Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen nov. 2002

Oppgave 1

a)

Vi har tre jordskjelvtyper, K, M og S, som har forskjellige sannsynlighetsfordelinger for skaden D de påfører en bygning.

Bruker her x som integrasjonsvariabel over mulige verdier for D.

$P(D > 1.0|K) = \int_1^\infty \lambda_K e^{-\lambda_K x} \mathrm{d}x = e^{-\lambda_K} = e^{-1.61} = 0.2$.
For de andre jordskjelvtypene blir integralet tilsvarende, og vi får:
$P(D > 1.0|M) = e^{-\lambda_M} = e^{-3.0} = 0.05$
$P(D > 1.0|S) = e^{-\lambda_S} = e^{-4.6} = 0.01$

b)

De tre jordskjelvtypene er distinkte og har forskjellig oppførsel, altså må området $D > 1.0$ undersøkes separat for de tre tilfellene. (Tegn gjerne et Venn-diagram.) Bruker loven om total sannsynlighet for å evaluere den totale sannsynligheten for skade.

$$\begin{aligned}
P(D > 1.0) &= P(D > 1.0 \cap K) + P(D > 1.0 \cap M) + P(D > 1.0 \cap S) \\
&= P(D > 1.0|K)P(K) + P(D > 1.0|M)P(M) + P(D > 1.0|S)P(S) \\
&= 0.2 \cdot 0.02 + 0.05 \cdot 0.20 + 0.01 \cdot 0.78 = 0.0218.
\end{aligned}$$

Merk at det selvfølgelig er forutsatt at det faktisk har inntruffet et jordskjelv.

Det neste spørsmålet er; gitt at vi ser at bygningen har falt sammen pga et jordskjelv (som ikke er observert), hva er sannsynligheten for at det ukjente jordskjelvet var av middels styrke? Poenget er å skrive om uttrykket for betinget sannsynlighet slik at bare kjente størrelser inngår.

$$\begin{aligned}
P(M|D > 1.0) &= \frac{P(D > 1.0 \cap M)}{P(D > 1.0)} = \frac{P(D > 1.0|M)P(M)}{P(D > 1.0)} \\
&= \frac{0.05 \cdot 0.20}{0.0218} = 0.459.
\end{aligned}$$

c)

Definerer hendelsene A: bygning A svikter, B: bygning B svikter.

$$\begin{aligned}
P(A \cap B) &= P(A \cap B \cap K) + P(A \cap B \cap M) + P(A \cap B \cap S) \\
&= P(B|A \cap K)P(A|K)P(K) + P(B|A \cap M)P(A|M)P(M) + P(B|A \cap S)P(A|S)P(S) \\
&= 0.5 \cdot 0.2 \cdot 0.02 + 0.15 \cdot 0.05 \cdot 0.2 + 0.02 \cdot 0.01 \cdot 0.78 \\
&= 0.002 + 0.0015 + 0.000156 = 0.0037
\end{aligned}$$

$$\begin{aligned}
P(K'|A \cap B') &= 1 - P(K|A \cap B') \\
P(K|A \cap B') &= \frac{P(A \cap B' \cap K)}{P(A \cap B')} = \frac{P(B'|A \cap K)P(A \cap K)}{P(A) - P(A \cap B)} \\
&= \frac{(1 - P(B|A \cap K))\,P(A \cap K)}{P(A) - P(A \cap B)} \\
&= \frac{0.5 \cdot 0.2 \cdot 0.02}{0.0218 - 0.0037} = \frac{0.002}{0.0181} = 0.11 \\
P(K'|A \cap B') &= 1 - 0.11 = 0.89
\end{aligned}$$

Oppgave 2

a)

X = absolutt styrke av vilårlig valgt jordskjelv. Har at $X \sim N(x; 4.2, 0.4)$. Har

$$\begin{aligned}
P(X > 5.4) &= P(\frac{X - 4.2}{0.4} > \frac{5.4 - 4.2}{0.4}) \\
&= 1 - \Phi(3) \\
&= 1 - 0.9987 \\
&= 0.0013.
\end{aligned}$$

Kaller vår grenseverdi k. Da er

$$\begin{gathered}
P(X > k) = 0.05 \\
\Updownarrow \\
P(\frac{X - 4.2}{0.4} > \frac{k - 4.2}{0.4}) = 0.05 \\
\Updownarrow \\
\frac{k - 4.2}{0.4} = 1.645 \\
\Updownarrow \\
\begin{aligned}
k &= 4.2 + 1.645 \cdot 0.4 \\
&= 4.858 \\
&\approx 4.9.
\end{aligned}
\end{gathered}$$

b)
Vi har følgende hypotese:

$$H_0 : \mu = 4.2$$
$$H_1 : \mu > 4.2.$$

Under H_0 er $\frac{\bar{X}-4.2}{\frac{s}{\sqrt{9}}} \sim$ t-fordelt med 8 frihetsgrader. Forkast dersom $\frac{\bar{X}-4.2}{\frac{s}{\sqrt{9}}} \geq t_{0.05,8} = 1.86$. Setter vi inn får vi $\frac{3.967-4.2}{\frac{s}{\sqrt{9}}} = -1.07 < 1.86$. Konklusjonen blir at man ikke forkaster hypotesen.

c)
X = antall jordskjelv i et tidsrom $[0, t]$, der t er gitt i år. $t = 1$. Da er

$$\begin{aligned} P(X = 8|\lambda = 5) &= \frac{(5 \cdot 1)^8}{8!} \cdot \exp(-5 \cdot 1) \\ &= 0.0653. \end{aligned}$$

$t = 0.5$. Da er $\lambda t = 2.5$, og

$$\begin{aligned} P(X \geq 5|\lambda t = 2.5) &= 1 - P(X \leq 4|\lambda t = 2.5) \\ &= 1 - 0.8912 \\ &= 0.1088. \end{aligned}$$

Ser på ekstremverdiene. Har da

$$\begin{aligned} P((\max X_i)_{i=1,2,\ldots,10} > 10) &= 1 - P((\max X_i)_{i=1,2,\ldots,10} \leq 10) \\ &= 1 - P(X_i \leq 10)^{10} \\ &= 1 - 0.9863^{10} \\ &= 1 - 0.871 \\ &= 0.129. \end{aligned}$$

d)
T = tid det tar fra å registrere et jordskjelv til første jordskjelv intreffer.

$$P(T>t) = \begin{cases} P(X=0) \quad \text{i intervallet [0,t]} & ,t>0 \\ 1 & ,t=0 \end{cases}$$
$$= \begin{cases} \frac{(\lambda t)^0 \cdot \exp(-\lambda t)}{0!} \quad \text{i intervallet [0,t]} & ,t>0 \\ 1 & ,t=0 \end{cases}$$
$$= \begin{cases} \exp(-\lambda t) \quad \text{i intervallet [0,t]} & ,t>0 \\ 1 & ,t=0 \end{cases}$$
$$\Updownarrow$$
$$F_T(t) = P(T \leq t)$$
$$= 1 - P(T>t)$$
$$= \begin{cases} 1-\exp(-\lambda t) & ,t \geq 0 \\ 0 & ,t<0. \end{cases}$$
$$\Updownarrow$$
$$f(t) = \frac{dF_T(t)}{dt}$$
$$= \begin{cases} \lambda \exp(-\lambda t) & ,t \geq 0 \\ 0 & ,t<0. \end{cases}$$

Ser da at T er eksponensialfordelt med parameter $\frac{1}{\lambda}$.
La X= 2λT. Har da

$$P(X \leq x) = P(2\lambda T \leq x)$$
$$= P(T \leq \frac{x}{2\lambda})$$
$$= 1 - \exp(-\frac{x}{2}), x \geq 0$$
$$\Updownarrow$$
$$f_X(x) = \frac{1}{2}\exp(-\frac{x}{2}), x \geq 0.$$

For n=2 får vi for tettheten i kjikvadratfordelinga

$$f_Y(y) = \frac{1}{2^{\frac{2}{2}}\Gamma(\frac{2}{2})} y^{\frac{2}{2}-1}\exp(-\frac{y}{2}) \qquad ,y \geq 0$$
$$= \frac{1}{2}\exp(-\frac{y}{2}) \qquad ,y \geq 0.$$

e)

Har $f_T(t) = \lambda \exp(-\lambda t), t \geq 0$. Lager en sannsynlighetsmaksimeringsfunksjon

$$\begin{aligned}
L(t_1, \ldots, t_n; \lambda) &= \Pi_{i=1}^{n} \lambda \exp(-\lambda t_i) \\
&= \lambda^n \exp(-\lambda \sum_{i=1}^{n} t_i) \\
\ln(L(t_1, \ldots, t_n); \lambda) &= n \ln(\lambda) - \lambda \sum_{i=1}^{n} t_i \\
\frac{dln(L)}{d\lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} t_i.
\end{aligned}$$

Setter $dln(L)/d\lambda = 0$. Får da

$$\begin{aligned}
\lambda &= \frac{n}{\sum_{i=1}^{n} t_i} \\
&= \frac{1}{\bar{t}}.
\end{aligned}$$

For $\lambda = \hat{\lambda}$;

$$\hat{\lambda} = \frac{1}{\bar{T}}.$$

Ser på $2\lambda n/\hat{\lambda}$. Vi har

$$\begin{aligned}
\frac{2\lambda n}{\hat{\lambda}} &= \frac{2\lambda n}{\frac{n}{\sum_{i=1}^{n} T_i}} \\
&= 2\lambda \sum_{i=1}^{n} T_i,
\end{aligned}$$

som er en χ^2-fordelt med $2n$ frihetsgrader (summen av n uavhengige χ^2-fordelte variable, hver med 2 frihetsgrader). Vi får

$$\begin{aligned}
P(\chi^2_{0.975,2n} < 2\lambda \sum_{i=1}^{n} T_i < \chi^2_{0.025,2n}) = 0.95 \\
P(\frac{\chi^2_{0.975,2n}}{2\sum_{i=1}^{n} T_i} < \lambda < \frac{\chi^2_{0.025,2n}}{2\sum_{i=1}^{n} T_i}) = 0.95.
\end{aligned}$$

For $n = 10$ og $\bar{t} = 1/4$, får vi $2\sum_{i=1}^{n} t_i = 2 \cdot 10 \cdot \frac{1}{4} = 5$, og intervallet blir $(\frac{9.591}{5}, \frac{34.17}{5}) = (1.92, 6.83)$.
Et konfidensitervall inneholder alle H_0-hypoteser som ikke blir forkasta I tosidig test, slik at vi ikke kan konkudere med at $\lambda \neq 5$ på 5% nivå. Har relativt god margin, slik at det er lite frunnlag for å konkudere med at $\lambda \neq 5$.

Oppgave 3

Modellen til laboratoriet er $Y = \alpha + \beta x + \epsilon$, hvor ϵ og dermed også Y er stokastiske variable, mens x er en variabel som laboranten har kontroll på. Parametrene α og β er ukjente og skal estimeres.

Minste kvadratsumsestimatorene A og B for hhv. α og β er

$$\begin{aligned} A &= \frac{1}{n}\sum_{i=1}^{n} Y_i - B\bar{x} = \bar{Y} - B\bar{x} \\ B &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}. \end{aligned}$$

a)

For å vise at estimatorene er forventningsrette, undersøker vi forventingsverdien;

$$\begin{aligned} E[B] &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\alpha+\beta x_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\alpha\sum_{i=1}^{n}(x_i-\bar{x}) + \beta\sum_{i=1}^{n}(x_i-\bar{x})x_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (1) \\ &= \frac{\beta\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \beta. \\ E[A] &= \frac{1}{n}\sum_{i=1}^{n}(\alpha+\beta x_i) - \beta\bar{x} = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha. \\ \mathrm{Var}[B] &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2\,\mathrm{Var}[Y_i]}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2} = \frac{\sigma^2\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}. \\ \mathrm{Var}[A] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[Y_i] + \mathrm{Var}[B]\bar{x}^2 \\ &= \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] \\ &= \frac{\frac{\sigma^2}{n}\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}. \end{aligned}$$

I beregningene har vi benyttet at $\sum_{i=1}^{n}(x_i-\bar{x}) = (\sum_{i=1}^{n}x_i) - n\bar{x} = 0$. Denne summen kan trekkes fra og legges til for å få uttrykket enklest mulig. (Dette kan kreve litt erfaring.) Alternativt kan en gjøre beregningen mer detaljert ved å utvide kvadratene

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n\bar{x}^2 = \sum_{i=1}^{n}x_i^2 - n\bar{x}^2,$$

og deretter forkorte der det er mulig. I beregningen av variansen til A har vi brukt opplysningen om at kovariansen mellom $\bar{Y}$ og B er null, ellers måtte kovariansen trekkes fra i uttrykket.

Både A og B blir normalfordelte, da de er lineærkombinasjoner av normalfordelte variable.

b)

Her undersøkes variansen til $Y_0 - \widehat{Y}_0$, for $\widehat{Y}_0 = A + Bx_0 = \bar{Y} - B\bar{x} + Bx_0$.

$$\begin{aligned}
\mathrm{Var}[Y_0 - \hat{Y}_0] &= \mathrm{Var}[Y_0] + \mathrm{Var}[\hat{Y}_0] \\
&= \sigma^2 + \mathrm{Var}[A + Bx_0] = \sigma^2 + \mathrm{Var}[\bar{Y} - B\bar{x} + Bx_0] \\
&= \mathrm{Var}[\bar{Y}] + \mathrm{Var}[B](x_0 - \bar{x})^2 + \sigma^2 \\
&= \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\} \\
&= \sigma^2 \left\{ 1 + \frac{1}{10} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_{i=1}^{10} - \bar{x})^2} \right\}.
\end{aligned}$$

Dette uttrykket er minimert for $\bar{x} = x_0$, som gir at serie 1 bør brukes selv om serie 2 gir minst varians for B. (Se uttrykket for variansen til B i oppgave a).) Variansen til A blir den samme for de to måleseriene.