Løsningsforslag ST1101/ST6101 kontinuasjonseksamen 2018

Oppgave 1

a) Definer hendelsene R_1, B_2, B_3 (rød kule i første trekning, blå kule i andre trekning, blå kule i tredje trekning). Vi skal finne $\mathrm{P}(R_1 \cap B_2 \cap B_3)$ for to ulike situasjoner. Generelt vet vi at

$$\mathrm{P}(R_1 \cap B_2 \cap B_3) = \mathrm{P}(B_3 | R_1 \cap B_2)\mathrm{P}(B_2 | R_1)\mathrm{P}(R_1).$$

Når vi trekker med tilbakelegging så trenger vi ikke ta hensyn til tidligere trekninger. Da er $\mathrm{P}(R_1) = \frac{10}{18}$, og $\mathrm{P}(B_2|R_1) = \frac{8}{18}$, og $\mathrm{P}(B_3|B_2 \cap R_1) = \frac{8}{18}$, og

$$\mathrm{P}(R_1 \cap B_2 \cap B_3) = \left(\frac{8}{18}\right)^2 \cdot \frac{10}{18} = 0.1097.$$

Når vi trekker uten tilbakelegging så er $\mathrm{P}(R_1) = \frac{10}{18}$, $\mathrm{P}(B_2|R_1) = \frac{8}{17}$ og $\mathrm{P}(B_3|B_2 \cap R_1) = \frac{7}{16}$. Dermed er

$$\mathrm{P}(R_1 \cap B_2 \cap B_3) = \frac{7}{16} \cdot \frac{8}{17} \cdot \frac{10}{18} = 0.1144.$$

Vi skal trekke 10 kuler uten tilbakelegging og finne sannsynligheten for at vi trekker minst 8 røde kuler. Da spiller ikke rekkefølgen på trekningene noen rolle. Hendelsen 'minst 8 røde kuler' tilsvarer unionen av de tre enkeltutfallene '8 røde kuler', '9 røde kuler' og '10 røde kuler'. Siden enkeltutfall er disjunkte hendelser, så er

$$\begin{aligned}\mathrm{P}(\text{'minst 8 røde kuler'}) &= \mathrm{P}(\text{'8 røde kuler'} \cup \text{'9 røde kuler'} \cap \text{'10 røde kuler'}) \\ &= \mathrm{P}(\text{'8 røde kuler'}) + \mathrm{P}(\text{'9 røde kuler'}) + \mathrm{P}(\text{'10 røde kuler'})\end{aligned}$$

La oss se først på P('8 røde kuler'). Det finnes

$$m = \binom{18}{10} = 43758$$

mulige utfall. For å finne antall gunstige utfall ser vi på situasjon der vi trekker 8 røde kuler. Dette utfallet kan oppstå på $g_8 = \binom{10}{8} \cdot \binom{8}{2}$ ulike måter. Dermed er

$$\mathrm{P}(\text{'8 røde kuler'}) = \frac{g_8}{m} = \frac{\binom{10}{8} \cdot \binom{8}{2}}{\binom{18}{10}}.$$

Tilsvarende kan hendelsen '9 røde kuler' oppstå på $g_9 = \binom{10}{9} \cdot \binom{8}{1}$ ulike måter, og '10 røde kuler' på $g_{10} = \binom{10}{10} \cdot \binom{8}{0}$ ulike måter.

Dermed er

$$\mathrm{P}(\text{'minst 8 røde kuler'}) == \frac{g_8}{m} + \frac{g_9}{m} + \frac{g_{10}}{m} = \frac{\binom{10}{8} \cdot \binom{8}{2}}{\binom{18}{10}} + \frac{\binom{10}{9} \cdot \binom{8}{1}}{\binom{18}{10}} + \frac{\binom{10}{10} \cdot \binom{8}{0}}{\binom{18}{10}} = 0.0306.$$

Oppgave 2

a)

$$P(X \leq 1 \cap Y < 1) = P(X \leq 1 \cap Y = 0) = P(Y = 0) = 0.1 + 0.2 = 0.3.$$

$$P(Y \leq 1 | X = 1) = \frac{P(Y \leq 1 \cap X = 1)}{P(X = 1)} = \frac{0.2 + 0.1}{0.2 + 0.1 + 0.1} = \frac{0.3}{0.4} = 0.75.$$

Marginale punktsannsynligheter for X:

$$\begin{aligned} P(X = 0) &= 0.1 + 0.2 + 0.3 = 0.6 \\ P(X = 1) &= 0.2 + 0.1 + 0.1 = 0.4 \end{aligned}$$

Marginale punktsannsynligheter for Y:

$$\begin{aligned} P(Y = 0) &= 0.1 + 0.2 = 0.3 \\ P(Y = 1) &= 0.2 + 0.1 = 0.3 \\ P(Y = 2) &= 0.3 + 0.1 = 0.4 \end{aligned}$$

X og Y er ikke uavhgengige, da (for eksempel)

$$P(X = 0) \cdot P(Y = 0) = 0.6 \cdot 0.3 = 0.18 \neq P_{X,Y}(0, 0) = 0.1.$$

b) Vi finner først $E(X)$ og $E(Y)$.

$$\begin{aligned} E(X) &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = 0.4 \\ E(Y) &= 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) + 2 \cdot P(Y = 2) = 1.1 \end{aligned}$$

Dermed er

$$E(2X + 3Y) = 2E(X) + 3E(Y) = 4.1.$$

Siden X og Y ikke er uavhengige, så er

$$\mathrm{Var}(2X + 3Y) = 2^2\mathrm{Var}(X) + 3^2\mathrm{Var}(Y) + 2 \cdot 2 \cdot 3 \cdot \mathrm{Covar}(X, Y)$$

Vi finner varians til X og Y samt kovariansen.

$$\begin{aligned} \mathrm{Var}(X) &= E(X^2) - E(X)^2 \\ &= \big(0^2 \cdot P(X = 0) + 1^2 \cdot P(X = 1)\big) - (0.4)^2 \\ &= 0.4 - 0.4^2 = 0.24 \\ \mathrm{Var}(Y) &= E(Y^2) - E(Y)^2 \\ &= \big(0^2 \cdot P(Y = 0) + 1^2 \cdot P(Y = 1) + 2^2 \cdot P(Y = 2)\big) - (1.1)^2 \\ &= 1.9 - 1.1^2 = 0.69 \\ \mathrm{Covar}(X, Y) &= E(XY) - E(X)E(Y) \\ &= \sum_{x=0}^{1} \sum_{y=0}^{2} x \cdot y \cdot P(X = x, Y = y) - E(X)E(Y) \\ &= 1 \cdot 1 \cdot P(X = 1, Y = 1) + 1 \cdot 2 \cdot P(X = 1, Y = 2) - 0.4 \cdot 1.1 = -0.14 \end{aligned}$$

Dermed er

$$\mathrm{Var}(2X + 3Y) = 2^2 \cdot 0.24 + 3^2 \cdot 0.69 - 2 \cdot 2 \cdot 3 \cdot 0.14 = 5.49.$$

Oppgave 3

a)

$$\mathrm{P}(X \leq 3) = F_X(3; 1.5) = 1 - \left(\frac{3}{2}\right)^{-1.5} = 0.4557$$

$$\mathrm{P}(3 < X \leq 4) = F_X(4; 1.5) - F_X(3; 1.5) = 1 - \left(\frac{4}{2}\right)^{-1.5} - \left(1 - \left(\frac{3}{2}\right)^{-1.5}\right) = 0.1908$$

$$\mathrm{P}(X > 4 | X > 3) = \frac{\mathrm{P}(X > 4 \cap X > 3)}{\mathrm{P}(X > 3)} = \frac{\mathrm{P}(X > 4)}{\mathrm{P}(X > 3)} = \frac{\left(\frac{4}{2}\right)^{-1.5}}{\left(\frac{3}{2}\right)^{-1.5}} = \left(\frac{4}{3}\right)^{-1.5} = 0.6495$$

b) Sannsynlighetstetthetsfunksjonen $f_X(x; \alpha)$ finner vi ved å derivere den kumulative fordelingsfunksjonen med hensyn på x.

$$f_X(x; \alpha) = \frac{d}{dx} F_X(x; \alpha) = \begin{cases} \frac{\alpha}{2}\left(\frac{x}{2}\right)^{-(\alpha+1)} & \text{for } x \geq 2 \\ 0 & \text{for } x < 2 \end{cases}$$

Vi antar først at vi har observasjoner $x_1, \ldots, x_n$ av det tilfeldige utvalget $X_1, \ldots, X_n$.

Sannsynlighetsmaksimeringsestimatet finner vi ved å finne α som maksimerer rimelighetsfunksjonen

$$L(\alpha; x_1, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i; \alpha) = \prod_{i=1}^{n} \frac{\alpha}{2}\left(\frac{x_i}{2}\right)^{-(\alpha+1)}.$$

I dette tilfellet er det enklere å maksimere logaritmen til rimelighetsfunksjonen

$$l(\alpha; x_1, \ldots, x_n) = \sum_{i=1}^{n}\left(\log(\alpha) - \log(2) - (\alpha+1)\log\left(\frac{x_i}{2}\right)\right).$$

Vi løser

$$\frac{d}{d\alpha} l(\alpha; x_1, \ldots, x_n) = \frac{n}{\alpha} - \sum_{i=1}^{n} \log\left(\frac{x_i}{2}\right) = 0$$

med hensyn på α og finner sannsynlighetsmaksimeringsestimatet $\widehat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log\left(\frac{x_i}{2}\right)}$. Ved å sette inn tilfeldige variabler $X_1, \ldots, X_n$ finner vi sannsynlighetsmaksimeringsestimatoren

$$\widehat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log\left(\frac{X_i}{2}\right)}.$$

c) Ved å bruke momentmetoden så løser vi ligningen

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} = \mathrm{E}(X) = \frac{2\alpha}{\alpha - 1}$$

med hensyn på α, og finner estimatoren

$$\widetilde{\alpha} = \frac{\overline{X}}{\overline{X} - 2}.$$

Oppgave 4

a) La X være tiden i minutter fra klokka 07:30 til Grete ankommer fabrikker. Vi vet at X er eksponensialfordelt med parameter $\lambda = \frac{1}{2}$.

Sannsynligheten for at Grete kommer til fabrikken før klokka 07:32 er

$$P(X \leq 2) = \int_0^2 \frac{1}{2} e^{-\frac{x}{2}}\, dx = -e^{-\frac{2}{2}} + e^{-\frac{0}{2}} = 1 - e^{-1} = 0.6321.$$

La $X_1, \ldots, X_{100}$ være stokastiske variabler (uavhengige) som måler tid til ankomst for 100 ansatte, slik at X_i er eksponensialfordelt med parameter $\lambda = \frac{1}{2}$ for $i = 1, \ldots, 100$. Da er

$$P(X_1 \leq 2 \cap \cdots \cap X_{100} \leq 2) = P(X_1 \leq 2) \cdot P(X_{100} \leq 2) = (0.6321)^{100} \approx 0.$$

Dersom vi repeterer et målbart eksperiment (her Gretes ankomsttid X) mange ganger, vil gjennomsnittsverdien over alle eksperimenter nærme seg forventningsverdien til eksperimentet. Siden $E(X) = 2$ minutter, så burde Gretes gjennomsnittlige ankomstid være omtrent klokka 07:32.

b) Antall desiliter iskaffe kan representeres med den stokastiske variabelen $V = X + Y$. Siden X og Y er uavhengige og normalfordelte, så er V normalfordelt med forventningsverdi $\mu_1 + \mu_2$ dl og standardavvik $\sigma = \sqrt{0.5^2 + 0.3^2} = 0.58$ dl. Grete ser på et tilfeldig utvalg at $n = 20$ kartonger; $V_1, \ldots, V_{20}$. Da vet vi at

$$\frac{1}{20} \sum_{i=1}^{20} V_i = \overline{V} \sim N\left(\mu_1 + \mu_2, \frac{\sigma^2}{20}\right).$$

Fra dette kan vi sette opp ligningen

$$P\left(-z_{\alpha/2} \leq \frac{\overline{V} - (\mu_1 + \mu_2)}{\sigma/\sqrt{20}} \leq z_{\alpha/2}\right) = 1 - \alpha,$$

der $z_{\alpha/2}$ er en kritisk verdi i standard normalfordelingen. For å finne et 95% konfidensintervall setter vi $\alpha = 0.05$, slik at $z_{\alpha/2} = z_{0.025} = 1.96$. Nå gjør vi de samme matematiske operasjonene på alle sidene av ulikhetstegnene slik at vi ender opp med $\mu_1 + \mu_2$ i midten.

$$\begin{aligned}
0.95 &= P\left(-1.96 \leq \frac{\overline{V} - (\mu_1 + \mu_2)}{\sigma/\sqrt{20}} \leq 1.96\right) \\
&= P\left(-1.96\frac{\sigma}{\sqrt{20}} \leq \overline{V} - (\mu_1 + \mu_2) \leq 1.96\frac{\sigma}{\sqrt{20}}\right) \\
&= P\left(-\overline{V} - 1.96\frac{\sigma}{\sqrt{20}} \leq -(\mu_1 + \mu_2) \leq -\overline{V} + 1.96\frac{\sigma}{\sqrt{20}}\right) \\
&= P\left(\overline{V} + 1.96\frac{\sigma}{\sqrt{20}} \geq \mu_1 + \mu_2 \geq \overline{V} - 1.96\frac{\sigma}{\sqrt{20}}\right) \\
&= P\left(\overline{V} - 1.96\frac{\sigma}{\sqrt{20}} \leq \mu_1 + \mu_2 \leq \overline{V} + 1.96\frac{\sigma}{\sqrt{20}}\right)
\end{aligned}$$

Vi setter inn tallverdier for σ og observert gjennomsnitt $\overline{v} = \frac{1}{20}\sum_{i=1}^{20} v_i = 5.0165$ og finner intervallet $[4.76, 5.27]$.

Dersom vi ser på veldig mange tilfeldige utvalg av 20 kartonger og regner ut et slikt intervall hver gang, så vil 95% av disse intervallene inneholde den sanne verdien for $\mu_1 + \mu_2$.

I dette tilfellet inneholder intervallet tallet 5 og det er rimelig å *anta* at maskinoperatøren snakker sant (men vi er ikke 100% sikre). Vi ville fått et smalere intervall (bedre presisjon) dersom Grete så på flere kartonger.

c) $\mathrm{P}(A) = \mathrm{P}(V \leq 4.5) = \mathrm{P}\left(Z \leq \frac{4.5-5}{0.58}\right) = \mathrm{P}(Z \leq -0.86) = 0.1949.$

W kan antas å være binomisk fordelt med parametere $p = 0.1949$ og $n = 50$ dersom

- Maskinens instilling er det samme hele tiden ($\mu_1 + \mu_2 = 5$), slik at $p = \mathrm{P}(A)$ er lik for alle iskaffekartongene.
- Mengen iskaffe i en gitt kartong er uavhengig av mengden som har blitt fylt i andre kartonger.
- De 50 kartongene kan antas å være tilfeldig valgt (studentene eller butikken har ikke valgt ut kartonger med noen spesiell hensikt eller sortering)

Dersom vi antar at W er binomisk fordelt så er

$$\begin{aligned}
&\mathrm{P}(W > 2) = 1 - \mathrm{P}(W = 0) + \mathrm{P}(W = 1) + \mathrm{P}(W = 2) \\
&= 1 - \binom{50}{0}(1-0.1949)^{50} - \binom{50}{1} \cdot 0.1949 \cdot (1-0.1949)^{49} - \binom{50}{2} \cdot 0.1949^2 \cdot (1-0.1949)^{48} \\
&= 1 - (1-0.1949)^{50} - 50 \cdot 0.1949 \cdot (1-0.1949)^{49} - 1225 \cdot 0.1949^2 \cdot (1-0.1949)^{48} \\
&= 0.9983
\end{aligned}$$

d) Antall desiliter melk Y er nå en konstant; $Y = 3$. Det vil si at $V = X + Y \sim N(\mu_1 + 3, \sigma_1^2)$. Hendelsen A inntreffer nå dersom $V < 4.5$ dl.

Vi løser
$$\mathrm{P}(V \leq 4.5) = \mathrm{P}(Z \leq \frac{4.5 - \mu_1 - 3}{0.5}) = 0.025.$$

Ved å slå opp i tabeller finner vi at $\frac{4.5-\mu_1-3}{0.5} = -1.96$ oppfyller kravet. Dette gir anbefalingen
$$\mu_1 = 1.5 - (-1.96 \cdot 0.5) = 2.48.$$