ST1201 Statistiske metoder

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen juni 2010

Oppgave 1

a)

$$X \sim \text{Geometrisk}(p),$$

vi må teste

$$\begin{aligned} H_0: &\quad p = \frac{1}{6} \\ H_1: &\quad p < \frac{1}{6} \end{aligned}$$

$$\begin{aligned} p &= P(X \geq 12 \mid H_0 \text{ er riktig}) \\ &= \sum_{x=12}^{\infty} P(X = x \mid H_0 \text{ er riktig}) \\ &= 0.1346 > 0.05 \end{aligned}$$

Dermed beholder vi H_0.

b)

$$Y \sim \text{Binomisk}(p),$$

vi må teste

$$\begin{aligned} H_0: &\quad p = \frac{1}{6} \\ H_1: &\quad p < \frac{1}{6} \end{aligned}$$

$$\begin{aligned} p &= P(Y \leq 1 \mid H_0 \text{ er riktig}) \\ &= \sum_{x=0}^{1} P(X = x \mid H_0 \text{ er riktig}) \\ &= 0.3813 > 0.05 \end{aligned}$$

Dermed beholder vi H_0.

Oppgave 2

a) $X \sim$ Binomisk(m, p)

$$\mathrm{E}X = mp$$

Setter X lik mp og får $\hat{p}_{MME} = \frac{1}{m}X$. Forventning og varians blir:

$$\mathrm{E}\hat{p}_{MME} = \mathrm{E}\frac{1}{m}X = \frac{1}{m}\mathrm{E}X = p$$
$$\mathrm{Var}\hat{p}_{MME} = \mathrm{Var}\frac{1}{m}X = \frac{1}{m^2}\mathrm{Var}X = \frac{1}{m^2}mp(1-p) = \frac{p(1-p)}{m}.$$

b) Har

$$L(p;x) = \binom{m}{x} p^x (1-p)^{m-x}$$
$$l(p;x) = \log L(p;x) = \log \binom{m}{x} + x \log p + (m-x)\log(1-p)$$
$$\frac{\partial l(p;x)}{\partial p} = \frac{x}{p} - \frac{m-x}{1-p}.$$

Løser

$$\frac{\partial l(p;x)}{\partial p} = 0$$
$$\frac{x}{p} = \frac{m-x}{1-p},$$

og får sannsynlighetsmaksimeringsestimatoren $\hat{p}_{MLE} = \frac{1}{m}X$, som er lik $\hat{p}_{MME}$, dvs. samme forventing 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)

$$\frac{\partial^2 l(p;x)}{\partial p^2} = -\frac{x}{p^2} - \frac{m-x}{(1-p)^2}.$$
$$\left\{-\mathrm{E}\left[\frac{\partial^2 l(p;x)}{\partial p^2}\right]\right\}^{-1} = \left\{\frac{mp}{p^2} + \frac{m-mp}{(1-p)^2}\right\}^{-1} = \left\{\frac{m}{p(1-p)}\right\}^{-1}$$
$$= \frac{p(1-p)}{m} = \mathrm{Var}\hat{p}_{MME} = \mathrm{Var}\hat{p}_{MLE},$$

dvs. momentestimatoren og sannsynlighetsmaksimeringsestimatoren er forventningsrette estimatorer med minimal varians.

e) Si

$$\tilde{p} = \frac{X+1}{m+2}.$$

Finner

$$\mathrm{E}(\tilde{p}-p)^2 = \mathrm{Var}\tilde{p} + (\mathrm{E}\tilde{p}-p)^2,$$

der

$$\begin{aligned}
\mathrm{E}\tilde{p} - p &= \frac{mp+1}{m+2} - p = \frac{1-2p}{m+2}, \\
\mathrm{Var}\tilde{p} &= \frac{mp(1-p)}{(m+2)^2},
\end{aligned}$$

dermed

$$\begin{aligned}
\mathrm{E}(\tilde{p}-p)^2 &= \frac{mp(1-p)}{(m+2)^2} + \frac{(1-2p)^2}{(m+2)^2} \\
&= \frac{mp(1-p)+(1-2p)^2}{(m+2)^2}
\end{aligned}$$

Ser på

$$\begin{aligned}
\mathrm{E}(\tilde{p}-p)^2 - \mathrm{E}(\hat{p}-p)^2 =& \frac{mp(1-p)+(1-2p)^2}{(m+2)^2} - \frac{p(1-p)}{m} \\
=& \frac{m-4p(1-p)(2m+1)}{m(m+2)^2} < 0
\end{aligned}$$

$m - 4p(1-p)(2m+1) < 0.$

Hvis vi løser

$$p^2(8m+4) - p(8m+4) + m = 0,$$

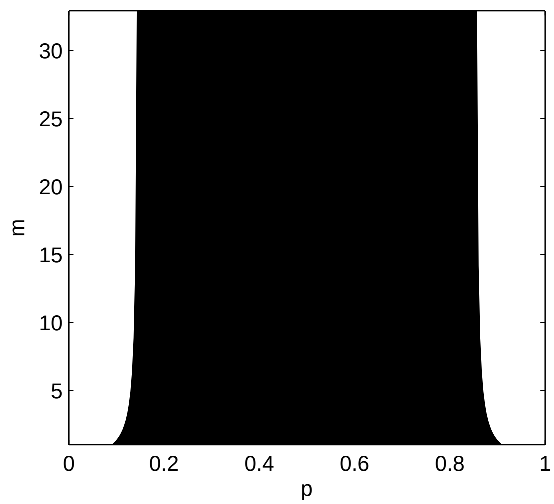
mhp p får vi

$$p = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{m+1}{2m+1}}.$$

Da vil

$$\begin{aligned}
\mathrm{E}(\tilde{p}-p)^2 \le \mathrm{E}(\hat{p}-p)^2, &\quad \text{for } \left|p-\frac{1}{2}\right| \le \frac{1}{2}\sqrt{\frac{m+1}{2m+1}} \\
\mathrm{E}(\tilde{p}-p)^2 > \mathrm{E}(\hat{p}-p)^2, &\quad \text{for } \left|p-\frac{1}{2}\right| > \frac{1}{2}\sqrt{\frac{m+1}{2m+1}}.
\end{aligned}$$

dvs. $\mathrm{E}(\tilde{p}-p)^2 \le \mathrm{E}(\hat{p}-p)^2$ i det svarte området:



f) $\tilde{p}$ er ikke forventingsrett så dette strider ikke mot Cramér-Rao's nedre grense.

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

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)

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

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

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) En estimator $\hat{\theta}_n$ er konsistent når $\mathrm{E}((\hat{\theta}_n-\theta)^2) \to 0$ når $n \to \infty$, der n er antall observasjoner.

$\hat{\beta}_1$ er ikke konsistente når $x_1 = \ldots = x_n$.

$\hat{\beta}_0$ er ikke konsistente når $x_1 = \ldots = x_n \neq 0$.