UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1120 — Innføring i anvendt statistikk – FASIT
Eksamensdag:	Fredag 2. juni 2006.
Tid for eksamen:	09.00 – 12.00.
Oppgavesettet er på 4 sider.	
Vedlegg:	Tabeller for χ^2, t og F fordelingen.
Tillatte hjelpemidler:	Godkjent kalkulator og formelsamling for STK1100/ STK1110 og for STK1120.

Kontroller at oppgavesettet er komplett
før du begynner å besvare spørsmålene.

Oppgave 1.

(a) Vi har

$$f(x_1, x_2, ..., x_n) = f(x_1)\prod_{i=2}^{n} f(x_i|x_{1:i-1})$$

og $x_i|x_{1:i-1} \sim N(a_1x_{i-1} + a_2x_{i-2}, \sigma^2)$ som gir

$$\begin{aligned} L(a_1, a_2, \sigma^2) =& f(x_1, x_2, ..., x_n; a_1, a_2, \sigma^2) \\ =& \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma}(x_i - a_1x_{i-1} - a_2x_{i-2})^2} \end{aligned}$$

Dette gir

$$\begin{aligned} l(a_1, a_2, \sigma^2) &= \sum_{i=1}^{n} [-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(x_i - a_1x_{i-1} - a_2x_{i-2})^2] \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a_1x_{i-1} - a_2x_{i-2})^2 \end{aligned}$$

(Fortsettes side 2.)

som viser at maksimering av $l(a_1, a_2, \sigma^2)$ (som er ekvivalent med maksimering av $L(a_1, a_2, \sigma^2)$) er ekvivalent med minimering av $\sum_{i=1}^n (x_i - a_1 x_{i-1} - a_2 x_{i-2})^2$

(b) Vi har

$$\begin{aligned}
\frac{\partial}{\partial a_1} S(a_1, a_2) =& -2 \sum_{i=1}^n (x_i - a_1 x_{i-1} - a_2 x_{i-2}) x_{i-1} \\
\frac{\partial}{\partial a_2} S(a_1, a_2) =& -2 \sum_{i=1}^n (x_i - a_1 x_{i-1} - a_2 x_{i-2}) x_{i-2}
\end{aligned}$$

som viser at $(\hat{a}_1, \hat{a}_2)$ må tilfredsstille likningene

$$\begin{aligned}
\sum_{i=1}^n (x_i - \hat{a}_1 x_{i-1} - \hat{a}_2 x_{i-2}) x_{i-1} =& 0 \\
\sum_{i=1}^n (x_i - \hat{a}_1 x_{i-1} - \hat{a}_2 x_{i-2}) x_{i-2} =& 0
\end{aligned}$$

som er ekvivalent med at

$$\begin{aligned}
\hat{a}_1 \sum_{i=1}^n x_{i-1}^2 + \hat{a}_2 \sum_{i=1}^n x_{i-1} x_{i-2} =& \sum_{i=1}^n x_{i-1} x_i \\
\hat{a}_1 \sum_{i=1}^n x_{i-1} x_{i-2} + \hat{a}_2 \sum_{i=1}^n x_{i-2}^2 =& \sum_{i=1}^n x_{i-2} x_i
\end{aligned}$$

Dette er et lineært likningssystem som enkelt kan løses med standard vektor/matrise beregninger.

$$\frac{\partial}{\partial \sigma^2} l(a_1, a_2, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - a_1 x_{i-1} - a_2 x_{i-2})^2$$

som gir at $\hat{\sigma}^2$ er gitt ved

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{a}_1 x_{i-1} - \hat{a}_2 x_{i-2})^2$$

(c) Vi har basert på normaltilnærmingen at $\hat{\theta} \pm z_{1-\alpha/2} s_{\hat{\theta}}$ er et tilnærmet $100(1-\alpha)\%$ konfidensintervall. For $\alpha = 0.05$, blir $<_{0.975}= 1.96$.

Her er $s_{\hat{a}_1} = \sqrt{0.0199} = 0.1411$ som gir at $[0.7110 + \pm 1.96 \times 0.1411] = [0.43450.9875]$ er et tilnærment 95% konfidensintervall for a_1.

Videre er $s_{\hat{a}_2} = \sqrt{0.0204} = 0.1428$ som gir at $[-0.2220 \pm 1.96 \times 0.1428] = [-0.5019, 0.0579]$ er et tilnærment 95% konfidensintervall for a_2

(Fortsettes side 3.)

Oppgave 2.

(a) Vi har

$$L(\theta) = \frac{n!}{x_1!x_2!x_3!x_4!}(\tfrac{1}{2} + \tfrac{1}{4}\theta)^{y_1}(\tfrac{1}{4}(1-\theta))^{y_2+y_3}(\tfrac{1}{4}\theta)^{y_4}$$

som gir

$$l(\theta) = K + y_1 \log(2+\theta) + (y_2+y_3)\log(1-\theta) + y_4 \log(\theta)$$

og

$$\frac{\partial}{\partial\theta} l(\theta) = \frac{y_1}{2+\theta} - \frac{y_2+y_3}{1-\theta} + \frac{y_4}{\theta}$$

Dermed må ML estimatet tilfredsstille likningen

$$\frac{y_1}{2+\hat{\theta}} - \frac{y_2+y_3}{1-\hat{\theta}} + \frac{y_4}{\hat{\theta}} = 0$$

som er ekvivalent med at

$$y_1(1-\theta)\theta - (y_2+y_3)(2+\theta)\theta + y_4(2+\theta)(1-\theta) = 0$$

som igjen er ekvivalent med at

$$n\theta^2 - (y_1 - 2*(y_2+y_3) - y_4)\theta - 2y_4 = 0$$

Da sannsynligheter må ligge mellom 0 og 1, er en negativ θ ikke mulig. Samtidig har vi at

$$\frac{\partial^2}{\partial\theta^2} l(\theta) = -\frac{y_1}{(2+\theta})^2) - \frac{y_2+y_3}{(1-\theta)^2} - \frac{y_4}{\theta^2}$$

som alltid er negativ for θ mellom 0 og 1, noe som viser at den positive verdien må være et maksimumspunkt.

(b) Vi estimerer forventningsskjevheten til å bli $0.62822 - 0.62682 = 0.0014$ som er forsvinnende liten i forhold til usikkerheten i $\hat{\theta}$, noe som tilsier at $\hat{\theta}$ er (tilnærmet) forventningsrett.

95% KI basert på normaltilnærming:

$$[\hat{\theta} - z_{0.975}s_{\theta^*}, \hat{\theta} + z_{0.975}s_{\theta^*}] = [0.5310, 0.7226]$$

95% KI basert på standard Bootstrap konfidensintervall:

$$[2\hat{\theta} - \theta^*_{(975)}, 2\hat{\theta} - \theta^*_{(25)}] = [0.52720.7296]$$

Svært like konfidensintervaller, som tilsier at normaltilnærmingen fungerer bra i dette tilfellet.

(c) En kan her enten bruke den generaliserte likelihood ratio testen eller Pearson's χ^2 test. Har $X^2 = \sum((y_i - E_i)^2/E_i) = 0.616$ og $-2log(\Lambda) = 0.302$. Frihetsgrader er $4 - 1 - 1 = 2$. Med $\chi^2_2(0.90) = 4.61$ ser vi at testobservatorene er langt fra signifikante og det er ingen grunn til å forkaste H_0.

(Fortsettes side 4.)

Oppgave 3.

Modell:

$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$

der α_i angir variasjoner som følge av gift, β_j variasjoner som følge av motgift og δ_{ij} interaksjon mellom de to faktorene. μ er det overordnede forventningsnivå mens ε_{ijk} er støledd som angir avvik fra modell (antas å være uavhengige med forventning null, konstant varians og normalfordelt).
Aktuelle hypoteser:

$$H_{AB} : \delta_{ij} = 0, \forall i, j \text{For testing om additive effekter av de to faktorer}$$
$$H_A : \alpha_i = 0, \forall i \text{For testing om de ulike giftstoffer har noen effekt}$$
$$H_B : \beta_j = 0, \forall j \text{For testing om de ulike motgiftstoffer har noen effekt}$$

Naturlig å teste H_{AB} først. Vi har at den observerte F verdi er mindre enn $\chi^2_6(0.10)$ som tilsvarer en P-verdi større enn 0.9.
H_A: Vi har at den observerte F verdi er større enn $\chi^2_2(0.995)$ som tilsvarer en P-verdi mindre enn 0.005.
H_B: Vi har at den observerte F verdi er større enn $\chi^2_3(0.995)$ som tilsvarer en P-verdi mindre enn 0.005.
Konklusjon: Det er ikke grunn til å forkaste H_{AB}, men god grunn til å forkaste både H_A og H_B, dvs additive effekter av faktorene.

SLUTT