ST1201 Statistiske metoder

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen desember 2010

Oppgave 1

a) Har $X \sim$ Binomial(n=300,p), $\mathrm{E}X = np$, $\mathrm{Var}X = np(1-p)$. Approksimerer p med X/n, og bruker normalfordelingsapproksimasjon:

$$P\left[-z_{\alpha/2} \leq \frac{X/n - p}{\sqrt{\frac{X/n(1-X/n)}{n}}} \leq z_{\alpha/2}\right] = 1-\alpha$$

med $x = 75$ blir approksimativt 0.95-konfidensintervall

$$\left[\frac{x}{n} - z_{\alpha/2}\sqrt{\frac{x/n(1-x/n)}{n}}, \frac{x}{n} + z_{\alpha/2}\sqrt{\frac{x/n(1-x/n)}{n}}\right] = [0.20, 0.30]$$

b) Har

$$p = \frac{\exp(\theta)}{1+\exp(\theta)}$$
$$\theta = \ln p - \ln(1-p)$$

Lar

$$\hat{\theta} = \ln \hat{p} - \ln(1-\hat{p}), \quad \hat{p} = X/n$$

Ved å bruke første ordens Taylor-utvikling får vi

$$\mathrm{E}\hat{\theta} \approx \ln \mathrm{E}\hat{p} - \ln(1-\mathrm{E}\hat{p}) \approx -1.10$$
$$\mathrm{Var}\hat{\theta} \approx \left(\frac{1}{\mathrm{E}\hat{p}} + \frac{1}{1-\mathrm{E}\hat{p}}\right)^2 \mathrm{Var}\hat{p} \approx 0.0033$$

dvs. approksimativt 0.95-konfidensintervall blir

$$\left[E\hat{\theta} - z_{\alpha/2}\sqrt{\mathrm{Var}\hat{\theta}}, E\hat{\theta} - z_{\alpha/2}\sqrt{\mathrm{Var}\hat{\theta}}\right] = [-1.44, -0.73]$$

Oppgave 2

a) La $X_1, X_2, \dots, X_n, X_{n+1}$ være tilfeldig utvalg med fordeling $N(\mu, \sigma^2)$, og

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Da er

$$\frac{X_{n+1} - \bar{X}}{\sqrt{\sigma^2(1+1/n)}} \sim N(0,1),$$
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Dermed er

$$\frac{X_{n+1} - \bar{X}}{S\sqrt{1+1/n}} = \frac{\frac{X_{n+1}-\bar{X}}{\sqrt{\sigma^2(1+1/n)}}}{\sqrt{\frac{(n+1)S^2}{\sigma^2}\frac{1}{1+n}}} \sim t_{n-1}$$

b) Har

$$\text{Prob}\left(-t_{0.1,n-1} < \frac{X_{n+1} - \bar{X}}{S\sqrt{1+1/n}} < t_{0.1,n-1}\right) = 0.8$$
$$\text{Prob}\left(\bar{X} - t_{0.1,n-1} \cdot S\sqrt{1+1/n} < X_{n+1} < \bar{X} + t_{0.1,n-1} \cdot S\sqrt{1+1/n}\right) = 0.8$$

dvs $k = t_{0.1,n-1} \cdot \sqrt{1+1/n}$.

Oppgave 3

a) La $X = (X_1, X_2, X_3) \sim \text{Multinomial}(N, p = (p_1, p_2, p_3)$

$$p(x_1, x_2, x_3) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$x_1 + x_2 + x_3 = N$ og $p_1 + p_2 + p_3 = 1$ antar $p_1 = q^2,\ p_2 = 2q(1-q),\ p_3 = (1-q)^2$:

$$p(x_1, x_2, x_3) = \frac{n!}{x_1!x_2!x_3!} q^{2x_1}(2q(1-q))^{x_2}(1-q)^{2x_3}$$
$$= \frac{n!}{x_1!x_2!x_3!} 2^{x_2} q^{2x_1+x_2}(1-q)^{x_2+2x_3}$$

$$\ln p(x_1, x_2, x_3) = \ln \frac{n!}{x_1!x_2!x_3!2^{x_2}} + (2x_1 + x_2)\ln q + (x_2 + 2x_3)\ln(1-q)$$

$$\frac{\partial}{\partial q} \ln p(x_1, x_2, x_3) = \frac{(2x_1 + x_2)}{q} - \frac{(x_2 + 2x_3)}{1-q}$$

$$\begin{aligned} \frac{2x_1 + x_2}{q} &= \frac{x_2 + 2x_3}{1-q} \\ q &= \frac{2x_1 + x_2}{2x_1 + 2x_2 + 2x_3} \end{aligned}$$

Estimatoren blir $\hat{q} = \frac{2X_1 + X_2}{2N}$.

b)
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.

c)

$$\frac{\partial^2}{\partial q^2} \ln p(x_1, x_2, x_3) = -\frac{2x_1 + x_2}{q^2} - \frac{x_2 + 2x_3}{(1-q)^2}$$

$$\begin{aligned} \mathrm{E}\left[\frac{2x_1 + x_2}{q^2} + \frac{x_2 + 2x_3}{(1-q)^2}\right] &= N\frac{2q^2 + 2q(1-q)}{q^2} + N\frac{2q(1-q) + 2(1-q)^2}{(1-q)^2} \\ &= N\frac{2}{q} + N\frac{2}{1-q} \\ &= N\frac{2}{q(1-q)} \end{aligned}$$

dvs

$$\mathrm{Var}(\hat{q}) \geq \frac{q(1-q)}{2N}$$

d) Ser at $\mathrm{E}\hat{q} = q$ og lar $\mathrm{Var}\hat{q} = \frac{q(1-q)}{2N}$. Anta $\hat{q} \sim N\left(q, \frac{q(1-q)}{2N}\right)$.

$$P\left(-z_{0.025} < \frac{\hat{q}-q}{\sqrt{\frac{q(1-q)}{2N}}} < z_{0.025}\right) = 0.95$$

Approksimerer med

$$P\left(\hat{q} - z_{0.025}\sqrt{\frac{\hat{q}(1-\hat{q})}{2N}} < q < \hat{q} + z_{0.025}\sqrt{\frac{\hat{q}(1-\hat{q})}{2N}}\right) = 0.95$$

Konfidensintervall:

$$\left[\hat{q} - z_{0.025}\sqrt{\frac{\hat{q}(1-\hat{q})}{2N}}, \hat{q} + z_{0.025}\sqrt{\frac{\hat{q}(1-\hat{q})}{2N}}\right].$$

e) Vi kan gjøre en goodness-of-fit test, og se på statistikken

$$D = \sum_{i=1}^{3} \frac{(X_i - n\hat{p}_i)^2}{n\hat{p}_i} \sim \chi^2_{n-2}.$$

Modellen er urimelig hvis

$$d = \sum_{i=1}^{3} \frac{(x_i - n\hat{p}_i)^2}{n\hat{p}_i} \geq \chi^2_{1-\alpha,n-2}.$$

Oppgave 4

Har

$$P(X \geq 100 \mid H_0) = 4\%,$$
$$P(X \geq 99 \mid H_0) = 6\%,$$

og vil lage en test som forkaster H_0 på nivå 5%.

Definer forkastningsreglen $0 \leq \phi(X) \leq 1$. Hvis $X = x$ og $\phi(x) = 0$ beholder vi H_0, hvis $\phi(x) = 1$ forkaster vi H_0, og hvis $0 < \phi(x) < 1$ utfører vi et Bernoulli-forsøk med sannsynlighet $\phi(x)$ for å forkaste H_0 og beholder H_0 ellers. La

$$\phi(X) = \begin{cases} 0 & \text{hvis } X < 99 \\ 0.5 & \text{hvis } X = 99 \\ 1 & \text{hvis } X \geq 100 \end{cases},$$

og U være en rettferdig mynt, dvs. Bernoulli med $p = 0.5$. Sannsynligheten for å forkaste er da:

$$\begin{aligned}
P(\text{Forkast } H_0 \mid H_0) &= 1 \times P(X \geq 100 \mid H_0) + 0.5 \times P(X = 99 \mid H_0) \\
&= 1 \times 4\% + 0.5 \times 2\% \\
&= 5\%.
\end{aligned}$$