - a$$

eller

$$P\left\{\beta - z_{a/2}\sqrt{Var[\widehat{\beta}]} \leq \widehat{\beta} \leq \beta - z_{a/2}\sqrt{Var[\widehat{\beta}]}\right\} = 1 - a$$

Ved å sette inn uttrykket for $Var[\widehat{\beta}]$ som vi regnet ut i punkt **b**) får vi at konfidensintervallet blir

$$\left[\beta - z_{a/2} \cdot \frac{\sigma_0}{\sqrt{n}} \cdot \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}{\bar{x}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right) - 1}}, \quad \beta + z_{a/2} \cdot \frac{\sigma_0}{\sqrt{n}} \cdot \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}{\bar{x}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right) - 1}}\right]$$