Litt om summering:

$$\sum_{i=1}^{n} x_i \neq n\, x_i \qquad \sum x_i = x_1 + x_2 + \dots + x_n$$

$$\sum_{i=1}^{n} a = n\, a$$

$$\sum x_i^2 \neq \left(\sum x_i\right)^2 \qquad 2^2 + 3^2 \neq (2+3)^2$$

$$\frac{\sum x_i}{\sum y_i} \neq \sum \frac{x_i}{y_i}$$

$$\sum \frac{1}{x_i} \neq \frac{1}{\sum x_i}$$

$$\sum_{i=1}^{n} (a x_i + b) = \sum a x_i + \sum_{i=1}^{n} b = a \sum x_i + n b$$

$$\sum_{i=1}^{n} a x_i + b = a \sum x_i + b$$

$$\ln\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \ln x_i \qquad \ln a \cdot b = \ln a + \ln b$$

Universitetet i Stavanger

EKSAMEN I: MOT150 MATEMATISK STATISTIKK

Del 1 av MOT100 MATEMATISK STATISTIKK OG STOKASTISKE PROSESSER A

Del 1 av MOT110 MATEMATISK STATISTIKK OG STOKASTISKE PROSESSER B

VARIGHET: 4 TIMER DATO: 3. OKTOBER 2006

TILLATTE HJELPEMIDLER: Kalkulator av typen HP30S; Tabeller og formler i statistikk (Tapir forlag); Rottman: Matematisk formelsamling.

OPPGAVESETTET BESTÅR AV 3 OPPGAVER PÅ 3 SIDER

Oppgave 1:

Det fins to ulike målemetoder for å fastslå verdien, μ, på en fysisk størrelse. Den første metoden gir resultater som er normalfordelte med forventning μ og standardavvik $\sigma_1 = 6$, mens den andre metoden gir resultater som er normalfordelte med forventning μ og standardavvik $\sigma_2 = 4$.

Anta først at $\mu = 105.4$

a) Hva er sannsynligheten for at den første målemetoden gir en verdi på under 110?
Hva er sannsynligheten for at den andre målemetoden gir en verdi på mellom 100 og 110?
Dersom man utfører en måling med hver metode, hva er sannsynligheten for at den høyeste av de to målingene er mindre enn 110?

I praksis er μ ukjent, og skal estimeres. Begge målemetodene er dyre å utføre i praksis, så man har bare anledning til å gjøre en måling med hver av metodene. La X_1 være resultatet av en måling med den første metoden, og X_2 resultatet av en måling med den andre metoden. Man ønsker å kombinere de to målingene til en best mulig estimator for μ.

b) Hvilke to krav ønsker man generelt at en god estimator skal oppfylle?
Bestem a og b slik at estimatoren $\hat{\mu} = aX_1 + bX_2$ oppfyller disse to kravene.

Oppgave 3:

La tiden X (målt i måneder) mellom etterfølgende feil i et mobilnett være en kontinuerlig stokastisk variabel med sannsynlighetstetthet

$$f_X(x) = \theta(x+1)^{-\theta-1} \quad \text{for } x > 0,$$

der $\theta > 0$.
Anta i første omgang at $\theta = 2$.

a) Vis at den kumulative fordelingsfunksjonen er gitt ved $F_X(x) = 1 - (x+1)^{-\theta}$.
Vis at $P(X > 1) = 0.25$.
Regn ut $P(0.5 < X < 1.5)$.
Dersom man observerer at det er gått 0.5 måned siden forrige feil, hva er sannsynligheten for at det vil gå mindre enn 1 måned til til neste feil? Dvs hva er $P(X < 1.5 | X > 0.5)$?

La Y være antall intervall mellom etterfølgende feil man må observere til første gang det går mer enn 1 måned mellom to feil. Man kan anta at lengdene på etterfølgende intervall mellom feil er uavhengige.

b) Forklar hvorfor Y er geometrisk fordelt.
Regn ut $P(Y < 4)$ og $E(Y)$.

I realiteten er verdien på parameteren θ ukjent, men kan estimeres ut fra observerte data $X_1, \dots X_n$ over tider mellom etterfølgende feil i nettverket. Vi antar at disse målingene er uavhengige og har kumulativ fordelingsfunksjon som angitt i starten av oppgaven.
Tre forslag til estimatorer for θ er

$$\hat{\theta}_1 = \frac{n}{\sum_{i=1}^n \ln(X_i+1)}, \quad \hat{\theta}_2 = \frac{n-1}{\sum_{i=1}^n \ln(X_i+1)} \quad \text{og} \quad \hat{\theta}_3 = \frac{\sum_{i=1}^n \ln(X_i+1)}{n}$$

c) Hvilken av estimatorene over er sannsynlighetsmaksimeringsestimatoren (SME)?
Begrunn svaret ved å utlede SME.
Beregn estimatet for θ når $n = 15$ målinger gav $\sum_{i=1}^n \ln(x_i+1) = 12.1$.

Ved bruk av transformasjonsformelen kan det vises at $2\theta \ln(X_i + 1)$ er χ^2_2-fordelt (du trenger ikke vise dette).

d) Forklar hvorfor resultatet gitt over gir at $2\theta \sum_{i=1}^n \ln(X_i+1)$ er χ^2_{2n}-fordelt.
Bruk dette som utgangspunkt, og utled et 95% konfidensintervall for θ. Regn ut intervallet numerisk når dataene er som i punkt c).

$f_X(x) = \theta(x+1)^{-\theta-1}$

$y = 2\theta \ln(x+1)$

$\frac{y}{2\theta} = \ln(x+1)$

$x+1 = e^{y/2\theta}$

$x = e^{y/2\theta} - 1 = w(y)$

$w'(y) = \frac{1}{2\theta} e^{y/2\theta}$

$f_Y(y) = f_X(w(y)) \cdot w'(y)$

$= \theta\left(e^{y/2\theta} - 1 + 1\right)^{-(\theta+1)} \cdot \frac{1}{2\theta} e^{y/2\theta}$

$= \theta e^{-\frac{(\theta+1)y}{2\theta}} \cdot \frac{1}{2\theta} e^{y/2\theta}$

$= \frac{1}{2} e^{\frac{[-(\theta+1)+1]y}{2\theta}} = \frac{1}{2} e^{-\frac{\theta y}{2\theta}} = \underline{\underline{\frac{1}{2} e^{-y/2}}}$

Dvs $Y \sim \chi^2_2$

Oppgave 3:

a)

$$
\begin{aligned}
F_X(x) &= \int_0^x f_X(u)du = \int_0^x \theta(u+1)^{-\theta-1}du = [-(u+1)^{-\theta}]_0^x = \underline{\underline{1-(x+1)^{-\theta}}} \\
P(X>1) &= 1-P(\leq 1) = 1-F_X(1) = 1-(1-(1+1)^{-2}) = \underline{\underline{0.25}} \\
P(0.5 \leq X \leq 1.5) &= P(X \leq 1.5) - P(X \leq 0.5) = 1-(1.5+1)^{-2} - [1-(0.5+1)^{-2}] = \underline{\underline{0.28}} \\
P(X < 1.5 | X > 0.5) &= \frac{P((X<1.5)\cap(X>0.5))}{P(X>0.5)} = \frac{P(0.5<X<1.5)}{1-P(X\leq 0.5)} = \frac{0.28}{(0.5+1)^{-2}} = \underline{\underline{0.63}}
\end{aligned}
$$

b)

- Uavhengige enkeltforsøk til første suksess - uavhengige intervall mellom feil i mobilnett.
- Sjekker i hvert forsøk om en bestemt hendelse, at tiden mellom to feil er minst 1 måned, inntreffer eller ikke.
- Sannsynligheten for at tiden er minst 1 måned er den samme, $p = 0.25$, i hvert enkeltforsøk.

Y="antall avstander mellom etterfølgende feil som må observeres inntil man første gang opplever et intervall som er minst 1 måned" er da geometrisk fordelt med $p = 0.25$ (spesialtilfelle av negativt binomisk fordeling der vi ser på antall forsøk til første hendelse).

$$
\begin{aligned}
P(Y<4) &= P(Y\leq 3) = P(Y=1)+P(Y=2)+P(Y=3) \\
&= 0.25 + 0.25\cdot(1-0.25) + 0.25\cdot(1-0.25)^2 = \underline{\underline{0.58}} \\
\mathrm{E}(Y) &= \frac{1}{p} = \frac{1}{0.25} = \underline{\underline{4}}
\end{aligned}
$$

$L(\theta) = \prod \theta(x_i+1)^{-\theta-1}$

$\ell(\theta) = \ln L(\theta) = \sum \ln[\theta(x_i+1)^{-\theta-1}]$

$= \sum(\ln\theta + \ln(x_i+1)^{-\theta-1})$

c)

1) $$L(\theta) = f(x_1,\ldots,x_n;\theta) = \prod_{i=1}^n f(x_i;\theta) = \prod_{i=1}^n \theta(x_i+1)^{-\theta-1} = \theta^n \prod_{i=1}^n (x_i+1)^{-\theta-1}$$

2) $$l(\theta) = \ln L(\theta) = \ln(\theta^n) + \ln(\prod_{i=1}^n (x_i+1)^{-\theta-1}$$

$\ln a\cdot b = \ln a + \ln b$

$\ln(\prod X_i) = \sum \ln X_i$

$\ln a^b = b\ln a$

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

$= \sum\ln\theta + \sum\ln(x_i+1)^{-\theta-1}$

$= n\ln\theta - (\theta+1)\sum\ln(x_i+1)$

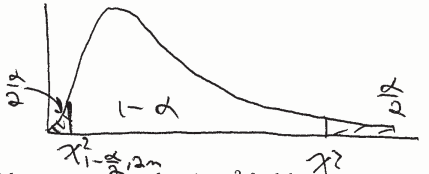
3) $$\frac{\partial l(\theta)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^n \ln(x_i+1) = 0 \;\Rightarrow\; n = \theta\sum_{i=1}^n \ln(x_i+1)$$

4) Dvs. SME blir: $\hat\theta = \underline{\underline{\frac{n}{\sum_{i=1}^n \ln(X_i+1)}}}$ og estimatet blir: $\hat\theta = \frac{15}{12.1} = \underline{\underline{1.24}}$

$\frac{\partial^2 \ell(\theta)}{\partial\theta^2} = -\frac{n}{\theta^2} < 0$ OK!

$2\theta\sum\ln(x_i+1) \sim \chi^2_{2n}$

$Y_i \sim \chi^2_2$



d) $2\theta\sum_{i=1}^n \ln(X_i+1) = \sum_{i=1}^n 2\theta\ln(X_i+1)$ Siden vi har at sum av uavhengige χ^2-fordelte variable er χ^2-fordelt med parameter ("frihetsgrader") lik summen av parametrene til enkeltvariablene (f.saml. side 34/35), så vil vi her ha (siden alle $2\theta\ln(X_i+1)$ er uavhengige og χ^2_2-fordelte) at $2\theta\sum_{i=1}^n \ln(X_i+1)$ er $\underline{\underline{\chi^2_{2n}\text{-fordelt}}}$

$\frac{2\theta n}{\hat\theta}$

$$
\begin{aligned}
P\left(\chi^2_{1-\alpha/2,2n} \leq 2\theta\sum_{i=1}^n \ln(X_i+1) \leq \chi^2_{\alpha/2,2n}\right) &= 1-\alpha \\
P\left(\frac{\chi^2_{1-\alpha/2,2n}}{2\sum_{i=1}^n \ln(X_i+1)} \leq \theta \leq \frac{\chi^2_{\alpha/2,2n}}{2\sum_{i=1}^n \ln(X_i+1)}\right) &= 1-\alpha
\end{aligned}
$$

D.v.s. et $(1-\alpha)100\%$-konfidensintervall for θ blir:

$P\left(\chi^2_{1-\frac{\alpha}{2},2n}\frac{\hat\theta}{2n} \leq \theta \leq \chi^2_{\frac{\alpha}{2},2n}\frac{\hat\theta}{2n}\right)$

$$\underline{\underline{\left[\frac{\chi^2_{1-\alpha/2,2n}}{2\sum_{i=1}^n \ln(X_i+1)}, \frac{\chi^2_{\alpha/2,2n}}{2\sum_{i=1}^n \ln(X_i+1)}\right]}}$$

Eventuelt kan man starte med at $2\theta\sum_{i=1}^n \ln(X_i+1) = 2\theta n(\sum_{i=1}^n \ln(X_i+1)/n) = 2n\theta/\hat\theta$ og får da intervallet uttrykt som:

$$\underline{\underline{\left[\frac{\hat\theta\cdot\chi^2_{1-\alpha/2,2n}}{2n}, \frac{\hat\theta\cdot\chi^2_{\alpha/2,2n}}{2n}\right]}}$$

Med $\alpha = 0.05$ og $n = 15$ får vi $\chi^2_{0.025,30} = 46.979$ og $\chi^2_{0.975,30} = 16.791$, og med $\sum_{i=1}^{15} \ln(x_i+1) = 12.1$ får vi følgende 95% konfidensintervall :

$$\left[\frac{16.791}{2\cdot 12.1}, \frac{46.979}{2\cdot 12.1}\right] = \underline{\underline{[0.69, 1.94]}}$$

$\hat\theta = \frac{n}{\sum\ln(x_i+1)}$

$2\theta\sum\ln(x_i+1) = \frac{2\theta n}{n}\sum\ln(x_i+1)$

$= 2\theta n\cdot\underbrace{\frac{\sum\ln(x_i+1)}{n}}_{1/\hat\theta} = \frac{2\theta n}{\hat\theta} \sim \chi^2_{2n}$

Dersom $h(\hat\theta,\theta) \sim t_{n-1}$ eller $N(0,1)$