Eksamen i STK1100 våren 2022 - løsningsforslag

Oppgave 1

a

La A være hendelsen at terningskasteren trekker terning 1 og la B være hendelsen at det blir en sekser. Da er $\mathrm{P}(A) = \frac{1}{2}$, $\mathrm{P}(B|A) = \frac{1}{5}$ og $\mathrm{P}(B|A') = \frac{1}{6}$. Vi får:

$$\begin{aligned}
\mathrm{P}(B) =& \mathrm{P}(B \cap A) + \mathrm{P}(B \cap A') \\
=& \mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A')\mathrm{P}(A') \\
=& \mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A')(1 - \mathrm{P}(A)) = \frac{1}{5} \cdot \frac{1}{2} + \frac{1}{6} \cdot \left(1 - \frac{1}{2}\right) = \frac{11}{60} \approx 0.183.
\end{aligned}$$

b

Vi har:

$$\begin{aligned}
\mathrm{P}(A|B') =& \frac{\mathrm{P}(A \cap B')}{\mathrm{P}(B')} \\
=& \frac{\mathrm{P}(B'|A)\mathrm{P}(A)}{1 - \mathrm{P}(B)} \\
=& \frac{(1 - \mathrm{P}(B|A))\mathrm{P}(A)}{1 - \mathrm{P}(B)} \\
=& \frac{\left(1 - \frac{1}{5}\right) \cdot \frac{1}{2}}{\left(1 - \frac{11}{60}\right)} = \frac{24}{49} \approx 0.490.
\end{aligned}$$

Oppgave 2

a

Her kan en bruke $\mathrm{E}(X_i) = M'_X(0)$, $\mathrm{E}(X_i^2) = M''_X(0)$ og $\mathrm{V}(X_i) = \mathrm{E}(X_i^2) - \mathrm{E}(X_i)^2$. En løsning som krever litt mindre regning er å bruke kumulantgenererende funksjon i stedet, dvs. $R_X(t) = \ln(M_X(t)) = \lambda(e^t - 1)$. Vi har:

$$\begin{aligned}
R'_X(t) &= \lambda e^t \\
R''_X(t) &= \lambda e^t.
\end{aligned}$$

Dermed er $\mathrm{E}(X_i) = R'_X(0) = \lambda e^0 = \lambda$ og $\mathrm{V}(X_i) = R''_X(0) = \lambda = \mathrm{E}(X_i)$.

b

Vi har:

$$\mathrm{E}(Y) = \mathrm{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \underbrace{\mathrm{E}(X_i)}_{=\lambda} = n\lambda.$$

Da $X_1, \ldots, X_n$ er uavhengige, får vi:

$$\mathrm{V}(Y) = \mathrm{V}(\sum_{i=1}^{n} X_i) \overset{uavh.}{=} \sum_{i=1}^{n} \underbrace{\mathrm{V}(X_i)}_{=\lambda} = n\lambda.$$

c

Da $X_1, \ldots, X_n$ er uavhengige, får vi:

$$\begin{aligned} M_Y(t) = M_{\sum_{i=1}^{n} X_i}(t) &\overset{uavh.}{=} M_{X_1}(t) \cdot \ldots M_{X_n}(t) \\ &= \prod_{i=1}^{n} e^{\lambda(e^t-1)} \\ &= e^{\sum_{i=1}^{n} \lambda(e^t-1)} = e^{n\lambda(e^t-1)}. \end{aligned}$$

Vi gjenkjenner $M_Y(t)$ som den momentgenererende funksjonen til Poisson-fordelingen med parameter $n\lambda$. Altså må Y være Poisson-fordelt med parameter $n\lambda$.

d

Likelihood-funksjonen er gitt ved:

$$L(\lambda) = p(x_1, \ldots, x_n; \lambda) \overset{uif}{=} \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} \frac{\lambda_i^x}{x_i!} e^{-\lambda} = e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{1}{x_i!}.$$

Da blir log-likelihood-funksjonen:

$$l(\lambda) = \ln(L(\lambda)) = -n\lambda + \ln(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i) = -n\lambda + \ln(\lambda) \cdot n\bar{x} - \sum_{i=1}^{n} \ln(x_i).$$

For å finne maksimum likelihood-estimatet $\widehat{\lambda}_{mle}$, må vi finne den verdien av λ som maksimerer $L(\lambda)$, som også maksimerer $l(\lambda)$. Det gjør vi ved å derivere $l(\lambda)$ m.h.p. λ og sette lik 0. Vi har:

$$l'(\lambda) = -n + \frac{n\bar{x}}{\lambda}$$

slik at

$$-n + \frac{n\bar{x}}{\widehat{\lambda}_{mle}} = 0 \rightarrow \widehat{\lambda}_{mle} = \bar{x}.$$

Altså er maksimum likelihood-estimatoren for λ gitt ved $\widehat{\lambda}_{mle} = \bar{X}$.

Momentestimatoren $\widehat{\lambda}_{mom}$ for λ finner en ved å avstemme det teoretiske førstemomentet $\mathrm{E}(X_i) = \lambda$ mot det empiriske førstemomentet $\bar{X}$ og løse for λ. Da får vi umiddelbart $\widehat{\lambda}_{mom} = \bar{X}$.

e

Vi har:

$$\sigma^2_{\widehat{\lambda}} = \mathrm{V}_{\widehat{\lambda}_{mle}} = \mathrm{V}(\bar{X}) = \frac{\mathrm{V}(X_i)}{n} = \frac{\lambda}{n}.$$

Dermed er standardfeilen gitt ved $\sigma_{\widehat{\lambda}} = \sqrt{\frac{\lambda}{n}}$. Denne kan estimeres med $\hat{\sigma}_{\widehat{\lambda}} = \sqrt{\frac{\widehat{\lambda}_{mle}}{n}} = \sqrt{\frac{\bar{x}}{n}}$. Når vi setter inn observerte tall får vi $\widehat{\lambda}_{mle} = 20.37$ og $\hat{\sigma}_{\widehat{\lambda}} = \sqrt{20.37/43} \approx 0.69$.

f

Fra Oppgave a vet vi at $\mathrm{E}(X_i) = \mathrm{V}(X_i) = \lambda$. Den empiriske variansen s^2 til tellingene er imidlertid mye høyere enn snittet, hvilket passer dårlig med antakelsen om lik forventning og varians. Det tyder på at Poisson-fordelingen ikke passer så godt i dette tilfellet.

Oppgave 3

a

Vi begynner med å finne den kumulative fordelingsfunksjonen til Y, og konstaterer at $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$:

$$\begin{aligned}
F_Y(y) = \mathrm{P}(Y \leq y) =& \mathrm{P}\left(\left(\frac{X-\mu}{\sigma}\right)^2 \leq y\right) \\
=& \mathrm{P}\left(-\sqrt{y} \leq \frac{X-\mu}{\sigma} \leq \sqrt{y}\right) \\
=& \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = \Phi(\sqrt{y}) - (1 - \Phi(\sqrt{y})) = 2\Phi(\sqrt{y}) - 1,
\end{aligned}$$

der $\Phi(\cdot)$ er den kumulative fordelingsfunksjonen til standard normalfordeling. Tettheten til Y finner vi så ved å derivere $F_Y(y)$ m.h.p. y. La $\phi(\cdot)$ være sannsynlighetstettheten til standard normalfordeling, dvs. $f_X(\cdot)$ med $\mu = 0$ og $\sigma = 1$. Vi får da:

$$\begin{aligned} f_Y(y) = F_Y'(y) =& 2\phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\ =& 2\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \cdot \frac{1}{2\sqrt{y}} \\ =& \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}. \end{aligned}$$

Dette er tettheten for $y > 0$, mens den er 0 for $y \leq 0$, og er altså *gamma* $\left(\frac{1}{2}, 2\right)$.

b

Vi har:

$$n\frac{\hat{\sigma}^2}{\sigma^2} = \frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} Y_i.$$

Videre er $Y_1, \ldots, Y_n$ uavhengige da $X_1, \ldots, X_n$ er det, med $Y_i \sim$ *gamma* $\left(\frac{1}{2}, 2\right)$ (fra Oppgave a). Dermed er $n\frac{\hat{\sigma}^2}{\sigma^2}$ en sum av uavhengige gammafordelte variabler med samme skalaparameter $\beta = 2$, og må derfor også være gammafordelt, nærmere bestemt $n\frac{\hat{\sigma}^2}{\sigma^2} \sim$ *gamma* $\left(\frac{n}{2}, 2\right)$.

c

Vi vet at $n\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_n$. Dermed har vi

$$\begin{aligned} & \mathrm{P}\left(\chi^2_{n,0.975} \leq n\frac{\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{n,0.025}\right) = 0.95 \\ \to & \mathrm{P}\left(\frac{\chi^2_{n,0.975}}{n\hat{\sigma}^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{n,0.025}}{n\hat{\sigma}^2}\right) = 0.95 \\ \to & \mathrm{P}\left(\hat{\sigma}^2\frac{n}{\chi^2_{n,0.975}} \geq \sigma^2 \geq \hat{\sigma}^2\frac{n}{\chi^2_{n,0.025}}\right) = 0.95. \end{aligned}$$

Et 95% konfidensintervall for σ^2 er dermed gitt ved

$$\left(\hat{\sigma}^2\frac{n}{\chi^2_{n,0.025}}, \hat{\sigma}^2\frac{n}{\chi^2_{n,0.975}}\right).$$

Oppgave 4

a
Når $\rho = 0$, har vi

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\left(\frac{x-\mu_1}{\sigma^1}\right)^2+\left(\frac{y-\mu_2}{\sigma^2}\right)^2\right)}, \quad (x,y) \in \mathbb{R}^2.$$

Videre har vi

$$\begin{aligned} f_X(x) \cdot f_Y(y) =& \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2} \\ =& \frac{1}{2\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\left(\frac{x-\mu_1}{\sigma^1}\right)^2+\left(\frac{y-\mu_2}{\sigma^2}\right)^2\right)}, \quad x \in \mathbb{R}, y \in \mathbb{R}. \end{aligned}$$

Altså er $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$, $\forall (x,y)$. Dermed må X og Y være uavhengige.

b
I den lineære regresjonsmodellen for $[Y|X = x]$, er Y gitt som en lineær kombinasjon av en normalfordelt variabel ε og en konstant $\beta_0 + \beta_1 x$, og må derfor også være normalfordelt. Videre har vi:

$$\begin{aligned} \mathrm{E}(Y|X=x) =& \mathrm{E}(\beta_0 + \beta_1 x + \varepsilon) = \beta_0 + \beta_1 x + \underbrace{\mathrm{E}(\varepsilon)}_{=0} = \beta_0 + \beta_1 x \\ \mathrm{V}(Y|X=x) =& \mathrm{V}(\beta_0 + \beta_1 x + \varepsilon) = \mathrm{V}(\varepsilon) = \psi^2. \end{aligned}$$

c
Den betingede fordelingen til $[Y|X = x]$ i den binormale fordelingen er en normalfordeling med en forventning som er en lineær funksjon av x og en konstant varians, akkurat som i den lineære regresjonsmodellen. For at det skal bli samme modell, må forventningen og variansen til $[Y|X = x]$ være den samme i de to, dvs.

$$\mathrm{E}(Y|X=x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 + \rho\frac{\sigma_2}{\sigma_1}x = \beta_0 + \beta_1 x$$

og

$$\mathrm{V}(Y|X=x) = \sigma_2^2(1-\rho^2) = \psi^2.$$

Altså kan den betingede fordelingen til $[Y|X = x]$ i den binormale fordelingen sees som en lineær regresjonsmodell med

$$\beta_0 = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 \quad \beta_1 = \rho\frac{\sigma_2}{\sigma_1} \quad \psi^2 = \sigma_2^2(1-\rho^2).$$