TMA4245 Statistikk
Eksamen mai 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsskisse

Oppgave 1

a) Når vi regner ut disse tre sannsynlighetene må man huske på at de mulige verdiene for X er $3, 4, \ldots$.

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) = 1 - P(X = 3) = 1 - \binom{3-1}{3-1} p^3 (1-p)^0 \\ &= 1 - p^3 = 1 - 0.1^3 = \underline{\underline{0.999}} \end{aligned}$$

$$\begin{aligned} P(X < 6) &= \sum_{x=3}^{5} P(X = x) = \sum_{x=3}^{5} \binom{x-1}{3-1} p^3 (1-p)^{x-3} \\ &= \binom{2}{2} p^3 (1-p)^0 + \binom{3}{2} p^3 (1-p)^1 + \binom{4}{2} p^3 (1-p)^2 \\ &= p^3 \left(1 + 3(1-p) + 6(1-p)^2\right) = 0.1^3 \left(1 + 2.7 + 4.4.86\right) = \underline{\underline{0.00856}} \end{aligned}$$

$$\begin{aligned} P(X \geq 6 | X > 3) &= \frac{P(X \geq 6 \cap X > 3)}{P(X > 3)} = \frac{P(X \geq 6)}{P(X > 3)} \\ &= \frac{1 - P(X < 6)}{P(X \geq 3)} = \frac{1 - 0.0.00856}{0.999} = \underline{\underline{0.9924}} \end{aligned}$$

b) For å finne SME må man starte med å finne rimelighetsfunksjonen. Siden $X_1, X_2, \ldots, X_n$ er uavhengige får vi at

$$L(p) = f(x_1, x_2, \ldots, x_n; p) = \prod_{i=1}^{n} f(x_i; p) = \prod_{i=1}^{n} \left[\binom{x_i - 1}{k - 1} p^k (1-p)^{x_i - k} \right].$$

Log-rimelighetsfunksjonen blir da

$$\begin{aligned} l(p) &= \ln L(p) = \sum_{i=1}^{n} \left[\ln \left(\binom{x_i - 1}{k - 1} \right) + k \ln p + (x_i - k) \ln(1-p) \right] \\ &= \sum_{i=1}^{n} \left[\ln \left(\binom{x_i - 1}{k - 1} \right) \right] + nk \ln p + \ln(1-p) \sum_{i=1}^{n} x_i - nk \ln(1-p). \end{aligned}$$

For å finne for hvilken verdi av p denne funksjonen har sitt maksimum deriverer vi med hensyn på p,

$$\begin{aligned} l'(p) &= 0 + \frac{nk}{p} + \frac{1}{1-p} \cdot (-1) \cdot \sum_{i=1}^{n} x_i - \frac{nk}{1-p} \cdot (-1) \\ &= \frac{nk}{p} - \frac{\sum_{i=1}^{n} x_i}{1-p} + \frac{nk}{1-p} \end{aligned}$$

Finner for hvilken verdi av p log-rimelighetsfunksjonen har sitt maksimum ved å løse ligningen $l'(p) = 0$ med hensyn på p,

$$\begin{aligned} l'(p) = 0 \quad &\Rightarrow \quad \frac{nk}{p} = \frac{\sum_{i=1}^{n} x_i - nk}{1-p} \\ &\Rightarrow \quad nk(1-p) = \left(\sum_{i=1}^{n} x_i - nk \right) p \\ &\Rightarrow \quad nk - nkp = p \sum_{i=1}^{n} x_i - nkp \\ &\Rightarrow \quad nk = p \sum_{i=1}^{n} x_i \\ &\Rightarrow \quad p = \frac{nk}{\sum_{i=1}^{n} x_i}. \end{aligned}$$

Sannsynlighetsmaksimeringsestimatoren blir dermed

$$\underline{\underline{\widehat{p} = \frac{nk}{\sum_{i=1}^{n} X_i} = \frac{k}{\overline{X}}}}.$$

c) La A og B være hendelsene at klokkene som inspiseres kommer fra henholdsvis produksjonslinje A og B. Vi har da oppgitt at $P(A) = P(B) = 0.5$ og at

$$P(X = x|A) = \binom{x-1}{k-1} p_A^k (1-p_A)^{x-k} \quad \text{og} \quad P(X = x|B) = \binom{x-1}{k-1} p_B^k (1-p_B)^{x-k}.$$

Oppgaven spør etter sannsynligheten $P(A|X = x)$. Ved å bruke Bayes' regel får man at

$$\begin{aligned} P(A|X = x) &= \frac{P(X = x|A)P(A)}{P(X = x)} = \frac{P(X = x|A)P(A)}{P(X = x|A)P(A) + P(X = x|B)P(B)} \\ &= \frac{\binom{x-1}{k-1} p_A^k (1-p_A)^{x-k} \cdot 0.5}{\binom{x-1}{k-1} p_A^k (1-p_A)^{x-k} \cdot 0.5 + \binom{x-1}{k-1} p_B^k (1-p_B)^{x-k} \cdot 0.5} \\ &= \underline{\underline{\frac{p_A^k (1-p_A)^{x-k}}{p_A^k (1-p_A)^{x-k} + p_B^k (1-p_B)^{x-k}}}}. \end{aligned}$$

Setter man inn de oppgitte tallene får man at

$$P(A|X = 5) = \frac{0.1^3 (1-0.1)^{5-3}}{0.1^3 (1-0.1)^{5-3} + 0.2^3 (1-0.2)^{5-3}} = \underline{\underline{0.1366}}.$$

Oppgave 2

a) I figur A ser det ut til at det er en ikke-lineær sammenheng mellom x og y, noe som ikke stemmer med regresjonsmodellen gitt i oppgaven.

I figur B ser det ut til at variabiliteten til y øker med økende x, mens man i regresjonsmodellen gitt i oppgaveteksten spesifiserer at variansen til Y er den samme for alle verdier av x. Dette datasettet passer dermed heller ikke regresjonsmodellen gitt i oppgaveteksten.

I figur C ser det ut til å være en lineær sammenheng mellom x og y og variabiliteten til y ser ut til å være den samme for alle verdier av x. Den oppgitte regresjonsmodellen ser ut til å være en god modell for dette datasettet.

Siden Y er tykkelsen på en bremsekloss må man rent fysisk nødvendigvis ha at $Y \geq 0$. Men ifølge regresjonsmodellen vil $E[Y|x]$ bli negativ når x er stor nok og $P(Y < 0)$ vil også bli stor når x er stor nok. Dermed kan ikke regresjonsmodellen være en rimelig modell for alle x. Den lineære regresjonsmodellen vil kun være gyldig for at intervall av x-verdier, $x \in [0, x_{\max}]$. Tilsvarende vil være tilfelle for de aller fleste regresjonsmodeller, modellen er gyldig kun for et intervall av x-verdier.

b) Minste kvadraters metode angir at $\widehat{\beta}$ er gitt ved

$$\widehat{\beta} = \underset{\widehat{\beta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} (Y_i - (k_0 - \widehat{\beta} x_i))^2 \right\}.$$

Vi finner minimum ved å sette den deriverte av kvadratsummen lik null,

$$\begin{aligned}
\frac{\partial}{\partial \widehat{\beta}} \sum_{i=1}^{n} (Y_i - (k_0 - \widehat{\beta} x_i))^2 &= \sum_{i=1}^{n} 2(Y_i - (k_0 - \widehat{\beta} x_i)) \cdot x_i \\
&= 2 \left[\sum_{i=1}^{n} Y_i x_i - k_0 \sum_{i=1}^{n} x_i + \widehat{\beta} \sum_{i=1}^{n} x_i^2 \right] = 0 \\
&\Rightarrow \widehat{\beta} \sum_{i=1}^{n} x_i^2 = k_0 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} Y_i x_i \\
&\Rightarrow \underline{\underline{\widehat{\beta} = \frac{k_0 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} Y_i x_i}{\sum_{i=1}^{n} x_i^2}}}.
\end{aligned}$$

Ved å bruke regneregler for forventningverdioperatoren får vi at (ved å huske at x_i'ene

er konstanter)

$$\begin{aligned}
\mathrm{E}\left[\widehat{\beta}\right] &= \mathrm{E}\left[\frac{k_0\sum_{i=1}^n x_i - \sum_{i=1}^n Y_i x_i}{\sum_{i=1}^n x_i^2}\right] \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\mathrm{E}\left[k_0\sum_{i=1}^n x_i - \sum_{i=1}^n Y_i x_i\right] \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\left[k_0\sum_{i=1}^n x_i - \sum_{i=1}^n \mathrm{E}\left[Y_i x_i\right]\right] \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\left[k_0\sum_{i=1}^n x_i - \sum_{i=1}^n x_i\mathrm{E}\left[Y_i\right]\right].
\end{aligned}$$

Ved å benytte modellantagelsen $\mathrm{E}[Y_i] = k_0 - \beta x_i$ får man dermed

$$\begin{aligned}
\mathrm{E}\left[\widehat{\beta}\right] &= \frac{1}{\sum_{i=1}^n x_i^2}\left[k_0\sum_{i=1}^n x_i - \sum_{i=1}^n x_i(k_0 - \beta x_i)\right] \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\left[k_0\sum_{i=1}^n x_i - k_0\sum_{i=1}^n x_i + \beta\sum_{i=1}^n x_i^2\right] \\
&= \beta\cdot\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} = \underline{\underline{\beta}}.
\end{aligned}$$

For å finne variansen til $\widehat{\beta}$ starter vi tilsvarende med å bruke regneregler for varians. Igjen må man huske på at x_i'ene er konstanter, og man må huske på at vi har antatt at Y_i'ene er uavhengige. Da får vi

$$\begin{aligned}
\mathrm{Var}\left[\widehat{\beta}\right] &= \mathrm{Var}\left[\frac{k_0\sum_{i=1}^n x_i - \sum_{i=1}^n Y_i x_i}{\sum_{i=1}^n x_i^2}\right] \\
&= \left(\frac{1}{\sum_{i=1}^n x_i^2}\right)^2\mathrm{Var}\left[k_0\sum_{i=1}^n x_i - \sum_{i=1}^n Y_i x_i\right] \\
&= \frac{1}{\left(\sum_{i=1}^n x_i^2\right)^2}\cdot\left[0 + (-1)^2\sum_{i=1}^n \mathrm{Var}\left[Y_i x_i\right]\right] \\
&= \frac{1}{\left(\sum_{i=1}^n x_i^2\right)^2}\cdot\left[\sum_{i=1}^n x_i^2\mathrm{Var}\left[Y_i\right]\right].
\end{aligned}$$

Ved å benytte modellantagelsen $\mathrm{Var}[Y_i] = \sigma^2$ får man dermed

$$\begin{aligned}
\mathrm{Var}\left[\widehat{\beta}\right] &= \frac{1}{\left(\sum_{i=1}^n x_i^2\right)^2}\cdot\left[\sum_{i=1}^n x_i^2\sigma^2\right] \\
&= \frac{\sigma^2\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} = \underline{\underline{\frac{\sigma^2}{\sum_{i=1}^n x_i^2}}}.
\end{aligned}$$

c) Fra forrige punkt har vi forventningverdi og varians til $\widehat{\beta}$ og siden det er oppgitt at $\widehat{\beta}$ er

normalfordelt kan vi lage en standard normalfordelt variabel

$$Z = \frac{\widehat{\beta} - \mathrm{E}\left[\widehat{\beta}\right]}{\sqrt{\mathrm{Var}\left[\widehat{\beta}\right]}} = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n x_i^2}}} \sim \mathrm{N}(0,1).$$

Siden verdien til variansen σ^2 er ukjent erstatter vi σ^2 i dette uttrykket med vår estimator $\widehat{\sigma}^2$ for σ^2. La oss kalle størrelsen vi da får for T. Vi kan vise hvilken sannsynlighetsfordeling T har ved å skrive

$$T = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}} = \frac{\frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n x_i^2}}}}{\sqrt{\frac{\frac{(n-1)\widehat{\sigma}^2}{\sigma^2}}{n-1}}} = \frac{Z}{\sqrt{\frac{V}{n-1}}}.$$

Siden $Z \sim \mathrm{N}(0,1)$, $V \sim \chi^2_{n-1}$, og Z og V er uavhengige siden $\widehat{\beta}$ og V er uavhengige, får vi at T er Student t-fordelt med $n-1$ frihetsgrader. Vi kan da lage konfidensintervall ved å starte med kvantiler i en Student t-fordeling. Siden en Student t-fordeling er symmetrisk om 0 har vi

$$P\left(-t_{\frac{\alpha}{2},n-1} \leq T \leq t_{\frac{\alpha}{2},n-1}\right) = 1-\alpha$$

$$P\left(-t_{\frac{\alpha}{2},n-1} \leq \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}} \leq t_{\frac{\alpha}{2},n-1}\right) = 1-\alpha$$

Løser så hver av ulikhetene inni sannsynlighetsuttrykket med hensyn på β. Vi får

$$\begin{aligned}
-t_{\frac{\alpha}{2},n-1} \leq \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}} \quad &\Leftrightarrow \quad -t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \leq \widehat{\beta} - \beta \\
&\Leftrightarrow \quad -\widehat{\beta} - t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \leq -\beta \\
&\Leftrightarrow \quad \widehat{\beta} + t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \geq \beta \\
&\Leftrightarrow \quad \beta \leq \widehat{\beta} + t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}
\end{aligned}$$

og tilsvarende for den andre ulikheten

$$\begin{aligned}
\frac{\widehat{\beta}-\beta}{\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}} \leq t_{\frac{\alpha}{2},n-1} \quad &\Leftrightarrow \quad \widehat{\beta}-\beta \leq t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \\
&\Leftrightarrow \quad -\beta \leq -\widehat{\beta}+t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \\
&\Leftrightarrow \quad \beta \geq \widehat{\beta}-t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \\
&\Leftrightarrow \quad \widehat{\beta}-t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \leq \beta.
\end{aligned}$$

Vi har dermed

$$P\left(\widehat{\beta}-t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}} \leq \beta \leq \widehat{\beta}+t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}\right) = 1-\alpha.$$

Et $(1-\alpha)\cdot 100\%$-konfidensintervall for β er dermed

$$\underline{\underline{\left[\widehat{\beta}-t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}, \widehat{\beta}+t_{\frac{\alpha}{2},n-1}\sqrt{\frac{\widehat{\sigma}^2}{\sum_{i=1}^n x_i^2}}\right]}}.$$

Hvis vi ønsker å utføre den tosidige hypotesetesten

$$H_0 : \beta = \beta_0 \quad \text{mot} \quad H_1 : \beta \neq \beta_0$$

med signifikansnivå α, kan vi benytte sammenhengen som alltid finnes mellom et konfidensintervall og en tilhørende tosidig hypotesetest. Vi skal forkaste H_0 hvis konfidensintervallet ikke inneholder verdien β_0.

Oppgave 3

a) Estimatoren $\widehat{\mu}$ er normalfordelt fordi den en en linær funksjon av X_1, X_2, X_3, som er uavhengige og normalfordelt.

Ved å benytte regneregler for forventingsverdi får vi at

$$\mathrm{E}\left[\widehat{\mu}\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i].$$

Ved å benytte at $\mathrm{E}[X_i] = \mu$ får vi da

$$\mathrm{E}\left[\widehat{\mu}\right] = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}\cdot n\mu = \underline{\underline{\mu}}.$$

Siden X_1, X_2, X_3 er uavhengige gir regneregler for varians at

$$\text{Var}\,[\widehat{\mu}] = \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \left(\frac{1}{n}\right)^2 \text{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}[X_i].$$

Ved å benytte at $\text{Var}[X_i] = \sigma^2$ gir dette

$$\text{Var}\,[\widehat{\mu}] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}\cdot n\sigma^2 = \underline{\underline{\frac{\sigma^2}{n}}}.$$

Hvis en estimator er forventingsrett innebærer det at dersom man gjentar forsøket uendelig mange ganger vil gjennomsnittet av de tilhørende estimatene være lik den sanne parameterverdien.

b) Det er oppgitt at man skal bruke $\widehat{\mu}$ som testobservator. Det er rimelig å forkaste H_0 dersom $\widehat{\mu} < k$, der den kritiske verdien k må bestemmes fra kravet

$$P(\text{Forkast } H_0 | H_0 \text{ er riktig}) = P\left(\widehat{\mu} < k | \mu = 3\right) = \alpha.$$

Ved å standardisere $\widehat{\mu}$ får vi

$$P\left(\widehat{\mu} < k | \mu = 3\right) = P\left(\frac{\widehat{\mu}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < \frac{k-\mu}{\sqrt{\frac{\sigma^2}{n}}}\middle| \mu = 3\right) = P\left(Z < \frac{k-3}{\sqrt{\frac{\sigma^2}{n}}}\right) = \alpha,$$

der $Z \sim \text{N}(0,1)$. Ved å tegne opp sannsynlighetstettheten i en standard normalfordeling ser vi da at vi må ha

$$\frac{k-3}{\sqrt{\frac{\sigma^2}{n}}} = -z_\alpha \quad \Rightarrow \quad k = 3 - z_\alpha\sqrt{\frac{\sigma^2}{n}}.$$

Innsatt tall får vi, ved å bruke at $z_\alpha = z_{0.05} = 1.645$,

$$k = 3 - 1.645\sqrt{\frac{0.4^2}{3}} = 2.62.$$

<u>Vi skal altså forkaste H_0 hvis $\widehat{\mu} < 2.62$.</u>

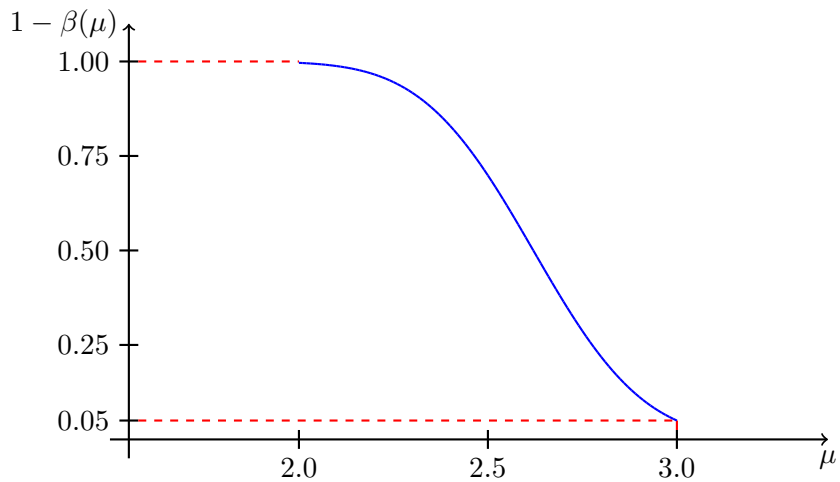
c) Teststyrken til testen er for $\mu < 3$ gitt ved

$$1 - \beta(\mu) = P(\text{Forkast } H_0 | \mu) = P\left(\widehat{\mu} < 3 - z_\alpha\sqrt{\frac{\sigma^2}{n}}\middle| \mu\right)$$

Standardiserer $\widehat{\mu}$ for å finne et uttrykk for sannsynligheten,

$$\begin{aligned} 1 - \beta(\mu) &= P\left(\frac{\widehat{\mu}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < \frac{3 - z_\alpha\sqrt{\frac{\sigma^2}{n}} - \mu}{\sqrt{\frac{\sigma^2}{n}}}\middle| \mu\right) \\ &= \Phi\left(\frac{3 - z_\alpha\sqrt{\frac{\sigma^2}{n}} - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right), \end{aligned}$$

der $\Phi(\cdot)$ er kumulativ fordeling i standard normalfordeling. Innsatt $n = 3$, $z_\alpha = 1.645$ og $\sigma^2 = 0.4^2$ blir styrkefunksjonen dermed seende slik ut.



Ønsker så å finne ut hvor stor n må være for at sannsynligheten for å forkaste H_0 skal være minst 0.9 når $\mu = 2.9$. Matematisk kan dette kravet uttrykkes som

$$P(\text{Forkast } H_0|\mu = 2.9) \geq 0.9 \quad \Leftrightarrow \quad 1 - \beta(2.9) \geq 0.9$$

Ved å bruke uttrykket vi fant for styrkefunksjonen (før vi satte inn for n) blir altså kravet

$$\Phi\left(\frac{3 - z_\alpha\sqrt{\frac{\sigma^2}{n}} - 2.9}{\sqrt{\frac{\sigma^2}{n}}}\right) \geq 0.9.$$

Ved å lage en skisse av sannsynlighetstettheten av en standard normalfordeling ser man at dette kravet er ekvivalent med

$$\begin{aligned}
\frac{3 - z_\alpha\sqrt{\frac{\sigma^2}{n}} - 2.9}{\sqrt{\frac{\sigma^2}{n}}} \geq z_{0.1} \quad &\Leftrightarrow \quad 3 - z_\alpha\sqrt{\frac{\sigma^2}{n}} - 2.9 \geq z_{0.1}\sqrt{\frac{\sigma^2}{n}} \\
&\Leftrightarrow \quad 3 - 2.9 \geq (z_\alpha + z_{0.1})\sqrt{\frac{\sigma^2}{n}} \\
&\Leftrightarrow \quad 0.1 \geq (1.645 + 1.282)\sqrt{\frac{\sigma^2}{n}} \\
&\Leftrightarrow \quad \frac{\sigma^2}{n} \leq \left(\frac{0.1}{1.645 + 1.282}\right)^2 \\
&\Leftrightarrow \quad \sigma^2\left(\frac{1.645 + 1.282}{0.1}\right)^2 \leq n \\
&\Leftrightarrow \quad n \geq 0.4^2\left(\frac{1.645 + 1.282}{0.1}\right)^2 = 137.08.
\end{aligned}$$

Siden antall poser som skal testes selvfølgelig må være et heltall må man følgelig teste minst 138 poser.

d) Starter med å utlede kumulativ fordelingsfunksjon for $X_{(2)}$,

$$\begin{aligned}
F_{X_{(2)}}(x) &= P(X_{(2)} \leq x) = P(\text{Minst to av } X_1, X_2, X_3 \leq x) \\
&= P(\text{Nøyaktig to av } X_1, X_2, X_3 \leq x) + P(\text{Nøyaktig tre av } X_1, X_2, X_3 \leq x) \\
&= \binom{3}{2} F_X(x)^2 (1 - F_X(x))^{3-2} + \binom{3}{3} F_X(x)^3 (1 - F_X(x))^{3-3} \\
&= 3F_X(x)^2(1 - F_X(x)) + F_X(x)^3 \\
&= 3F_X(x)^2 - 3F_X(x)^3 + F_X(x)^3 \\
&= 3F_X(x)^2 - 2F_X(x)^3.
\end{aligned}$$

Finner så sannsynlighetstettheten til $X_{(2)}$ ved å derivere,

$$\begin{aligned}
f_{X_{(2)}}(x) &= F'_{x_{(2)}}(x) = 3 \cdot 2F_X(x) \cdot f_X(x) - 2 \cdot 3F_X(x)^2 f_X(x) \\
&= \underline{\underline{6F_X(x) f_X(x) \left[1 - F_X(x)\right]}},
\end{aligned}$$

der $F_X(x)$ og $f_X(x)$ er henholdsvis kumulativ fordeling og sannsynlighetstettheten i en normalfordeling med forventningsverdi μ og varians σ^2.

Man kan tenke seg å undersøke om $\widetilde{\mu} = X_{(2)}$ er en forventningsrett estimator for μ ved å regne ut $\mathrm{E}\left[\widetilde{\mu}\right]$, men man vil da ende opp med et integral som er svært vanskelig å evaluere analytisk. Det som gir enklere regning er å vise at $f_{X_{(x)}}(x)$ er symmetrisk om $x = \mu$. Man må altså vise at $f_{X_{(2)}}(\mu + \delta) = f_{X_{(2)}}(\mu - \delta)$ for alle $\delta > 0$.

Siden $f_X(x)$ er sannsynlighetstettheten til en normalfordeling med forventingsverdi μ vet vi at $f_X(x)$ er symmetrisk om μ, dvs.

$$f_X(\mu + \delta) = f_X(\mu - \delta) \quad \text{for alle } \delta > 0. \tag{3.1}$$

Når X er normalfordelt med forventningsverdi μ har vi også at $P(X > \mu + \delta) = P(X < \mu - \delta)$ for alle $\delta > 0$. Siden $P(X > \mu + \delta) = 1 - P(X \leq \mu + \delta) = 1 - F_X(\mu + \delta)$ og $P(X < \mu - \delta) = P(X \leq \mu - \delta) = F_X(\mu - \delta)$ har vi dermed at

$$F_X(\mu - \delta) = 1 - F_X(\mu + \delta) \quad \text{for alle } \delta > 0. \tag{3.2}$$

Ved å bruke (3.1) og (3.2) og uttrykket vi fant for $f_{X_{(2)}}(x)$ får vi da at

$$\begin{aligned}
f_{X_{(2)}}(\mu + \delta) &= 6F_X(\mu + \delta) f_X(\mu + \delta) \left[1 - F_X(\mu + \delta)\right] \\
&= 6\left[1 - F_X(\mu - \delta)\right] f_X(\mu - \delta) F_X(\mu - \delta) \\
&= 6F_X(\mu - \delta) f_X(\mu - \delta) \left[1 - F_X(\mu - \delta)\right] \\
&= f_{X_{(2)}}(\mu - \delta).
\end{aligned}$$

Vi har dermed vist at $f_{X_{(2)}}(x)$ er symmetrisk om μ, og dermed blir

$$\underline{\underline{\mathrm{E}\left[\widetilde{\mu}\right] = \mathrm{E}\left[X_{(2)}\right] = \mu}}.$$