Løsningsforslag

Oppgave 1

a) $\frac{X_i-\mu_0}{\sigma}, i = 1, \ldots, 5$ er uavhengige og standard $N(0,1)$ fordelte. Da er $\frac{(X_i-\mu_0)^2}{\sigma^2}, i = 1, \ldots, 5$ uavhengige og χ^2-fordelte med en frihetsgrad. Da er summen χ^2-fordelt med antall frihetsgrader lik summen av antall frihetsgrader til hvert av leddene. Altså er $\frac{1}{\sigma^2}\sum_{i=1}^5 (X_i - \mu_0)^2 \sim \chi^2$-fordelt med 5 frihetsgrader.

Da vet vi at $E[\frac{1}{\sigma^2}\sum_{i=1}^5 (X_i - \mu_0)^2] = 5$, dvs antall frihetsgrader, eller $E[\sum_{i=1}^5 (X_i - \mu_0)^2] = 5\sigma^2$ slik at $\frac{1}{5}\sum_{i=1}^5 (X_i - \mu_0)^2$ er en forventingsrett estimator for σ^2. Variansen til en χ^2-fordelt variabel er to ganger antall frihetsgrader, slik at

$$Var(\frac{1}{5}\sum_{i=1}^5 (X_i - \mu_0)^2) = Var(\frac{\sigma^2}{5}\sum_{i=1}^5 (X_i - \mu_0)^2)/\sigma^2) = \frac{\sigma^4}{25} \times 2 \times 5 = \frac{2\sigma^4}{5}$$

b) Hvis $\sigma^2 > \sigma_0^2$, kan en forvente store verdier av $\sum_{i=1}^5 (X_i - \mu_0)^2$. Det er derfor rimelig å forkaste for store verdier av denne observatoren, dvs forkastningsområdet har formen $\{\sum_{i=1}^5 (X_i - \mu_0)^2] > k\}$.

Kravet om nivå α betyr at $\alpha = P(forkasteH_0|H_0riktig) = P(\sum_{i=1}^5 (X_i - \mu_0)^2] > k|\sigma^2 = \sigma_0^2) = P(\sum_{i=1}^5 (X_i - \mu_0)^2]/\sigma_0^2 > k/\sigma_0^2|\sigma^2 = \sigma_0^2)$ Derfor er $k/\sigma_0^2 = \chi^2_{1-\alpha,5}$ der $\chi^2_{1-\alpha,5}$ er $1-\alpha$ kvantilen i χ^2-fordelingen med 5 frihetsgrader.

En test med nivå α en blir derfor: Forkast H_0 når $\frac{1}{\sigma_0^2}\sum_{i=1}^5 (X_i - \mu_0)^2 > \chi^2_{1-\alpha,5}$.

c) Med tallene i oppgave er $\sum_{i=1}^5 (x_i - 100)^2 = 19.4$. Siden $\sigma_0^2 = 2$ er p-verdien $P(\chi_5^2 > 19.4/2 = 9.7)$ der χ_5^2 er en χ^2-fordelt tilfeldig variabel med 5 frihetsgrader.

I dette tilfellet er $P(\chi_5^2 \leq 9.24) = 0.9$ og $P(\chi_5^2 \leq 11.07) = 0.95$, slik at p-verdien ligger mellom 0.05 og 0.1.

Ovenfor har vi beregnet sannsynligheten for at en χ_5^2-fordelt variabel skal være større en den realiserte verdien av testobservatoren, som er en måte å definere p-verdier . En alternativ ekvivalent definisjon fås ved å ta utgangspunkt i at en test med nivå α gir forkastning hvis $9.7 > \chi^2_{1-\alpha,5}$. Høre side i ulikheten er avtagende i α, slik at p-verdien er minste signifikansnivå som gjør at man forkaster nullhypotesen.

Oppgave 2

a) Antagelser:

i) $x_1, \ldots, x_{12}$ er faste gitte tall

ii) $E[e_i] = 0, i = 1, \ldots, 12$ slik at sammenhengen mellom $E[Y]$ og x er lineær.

iii) $Var(e_i) = \sigma^2, i = 1, \ldots, 12$, dvs uavhengig av $x_1, \ldots, x_{12}$.

iv) $e_i, i = 1, \ldots, 12$ er uavhengige.

Dessuten antas ofte

v) $e_i \sim N(0, \sigma^2), i = 1, \ldots, 12$.

Egenskaper: Fra antagelsene i)-iv) kan det vises at estimatorene $\hat{\beta}_0$ og $\hat{\beta}_1$ er forventningsrette og at $Var(\hat{\beta}_0)$, $Var(\hat{\beta}_1$ og $Cov(\hat{\beta}_0, \hat{\beta}_1)$ kan utrykkes ved σ^2 og verdiene av den uavhengige variabelen, $x_1, \ldots, x_{12}$.

b)

$$\hat{\beta}_1 = \frac{\sum y_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{\sum y_i x_i - (\sum y_i)(\sum x_i)/12}{\sum x_i^2 - (\sum x_i)(\sum x_i)/12} =$$
$$\frac{381.526 - 16.078 \times 280.869/12}{21.873 - 16.078 \times 16.078/12} = \frac{5.208}{0.331} = 15.728$$

Konfidensintervall: Med konfidenskoiffisient $1 - \alpha$ har konfidensintervallet grensene

$$\hat{\beta}_1 \pm t_{1-\alpha/2,n-2} s / \sqrt{\sum(x_i - \bar{x})^2}$$

der $n = 12$ og $t_{1-\alpha/2,n-2}$ er $1 - \alpha/2$-fraktilen i t-fordelingen med $n - 2$ frihetsgrader og $s^2 = RSS/(n - 2)$.

Siden $t_{0.975,10} = 2.23$,har 95%-konfidensintervallet grensene
$15.528 \pm 2.23 * \sqrt(41.645/10)/\sqrt((21.873 - 16.078 * 16.078/12))$ dvs. $(7.82, 23.64)$

c)

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_{11} \\ Y_{12} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_1 & z_1 \\ x_2 & z_2 \\ x_3 & z_3 \\ \vdots & \vdots \\ x_{11} & z_{11} \\ x_{12} & z_{12} \end{pmatrix} = \begin{pmatrix} 1.44 & 8 \\ 1.50 & 10 \\ 1.24 & 6 \\ \vdots & \vdots \\ 1.55 & 12 \\ 1.45 & 9 \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{11} \\ e_{12} \end{pmatrix}$$

og

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

d)

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 21.873 & 144.247 \\ 144.247 & 976.0 \end{pmatrix},$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} 1.804 & -0.267 \\ -0.267 & 0.040 \end{pmatrix},$$

$$(\mathbf{X}^T\mathbf{y}) = \begin{pmatrix} 381.526 \\ 2534.908 \end{pmatrix}$$

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} 1.804 & -0.267 \\ -0.267 & 0.040 \end{pmatrix} \begin{pmatrix} 381.526 \\ 2534.908 \end{pmatrix} = \begin{pmatrix} -- \\ -0.471 \end{pmatrix}$$

Estimert standardfeil: $\sqrt{28.3/10} \times \sqrt{0.04} = 0.336$

Oppgave 3

a) Se Rice for utledning:
ME:$\tilde{\lambda} = \sum_i X_i/n$
SME:$\hat{\lambda} = \sum_i X_i/n$
Numerisk:$\hat{\lambda} = (0 \times 9 + 1 \times 7 + 2 \times 3 + 3 \times 1)/20 = 16/20 = 0.8$

b) Se notat om konfidensintervall for utledning: Tilnærmet fordeling til $\hat{\lambda}$ for store verdier av n, dvs $n\lambda > 10$,

$$\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{\hat{\lambda}}} \text{ tilnærmet } N(0,1)$$

90% konfidensintervall har grenser: $\hat{\lambda} \pm 1.96\sqrt{\hat{\lambda}/n}$
Numerisk: $0.8 \pm 1.96\sqrt{0.8/20} = 0.8 \pm 1.96 \times 0.2 = 0.8 \pm 0.39$.

c) La $Y_i = 1$ når $X_i = 0$ og $Y_i = 0$ når $X_i = 1, 2, \cdots, 20$. Dette definerer $n = 20$ Bernoulli forsøk: (i) To kjennetegn, $Y_i = 1$ eller $Y_i = 0$, (ii) uavhengighet siden $X_1, \cdots, X_n$ er uavhengige, (iii) konstant suksessannnsynlighet $P(Y_i = 1) = P(X_i = 0) = \exp(-\lambda)$. Siden antallet ulykkesfrie måneder, Y, er antallet suksesser, $Y_i = 1$, er Y binomisk fordelt

Punksannsynlighet :

$$P(Y_i = y) = \begin{pmatrix} n \\ y \end{pmatrix} (\exp(-\lambda))^y (1 - \exp(-\lambda))^{n-y}, \quad y = 0, \ldots, n$$

Momentestimatoren finnes fra ligningen $\frac{\sum_{i=1}^n Y_i}{n} = \exp(-\lambda)$, dvs. $\tilde{\tilde{\lambda}} = -\log(Y/n)$.
Numerisk: $\tilde{\tilde{\lambda}} = -\log(9/20) = 0.7985 \approx 0.8$. Log likelihood:

$$l(\lambda) = \log(\begin{pmatrix} n \\ y \end{pmatrix}) - y\lambda + (n - y)\log(1 - \exp(-\lambda))$$

SME finnes fra ligningen

$$y + (n - y)\frac{1}{1 - \exp(-\lambda)}(-\exp(-\lambda)) = 0, \; y = n\exp(-\lambda)$$

som gir estimatoren $\hat{\hat{\lambda}} = -\log(Y/n)$.
Numerisk: $\hat{\hat{\lambda}} = -\log(9/20) == 0.7985 \approx 0.8$.

d) Fra sentralgrenseteoremet vil for store verdier av n

$$\frac{\sqrt{n}(Y/n - \exp(-\lambda))}{\sqrt{\exp(-\lambda)(1 - \exp(-\lambda))}} \text{ være tilnærmet } N(0,1)$$

fordelt. Da er også

$$\frac{\sqrt{n}(Y/n - \exp(-\lambda))}{\sqrt{\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))}} \text{ tilnærmet } N(0,1)$$

fordelt slik at

$$\begin{aligned}
0.9 &\approx P(-1.96 < \frac{\sqrt{n}(Y/n - \exp(-\lambda))}{\sqrt{\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))}} < 1.96) \\
&= P(Y/n - 1.96\sqrt{\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))}/\sqrt{n} < \exp(-\lambda) \\
&< Y/n + 1.96\sqrt{\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))}/\sqrt{n}) \\
&= P(-\log(Y/n + 1.96\sqrt{\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))}/\sqrt{n}) < \lambda \\
&< -\log(Y/n - 1.96\sqrt{\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))}/\sqrt{n}))
\end{aligned}$$

Numerisk: Det tilnærmede konfidensintervallet blir derfor $(-\log(0.45{+}1.96{\times}\sqrt{0.45 \times 0.55}/\sqrt{20}), -\log(0.45{-}1.96{\times}\sqrt{0.45 \times 0.55}/\sqrt{20}) = (0.40, 1.46)$.