

Løsningsskisse for eksamen i TMA4245 vår 2021

14. mai 2021

Oppgave 1

La \bar{x} , \tilde{x} og s^2 betegne henholdsvis gjennomsnitt, median og empirisk varians.

Variant A: $\bar{x} = 3.8571 \approx 3.9$, $\tilde{x} = 4$, og $s^2 = 2.8095 \approx 2.8$.

Variant B: $\bar{x} = 3.1429 \approx 3.1$, $\tilde{x} = 3$, og $s^2 = 1.8095 \approx 1.8$.

Variant C: $\bar{x} = 4.1429 \approx 4.1$, $\tilde{x} = 4$, og $s^2 = 2.4762 \approx 2.5$.

Variant D: $\bar{x} = 2.7143 \approx 2.7$, $\tilde{x} = 2$, og $s^2 = 1.9048 \approx 1.9$.

Oppgave 2

I løsningsforslaget under betegner k et forskjellig tall i hver variant av oppgaven.

- $P(A \cap B) = P(X = k) = 0 \neq f_X(k)$ siden X er kontinuerlig fordelt.
- Siden $A \cap B = \{X = k\} \neq \emptyset$, er ikke A og B disjunkte.
- Siden $0 = P(A \cap B) \neq P(A)P(B) > 0$ er A og B avhengige hendelser.

Oppgave 3

Antall mulige utfall blir $\binom{52}{13} = 2598960$.

Variant A: Antall måter å trekke 5 kort i farge hjerter blir $\binom{13}{5} = 1287$.

Variant B: Farge kan velges på 4 måter. Antall straights i hver farge som blir 10. Totalt antall måter vi kan trekke straight flush på blir dermed 40.

Variant C: Antall måter å trekke 5 kort som alle er billedkort blir $\binom{16}{4} = 4368$

Variant D: Antall måter å trekke 5 kort hvorav 4 ess er lik antall måter kortet som ikke er ess kan velges på lik 48.

Oppgave 4

Variant A: $P(X < 4) = P(X = 2) + P(X = 3) = 1/36 + 2/36 = 1/12 = 0.083$. $P(Z \leq 35) = P(X \leq 3) = P(X < 4) = 0.083$. $P(Z = 25) = P(X = 2) = 1/36 = 0.028$. $P(Z = 25|X < 4) = P(X = 2|X < 4) = P(X = 2)/P(X < 4) = 1/3 = 0.333$.

Variant B: $P(X > 10) = P(X = 11) + P(X = 12) = 2/36 + 1/36 = 0.083$. $P(Z \geq 65) = P(X \geq 11) = P(X > 10) = 0.083$. $P(Z = 70) = P(X = 12) = 1/36 = 0.028$. $P(Z = 70|X > 10) = P(X = 12|X > 10) = P(X = 12)/P(X > 10) = 1/3 = 0.333$.

Variant C: $P(X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = (1 + 2 + 3)/36 = 1/6 = 0.167$. $P(Z < 20) = P(X < 5) = P(X \leq 4) = 0.167$. $P(Z = 14) = P(X = 2) = 0.028$. $P(Z = 14|X \leq 4) = P(X = 2)/P(X \leq 4) = 1/6 = 0.167$.

Variant D: $P(X \leq 4) = 0.167$. $P(Z < 50) = P(X < 5) = P(X \leq 4) = 0.167$. $P(Z = 20) = P(X = 2) = 1/36 = 0.028$.
 $P(Z = 20 | X \leq 4) = P(X = 2) / P(X \leq 4) = 1/6 = 0.167$.

Oppgave 5

Vi viser her eksakte svar til tre desimaler. Bruk av tabell gir litt andre svar som også godkjennes som korrekte.

Variant A: $P(X > 3) = 0.691$, $P(X + Y > 3) = 0.500$ og $P(X + Y > 3 | Y = 3) = 0.933$.

Variant B: $P(X > 3) = 0.434$, $P(X + Y > 3) = 0.242$ og $P(X + Y > 3 | Y = 2) = 0.691$.

Variant C: $P(X > 3) = 0.401$, $P(X + Y > 3) = 0.421$ og $P(X + Y > 3 | Y = 1) = 0.500$.

Variant D: $P(X > 3) = 0.159$, $P(X + Y > 3) = 0.159$ og $P(X + Y > 3 | Y = 3) = 0.401$.

Oppgave 6

Variant A: $E[X + Y] = 1.0$ og $\text{Var}[X + Y] = 4.0$.

Variant B: $E[X + Y] = 0.0$ og $\text{Var}[X + Y] = 3.0$.

Variant C: $E[X + Y] = 3.0$ og $\text{Var}[X + Y] = 5.0$.

Variant D: $E[X + Y] = -1.0$ og $\text{Var}[X + Y] = 3.5$.

Oppgave 7a

- Kumulativ fordelingsfunksjon blir

$$F_Y(y) = P(Y \leq y) = \int_{y_0}^y \frac{\lambda y_0^\lambda}{u^{\lambda+1}} du = -\lambda y_0^\lambda \frac{1}{\lambda u^\lambda} \Big|_{y_0}^y = 1 - \left(\frac{y_0}{y}\right)^\lambda$$

for $y \geq y_0$.

- Dermed blir

$$\begin{aligned} P(\min(Y_1, Y_2, \dots, Y_n) < k) &= 1 - P(\min(Y_1, Y_2, \dots, Y_n) \geq k) \\ &= 1 - P(Y_1 \geq k \cap \dots \cap Y_n \geq k) \\ &= 1 - (P(Y_i \geq k))^n \\ &= 1 - (1 - F_Y(k))^n \\ &= 1 - \left(\frac{y_0}{k}\right)^{n\lambda} \end{aligned}$$

Innsatt blir sannsynligheten

Variant A: 0.28

Variant B: 0.3

Variant C: 0.17

Variant D: 0.6

Oppgave 7b

- Likelihoodfunksjonen blir

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda y_0^\lambda}{y_i^{\lambda+1}} = \lambda^n y_0^{n\lambda} \prod_{i=1}^n y_i^{-\lambda-1}$$

og log-likelihoodet

$$l(\lambda) = n \ln \lambda + n\lambda \ln y_0 - (\lambda + 1) \sum_{i=1}^n \ln y_i.$$

I maksimum er $dl/d\lambda = 0$ slik at

$$\frac{n}{\lambda} + n \ln y_0 - \sum_{i=1}^n \ln y_i = 0,$$

Løses vi ligningen finner vi at SME av λ kan skrives på formen

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n \ln(Y_i/y_0)}$$

- Estimatet blir

Variant A: 2.137

Variant B: 2.508

Variant C: 3.243

Variant D: 2.268

Oppgave 7c

- Tettheten til X blir

$$\begin{aligned} f_X(x) &= f_Y(y_0 e^{x/(2\lambda)}) \left| \frac{dy}{dx} \right| \\ &= \frac{\lambda y_0^\lambda}{(y_0 e^{x/(2\lambda)})^{\lambda+1}} y_0 e^{x/(2\lambda)} \frac{1}{2\lambda} \\ &= \frac{1}{2} e^{-\frac{1}{2}x} \end{aligned}$$

som er tettheten til en kji-kvadratfordelt variabel med 2 frihetsgrader.

Videre blir $\sum_{i=1}^n 2\lambda \ln(Y_i/y_0)$ kji-kvadratfordelt med $2n$ frihetsgrader siden hvert ledd i summen er uavhengig kji-kvadrat med 2 frihetsgrader.

- Dermed er

$$P\left(\chi_{1-\alpha/2, 2n}^2 < \sum_{i=1}^n 2\lambda \ln(Y_i/y_0) < \chi_{\alpha/2, 2n}^2\right) = 1 - \alpha$$

slik at

$$\left(\frac{\chi_{1-\alpha/2, 2n}^2}{2 \sum_{i=1}^n \ln(y_i/y_0)}, \frac{\chi_{\alpha/2, 2n}^2}{2 \sum_{i=1}^n \ln(y_i/y_0)}\right) = \left(\frac{\hat{\lambda} \chi_{1-\alpha/2, 2n}^2}{2n}, \frac{\hat{\lambda} \chi_{\alpha/2, 2n}^2}{2n}\right)$$

er et $(1 - \alpha)$ konfidensintervall for λ . For $\alpha = 0.05$ blir intervallet innsatt observasjonene i oppgaven

Variant A: (1.025, 3.651)

Variant B: (1.083, 4.521)

Variant C: (1.190, 6.307)

Variant D: (1.172, 3.721)

Oppgave 7d

- Den alternative hypotesen H_1 , utsagnet

$$\mu^* = 2^{1/\lambda} y_0 > \mu_0^*$$

er ekvivalent med

$$\frac{1}{\lambda} \ln 2 + \ln y_0 > \ln \mu_0^*$$

eller

$$\lambda < \frac{\ln(2)}{\ln(\mu_0^*/y_0)} = \lambda_0$$

hvor λ_0 ved innsetting er som gitt i oppgaven. Å teste om medianlønnen μ^* er større enn μ_0^* er dermed det samme som å teste om $\lambda < \lambda_0$.

- Under H_0 er testobservatoren

$$W = \sum_{i=1}^n 2\lambda_0 \ln(Y_i/y_0) = \frac{2n\lambda_0 \sum_{i=1}^n \ln(Y_i/y_0)}{n} = \frac{2n\lambda_0}{\hat{\lambda}}$$

som tidligere vist kji-kvadrat med $2n$ frihetsgrader. Under H_1 vil $\hat{\lambda}$ tendere til å bli liten og W dermed stor. Vi forkaster derfor H_0 hvis

$$W > \chi_{\alpha, 2n}^2$$

Observert og kritisk verdi og testens konklusjons blir:

Variant A: $W = 9.36$, $\chi_{0.05, 20}^2 = 31.410$, Beholder H_0

Variant B: $W = 8.6768$, $\chi_{0.05, 16}^2 = 26.296$, Beholder H_0

Variant C: $W = 4.588$, $\chi_{0.05, 12}^2 = 21.026$, Beholder H_0

Variant D: $W = 32.9038$, $\chi_{0.05, 24}^2 = 36.415$, Beholder H_0

Oppgave 8a

La Z betegne en standardnormalfordelt stokastisk variabel. Vi vet at $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/\sqrt{n})$ slik at for $\sigma^2 = 1$ har vi $\bar{X} \sim \mathcal{N}(\mu, 1/\sqrt{n})$.

- Type I feil:**

$$\begin{aligned} \text{P(Type I feil)} &= \text{P(Forkast } H_0 | \mu = 5) \\ &= \text{P}(\bar{X} \leq 5 - z_\alpha/\sqrt{n} | \mu = 5) \\ &= \text{P}\left(\frac{\bar{X} - 5}{1/\sqrt{n}} < \frac{5 - 5 - z_\alpha/\sqrt{n}}{1/\sqrt{n}} \mid \mu = 5\right) \\ &= \text{P}(Z \leq -z_\alpha) = \alpha. \end{aligned}$$

Variant A: Tallverdi: $\text{P}(\text{Type I feil}) = \alpha = 0.05$.

Variant B: Tallverdi: $\text{P}(\text{Type I feil}) = \alpha = 0.10$.

Variant C: Tallverdi: $\text{P}(\text{Type I feil}) = \alpha = 0.05$.

Variant D: Tallverdi: $P(\text{Type I feil}) = \alpha = 0.10$.

• **Styrke:**

$$\begin{aligned} \text{Styrke} &= P(\text{Forkast } H_0 | \mu = 4.5) \\ &= P(\bar{X} \leq 5 - z_\alpha / \sqrt{n} | \mu = 4.5) \\ &= P\left(\frac{\bar{X} - 4.5}{1/\sqrt{n}} < \frac{5 - 4.5 - z_\alpha / \sqrt{n}}{1/\sqrt{n}} \mid \mu = 4.5\right) \\ &= P(Z \leq 0.5\sqrt{n} - z_\alpha). \end{aligned}$$

Variant A: Tallverdi: $P(Z \leq 0.5\sqrt{10} - 1.645) = 0.475$.

Variant B: Tallverdi: $P(Z \leq 0.5\sqrt{10} - 1.282) = 0.618$.

Variant C: Tallverdi: $P(Z \leq 0.5\sqrt{20} - 1.645) = 0.723$.

Variant D: Tallverdi: $P(Z \leq 0.5\sqrt{20} - 1.282) = 0.830$.

Oppgave 8b

La Z betegne en standardnormalfordelt stokastisk variabel. Anta at utvalg 1 består av X_1, \dots, X_n og at utvalg 2 består av Y_1, \dots, Y_n . Som i oppgave 8a vet vi da at $\bar{X} \sim \mathcal{N}(\mu, 1/\sqrt{n})$ og $\bar{Y} \sim \mathcal{N}(\mu, 1/\sqrt{n})$. I tillegg vet vi at \bar{X} og \bar{Y} er uavhengige.

• **Type I feil:**

$$\begin{aligned} P(\text{Type I feil}) &= P(\text{Forkast } H_0 | H_0) \\ &= P((\bar{X} \leq 5 - z_\alpha / \sqrt{n}) \cup (\bar{Y} \leq 5 - z_\alpha / \sqrt{n}) | \mu = 5) \\ &= P(\bar{X} \leq 5 - z_\alpha / \sqrt{n} | \mu = 5) + P(\bar{Y} \leq 5 - z_\alpha / \sqrt{n} | \mu = 5) \\ &\quad - P((\bar{X} \leq 5 - z_\alpha / \sqrt{n}) \cap (\bar{Y} \leq 5 - z_\alpha / \sqrt{n}) | \mu = 5) \\ &= P(Z \leq -z_\alpha) + P(Z \leq -z_\alpha) - P(\bar{X} \leq 5 - z_\alpha / \sqrt{n} | \mu = 5)P(\bar{Y} \leq 5 - z_\alpha / \sqrt{n} | \mu = 5) \\ &= \alpha + \alpha - P(Z \leq -z_\alpha)P(Z \leq -z_\alpha) \\ &= 2\alpha - \alpha^2. \end{aligned}$$

Variant A: Tallverdi: $P(\text{Type I feil}) = 0.0975$.

Variant B: Tallverdi: $P(\text{Type I feil}) = 0.190$.

Variant C: Tallverdi: $P(\text{Type I feil}) = 0.0975$.

Variant D: Tallverdi: $P(\text{Type I feil}) = 0.190$.

- **Forklaring:** Hvert av de to utvalgene gir opphav til en hypotesetest der sannsynligheten for å forkaste H_0 er α når H_0 er sann, men med forskerens fremgangsmåte kan vi forkaste H_0 selv om bare et av utvalgene tilfredstiller forkastningsregelen. Dermed må det være høyere sannsynlighet for å forkaste H_0 med forskerens fremgangsmåte enn ved å gjøre én hypotesetest.

- **Ny beslutningsregel:** Vi ønsker at $P(\text{Forkast } H_0 | H_0) = \alpha$ slik at vi må ha

$$\begin{aligned}
 \alpha &= P(\text{Forkast } H_0 | H_0) = P((\bar{X} \leq k) \cup (\bar{Y} \leq k) | \mu = 5) \\
 &= P(\bar{X} \leq k | \mu = 5) + P(\bar{Y} \leq k | \mu = 5) - P((\bar{X} \leq k) \cap (\bar{Y} \leq k) | \mu = 5) \\
 &= P\left(Z \leq \frac{k-5}{1/\sqrt{n}}\right) + P\left(Z \leq \frac{k-5}{1/\sqrt{n}}\right) - P(\bar{X} \leq k | \mu = 5)P(\bar{Y} \leq k | \mu = 5) \\
 &= 2P\left(Z \leq \frac{k-5}{1/\sqrt{n}}\right) - P\left(Z \leq \frac{k-5}{1/\sqrt{n}}\right)^2 \\
 &= 1 - \left(1 - P\left(Z \leq \frac{k-5}{1/\sqrt{n}}\right)\right)^2
 \end{aligned}$$

Vi kan løse denne likheten for sannsynligheten for å finne

$$\begin{aligned}
 \left(1 - P\left(Z \leq \frac{k-5}{1/\sqrt{n}}\right)\right)^2 &= 1 - \alpha \\
 P\left(Z \leq \frac{k-5}{1/\sqrt{n}}\right) &= 1 - \sqrt{1 - \alpha}
 \end{aligned}$$

For både $\alpha = 0.05$ og $\alpha = 0.10$ vil vi ha $1 - \sqrt{1 - \alpha} \approx \alpha/2$ slik at vi trenger

$$\begin{aligned}
 (k-5)\sqrt{n} &\approx -z_{\alpha/2} \\
 k &\approx 5 - z_{\alpha/2}/\sqrt{n}.
 \end{aligned}$$

Variant A: Tallverdi: $k \approx 5 - z_{0.025}/\sqrt{10} = 4.38$.

Variant B: Tallverdi: $k \approx 5 - z_{0.05}/\sqrt{10} = 4.48$.

Variant C: Tallverdi: $k \approx 5 - z_{0.025}/\sqrt{20} = 4.56$.

Variant D: Tallverdi: $k \approx 5 - z_{0.05}/\sqrt{20} = 4.63$.

Oppgave 9

La modellen være $Y_i = u_i\beta + \epsilon_i$ der $u_i = f(x_i)$ er en transformasjon av kovariaten x_i . Med den notasjonen kan vi skrive en generell løsning for alle varianter av oppgaven. Vi har estimatoren $\hat{\beta} = \frac{\sum_{i=1}^n u_i Y_i}{\sum_{i=1}^n u_i^2}$.

- **Forventningsverdi og varians:** Vi har en lineærkombinasjon av Y_i og kan skrive

$$E[\hat{\beta}] = \frac{\sum_{i=1}^n u_i E[Y_i]}{\sum_{i=1}^n u_i^2} = \frac{\sum_{i=1}^n u_i \cdot u_i \beta}{\sum_{i=1}^n u_i^2} = \beta \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n u_i^2} = \beta$$

slik at $\hat{\beta}$ er forventningsrett for β . Vi bruker at Y_1, \dots, Y_n er uavhengige til å skrive variansen som

$$\text{Var}[\hat{\beta}] = \sum_{i=1}^n \text{Var}\left[\frac{u_i Y_i}{\sum_{i=1}^n u_i^2}\right] = \sum_{i=1}^n \left(\frac{u_i^2 \text{Var}[Y_i]}{(\sum_{i=1}^n u_i^2)^2}\right) = \sigma^2 \frac{\sum_{i=1}^n u_i^2}{(\sum_{i=1}^n u_i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^n u_i^2}.$$

Variant A: $\text{Var}[\hat{\beta}] = \sigma^2 / \sum_{i=1}^n (\ln(x_i))^2$.

Variant B: $\text{Var}[\hat{\beta}] = \sigma^2 / \sum_{i=1}^n x_i^6$.

Variant C: $\text{Var}[\hat{\beta}] = \sigma^2 / \sum_{i=1}^n e^{2x_i}$.

Variant D: $\text{Var}[\hat{\beta}] = \sigma^2 / \sum_{i=1}^n x_i$.

- **Prediksjonsintervall:** Siden Y_1, Y_2, \dots, Y_n er uavhengige og normalfordelte er estimatoren $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 / \sum_{i=1}^n u_i^2)$ der forventningsverdi og varians ble regnet ut i punktet over. La Y være en ny observasjon som er uavhengig av de n tidligere observasjonene med kovariatverdi x og transformert kovariatverdi $u = f(x)$. Siden variansen er kjent kan vi basere konfidensintervallet på den stokastiske variabelen

$$Z = \frac{Y - u\hat{\beta}}{\sigma\sqrt{1 + u^2/\sum_{i=1}^n u_i^2}} \sim \mathcal{N}(0, 1).$$

Dette ser vi ved at $E[Y - u\hat{\beta}] = 0$ og $\text{Var}[Y - u\hat{\beta}] = \text{Var}[Y] + u^2\text{Var}[\hat{\beta}] = \sigma^2(1 + u^2/\sum_{i=1}^n u_i^2)$.

Basert på $P(-z_{0.025} < Z < z_{0.025}) = 0.95$ får vi et 95% prediksjonsintervall for observert ny verdi y ,

$$u\hat{\beta} - z_{0.025}\sigma\sqrt{1 + \frac{u^2}{\sum_{i=1}^n u_i^2}} < y < u\hat{\beta} + z_{0.025}\sigma\sqrt{1 + \frac{u^2}{\sum_{i=1}^n u_i^2}}.$$

Variant A:

$$\ln(x)\hat{\beta} - z_{0.025}\sigma\sqrt{1 + \frac{(\ln(x))^2}{\sum_{i=1}^n (\ln(x_i))^2}} < y < \ln(x)\hat{\beta} + z_{0.025}\sigma\sqrt{1 + \frac{(\ln(x))^2}{\sum_{i=1}^n (\ln(x_i))^2}}$$

Variant B:

$$x^3\hat{\beta} - z_{0.025}\sigma\sqrt{1 + \frac{x^6}{\sum_{i=1}^n x_i^2}} < y < x^3\hat{\beta} + z_{0.025}\sigma\sqrt{1 + \frac{x^6}{\sum_{i=1}^n x_i^2}}$$

Variant C:

$$e^x\hat{\beta} - z_{0.025}\sigma\sqrt{1 + \frac{e^{2x}}{\sum_{i=1}^n e^{2x_i}}} < y < e^x\hat{\beta} + z_{0.025}\sigma\sqrt{1 + \frac{e^{2x}}{\sum_{i=1}^n e^{2x_i}}}$$

Variant D:

$$\sqrt{x}\hat{\beta} - z_{0.025}\sigma\sqrt{1 + \frac{x}{\sum_{i=1}^n x_i}} < y < \sqrt{x}\hat{\beta} + z_{0.025}\sigma\sqrt{1 + \frac{x}{\sum_{i=1}^n x_i}}$$

- **Tallverdier:**

Variant A: $\ln(x)\hat{\beta} = \ln(2) \cdot 4/9.3 = 0.298$ og $z_{0.025}\sigma\sqrt{1 + \frac{(\ln(x))^2}{\sum_{i=1}^n (\ln(x_i))^2}} = 1.960 \cdot \sqrt{2}\sqrt{1 + \ln(2)^2/9.3} = 2.843$ gir
 $-2.55 < y < 3.14$.

Variant B: $x^3\hat{\beta} = 2^3 \cdot 3/10 = 2.400$ og $z_{0.025}\sigma\sqrt{1 + \frac{x^6}{\sum_{i=1}^n x_i^2}} = 1.960 \cdot \sqrt{2}\sqrt{1 + 2^6/10} = 7.540$ gir
 $-5.14 < y < 9.94$.

Variant C: $e^x\hat{\beta} = e^2 \cdot 5/15 = 2.463$ og $z_{0.025}\sigma\sqrt{1 + \frac{e^{2x}}{\sum_{i=1}^n e^{2x_i}}} = 1.960 \cdot \sqrt{2}\sqrt{1 + e^{2 \cdot 2}/15} = 5.971$ gir
 $-3.51 < y < 8.43$.

Variant D: $\sqrt{x}\hat{\beta} = \sqrt{2} \cdot 100/50 = 2.828$ og $z_{0.025}\sigma\sqrt{1 + \frac{x}{\sum_{i=1}^n x_i}} = 1.960 \cdot \sqrt{2}\sqrt{1 + 2/50} = 2.827$ gir
 $0.00 < y < 5.66$.