

Institutt for matematiske fag

Eksamensoppgave i **Løsningsskisse TMA4240/TMA4245**

Faglig kontakt under eksamen: Håkon Tjelmeland

Tlf: 48 22 18 96

Eksamensdato: 10. august 2017

Eksamenstid (fra–til): 09.00-13.00

Hjelpemiddelkode/Tillatte hjelpemidler:

Tabeller og formler i statistikk, Akademika,

K. Rottmann: *Matematisk formelsamling,*

Kalkulator Casio fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College eller HP30S,
Gult stemplet A5-ark med egne håndskrevne notater.

Annen informasjon:

Alle svar skal begrunnes og besvarelsen skal inneholde naturlig mellomregning.

Målform/språk: bokmål

Antall sider: 9

Antall sider vedlegg: 0

Kontrollert av:

Dato

Sign

Oppgave 1

a) Kan bestemme c ut fra kravet

$$\int_{-\infty}^{\infty} f(x; \theta) dx = 1.$$

Siden $f(x; \theta) = 0$ for $x < \theta$ får vi

$$\begin{aligned} \int_{-\infty}^{\infty} f(x; \theta) dx &= \int_{\theta}^{\infty} c \exp\{-(x - \theta)\} dx \\ &= c [-\exp\{-(x - \theta)\}]_{x=\theta}^{\infty} \\ &= c(-0 + e^0) = \underline{\underline{c = 1}}. \end{aligned}$$

Sannsynligheten det spørres etter er

$$\begin{aligned} P(X > \theta + 1) &= \int_{\theta+1}^{\infty} f(x; \theta) dx = \int_{\theta+1}^{\infty} \exp\{-(x - \theta)\} dx \\ &= [-\exp\{-(x - \theta)\}]_{x=\theta+1}^{\infty} \\ &= -0 + \exp\{-(\theta + 1 - \theta)\} = \underline{\underline{e^{-1} = 0.3679}}. \end{aligned}$$

b) For å finne sannsynlighetsmaksimeringsestimatorens starter vi med å finne rimelighetsfunksjonen $L(\theta)$. Siden observasjonene er uavhengige, og ved å huske på at hvilken formel som gjelder for $f(x; \theta)$ avhenger av om $x < \theta$ eller $x \geq \theta$, får vi at

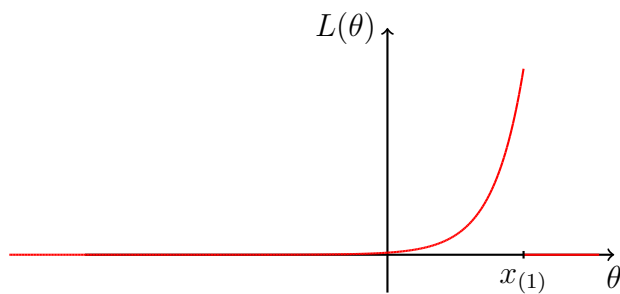
$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(x_i; \theta) = \begin{cases} \prod_{i=1}^n e^{-(x_i - \theta)} & \text{hvis } x_1, \dots, x_n \geq \theta, \\ 0 & \text{ellers} \end{cases} \\ &= \begin{cases} \exp\{-\sum_{i=1}^n x_i + n\theta\} & \text{hvis } \min\{x_1, \dots, x_n\} \geq \theta, \\ 0 & \text{ellers} \end{cases} \\ &= \begin{cases} \exp\{-\sum_{i=1}^n x_i + n\theta\} & \text{hvis } \theta \leq x_{(1)}, \\ 0 & \text{ellers} \end{cases} \end{aligned}$$

der vi har benyttet at alle $x_1, \dots, x_n \geq \theta$ hvis og bare hvis $\min\{x_1, \dots, x_n\} \geq \theta$ og at $x_{(1)} = \min\{x_1, \dots, x_n\}$. Et plott av rimelighetsfunksjonen $L(\theta)$ er vist i figur 1. Vi ser at $L(\theta)$ har sitt maksimum for

$$\theta = x_{(1)} = \min\{x_1, \dots, x_n\}.$$

Sannsynlighetsmaksimeringsestimatorens blir dermed

$$\underline{\underline{\hat{\theta} = X_{(1)} = \min\{X_1, \dots, X_n\}}}.$$



Figur 1: Skisse av rimelighetsfunksjonen $L(\theta)$ som funksjon av θ .

- c) For å finne sannsynlighetstettheten for $W = \min\{X_1, X_2, \dots, X_n\}$ lønner det seg først å finne den tilhørende kumulative fordeling $F_W(w)$,

$$\begin{aligned}
 F_W(w) &= P(W \leq w) = P(\min\{X_1, X_2, \dots, X_n\} \leq w) \\
 &= 1 - P(\min\{X_1, X_2, \dots, X_n\} > w) \\
 &= 1 - P(X_1 > w \cap X_2 > w \cap \dots \cap X_n > w) \\
 &= 1 - P(X_1 > w) \cdot P(X_2 > w) \cdot \dots \cdot P(X_n > w) \\
 &= 1 - (1 - P(X_1 \leq w)) \cdot (1 - P(X_2 \leq w)) \cdot \dots \cdot (1 - P(X_n \leq w)) \\
 &= 1 - (1 - F_X(w)) \cdot (1 - F_X(w)) \cdot \dots \cdot (1 - F_X(w)) \\
 &= 1 - (1 - F_X(w))^n.
 \end{aligned}$$

Kan da finne sannsynlighetstettheten til W ved å derivere denne med hensyn på w , men først finner ved kumulativ fordeling for X . For $x \geq \theta$ får vi

$$F_X(x) = \int_{-\infty}^x f(x) dx = \int_{\theta}^x e^{-(x-\theta)} dx = \left[-e^{-(x-\theta)} \right]_{x=\theta}^x = -e^{-(x-\theta)} + e^0 = 1 - e^{-(x-\theta)},$$

mens $F_X(x) = 0$ for $x < \theta$. Sannsynlighetstettheten til W blir da

$$\begin{aligned}
 f_W(w) &= F'_W(w) = -n (F_X(w))^{n-1} \cdot (-f_X(w)) \\
 &= \begin{cases} n (1 - F_X(w))^{n-1} f_X(w) & \text{for } w \geq \theta, \\ 0 & \text{ellers} \end{cases} \\
 &= \begin{cases} n (1 - 1 + e^{-(w-\theta)})^{n-1} e^{-(w-\theta)} & \text{for } w \geq \theta, \\ 0 & \text{ellers} \end{cases} \\
 &= \underline{\underline{\begin{cases} ne^{-n(w-\theta)} & \text{for } w \geq \theta, \\ 0 & \text{ellers.} \end{cases}}}
 \end{aligned}$$

Sannsynligheten det spørres etter blir da

$$\begin{aligned}
 P(W > \theta + 1) &= \int_{\theta+1}^{\infty} f_W(w)dw = \int_{\theta+1}^{\infty} ne^{-n(w-\theta)}dw \\
 &= \left[-e^{-n(w-\theta)}\right]_{w=\theta+1}^{\infty} \\
 &= -0 + e^{-n(\theta+1-\theta)} = \underline{\underline{e^{-n}}}.
 \end{aligned}$$

Oppgave 2

- a) Finner forventingsverdien til $\hat{\mu}$ ved å bruke regneregler for forventingsverdi, samt at vi vet at $E[X_i] = E[Y_i] = \mu$ for alle i ,

$$\begin{aligned}
 E[\hat{\mu}] &= E\left[\frac{1}{2}(\bar{X} + \bar{Y})\right] = \frac{1}{2}E[\bar{X} + \bar{Y}] \\
 &= \frac{1}{2}(E[\bar{X}] + E[\bar{Y}]) \\
 &= \frac{1}{2}\left(E\left[\frac{1}{7}\sum_{i=1}^7 X_i\right] + E\left[\frac{1}{6}\sum_{i=1}^6 Y_i\right]\right) \\
 &= \frac{1}{2}\left(\frac{1}{7}E\left[\sum_{i=1}^7 X_i\right] + \frac{1}{6}E\left[\sum_{i=1}^6 Y_i\right]\right) \\
 &= \frac{1}{2}\left(\frac{1}{7}\sum_{i=1}^7 E[X_i] + \frac{1}{6}\sum_{i=1}^6 E[Y_i]\right) \\
 &= \frac{1}{2}\left(\frac{1}{7}\sum_{i=1}^7 \mu + \frac{1}{6}\sum_{i=1}^6 \mu\right) \\
 &= \frac{1}{2}\left(\frac{1}{7} \cdot 7\mu + \frac{1}{6} \cdot 6\mu\right) = \mu.
 \end{aligned}$$

Dermed har vi vist at $\hat{\mu}$ er forventingsrett.

Ved å benytte regneregler for varians, at X_i 'ene og Y_i 'ene alle er uavhengige,

og at $\text{Var}[X_i] = \text{Var}[Y_i] = \sigma^2$ for alle i , får vi at variansen til $\hat{\mu}$ blir

$$\begin{aligned}
 \text{Var}[\hat{\mu}] &= \text{Var}\left[\frac{1}{2}(\bar{X} + \bar{Y})\right] = \left(\frac{1}{2}\right)^2 \text{Var}[\bar{X} + \bar{Y}] \\
 &= \frac{1}{4}(\text{Var}[\bar{X}] + \text{Var}[\bar{Y}]) \\
 &= \frac{1}{4}\left(\text{Var}\left[\frac{1}{7}\sum_{i=1}^7 X_i\right] + \text{Var}\left[\frac{1}{6}\sum_{i=1}^6 Y_i\right]\right) \\
 &= \frac{1}{4}\left(\left(\frac{1}{7}\right)^2 \text{Var}\left[\sum_{i=1}^7 X_i\right] + \left(\frac{1}{6}\right)^2 \text{Var}\left[\sum_{i=1}^6 Y_i\right]\right) \\
 &= \frac{1}{4}\left(\frac{1}{49}\sum_{i=1}^7 \text{Var}[X_i] + \frac{1}{36}\sum_{i=1}^6 \text{Var}[Y_i]\right) \\
 &= \frac{1}{4}\left(\frac{1}{49}\sum_{i=1}^7 \sigma^2 + \frac{1}{36}\sum_{i=1}^6 \sigma^2\right) \\
 &= \frac{1}{4}\left(\frac{7\sigma^2}{49} + \frac{6\sigma^2}{36}\right) = \frac{1}{4}\left(\frac{1}{7} + \frac{1}{6}\right)\sigma^2 = \frac{1}{4} \cdot \frac{6+7}{42}\sigma^2 = \underline{\underline{\frac{13}{168}\sigma^2}}.
 \end{aligned}$$

- b) Når $\mu_A = \mu_B$ og $\sigma_A = \sigma_B$ kommer alle $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n$ fra samme populasjon og man får dermed en forventingsrett estimator med mindre varians ved å estimere μ ved gjennomsnittet av disse variablene, dvs

$$\underline{\underline{\mu^* = \frac{1}{13}\left(\sum_{i=1}^7 X_i + \sum_{i=1}^6 Y_i\right)}}.$$

For å sjekke at den er forventingsrett går man tilsvarende som for $\hat{\mu}$ over,

$$\begin{aligned}
 \text{E}[\mu^*] &= \text{E}\left[\frac{1}{13}\left(\sum_{i=1}^7 X_i + \sum_{i=1}^6 Y_i\right)\right] = \frac{1}{13}\text{E}\left[\sum_{i=1}^7 X_i + \sum_{i=1}^6 Y_i\right] \\
 &= \frac{1}{13}\left(\text{E}\left[\sum_{i=1}^7 X_i\right] + \text{E}\left[\sum_{i=1}^6 Y_i\right]\right) \\
 &= \frac{1}{13}\left(\sum_{i=1}^7 \text{E}[X_i] + \sum_{i=1}^6 \text{E}[Y_i]\right) \\
 &= \frac{1}{13}\left(\sum_{i=1}^7 \mu + \sum_{i=1}^6 \mu\right) \\
 &= \frac{1}{13}(7\mu + 6\mu) = \mu.
 \end{aligned}$$

Variansen til μ^* finner man også tilsvarende som for $\hat{\mu}$,

$$\begin{aligned} \text{Var}[\mu^*] &= \text{Var}\left[\frac{1}{13}\left(\sum_{i=1}^7 X_i + \sum_{i=1}^6 Y_i\right)\right] = \left(\frac{1}{13}\right)^2 \text{Var}\left[\sum_{i=1}^7 X_i + \sum_{i=1}^6 Y_i\right] \\ &= \frac{1}{13^2} \left(\text{Var}\left[\sum_{i=1}^7 X_i\right] + \text{Var}\left[\sum_{i=1}^6 Y_i\right]\right) \\ &= \frac{1}{13^2} \left(\sum_{i=1}^7 \text{Var}[X_i] + \sum_{i=1}^6 \text{Var}[Y_i]\right) \\ &= \frac{1}{13^2} \left(\sum_{i=1}^7 \sigma^2 + \sum_{i=1}^6 \sigma^2\right) \\ &= \frac{1}{13^2} \left(7\sigma^2 + 6\sigma^2\right) = \frac{\sigma^2}{13}. \end{aligned}$$

Siden $\text{Var}[\hat{\mu}] = \frac{13}{168}\sigma^2 \approx 0.0774\sigma^2$ og $\text{Var}[\mu^*] = \frac{\sigma^2}{13} \approx 0.0769\sigma^2$ ser vi at μ^* har mindre varians enn $\hat{\mu}$.

- c) For å utlede et konfidensintervall for $\delta = \mu_A - \mu_B$ er det naturlig å ta utgangspunkt i

$$T = \frac{\bar{X} - \bar{Y} - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

som er tilnærmet t -fordelt med

$$\nu = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{\left(\frac{s_A^2}{n_A}\right)^2}{n_A-1} + \frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B-1}}$$

frihetsgrader. Vi har dermed at

$$P\left(-t_{\frac{\alpha}{2}, \nu} \leq \frac{\bar{X} - \bar{Y} - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \leq t_{\frac{\alpha}{2}, \nu}\right) = 1 - \alpha.$$

Må så løse hver av ulikhetene inne i sannsynlighetsuttrykket over med hensyn

på $\delta = \mu_A - \mu_B$. Starter med den venstre ulikheten,

$$\begin{aligned} -t_{\frac{\alpha}{2}, \nu} &\leq \frac{\bar{X} - \bar{Y} - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \\ \Leftrightarrow -t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} &\leq \bar{X} - \bar{Y} - (\mu_A - \mu_B) \\ \Leftrightarrow -\bar{X} - \bar{Y} - t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} &\leq -(\mu_A - \mu_B) \\ \Leftrightarrow \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} &\geq \mu_A - \mu_B \\ \Leftrightarrow \mu_A - \mu_B &\leq \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \end{aligned}$$

Den høyre ulikheten løses tilsvarende,

$$\begin{aligned} \frac{\bar{X} - \bar{Y} - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} &\leq t_{\frac{\alpha}{2}, \nu} \\ \Leftrightarrow \bar{X} - \bar{Y} - (\mu_A - \mu_B) &\leq t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \\ \Leftrightarrow -(\mu_A - \mu_B) &\leq -(\bar{X} - \bar{Y}) + t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \\ \Leftrightarrow \mu_A - \mu_B &\geq \bar{X} - \bar{Y} - t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \\ \Leftrightarrow \bar{X} - \bar{Y} - t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} &\leq \mu_A - \mu_B \end{aligned}$$

Ved å sette de to ulikhetene sammen igjen inne i sannsynlighetsuttrykket får man dermed at

$$P \left(\bar{X} - \bar{Y} - t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \leq \mu_A - \mu_B \leq \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \right) = 1 - \alpha,$$

slik at et $(1 - \alpha) \cdot 100\%$ konfidensintervall for $\delta = \mu_A - \mu_B$ er gitt ved

$$\left[\bar{X} - \bar{Y} - t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}, \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \right].$$

Innsatt de oppgitte tallene får vi ved å benytte $t_{0.05,9} = 1.833$ at intervallet blir

$$\left[15.22 - 14.56 - 1.833\sqrt{\frac{0.32}{7-1} + \frac{0.47}{6-1}}, 15.22 - 14.56 - 1.833\sqrt{\frac{0.32}{7-1} + \frac{0.47}{6-1}} \right]$$

$$= \underline{\underline{[0.3803, 0.9397]}}.$$

- d) Vi ønsker å finne ut om resultatene tyder på at metode A gir bedre utmattingsfasthet enn metode B. Hvis metode A er bedre enn metode B vil $\mu_A > \mu_B$ og vi velger derfor dette som vår alternative hypotese, dvs vi skal teste

$$H_0 : \mu_A = \mu_B \quad \text{mot} \quad H_1 : \mu_A > \mu_B.$$

Som testobservator bruker vi

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

som, når H_0 er riktig, er tilnærmet t -fordelt med ν frihetsgrader. Antall frihetsgrader ν er gitt ved samme formel som i forrige punkt. Dersom H_1 er riktig vil T tendere til å være stor. Vi forkaster derfor H_0 dersom $T > k$, der kritisk verdi k bestemmes fra det generelle kravet

$$P(\text{Forkast } H_0 | H_0 \text{ er riktig}) = \alpha.$$

I vår situasjon blir kravet

$$P(T > k | H_0 \text{ er riktig}) = \alpha \quad \Rightarrow \quad k = t_{\alpha, \nu}.$$

Dvs. vi forkaster H_0 dersom $T > t_{\alpha, \nu}$.

Innsatt observerte verdier får vi at observert verdi for testobservatoren T blir

$$t_{\text{obs}} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{15.22 - 14.56}{\sqrt{\frac{0.32}{7} + \frac{0.47}{6}}} = 4.325,$$

mens kritisk verdi når $\alpha = 0.05$ er $k = t_{0.05,9} = 1.833$. Vi har dermed at $t_{\text{obs}} > t_{\alpha, \nu}$ og vi forkaster H_0 .

Oppgave 3

a) La Y være antall infiserte bokser i en k -gruppe. Da er Y binomisk fordelt fordi

- man undersøker k bokser,
- hver boks er enten infisert eller ikke infisert,
- for hver boks er det samme sannsynlighet p for at den er infisert, og
- de ulike boksene er infisert eller ikke infisert uavhengig av hverandre.

Dermed er Y binomisk fordelt med k forsøk og sannsynlighet p for suksess.

En blanding av innholdet i k bokser gir positivt testresultat hvis og bare hvis minst en av de k boksene er infisert, dvs. hvis $Y \geq 1$,

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{k}{0} p^0 (1-p)^{k-0} = \underline{\underline{1 - (1-p)^k}}.$$

b) Betinget sannsynlighet for en hendelse A gitt en annen hendelse B er definert som

$$\underline{\underline{P(A|B) = \frac{P(A \cap B)}{P(B)}}}.$$

La Y være antall i en k -gruppe som er infisert og la A være hendelsen at en bestemt boks i denne blandingen er infisert. Da har vi at

$$\underline{\underline{P(A|Y \geq 1) = \frac{P(A \cap (Y \geq 1))}{P(Y \geq 1)} = \frac{P(A)}{P(Y \geq 1)} = \frac{p}{1 - (1-p)^k}}}.$$

c) For $i = 1, 2, \dots, m$ la

$$Z_i = \begin{cases} 1 & \text{hvis blandingen fra eske } i \text{ er positiv,} \\ 0 & \text{ellers.} \end{cases}$$

Fra a) vet vi da at

$$P(Z_i = 1) = 1 - (1-p)^k$$

slik at vi får

$$E[Z_i] = \sum_{z_i=0}^1 z_i P(Z_i = z_i) = 0 \cdot P(Z_i = 0) + 1 \cdot P(Z_i = 1) = P(Z_i = 1) = 1 - (1-p)^k.$$

Dessuten har vi at

$$X = \sum_{i=1}^m (1 + Z_i \cdot k).$$

Ved å bruke regneregler for forventningsverdi får vi da

$$\begin{aligned}
 E[X] &= E \left[\sum_{i=1}^m (1 + Z_i \cdot k) \right] \\
 &= \sum_{i=1}^m E[1 + Z_i \cdot k] \\
 &= \sum_{i=1}^m (1 + k \cdot E[Z_i]) \\
 &= \sum_{i=1}^m (1 + k (1 - (1 - p)^k)) \\
 &= m (1 + k (1 - (1 - p)^k)) \\
 &= \underline{\underline{m + mk (1 - (1 - p)^k)}}.
 \end{aligned}$$

Dersom man tester alle boksene enkeltvis vil man trenge mk tester. Man får dermed at den benyttede fremgangsmåten er å foretrekke dersom (for $k = 4$)

$$\begin{aligned}
 E[X] &= m + 4m (1 - (1 - p)^4) < 4m \\
 \Leftrightarrow 4m (1 - (1 - p)^4) &< 3m \\
 \Leftrightarrow 1 - (1 - p)^4 &< \frac{3}{4} \\
 \Leftrightarrow -(1 - p)^4 &< -\frac{1}{4} \\
 \Leftrightarrow (1 - p)^4 &> \frac{1}{4} \\
 \Leftrightarrow 1 - p &> \frac{1}{\sqrt{2}} \\
 \Leftrightarrow p - 1 &< -\frac{1}{\sqrt{2}} \\
 \Leftrightarrow p &< 1 - \frac{1}{\sqrt{2}} = 0.2929
 \end{aligned}$$

Den benyttede fremgangsmåten er dermed å foretrekke dersom $p < 0.2929$.