Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

LØSNINGSFORSLAG TIL EKSAMENSOPPGAVENE I

EMNE TMA4245 STATISTIKK

3. juni 2010
Tid: 09:00–13:00

Oppgave 1

a) La S være hendelsen at individet er sykt og la aa, Aa og AA være hendelsene at individet er av hver av de tre respektive genotypene. Lov om totalsannsynlighet gir da

$$\begin{aligned} P(S) &= P(S|aa)P(aa) + P(S|Aa)P(Aa) + P(S|AA)P(AA) \\ &= 0.6 \cdot 0.0001 + 0.0198 \cdot 0.02 + 0.9801 \cdot 0.01 \\ &= 0.0103. \end{aligned} \tag{1}$$

b) Den betingede sannsynligheten for at et sykt individ er av type aa blir i følge Bayes teorem

$$\begin{aligned} P(aa|S) &= \frac{P(S|aa)P(aa)}{P(S)} \\ &= \frac{0.6 \cdot 0.0001}{0.10257} \\ &= 0.00584 \end{aligned} \tag{2}$$

På tilsvarende måte får vi

$$P(Aa|S) = 0.0386 \tag{3}$$

og

$$P(AA|S) = 0.956 \tag{4}$$

Oppgave 2

a) For $t \geq 0$,

$$F_T(t) = \int_{-\infty}^{t} f_T(x)\,dx = \int_0^t 2\lambda x e^{-\lambda x^2}\,dx = (-e^{-\lambda x^2})_0^t = 1 - e^{-\lambda t^2} \tag{5}$$

Det gir

$$\begin{aligned} P(20 < T \leq 30) &= F_T(30) - F_T(20) = e^{-1.5\cdot 10^{-3} 20^2} - e^{-1.5\cdot 10^{-3} 30^2} \\ &= e^{-0.6} - e^{-1.35} = 0.55 - 0.26 = 0.29 \end{aligned} \tag{6}$$

b) Anta uavhengige observasjoner $t_1, \ldots, t_n$ (utfall av TS $T_1, \ldots, T_n$). Rimelighetsfunksjonen (RF) blir,

$$L(t_1, \ldots, t_n|\lambda) = \prod_{i=1}^{n} f_T(t_i) = 2^n \lambda^n (\prod_{i=1}^{n} t_i) e^{-\lambda \sum_{i=1}^{n} t_i^2} \tag{7}$$

Greit å bruke log(RF):

$$l(\lambda) = \log L(t_1, \ldots, t_n|\lambda) = \prod_{i=1}^{n} f_T(t_i) = n \log \lambda + \log\big(2^n \prod_{i=1}^{n} t_i\big) - \lambda \sum_{i=1}^{n} t_i^2 \tag{8}$$

Bestemmer makspunktet λ^* ved å løse ligningen,

$$\frac{dl(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i^2 = 0 \tag{9}$$

Det gir

$$\lambda^* = \frac{n}{\sum_{i=1}^{n} t_i^2}\,, \tag{10}$$

som åpenbart er et makspunkt siden $d^2 l(\lambda)/d\lambda^2 = -n/\lambda^2 < 0$.

SME blir derfor

$$\Lambda^* = \frac{n}{\sum_{i=1}^{n} T_i^2}\,, \tag{11}$$

c) Fra 'Tabeller og formler...' er sannnsynlighetstettheten (ST) til en χ^2-fordelt stokastisk variabel X med $2n$ frihetsgrader ($n = 1, 2, \dots$) gitt som

$$\begin{aligned} f_X(x) &= c_n x^{n-1} e^{-x/2}\,, \quad x \geq 0\,, \\ &= 0\,, \qquad \text{ellers}\,, \end{aligned} \tag{12}$$

hvor $c_n = 1/(2^n(n-1)!)$. Siden $f_X(x)$ er en ST, må det gjelde at

$$c_n \int_0^\infty x^{n-1} e^{-x/2}\,dx = 1\,, \;\; n = 1, 2, \dots. \tag{13}$$

Dermed er

$$\begin{aligned} \mathrm{E}[X^{-1}] &= c_n \int_0^\infty x^{-1} x^{n-1} e^{-x/2}\,dx = c_n \int_0^\infty x^{(n-1)-1} e^{-x/2}\,dx \\ &= \frac{c_n}{c_{n-1}}\, c_{n-1} \int_0^\infty x^{(n-1)-1} e^{-x/2}\,dx = \frac{c_n}{c_{n-1}} = \frac{1}{2(n-1)}\,, \;\; n = 2, 3 \dots. \end{aligned} \tag{14}$$

Tilsvarende,

$$\begin{aligned} \mathrm{E}[X^{-2}] &= c_n \int_0^\infty x^{-2} x^{n-1} e^{-x/2}\,dx = c_n \int_0^\infty x^{(n-2)-1} e^{-x/2}\,dx \\ &= \frac{c_n}{c_{n-2}}\, c_{n-2} \int_0^\infty x^{(n-2)-1} e^{-x/2}\,dx = \frac{c_n}{c_{n-2}} = \frac{1}{4(n-1)(n-2)}\,, \;\; n = 3, 4 \dots. \end{aligned} \tag{15}$$

ST til Y:

$$f_Y(y) = f_T\big(\sqrt{y/(2\lambda)}\big) \frac{1}{|\frac{dy}{dt}|} = 2\lambda \sqrt{y/(2\lambda)}\, e^{-\lambda y/(2\lambda)} \frac{1}{4\lambda\sqrt{y/(2\lambda)}} = \frac{1}{2}\, e^{-\frac{y}{2}}\,, \tag{16}$$

som er ST til en χ^2-fordelt stokastisk variabel med 2 frihetsgrader.

Vi vet nå at $2\lambda T_i^2$ er χ^2-fordelt med 2 frihetsgrader for hver $i = 1, \dots, n$. Siden T_i'ene er uavhengige medfører det at $\sum_{i=1}^n 2\lambda T_i^2$ er χ^2-fordelt med $2n$ frihetsgrader. Dette resultatet sammen med ligning (3) i oppgavesettet gir

$$\mathrm{E}[\Lambda^*] = \mathrm{E}\Big[\frac{n}{\sum_{i=1}^n T_i^2}\Big] = \mathrm{E}\Big[\frac{2\lambda n}{\sum_{i=1}^n 2\lambda T_i^2}\Big] = 2\lambda n\, \mathrm{E}\Big[\frac{1}{\sum_{i=1}^n 2\lambda T_i^2}\Big] = 2\lambda n \cdot \frac{1}{2(n-1)} = \frac{n}{n-1}\,\lambda\,, \tag{17}$$

som ikke er forventningsrett (bare asymptotisk når $n \to \infty$). Ser at estimatoren

$$\hat{\Lambda} = \frac{n-1}{n}\,\Lambda^* = \frac{n-1}{\sum_{i=1}^n T_i^2}\,, \tag{18}$$

blir forventningsrett.

Vi benytter ligning (3) i oppgavesettet på nytt, og beregner

$$\mathrm{E}[\hat{\Lambda}^2] = \big(2\lambda(n-1)\big)^2 \mathrm{E}\Big[\frac{1}{\big(\sum_{i=1}^n 2\lambda T_i^2\big)^2}\Big] = \frac{4(n-1)^2\lambda^2}{4(n-1)(n-2)} = \frac{n-1}{n-2}\,\lambda^2\,, \tag{19}$$

Dermed følger det at

$$\mathrm{Var}[\hat{\Lambda}] = \frac{n-1}{n-2}\,\lambda^2 - \lambda^2 = \frac{1}{n-2}\,\lambda^2\,. \tag{20}$$

d) Vi har allerede sett at $X = \sum_{i=1}^n 2\lambda T_i^2$ er χ^2-fordelt med $2n$ frihetsgrader. Dermed blir

$$\mathrm{Prob}(\chi^2_{1-\frac{\alpha}{2},2n} < X \le \chi^2_{\frac{\alpha}{2},2n}) = 1-\alpha\,. \tag{21}$$

Et $100(1-\alpha)\%$ konfidensintervall for λ blir dermed,

$$\Big(\frac{\chi^2_{1-\frac{\alpha}{2},2n}}{2\sum_{i=1}^n t_i^2}, \frac{\chi^2_{\frac{\alpha}{2},2n}}{2\sum_{i=1}^n t_i^2}\Big)\,. \tag{22}$$

Observasjonene gir $2\sum_{i=1}^5 t_i^2 = 5329.76$. Med $\alpha = 0.05$ (og $n = 5$):

$$\Big(\frac{3.247}{5329.76}, \frac{20.483}{5329.76}\Big) = (0.61\cdot 10^{-3}, 3.84\cdot 10^{-3})\,. \tag{23}$$

e) I utgangspunktet kunne en tenke seg å teste om $\hat{\lambda} = (n-1)/(\sum_{i=1}^n t_i^2) > \lambda_\alpha$ hvor λ_α er definert ved at

$$\mathrm{Prob}(\hat{\Lambda} > \lambda_\alpha | \lambda = \lambda_0) = \alpha\,. \tag{24}$$

Problemet er at vi ikke kjenner fordelingen til $\hat{\Lambda}$. Men vi vet at under H_0 er $2\lambda_0\sum_{i=1}^n T_i^2$ χ^2-fordelt med $2n$ frihetsgrader. Kan omforme:

$$\begin{aligned}\mathrm{Prob}(\hat{\Lambda} > \lambda_\alpha | \lambda = \lambda_0) &= \mathrm{Prob}\Big(\frac{n-1}{\sum_{i=1}^n T_i^2} > \lambda_\alpha | \lambda = \lambda_0\Big)\\ &= \mathrm{Prob}\Big(2\lambda_0 \sum_{i=1}^n T_i^2 < \frac{2\lambda_0(n-1)}{\lambda_\alpha}\Big) = \alpha\,.\end{aligned} \tag{25}$$

Dermed er

$$\frac{2\lambda_0(n-1)}{\lambda_\alpha} = \chi^2_{1-\alpha,2n}\,. \tag{26}$$

Det følger at vi kan bruke $2\lambda_0\sum_{i=1}^n T_i^2$ som testobservator med kritisk område $(0, \chi^2_{1-\alpha,2n})$ For signifikansnivå $\alpha = 0.05$ og med de oppgitte observasjonene, finner vi $\chi^2_{0.95,10} = 3.94$, $2\lambda_0 = 3\cdot 10^{-3}$, og $2\lambda_0\sum_{i=1}^n t_i^2 = 3\cdot 10^{-3}\cdot 1227.75 = 3.68 < 3.94$. Dermed må H_0 forkastes.

f)

$$\begin{aligned}\text{Prob(Ikke reklamasjon)} &= \text{Prob}\big(T_1 > a \cap \ldots \cap T_5 > a\big) = \text{Prob}(T > a)^5 \\ &= \big(1 - F_T(a)\big)^5 = e^{-5\lambda a^2} \geq 0.95\,. \end{aligned} \tag{27}$$

Dette gir

$$a^2 \leq \frac{-\ln 0.95}{0.0075} = 6.84 \quad \Rightarrow \quad a \leq 2.62\,\text{uker.} \tag{28}$$

g) Med uavhengige levetider og a og λ som i punkt f), dvs. konstante, blir U binomisk fordelt. $\text{E}[U] = np = 1000 \cdot 0.05 = 50$ og $\text{Var}[U] = np(1-p) = 1000 \cdot 0.05 \cdot 0.95 = 47.5$

$$\text{Prob}(U \leq 60) = \text{Prob}\Big(\frac{U - 50}{\sqrt{47.5}} \leq \frac{60 - 50}{\sqrt{47.5}}\Big) \approx \Phi\Big(\frac{10.5}{6.89}\Big) = \Phi(1.52) = 0.936\,. \tag{29}$$

Merk: Tallet 10.5 gir bedre tilnærmelse enn $10 = 60 - 50$.