TMA4245 Statistikk
Vår 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer 10, blokk II
Løsningsskisse

Oppgave 1

a) The probability is $\int_{0.5}^{0.9} 6x(1-x)\,dx = \int_{0.5}^{0.9}(6x-6x^2)\,dx = [3x^2-2x^3]_{0.5}^{0.9} = 0.472$.

b) The likelihood function is given by

$$L(\beta) = \prod_{i=1}^{n} \beta(\beta+1)x_i(1-x_i)^{\beta-1} = \beta^n(\beta+1)^n \left(\prod_{i=1}^{n} x_i\right) \prod_{i=1}^{n}(1-x_i)^{\beta-1},$$

and the log likelihood

$$\ln L(\beta) = n\ln\beta + n\ln(\beta+1) + \sum_{i=1}^{n}\ln x_i + (\beta-1)\sum_{i=1}^{n}\ln(1-x_i),$$

which has derivative

$$(\ln L)'(\beta) = \frac{n}{\beta} + \frac{n}{\beta+1} + \sum_{i=1}^{n}\ln(1-x_i).$$

$(\ln L)'$ is decreasing on $(0,\infty)$ and the sum of two first terms tends to ∞ when $\beta \to 0^+$ and to 0 when $\beta \to \infty$, so that $(\ln L)'$ will have a single zero (the third term is negative) for $\beta > 0$ and be positive left of the zero and negative right of the zero. This means that L has its maximum at this zero. Solving for the zero,

$$\beta^2 \sum_{i=1}^{n}\ln(1-x_i) + \Big(2n + \sum_{i=1}^{n}\ln(1-x_i)\Big)\beta + n = 0,$$

we get

$$\beta = \frac{-2n - \sum_{i=1}^{n}\ln(1-x_i) \pm \sqrt{4n^2 + (\sum_{i=1}^{n}\ln(1-x_i))^2}}{2\sum_{i=1}^{n}\ln(1-x_i)}$$
$$= -\frac{n}{\sum_{i=1}^{n}\ln(1-x_i)} - \frac{1}{2} \pm \sqrt{\Big(\frac{n}{\sum_{i=1}^{n}\ln(1-x_i)}\Big)^2 + \frac{1}{4}}.$$

We choose the larger zero since $(\ln L)'$ has only one zero for positive arguments (the other we found must be negative), and get the maximum likelihood estimator

$$\sqrt{\Big(\frac{n}{\sum_{i=1}^{n}\ln(1-X_i)}\Big)^2 + \frac{1}{4}} - \frac{n}{\sum_{i=1}^{n}\ln(1-X_i)} - \frac{1}{2} = \sqrt{\frac{1}{\big(\overline{\ln(1-X)}\big)^2} + \frac{1}{4}} - \frac{1}{\overline{\ln(1-X)}} - \frac{1}{2}.$$

For $n = 100$ and $\sum_{i=1}^{n} \ln(1 - x_i) = -104.0$ the estimate is $\sqrt{1/1.04^2 + 1/4} + 1/1.04 - 1/2 = 1.545$.

(The discussion of actual attainment of maximum at the zero and of which zero to be chosen, is not required.)

Oppgave 2

a) Antagelser for at X er binomisk fordelt:

- Gjør n forsøk: Spør n personer.
- Registrerer suksess eller fiasko i hvert forsøk: Får svaret JA eller ikke JA (nei eller vet ikke) i hvert forsøk.
- P(suksess) lik i alle forsøk: Sannsynlighet for JA er p for alle som blir spurt.
- Forsøka er uavhengige: Rimelig å anta at de som blir spurt svarer uavhengig av hverandre.

$P(X \geq 18) = 1 - P(X < 18) = 1 - P(X \leq 17) \overset{\text{tabell}}{=} 1 - 0.965 = 0.035.$

$P(10 < X < 15) = P(X \leq 14) - P(X \leq 10) \overset{\text{tabell}}{=} 0.584 - 0.048 = 0.536$

b)
- $E(\hat{P}) = p$ og $\text{Var}(\hat{P}) = \frac{1}{4}(\frac{1}{n_1} + \frac{1}{n_2})p(1-p)$.
- $E(P^*) = p$ og $\text{Var}(P^*) = \frac{1}{n_1+n_2}p(1-p)$.

Egenskaper for god estimator: forventningsrett og liten varians. Begge estimatorene er forventningsrette, men P^* har minst varians, vi velger derfor P^*.

La $\alpha = 0.05$. Siden $\frac{\hat{P}-p}{\sqrt{\frac{1}{2n}\hat{P}(1-\hat{P})}}$ er tilnærmet standardnormalfordelt får vi:

$$
\begin{aligned}
P\left(-z_{\frac{\alpha}{2}} < \frac{\hat{P} - p}{\sqrt{\frac{1}{2n}\hat{P}(1-\hat{P})}} < z_{\frac{\alpha}{2}}\right) &\approx 1 - \alpha \\
P\left(\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{1}{2n}\hat{P}(1-\hat{P})} < p < \hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{1}{2n}\hat{P}(1-\hat{P})}\right) &\approx 1 - \alpha
\end{aligned}
$$

Et tilnærmet 95% konfidensintervall for p blir da:

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

c) Vi har at

$$
Y = X_3 - n\hat{P} = X_3 - n\frac{X_1 + X_2}{2n} = X_3 - \frac{1}{2}X_1 - \frac{1}{2}X_2.
$$

Siden n er stor og p ikke nær 0 og 1, vil vi ha at $np > 5$ og $n(1-p) > 5$, slik at vi kan bruke normaltilnærming til binomisk fordeling. Vi kan dermed anta at X_1, X_2 og X_3 alle er tilnærmet normalfordelt, de er uavhengige, og lineærkombinasjonen Y er dermed også tilnærmet normalfordelt.

$\text{Var}(Y) = \text{Var}(X_3 - n\hat{P}) \stackrel{\text{uavh.}}{=} \text{Var}(X_3) + n^2\text{Var}(\hat{P}) \stackrel{b)}{=} np(1-p) + n^2\frac{1}{2n}p(1-p) = \frac{3}{2}np(1-p)$.

Har da at

- $X_3 - n\hat{P}$ er tilnærmet normalfordelt
- $\text{Var}(X_3 - n\hat{P}) = \frac{3}{2}np(1-p)$
- $\text{E}(X_3 - n\hat{P}) = E(X_3) - nE(\hat{P}) = np - np = 0$

Vi får da et prediksjonsintervall ved:

$$P\left(-z_{\frac{\alpha}{2}} < \frac{X_3 - n\hat{P}}{\sqrt{\frac{3}{2}np(1-p)}} < z_{\frac{\alpha}{2}}\right) \approx 1-\alpha$$

$$P\left(n\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{3}{2}np(1-p)} < X_3 < n\hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{3}{2}np(1-p)}\right) \approx 1-\alpha$$

Siden n er stor, vil variansen til $\hat{P}$ være liten, og $\hat{P}$ være en god estimator for p. Vi kan derfor erstatte p med estimatet $\hat{p}$ i uttrykket for intervallgrensene.

Intervallet blir: $[n\hat{p} - z_{0.025}\sqrt{\frac{3}{2}n\hat{p}(1-\hat{p})}, n\hat{p} + z_{0.025}\sqrt{\frac{3}{2}n\hat{p}(1-\hat{p})}]$

Innsatt verdier blir intervallet $[633, 704]$.

Oppgave 3

a) $T \sim \text{eksp}(\frac{z}{\mu}) \quad \text{E}(T) = \frac{\mu}{z}$

$\mu = 1000, \quad z = 2.0$

$P(T \leq 1000) = \int_0^{1000} \frac{z}{\mu}e^{-\frac{z}{\mu}x}dx = \int_0^{1000} \frac{1}{500}e^{-\frac{x}{500}}dx = [-e^{-\frac{x}{500}}]_0^{1000} = 1 - e^{-2} = \underline{\underline{0.86}}$

$P(T \leq 1000) = 0.5 \quad \Leftrightarrow \quad 1 - e^{-\frac{1000z}{1000}} = 0.5$

$e^{-z} = 0.5 \quad \Leftrightarrow \quad z = -\ln 0.5 = \underline{\underline{0.69}}$

$z_1 = 1.0, \quad z_2 = 2.0$

$P(T_2 \geq T_1) =?$

Finner simultanfordelingen til T_1 og T_2:

$f(t_1, t_2) = \frac{z_1}{\mu}e^{-\frac{z_1}{\mu}t_1}\frac{z_2}{\mu}e^{-\frac{z_2}{\mu}t_2}$ siden T_1 og T_2 er uavhengige.

$$\begin{aligned}P(T_2 \geq T_1) &= \int_0^\infty \int_{t_1}^\infty f(t_1,t_2)dt_2dt_1 = \tfrac{z_1z_2}{\mu^2}\int_0^\infty \int_{t_1}^\infty e^{-\frac{z_1}{\mu}t_1}e^{-\frac{z_2}{\mu}t_2}dt_2dt_1 \\ &= \tfrac{z_1z_2}{\mu^2}\int_0^\infty [-\tfrac{\mu}{z_2}e^{-\frac{z_1}{\mu}t_1-\frac{z_2}{\mu}t_2}]_{t_1}^\infty dt_1 = \tfrac{z_1z_2}{\mu^2}\tfrac{\mu}{z_2}\int_0^\infty e^{-\frac{z_1}{\mu}t_1-\frac{z_2}{\mu}t_2}dt_1 \\ &= \tfrac{z_1}{\mu}[-\tfrac{\mu}{z_1+z_2}e^{-(\frac{z_1+z_2}{\mu})t_1}]_0^\infty = \tfrac{z_1}{z_1+z_2} = \tfrac{1.0}{1.0+2.0} = \underline{\underline{\tfrac{1}{3}}}\end{aligned}$$

b) SME for μ:

$f(t_1,\dots,t_n;\mu,z_1,\dots,z_n)=\prod_{i=1}^n \frac{z_i}{\mu}e^{-\frac{z_i}{\mu}t_i}$

$L(\mu;t_1,\dots,t_n,z_1,\dots,z_n)=\prod_{i=1}^n \frac{z_i}{\mu}e^{-\frac{z_i}{\mu}t_i}$

$l(\mu)=\ln L(\mu)=\sum_{i=1}^n \ln z_i - n\ln\mu - \sum_{i=1}^n \frac{z_i}{\mu}t_i$

$\frac{\partial l}{\partial \mu}=-\frac{n}{\mu}+\sum_{i=1}^n \frac{z_i t_i}{\mu^2}=0$

$n=\sum_{i=1}^n \frac{z_i t_i}{\mu}$

$\mu=\frac{1}{n}\sum_{i=1}^n z_i t_i$ Dermed er SME $\widehat{\mu}=\frac{1}{n}\sum_{i=1}^n z_i T_i$.

$$\mathrm{E}(\widehat{\mu})=\mathrm{E}(\frac{1}{n}\sum_{i=1}^n z_iT_i)=\frac{1}{n}\sum_{i=1}^n z_i\mathrm{E}(T_i)=\frac{1}{n}\sum_{i=1}^n z_i\frac{\mu}{z_i}=\frac{1}{n}\sum_{i=1}^n \mu=\underline{\underline{\mu}}$$

Dvs. estimatoren er forventningsrett.

$$\mathrm{Var}(\hat{\mu})=\mathrm{Var}(\frac{1}{n}\sum_{i=1}^n z_iT_i)=\frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(z_iT_i)=\frac{1}{n^2}\sum_{i=1}^n z_i^2\mathrm{Var}(T_i)$$
$$=\frac{1}{n^2}\sum_{i=1}^n z_i^2\frac{\mu^2}{z_i^2}=\frac{1}{n^2}\sum_{i=1}^n \mu^2=\underline{\underline{\frac{\mu^2}{n}}}$$

c) MGF for T_i: $M_{T_i}(t)=\frac{\frac{z_i}{\mu}}{\frac{z_i}{\mu}-t}$ (Funnet i tabell.)

$V=\frac{2n\widehat{\mu}}{\mu}=\frac{2\sum_{i=1}^n z_iT_i}{\mu}=\sum_{i=1}^n \frac{2z_i}{\mu}T_i$

$M_{\frac{2z_i}{\mu}T_i}(t)=\frac{\frac{z_i}{\mu}}{\frac{z_i}{\mu}-\frac{2z_i}{\mu}t}=(1-2t)^{-1}$ (Bruker at $M_{aX}(t)=M_X(at)$)

$M_V(t)=\prod_{i=1}^n(1-2t)^{-1}=(1-2t)^{-n}$

(Bruker at $M_{\sum_{i=1}^n X_i}(t)=\prod_{i=1}^n M_{X_i}(t)$)

$(1-2t)^{-n}$ er MGF for kji-kvadratfordelingen med $2n$ frihetsgrader. V har samme MGF som kji-kvadratfordelingen med $2n$ frihetsgrader, derfor er $V\sim\chi^2_{2n}$.

d) $(1-\alpha)100\%$ konfidensintervall for μ:

Bruker at $V=\frac{2n\widehat{\mu}}{\mu}\sim\chi^2_{2n}$.

$$P(z_{1-\alpha/2,2n}\le V\le z_{\alpha/2,2n})=1-\alpha$$
$$P(z_{1-\alpha/2,2n}\le \frac{2n\widehat{\mu}}{\mu}\le z_{\alpha/2,2n})=1-\alpha$$
$$P(\frac{z_{1-\alpha/2,2n}}{2n\widehat{\mu}}\le\frac{1}{\mu}\le\frac{z_{\alpha/2,2n}}{2n\widehat{\mu}}\le\frac{1}{\mu})=1-\alpha$$
$$P(\frac{2n\widehat{\mu}}{z_{\alpha/2,2n}}\le\mu\le\frac{2n\widehat{\mu}}{z_{1-\alpha/2,2n}})=1-\alpha$$

Det gir konfidensintervallet $\underline{\underline{[\frac{2n\widehat{\mu}}{z_{\alpha/2,2n}},\frac{2n\widehat{\mu}}{z_{1-\alpha/2,2n}}]}}$

$\alpha = 0.10$, $n = 10$, $\widehat{\mu} = 1270.38$

$z_{1-\alpha/2,2n} = z_{0.95,20} = 10.85$, $z_{\alpha/2,2n} = z_{0.05,20} = 31.41$

Innsatt disse tallverdiene blir konfidensintervallet $\underline{\underline{[808.90, 2341.71]}}$