Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen mai 2003

Oppgave 1

a)

$$\mathrm{P}(X \leq 120) = \mathrm{P}(\frac{X-115}{5} \leq \frac{120-115}{5}) = \Phi(1) = 0.841.$$

$$\begin{aligned}\mathrm{P}(120 < X \leq 125) &= \mathrm{P}(X \leq 125) - \mathrm{P}(X \leq 120) \\ &= P(\frac{X-115}{5} \leq \frac{125-115}{5}) - \mathrm{P}(X \leq 120) \\ &= \Phi(2) - 0.841 \;=\; 0.977 - 0.841 \;=\; 0.136.\end{aligned}$$

A og B er disjunkte siden de to hendelsene ikke kan inntreffe samtidig.

$\mathrm{P}(A \mid B) = \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \dfrac{\mathrm{P}(\emptyset)}{\mathrm{P}(B)} = \dfrac{0}{\mathrm{P}(B)} = 0.$
Siden $\mathrm{P}(A) = 0.841 \neq \mathrm{P}(A \mid B)$ er A og B ikke uavhengige.

b)

Sjekker kriteriene for Bernoulli-forsøk:

- Vi undersøker navnet til n jenter, dvs. gjør n forsøk
- I hvert forsøk er navnet enten Maud eller ikke Maud
- Sannsynligheten for at navnet er Maud er $p = 0.02$ og lik i alle forsøk
- Det er rimelig å anta at forsøkene er uavhengige

Dermed er de fire kriteriene for et Bernoulli-forsøk oppfylt, og det er rimelig å anta at Z er binomisk fordelt.

$$\mathrm{P}(C) = \mathrm{P}(Z \geq 1) = 1 - \mathrm{P}(Z = 0) = 1 - \binom{15}{0} 0.02^0 (1-0.02)^{15} = 1 - 0.739 = 0.261$$

$$\begin{aligned}\mathrm{P}(D \mid C) &= \mathrm{P}(Z = 2 \mid Z \geq 1) \\ &= \frac{\mathrm{P}(Z = 2 \cap Z \geq 1)}{\mathrm{P}(Z \geq 1)} = \frac{\mathrm{P}(Z=2)}{\mathrm{P}(Z \geq 1)} = \frac{\binom{15}{2} 0.02^2 (1-0.02)^{13}}{\mathrm{P}(Z \geq 1)} = \frac{0.032}{0.261} = 0.123.\end{aligned}$$

Definer hendelsene

M: navnet på tilfeldig valgt elev er Maud
J: tilfeldig valgt elev er jente
G: tilfeldig valgt elev er gutt

Bruker lov om total sannsynlighet:

$$\mathrm{P}(M) = \mathrm{P}(M \cap J) + \mathrm{P}(M \cap G) = \mathrm{P}(M \mid J)\mathrm{P}(J) + \mathrm{P}(M \mid G)\mathrm{P}(G) = 0.02 \cdot \frac{15}{25} + 0 \cdot \frac{15}{25} = 0.012.$$

Oppgave 2

a)

Det er mest rimelig med en venstresidig hypotesetest:

$$\begin{aligned} H_0: &\quad \mu = 16, \\ H_1: &\quad \mu < 16. \end{aligned}$$

Begrunnelse: forhandleren sier at bilen kan forventes å kjøre **minst** 16 km pr liter. Vi vil avsløre ev. feil i markedsføringen.

NB: Hypotetesetesten skal være uavhengig av målingene. En bør altså ikke velge alternativ hypotese på grunnlag av $\overline{x}$.

$\overline{X}$ er normalfordelt med forventning μ og varians σ^2/n. Variansen er ukjent, derfor kreves T-fordeling med $\nu = n-1 = 19$ frihetsgrader. Gjennomfører testen med $\alpha = 0.05$. Testobservator: $T_{\text{obs}} = \frac{\overline{X}-\mu}{S/\sqrt{n}}$. Observert verdi:

$$t_{\text{obs}} = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{15.56 - 16}{0.94/\sqrt{20}} = -2.093.$$

Fra tabell over kvantiler i T-fordelingen; $-t_{0.05,19} = -1.729$. Altså: $t < -t_{0.05,19}$, dermed skal H_0 forkastes.

Hvis vi hadde valgt å bruke en normalfordelingshypotese, ville kvantilen $-z_{0.05} = -1.645$ gitt samme konklusjon. Imidlertid bør vi da argumentere for at avstanden til denne kvantilen er så stor at høyere varians i T-fordelingen ikke ville påvirket resultatet. Å sammenlikne med denne kvantilen kan ikke regnes som fullgodt svar.

b)

P-verdien finnes ved å lete opp verdien på testobservatoren fra a) i tabell. For T-fordeling med $\nu = 19$, finner vi $t_{0.025,19} = 2.093$. Ettersom T-fordelingen er symmetrisk, har vi at $\mathrm{P}(T > t_{\alpha,\nu}) = \mathrm{P}(T < -t_{\alpha,\nu})$. Dermed; $p = \alpha = 0.025 = 2.5\%$.

Testobservatoren er normalfordelt hvis $\sigma = s$. Dette bør være tilnærmet oppfylt for å bruke normalfordeling. Hvis en ikke har ekstra informasjon om σ, er det ikke anbefalt å tilnærme student-fordelingen med en normalfordeling når $n < 30$, da s ikke er et godt nok estimat.

Under normalfordelingen får vi p-verdi

$$\mathrm{P}(z \leq -2.09) = \mathrm{P}(z \geq 2.09) = 1 - \mathrm{P}(z < 2.09) = 1 - \Phi(2.09) = 1 - 0.9817 = 0.0183 = 1.8\%.$$

c)

Antar $H_1' : \mu = \mu_1 = 15.5$ og $\sigma = s$. Teststyrken er sannsynligheten for å forkaste H_0 under H_1', dvs

$$\mathrm{P}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < -1.645 \mid \mu = \mu_1 = 15.5\right).$$

For å få en normalfordelt variabel, flytter vi alt utenom $\overline{X}$, som er stokastisk, over på høyre side. Deretter trekker vi fra sann forventningsverdi μ_1 og dividerer med standardavviket på begge sider.

$$\begin{aligned} &\mathrm{P}(\overline{X} < -1.645 \cdot \sigma/\sqrt{n} + \mu_0) \\ &= \mathrm{P}(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} < \frac{-1.645 \cdot \sigma/\sqrt{n} + \mu_0 - \mu_1}{\sigma/\sqrt{n}}) \\ &= \mathrm{P}(Z < 0.7338) \approx 0.767. \end{aligned}$$

Hvis vi ikke kunne bruke normalfordelingsantakelsen, ville teststyrken blitt svakere. Her er det forutsatt at vi er ganske sikre på variansen, f.eks. på grunnlag av data fra produsent.

Generelt må antall observasjoner økes for å oppnå økt teststyrke. (Dette er fullgodt svar.) Mulig tillegg: Hvis en har mulighet til å gjennomføre forsøket på en måte slik at variansen blir mindre, f.eks. kjøre bilene under mer kontrollerte former i et laboratorium, ville også teststyrken økes. Eventuelt kan en øke signifikansnivået α f.eks. til 0.1, og dermed øke teststyrken, men dette er sjelden aktuelt i praksis.

Oppgave 3

a)

Kumulativ fordelingsfunksjon finnes ved å integrere sannsynlighetstettheten.

$$F(t) = \int_{-\infty}^{t} f(\tau)\mathrm{d}\tau = \int_{0}^{t} \frac{1}{\beta} e^{-\tau/\beta}\mathrm{d}\tau = 1 - e^{-t/\beta},$$

for $t \geq 0$, og $F(t) = 0$ for $t < 0$.

For $\beta = 5$:

$\mathrm{P}(T_1 < 3) = F(3) = 1 - e^{-3/5} = 0.4512.$

$\mathrm{P}(2 < T_1 < 4) = P(T_1 < 4) - P(T_1 \leq 2) = F(4) - F(2) = e^{-2/5} - e^{-4/5} = 0.2210.$

b)

Apparatet fungerer bare hvis k uavhengige komponenter fungerer. Altså er levetiden X lik levetiden til komponenten som bryter sammen først. Dvs.

$$X = \mathrm{MIN}(T_1, T_2, T_3, \ldots, T_k).$$

Dermed er

$$\begin{aligned} \mathrm{P}(X > x) &= \mathrm{P}(T_1 > x \cap T_2 > x \cap T_3 > x \ldots \cap T_k > x) \\ &= \mathrm{P}(T_1 > x)\mathrm{P}(T_2 > x)\mathrm{P}(T_3 > x) \cdots \mathrm{P}(T_k > x). \end{aligned}$$

Den siste likheten kommer av at $T_1, T_2, \ldots, T_k$ er uavhengige.

Videre, $\mathrm{P}(T_j > x) = 1 - F(x) = e^{-x/\beta}$ for alle $j = 1, 2, 3, \ldots, k$. Dermed har vi

$$\mathrm{P}(X > x) = \left(e^{-x/\beta}\right)^k = e^{-kx/\beta} = e^{-\frac{x}{\beta/k}}.$$

Dette tilsvarer en kumulativ eksponensialfordeling med parameter β/k, dvs

$$F_X(x) = 1 - e^{-\frac{x}{\beta/k}},\ x \geq 0.$$

Forventningsverdien til en eksponensialfordeling med parametrisering som gitt i formelsamlingen, er parameteren selv. Altså er forventningsverdien til X lik β/k_i. Med $k = 4$ og $\beta = 5$, er apparatets forventede levealder $5/4 = 1.25$. (Tidsenhet ikke oppgitt i oppgaven.)

c)

Forventningsverdi $\widehat{\beta}$:

$$\begin{aligned}
\mathrm{E}(\widehat{\beta}) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i k_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) k_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\beta}{k_i} k_i = \beta.
\end{aligned}$$

Forventningsverdi $\widetilde{\beta}$:

$$\begin{aligned}
\mathrm{E}(\widetilde{\beta}) &= \mathrm{E}\left(\frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} k_i^{-1}}\right) \\
&= \frac{\sum_{i=1}^{n} \mathrm{E}(X_i)}{\sum_{i=1}^{n} k_i^{-1}} \\
&= \frac{\sum_{i=1}^{n} \frac{\beta}{k_i}}{\sum_{i=1}^{n} k_i^{-1}} = \beta.
\end{aligned}$$

Dermed er begge estimatorene forventningsrette. Beregner variansene:

Varians til $\widehat{\beta}$:

$$\begin{aligned}
\mathrm{Var}(\widehat{\beta}) &= \frac{1}{n^2}\sum_{i=1}^{n} k_i^2 \mathrm{Var}(X_i) \quad \text{(pga. uavhengighet)} \\
&= \frac{1}{n^2}\sum_{i=1}^{n} k_i^2 \left(\frac{\beta}{k_i}\right)^2 \\
&= \frac{1}{n}\beta^2.
\end{aligned}$$

Varians til $\widetilde{\beta}$:

$$\begin{aligned}
\mathrm{Var}(\widetilde{\beta}) &= \frac{\sum_{i=1}^{n} \mathrm{Var}(X_i)}{\left(\sum_{i=1}^{n} k_i^{-1}\right)^2} \\
&= \frac{\sum_{i=1}^{n} \left(\frac{\beta}{k_i}\right)^2}{\left(\sum_{i=1}^{n} k_i^{-1}\right)^2} \\
&= \beta^2 \frac{\sum_{i=1}^{n} \left(k_i^{-1}\right)^2}{\left(\sum_{i=1}^{n} k_i^{-1}\right)^2}.
\end{aligned}$$

d)

Sannsynlighetsmaksimeringsestimator (SME):
Begynner med likelihood-funksjonen

$$\mathrm{L}(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} f_X(X_i) = \prod_{i=1}^{n} \left(\frac{k_i}{\beta}\right) e^{-k_i X_i/\beta} = \frac{1}{\beta^n} e^{-\sum_{i=1}^{n} X_i k_i/\beta} \prod_{i=1}^{n} k_i.$$

Tar logaritmen til denne og deriverer mhp β.

$$\Lambda = \log(\mathrm{L}) = -n\log(\beta) - \beta^{-1} \sum_{i=1}^{n} X_i k_i + \log(\prod_{i=1}^{n} k_i).$$

$$\frac{\mathrm{d}\Lambda}{\mathrm{d}\beta} = -\frac{n}{\beta} + \beta^{-2} \sum_{i=1}^{n} X_i k_i.$$

Setter til slutt den deriverte lik null for $\beta = \overline{\beta}$ og løser ut for $\overline{\beta}$.

$$-n\overline{\beta}^{-1} + \overline{\beta}^{-2} \sum_{i=1}^{n} X_i k_i = 0$$

$$\overline{\beta} = \frac{1}{n} \sum_{i=1}^{n} X_i k_i = \widehat{\beta}.$$

Så sjekker vi påstanden at $\widehat{\beta}$ har minst varians, ved å sette $r_i = 1/k_i$ som foreslått i oppgaveteksten.

$$\mathrm{Var}(\widehat{\beta}) = \frac{1}{n}\beta^2.$$

$$\mathrm{Var}(\widetilde{\beta}) = \frac{\sum_{i=1}^{n} r_i^2}{\left(\sum_{i=1}^{n} r_i\right)^2}\beta^2.$$

SME har minst varians hvis $\frac{1}{n} \leq \frac{\sum_{i=1}^{n} r_i^2}{\left(\sum_{i=1}^{n} r_i\right)^2}$, dvs.

$$\frac{1}{n}\left(\sum_{i=1}^{n} r_i\right)^2 \leq \sum_{i=1}^{n} r_i^2$$

$$\left(\frac{1}{n}\sum_{i=1}^{n} r_i\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n} r_i^2$$

$$\frac{1}{n}\sum_{i=1}^{n} r_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} r_i\right)^2 \geq 0.$$

Det er oppgitt at nederste linje stemmer, dermed har SME alltid varians mindre enn eller lik variansen til den andre estimatoren.

e)

Setter $Y = 2k_iX_i/\beta$. Vi kan benytte variabelskifte til å sette inn i sannsynlighetstettheten til X_i, som er eksponensialfordelt med parameter β/k_i. La $X_i = w(y) = \frac{\beta}{2k_i}y$. Da er

$$\begin{aligned} f_Y(y) &= w'(y)f_X(w(y)) = \frac{\beta}{2k_i}f_X\left(\frac{\beta y}{2k_i}\right) \\ &= \frac{\beta}{2k_i}\frac{k_i}{\beta}e^{-\frac{\beta y}{2k_i}\frac{k_i}{\beta}} = \frac{1}{2}e^{-y/2}. \end{aligned}$$

Dette er en χ^2-fordeling med 2 frihetsgrader.

Alternativt, sett inn i momentgenererende funksjon for eksponensialfordelingen:

$$\begin{aligned} M_X(t) &= \frac{1}{1-\beta t/k_i},\ t < \frac{1}{\beta/k_i}. \\ M_Y(t) &= M_X\left(\frac{\beta}{2k_i}t\right) = \frac{1}{1-2t},\ t < \frac{1}{2}. \end{aligned}$$

Dette er momentgenererende funksjon for χ^2-fordelingen med $\nu = 2$ frihetsgrader.

Ser deretter at $2n\widehat{\beta}/\beta = \sum_{i=1}^{n} 2k_iX_i/\beta$ er en lineærkombinasjon av χ^2-fordelte variable, som også er χ^2-fordelt. Med to frihetsgrader fra hver X_i, er det totalt $\nu = 2n$ frihetsgrader.

Et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\chi^2 = 2n\widehat{\beta}/\beta$ er $\chi^2_{1-\alpha/2,\nu} < \chi^2 < \chi^2_{\alpha/2,\nu}$, altså

$$\begin{aligned} \chi^2_{1-\alpha/2,\nu} &< 2n\widehat{\beta}/\beta < \chi^2_{\alpha/2,\nu} \\ \frac{\chi^2_{1-\alpha/2,\nu}}{2n\widehat{\beta}} &< \frac{1}{\beta} < \frac{\chi^2_{\alpha/2,\nu}}{2n\widehat{\beta}} \\ \frac{2n\widehat{\beta}}{\chi^2_{\alpha/2,\nu}} &< \beta < \frac{2n\widehat{\beta}}{\chi^2_{1-\alpha/2,\nu}}. \end{aligned}$$

For $n = 8$ og $\alpha = 0.05$, finner vi kvantilene i tabell: $\chi^2_{0.975,16} = 6.908$ og $\chi^2_{0.025,16} = 28.845$. Setter inn dette og $\widehat{\beta} = 8.3$ og får konfidensintervall

$$\begin{aligned} \frac{2\cdot 8\cdot 8.3}{28.845} &< \beta < \frac{2\cdot 8\cdot 8.3}{6.908} \\ 4.604 &< \beta < 19.224. \end{aligned}$$