Løsningsforslag øving 4

Oppgaver fra boka:

8.13 / 8.2: 13

Når man skal regne ut s^2 kan det lønne seg å begynne med å skrive om s^2 (se også side 199) f.eks. til:

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2) = \frac{1}{n-1}(\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2) \\ &= \frac{1}{n-1}(\sum_{i=1}^{n}x_i^2-2\bar{x}n\bar{x}+n\bar{x}^2) = \frac{1}{n-1}(\sum_{i=1}^{n}x_i^2-n\bar{x}^2) \end{aligned}$$

For de oppgitte dataene får vi $\bar{x}=\sum_{i=1}^{20}x_i/20 = 53.3/20 = 2.665$ og $\sum_{i=1}^{20}x_i^2 = 148.55$ som gir

$$s^2 = \frac{1}{20-1}(148.55-20\cdot 2.665^2) = 0.342$$

Empirisk standardavvik (=estimert standardavvik) blir da: $s=\sqrt{0.342}=\underline{\underline{0.585}}$. (Alternativt kan man finne dette direkte på kalkulatoren...)

8.14 / 8.2: 14

a) Legger til en konstant c til alle data $x_1,\dots,x_n$: Dataene blir da $x_1+c,\dots,x_n+c$, mens gjennomsnittet blir $\bar{x}+c$. Får da:

$$S^2_{X+c} = \frac{1}{n-1}\sum_{i=1}^{n}((x_i+c)-(\bar{x}+c))^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \underline{\underline{S^2_X}}$$

b) Multipliserer alle data $x_1,\dots,x_n$ med en konstant c: Dataene blir da $cx_1,\dots,cx_n$, mens gjennomsnittet blir $c\bar{x}$. Får da:

$$S^2_{cX} = \frac{1}{n-1}\sum_{i=1}^{n}(cx_i-c\bar{x})^2 = \frac{c^2}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \underline{\underline{c^2S^2_X}}$$

8.21 / 8.5: 5

Siden $\sigma_{\bar{X}} = \mathrm{SD}(\bar{X}) = \sqrt{\mathrm{Var}(\bar{X})} = \sqrt{\sigma^2/n} = \sigma/\sqrt{n}$ og $\mu_{\bar{X}} = \mathrm{E}(\bar{X}) = \mu$ får vi med $n=40$, $\sigma=15$ og $\mu=240$ (som gjelder dersom maskinen er korrekt innstilt!) at maskinen ansees for å være korrekt innstilt dersom målt gjennomsnitt av 40 er i intervallet

$$[\mu_{\bar{X}}-2\sigma_{\bar{X}},\mu_{\bar{X}}+2\sigma_{\bar{X}}] = [240-2\cdot 15/\sqrt{40}, 240+2\cdot 15/\sqrt{40}] = [235.3, 244.7]$$

Siden 236 er i dette intervallet gjorde de $\underline{\underline{\text{rett}}}$ i henhold til kravet sitt.

Merk forøvrig at fra tommelfingerregelen om at ofte vil rundt 95% av målingene ligge i et intervall pluss/minus to standardavvik fra forventningsverdien får vi her at dersom maskinen er korrekt innstilt er det rundt 95% sannsynlighet for å få en verdi i det aktuelle intervallet.

9.81 / 9.15:1

Kanskje litt uklart formulert oppgavetekst i denne oppgave. Når man sier at $x_1,\dots,x_n$ er en Bernoulli-prosess (eller utfall i en binomisk forsøksrekke) med parameter p, betyr dette at $x_1,\dots,x_n$ er utfall av 0-1 variable $X_1,\dots,X_n$ med fordeling

$$f(x) = p^x(1-p)^{(1-x)} \qquad ,x=0,1$$

(dvs $P(X=1)=p$ og $P(X=0)=1-p$). Vi får da

$$\begin{aligned} L(p;x_1,\dots,x_n) &\overset{\text{uavh.}}{=} \prod_{i=1}^{n}f(x_i,p) = \prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i} \\ l(p;x_i,\dots,x_n) &= \ln L(p;x_i,\dots,x_n) = \sum_{i=1}^{n}x_i\ln p + \left(n-\sum_{i=1}^{n}x_i\right)\ln(1-p) \\ \frac{\partial l(p;x_1,\dots,x_n)}{\partial p} &= \frac{1}{p}\sum_{i=1}^{n}x_i - \frac{1}{1-p}\left(n-\sum_{i=1}^{n}x_i\right) = 0 \\ &\Rightarrow \quad (1-p)\sum_{i=1}^{n}x_i - p\left(n-\sum_{i=1}^{n}x_i\right) = 0 \\ & \qquad \sum_{i=1}^{n}x_i - p\sum_{i=1}^{n}x_i - pn + p\sum_{i=1}^{n}x_i = 0 \\ &\Rightarrow \quad \underline{\underline{\hat{p} = \frac{1}{n}\sum_{i=1}^{n}X_i}} \end{aligned}$$

Sjekker at vi har funnet maksimum:

$$\frac{\partial^2 l(p;x_1,\dots,x_n)}{\partial p^2} = -\frac{1}{p^2}\sum_{i=1}^{n}x_i - \frac{1}{(1-p)^2}\left(n-\sum_{i=1}^{n}x_i\right) < 0 \qquad \text{dvs maksimum!}$$

Vi kan også undersøke om estimatoren er forventningsrett. Merk først at $\mathrm{E}(X)=\sum_x xf(x) = 0(1-p)+1p=p$. Dermed:

$$E[\hat{p}] = E[\frac{1}{n}\sum_{i=1}^{n}X_i] = \frac{1}{n}\sum_{i=1}^{n}EX_i = \frac{1}{n}\sum_{i=1}^{n}p = \underline{p}$$

Vi ser at $E[\hat{p}]=p$. Det betyr at estimatoren er forventningsrett.

9.82 / 9.15: 2

a)

$$\begin{aligned} L(\alpha,\beta;x_1,..x_n) &= \prod_{i=1}^{n} f(x_i;\alpha,\beta) = \prod_{i=1}^{n} \alpha\beta x_i^{\beta-1} e^{-\alpha x_i^\beta} \\ &= (\alpha\beta)^n (\prod_{i=1}^{n} x_i^{\beta-1}) e^{-\sum_{i=1}^n \alpha x_i^\beta} = \underline{\underline{(\alpha\beta)^n e^{-\alpha\sum_{i=1}^n x_i^\beta} \prod_{i=1}^{n} x_i^{\beta-1}}} \end{aligned}$$

b)

$$\begin{aligned} l(\alpha,\beta;x_1,..x_n) &= \ln L(\alpha,\beta;x_1,..x_n) = n\ln(\alpha) + n\ln(\beta) - \alpha\sum_{i=1}^{n} x_i^\beta + \sum_{i=1}^{n}\ln(x_i^{\beta-1}) \\ &= n\ln(\alpha) + n\ln(\beta) - \alpha\sum_{i=1}^{n} x_i^\beta + (\beta-1)\sum_{i=1}^{n}\ln(x_i) \end{aligned}$$

Siden vi her skal optimere over to parametre, α og β må vi løse ligningene $\frac{\partial l(\alpha,\beta)}{\partial\alpha} = 0$ og $\frac{\partial l(\alpha,\beta)}{\partial\beta} = 0$ samtidig, dvs løse ligningsettet:

$$\begin{aligned} \frac{\partial l(\alpha,\beta)}{\partial\alpha} &= \frac{n}{\alpha} - \sum_{i=1}^{n} x_i^\beta = 0 \\ \frac{\partial l(\alpha,\beta)}{\partial\beta} &= \frac{n}{\beta} - \alpha\sum_{i=1}^{n} x_i^\beta \ln(x_i) + \sum_{i=1}^{n}\ln(x_i) = 0 \end{aligned}$$

Første ligning gir et enkelt uttrykk for α ($\alpha = n/\sum_{i=1}^n x_i^\beta$) som vi kan sette inn i andre ligning og så prøve å løse denne med hensyn på β. Denne ligningen lar seg imidlertid ikke løse analytisk, numeriske metoder må brukes for å finne en løsning.

7.7 / 7.3: 7

$$f(v) = \begin{cases} kv^2\exp(-bv^2) & ,v>0 \\ 0 & ,\text{ellers} \end{cases}$$

$W = \frac{1}{2}mV^2$, $v = \sqrt{\frac{2w}{m}}$, $\frac{dv}{dw} = \frac{1}{2\sqrt{\frac{2w}{m}}}\cdot\frac{2}{m} = \underline{\frac{1}{\sqrt{2mw}}}$

$$g(w) = f(\sqrt{\frac{2w}{m}})|\frac{dv}{dw}| = k\,\frac{2w}{m}\exp(-b\frac{2w}{m})\underline{\frac{1}{\sqrt{2mw}}} = \frac{\sqrt{2}k}{m^{3/2}}w^{1/2}\exp(-\frac{2b}{m}w) \qquad ,w>0$$

Hvis vi skriver om uttrykket litt:

$$g(w) = \frac{\sqrt{2}k}{m^{3/2}}w^{\frac{3}{2}-1}\exp(-\frac{2b}{m}w)$$

ser vi at $\underline{\underline{\text{dette er en gammafordeling med } \alpha = 3/2 \text{ og } \beta = \frac{m}{2b}}}$.

At det er en gammafordelingen ser vi fra formen på leddene hvor variabelen w inngår. Hvilken verdi konstanten k har kan vi nå bestemme ved å sammenligne konstantleddet i vårt uttrykk med konstanleddet i en gammafordeligne med $\alpha = 3/2$ og $\beta = \frac{m}{2b}$. Vi får da at:

$$\frac{\sqrt{2}k}{m^{3/2}} = \Big((\frac{m}{2b})^{\frac{3}{2}}\Gamma(3/2)\Big)^{-1}$$

Dvs vi får at konstanten k må være:

$$k = \frac{m^{3/2}/\sqrt{2}}{(\frac{m}{2b})^{\frac{3}{2}}\Gamma(3/2)} = \frac{2^{-1/2}(2b)^{3/2}}{(\frac{3}{2}-1)\Gamma(1/2)} = \frac{4b^{3/2}}{\sqrt{\pi}}.$$

7.9 / 7.3: 9

$$f(x) = \begin{cases} \frac{32}{(x+4)^3} & ,x>0 \\ 0 & ,\text{ellers} \end{cases}$$

a) Med $Y = X+4$ får vi $X = Y-4 = w(Y)$ og $w'(Y) = 1$ som gir

$$g(y) = f(y-4)\cdot 1 = \frac{32}{(y)^3} \quad ,y>4$$

b)

$$P(Y>8) = \int_8^\infty g(y)dy = \int_8^\infty \frac{32}{(y)^3} = [-\frac{16}{(y)^2}]_8^\infty = \frac{16}{(8)^2} = \underline{\underline{0.25}}$$

Oppgave 1:

a)

$$\begin{aligned} \mathrm{E}(\hat\mu_1) &= \frac{1}{25}[\sum_{i=1}^{10}\mathrm{E}(X_i) + \sum_{i=1}^{15}\mathrm{E}(Y_i)] = \frac{1}{25}[\sum_{i=1}^{10}\mu + \sum_{i=1}^{15}\mu] = \frac{1}{25}[10\mu + 15\mu] = \underline{\underline{\mu}} \\ \mathrm{E}(\hat\mu_2) &= \frac{6}{7}\mathrm{E}(\bar X) + \frac{1}{7}\mathrm{E}(\bar Y) = \frac{6}{7}\mu + \frac{1}{7}\mu = \underline{\underline{\mu}} \end{aligned}$$

Dvs begge estimatorene er forventningsrette.

$$\begin{aligned} \mathrm{Var}(\hat\mu_1) &= \mathrm{Var}(\frac{1}{25}[\sum_{i=1}^{10}X_i + \sum_{i=1}^{15}Y_i]) = \frac{1}{25^2}\mathrm{Var}[\sum_{i=1}^{10}X_i + \sum_{i=1}^{15}Y_i] \\ &\overset{\text{uavh.}}{=} \frac{1}{25^2}[\sum_{i=1}^{10}\mathrm{Var}(X_i) + \sum_{i=1}^{15}\mathrm{Var}(Y_i)] = \frac{1}{25^2}[10\cdot 0.01 + 15\cdot 0.09] = \underline{0.00232} \\ \mathrm{Var}(\hat\mu_2) &= \mathrm{Var}(\frac{6}{7}\bar X + \frac{1}{7}\bar Y) \overset{\text{uavh.}}{=} (\frac{6}{7})^2\mathrm{Var}(\bar X) + (\frac{1}{7})^2\mathrm{Var}(\bar Y) \\ &= (\frac{6}{7})^2\frac{0.01}{10} + (\frac{1}{7})^2\frac{0.09}{15} = \underline{0.000857} \end{aligned}$$

Siden begge estimatorene er forventningsrette og $\mathrm{Var}(\hat\mu_1) > \mathrm{Var}(\hat\mu_2)$ er $\underline{\underline{\hat\mu_2 \text{ den beste estimatoren}}}$.

b) Siden vi ønsker en forventningsrett estimator må:

$$\mathrm{E}(\hat{\mu}) = a\mathrm{E}(\bar{X}) + b\mathrm{E}(\bar{Y}) = a\mu + b\mu = \mu$$

Dvs vi må ha $a + b = 1$. Videre ønsker vi en estimator med minst mulig varians. Estimatorens varians blir:

$$\begin{aligned}\mathrm{Var}(\hat{\mu}) &\overset{\text{uavh.}}{=} a^2\mathrm{Var}(\bar{X}) + b^2\mathrm{Var}(\bar{Y}) = a^2\frac{0.01}{10} + b^2\frac{0.09}{15} \\ &= 0.001a^2 + 0.006b^2\end{aligned}$$

Dvs for å finne den forventningsrette estimatoren med minst mulig varians, må vi finne de verdiene a og b der $a + b = 1$ og som minimerer $0.001a^2 + 0.006b^2$. Setter vi inn $b = 1 - a$ i dette uttrykker får vi at vi må minimere:

$$\begin{aligned}V(a) &= 0.001a^2 + 0.006(1-a)^2 = 0.007a^2 - 0.012a + 0.006 \\ V'(a) &= 0.014a - 0.012 = 0 \\ &\Rightarrow\ a = \frac{0.012}{0.014} = \underline{\underline{\frac{6}{7}}} \\ &\Rightarrow\ b = 1 - \frac{6}{7} = \underline{\underline{\frac{1}{7}}}\end{aligned}$$

Dvs vi ser at $\hat{\mu}_2$ fra punkt a) er den beste forventningsrette estimatoren i dette tilfellet!

Oppgave 2:

a) Med $Y = 2X/\beta$ får vi $X = \beta Y/2 = w(Y)$ og dermed $w'(Y) = \beta/2$ slik at transformasjonsformelen gir:

$$g(y) = f(w(y))|w\prime(y)| = \frac{1}{\beta}e^{-(\beta y/2)/\beta}\frac{\beta}{2} = \underline{\underline{\frac{1}{2}e^{-y/2}}}$$

Dersom vi setter $\nu = 2$ i sannsynlighetstettheten til χ^2-fordelingen får vi:

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2} \overset{\nu=2}{=} \frac{1}{2^{2/2}\Gamma(2/2)}x^{2/2-1}e^{-x/2} = \frac{1}{2}e^{-x/2}$$

dvs vi ser at fordelingen til Y er en χ^2_2-fordeling.

b) Resultatet sier et en sum av uavhengige χ^2-fordelte variable er χ^2-fordelt med parameter ("frihetsgrader") lik summen av parametrene i fordelingene til enkeltvariablene. Dvs i vårt tilfelle vil $Z = \sum_{i=1}^n Y_i = 2(\sum_{i=1}^n X_i)/\beta$ være χ^2-fordelt med parameter $\sum_{i=1}^n 2 = 2n$, dvs Z er $\underline{\underline{\chi^2_{2n}\text{-fordelt}}}$.

c) $Z = 2V/\beta = w(V)$ og dermed er $w'(V) = 2/\beta$. Siden Z er χ^2_{2n}-fordelt får vi nå:

$$\begin{aligned}f_V(v) &= f_Z(w(v))|w'(v)| = \frac{1}{2^{2n/2}\Gamma(2n/2)}(2v/\beta)^{2n/2-1}e^{-(2v/\beta)/2}\frac{2}{\beta} \\ &= \frac{1}{2^n\Gamma(n)}\frac{2^{n-1}v^{n-1}}{\beta^{n-1}}e^{-v/\beta}\frac{2}{\beta} \\ &= \underline{\underline{\frac{1}{\beta^n\Gamma(n)}v^{n-1}e^{-v/\beta}}}\end{aligned}$$

og sammenligner vi dette med gammafordelingen

$$f(x) = \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}$$

ser vi at $V = \sum_{i=1}^n X_i$ er gammafordelt med $\alpha = n$ og $\beta = \beta$.

Vi har altså her vist resultatet at en sum av uavhengige identisk eksponensialfordelte variable er gammafordelt med $\alpha = n$ og $\beta = \beta$.

Oppgave 3:

La X_1 betegne resultatet av en måling med metode 1 og X_2 betegne resultatet av en måling med metode 2.

a)

$$\begin{aligned}P(X_1 < 4.0) &= P(Z < \frac{4.0 - 4.3}{0.5}) = P(Z < -0.60) = \underline{\underline{0.2743}} \\ P(X_2 > \mu + \sigma_2) &= 1 - P(X_2 \le \mu + \sigma_2) = 1 - P(Z \le \frac{\mu + \sigma_2 - \mu}{\sigma_2}) \\ &= 1 - P(Z \le 1) = 1 - 0.8413 = \underline{\underline{0.1587}}\end{aligned}$$

b)

$$\begin{aligned}L(\mu) &= \prod_{i=1}^n f(x_i;\mu) = f(x_1;\mu)f(x_2;\mu) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1}e^{-(x_1-\mu)^2/(2\sigma_1^2)}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-(x_2-\mu)^2/(2\sigma_2^2)} \\ &= \frac{1}{2\pi\sigma_1\sigma_2}e^{-(x_1-\mu)^2/(2\sigma_1^2)-(x_2-\mu)^2/(2\sigma_2^2)} \\ l(\mu) &= \ln L(\mu) = \ln(1) - \ln(2\pi\sigma_1\sigma_2) - \frac{(x_1-\mu)^2}{2\sigma_1^2} - \frac{(x_2-\mu)^2}{2\sigma_2^2}\end{aligned}$$

Deriverer denne med hensyn på μ og setter lik null:

$$\begin{aligned}\frac{\partial l(\mu)}{\partial\mu} = -\frac{2(x_1-\mu)(-1)}{2\sigma_1^2} - \frac{2(x_2-\mu)(-1)}{2\sigma_2^2} &= 0 \\ \frac{x_1-\mu}{\sigma_1^2} + \frac{x_2-\mu}{\sigma_2^2} &= 0 \\ \frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} &= \frac{\mu}{\sigma_1^2} + \frac{\mu}{\sigma_2^2} \\ \sigma_2^2x_1 + \sigma_1^2x_2 &= \sigma_2^2\mu + \sigma_1^2\mu \\ \hat{\mu} &= \underline{\underline{\frac{\sigma_2^2X_1 + \sigma_1^2X_2}{\sigma_1^2 + \sigma_2^2}}}\end{aligned}$$

c)

$$\begin{aligned}
\mathrm{E}(\hat{\mu}) &= \mathrm{E}\left(\frac{\sigma_2^2 X_1 + \sigma_1^2 X_2}{\sigma_1^2+\sigma_2^2}\right) = \frac{\sigma_2^2 \mathrm{E}(X_1) + \sigma_1^2 \mathrm{E}(X_2)}{\sigma_1^2+\sigma_2^2} = \frac{\sigma_2^2\mu + \sigma_1^2\mu}{\sigma_1^2+\sigma_2^2} = \underline{\underline{\mu}} \\
\mathrm{Var}(\hat{\mu}) &= \mathrm{Var}\left(\frac{\sigma_2^2 X_1 + \sigma_1^2 X_2}{\sigma_1^2+\sigma_2^2}\right) = \mathrm{Var}\left(\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}X_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}X_2\right) \\
&= \left(\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^2 \mathrm{Var}(X_1) + \left(\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2 \mathrm{Var}(X_2) = \frac{\sigma_2^4\sigma_1^2 + \sigma_1^4\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2} \\
&= \frac{\sigma_1^2\sigma_2^2(\sigma_2^2+\sigma_1^2)}{(\sigma_1^2+\sigma_2^2)^2} = \underline{\underline{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}}
\end{aligned}$$

d) Siden $\hat{\mu}$ er en lineærkombinasjon av de to uavhengige og normalfordelte variablene X_1 og X_2 er også $\hat{\mu}$ normalfordelt. Med forventning og varians som regnet ut i c) får vi at

$$\frac{\hat{\mu}-\mu}{\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}} \sim N(0,1)$$

og et konfidensintervall finner vi da ved:

$$\begin{aligned}
P(-z_{\alpha/2} \le \frac{\hat{\mu}-\mu}{\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}} \le z_{\alpha/2}) &= 1-\alpha \\
P(-z_{\alpha/2}\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \le \hat{\mu}-\mu \le z_{\alpha/2}\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}) &= 1-\alpha \\
P(\hat{\mu} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \le \mu \le \hat{\mu} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}) &= 1-\alpha
\end{aligned}$$

Med $\alpha = 0.05$ får vi at $z_{\alpha/2} = z_{0.025} = 1.96$ og et 95% konfidensintervall er gitt ved

$$\underline{\underline{\left[\hat{\mu} - 1.96\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}},\ \hat{\mu} + 1.96\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right]}}$$

Innsatt observerte tall får vi at $\hat{\mu} = \frac{0.2^2 \cdot 4.41 + 0.5^2 \cdot 4.19}{0.5^2+0.2^2} = 4.220$ og $\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} = \frac{0.5^2\cdot 0.2^2}{0.5^2+0.2^2} = 0.0345$ som gir konfidensintervallet:

$$[4.220 - 1.96\cdot\sqrt{0.0345},\ 4.220 + 1.96\cdot\sqrt{0.0345}] = \underline{\underline{[3.86, 4.58]}}$$