ST1201 Statistiske metoder

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen desember 2009

Oppgave 1

La p være sannsynligheten for å få en 6'er i ett kast med en terning, og $H_0 : p = 1/6$ mot $H_1 : p < 1/6$.

$$\text{P-verdi} = \text{P(ingen 6-ere på n kast}|H_0 \text{ sann)} = \left(\frac{5}{6}\right)^n$$

Løser

$$\left(\frac{5}{6}\right)^n < \alpha = 0.05$$
$$n > \frac{\ln(0.05)}{\ln(5/6)} = 16.43$$

Trenger 17 kast.

Oppgave 2

a) Vi har

$$\mathrm{E}X = \lambda.$$

Momentestimatoren er dermed

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

som har forventning

$$\mathrm{E}\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}x_i = \lambda,$$

og varians

$$\mathrm{Var}\hat{\lambda} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}x_i = \frac{1}{n}\lambda.$$

b)

$$L(\lambda) = \prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}\mathrm{e}^{\lambda} = \frac{\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}\mathrm{e}^{n\lambda}$$

$$\ln L(\lambda) = \left(\sum_{i=1}^{n}x_i\right)\ln\lambda - \ln\prod_{i=1}^{n}x_i! + n\lambda$$

$$\frac{\partial \ln L(\lambda)}{\partial\lambda} = \frac{\sum_{i=1}^{n}x_i}{\lambda} + n$$

Vi setter $\frac{\partial \ln L(\lambda)}{\partial\lambda}$ lik 0 og får

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}x_i,$$

som er lik momentestimatoren i oppgave a) og dermed har samme forventningsverdi og varians.

c) 1. Cramér-Rao's ulikhet gir en nedre grense på variansen til en forventningsrett estimator av θ, dermed kan man sjekke om man har funnet den forventningsrette estimatoren med minst varians.
2. $f_Y(y;\theta)$ er en kontinuerlig pdf med kontinuerlige første- og andrederiverte, estimatoren $\hat{\theta}$ er forventningsrett og $\{y : f_Y(y;\theta) \neq 0\}$ er uavhengig av θ.

d) Fra b) har vi

$$\frac{\partial \ln f(x;\lambda)}{\partial\lambda} = \frac{x}{\lambda} + 1,$$

dermed

$$\frac{\partial^2 \ln f(x;\lambda)}{\partial\lambda^2} = -\frac{x}{\lambda^2},$$

og

$$\mathrm{E}\left[\frac{\partial^2 f(x;\lambda)}{\partial\lambda^2}\right] = -\frac{\mathrm{E}x}{\lambda^2} = -\frac{\lambda}{\lambda^2} = -\frac{1}{\lambda}.$$

Dermed er Cramér-Rao's nedere grense

$$\left\{-n\mathrm{E}\left[\frac{\partial^2 \ln f(x;\lambda)}{\partial\lambda^2}\right]\right\}^{-1} = \frac{\lambda}{n}.$$

Dermed er $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$ optimal i betydningen at ingen forventingsrett estimator kan estimere λ med høyere presisjon.

Oppgave 3

H_0 : antall mål laget av hjemmelaget er uavhengig av antall mål aget av bortelaget

	0 mål bortelag	1 mål bortelag	sum
0 mål hjemmelag	41	32	73
1 mål hjemmelag	51	65	116
sum	92	97	189

Har

$$\frac{(41-35.5)^2}{35.5} + \frac{(32-37.5)^2}{37.5} + \frac{(51-56.5)^2}{56.5} + \frac{(65-59.5)^2}{59.5} = 2.70 < \chi^2_{0.05,1} = 3.841 \quad (1)$$

Kan forkaste påstanden om at antall mål laget av hjemmelaget er uavhengig av mål laget av bortelaget.

Oppgave 4

La p være sannsynlighenten for at Kåres metode er best, og $H_0 : p = 0.5$ mot $H_1 : p > 0.5$.

$$\begin{aligned}\text{P-verdi} &= \mathrm{P}(\text{Kåres metode er bedre for 15 eller flere av de } 20 | H_0 \text{ sann}) \\ &= \sum_{i=15}^{20} \binom{20}{i} 0.5^{20} = 0.021 < \alpha = 0.05\end{aligned}$$

Forkaster H_0 på nivå $\alpha = 0.05$.

Oppgave 5

a) Har $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, hvor $\epsilon_i \sim N(0, \sigma^2)$, uavhengige ϵ_i og

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x}.$$

Da er

$$
\begin{aligned}
\mathrm{E}\hat{\beta}_1 &= \mathrm{E}\left(\frac{\sum_{i=1}^n (x_i-\bar{x})Y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)\\
&= \frac{\sum_{i=1}^n (x_i-\bar{x})\mathrm{E}Y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\\
&= \frac{\sum_{i=1}^n (x_i-\bar{x})(\beta_0+\beta_1 x_i)}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\\
&= \frac{\beta_0\sum_{i=1}^n (x_i-\bar{x}) + \beta_1\sum_{i=1}^n (x_i-\bar{x})x_i)}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\\
&= \frac{\beta_1\left(\sum_{i=1}^n x_i - n\bar{x}^2\right)}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\\
&= \beta_1
\end{aligned}
$$

og

$$
\begin{aligned}
\mathrm{Var}\hat{\beta}_1 &= \mathrm{Var}\left(\frac{\sum_{i=1}^n (x_i-\bar{x})Y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)\\
&= \frac{1}{\left(\sum_{i=1}^n (x_i-\bar{x})^2\right)^2}\mathrm{Var}\left(\sum_{i=1}^n (x_i-\bar{x})Y_i\right)\\
&= \frac{1}{\left(\sum_{i=1}^n (x_i-\bar{x})^2\right)^2}\sum_{i=1}^n (x_i-\bar{x})^2\mathrm{Var}Y_i\\
&= \frac{1}{\left(\sum_{i=1}^n (x_i-\bar{x})^2\right)^2}\sum_{i=1}^n (x_i-\bar{x})^2\sigma^2\\
&= \frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}.
\end{aligned}
$$

Siden $\hat{\beta}_1$ er en lineærkombinasjon av normalfordelte variable er også $\hat{\beta}_1$ normalfordelt:

$$
\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right).
$$

$$
\begin{aligned}
\mathrm{E}\hat{\beta}_0 &= \mathrm{E}\bar{Y} - E\hat{\beta}_1\bar{x} \\
&= \beta_0 + \beta_1\bar{x} - \beta_1\bar{x} \\
&= \beta_0 \\
\mathrm{Var}\hat{\beta}_0 &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - \bar{x}\frac{\sum_{i=1}^n (x_i-\bar{x})Y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}\right) \\
&= \mathrm{Var}\left(\frac{\sum_{i=1}^n Y_i \sum_{j=1}^n (x_j-\bar{x})^2 - n\bar{x}\sum_{i=1}^n (x_i-\bar{x})Y_i}{n\sum_{i=1}^n (x_i-\bar{x})^2}\right) \\
&= \mathrm{Var}\left(\frac{\sum_{i=1}^n \left[\left(\sum_{j=1}^n (x_j-\bar{x})^2 - n\bar{x}(x_i-\bar{x})\right) Y_i\right]}{n\sum_{i=1}^n (x_i-\bar{x})^2}\right) \\
&= \frac{\sum_{i=1}^n \left[\left(\sum_{j=1}^n x_j^2 - n\bar{x}x_i\right)^2 \mathrm{Var}Y_i\right]}{\left(n\sum_{i=1}^n (x_i-\bar{x})^2\right)^2} \\
&= \frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}
\end{aligned}
$$

$\hat{\beta}_0$ er også en lineærkombinasjon av normalfordelte variable dermed er $\hat{\beta}_1$ normalfordelt:

$$
\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i-\bar{x})^2}\right).
$$

b)

$$
\begin{aligned}
\mathrm{Prob}\left(-z_{\alpha/2} < \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\mathrm{Var}\beta_1}} < z_{\alpha/2}\right) &= 1-\alpha \\
\mathrm{Prob}\left(\hat{\beta}_1 - z_{\alpha/2}\sqrt{\mathrm{Var}\beta_1} < \beta_1 < \hat{\beta}_1 - z_{\alpha/2}\sqrt{\mathrm{Var}\beta_1}\right) &= 1-\alpha
\end{aligned}
$$

dvs konfidensintervallet er

$$
\left[\hat{\beta}_1 - z_{\alpha/2}\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}, \hat{\beta}_1 - z_{\alpha/2}\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}\right]
$$

c) Når $x_1 = \ldots = x_n = 0$ er $\hat{\beta}_0$ konsistent, men ikke $\hat{\beta}_1$. Vi har da bare observasjoner for $Y(x=0) = \beta_0 + \epsilon$ og kan derfor bare estimere β_0.