Løsningsforslag øving 15

2.6

I eksempel 2.2 kaster vi en mynt tre ganger, og får da enten kron eller mynt for hvert av de tre kastene. Akkurat som i eksempel 2.1 er det mulig å sette utfallsrommet for et enkelt kast til å være $\Omega = \mathbb{R}$, som resulterer i utfallsrommet $\Omega = \mathbb{R}^3$ for tre kast. Da har vi $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$, hvor resultatet fra hvert av de tre myntkastene er enten $\omega_i = 0$ eller $\omega_i = 1$, for $i = 1, 2, 3$. De mulige utfallene kan tegnes i $\mathbb{R}^3$, og vil sammen forme alle de forskjellige hjørnene i enhetskuben. Nå lar vi $\mathcal{E}$ være den minste hendelsesfamilien som inneholder alle åpne bokser. Vi vil vise at $\{\boldsymbol{\omega}\}$ er en hendelse i denne hendelsesfamilien, altså at $\{\boldsymbol{\omega}\} \in \mathcal{E}$. For å vise dette viser vi først at en hendelsesfamilie er lukket under tellbart snitt i tillegg til tellbar union. Anta at A og B er to hendelse. Da vet vi at også at både A^c, B^c og $A \cup B$ er hendelser. Da vet vi også at $(A^c \cup B^c)^c = (A \cap B)$ er en hendelse. Dette betyr at for alle hendelser A og B vet vi også at $A \cap B$ er en hendelse. Dette kan generaliseres til tellbar union og tellbart snitt. Definer nå boksen $B_i = (\omega_1 - t_i, \omega_1 + t_i) \times (\omega_2 - t_i, \omega_2 + t_i) \times (\omega_3 - t_i, \omega_3 + t_i)$, hvor t_i er et positivt tall. Vi lar rekken $t_1, t_2, \ldots$ konvergere mot null slik at $t_i > t_{i+1}$ for alle i, og $t_i \to 0$. Siden $\mathcal{E}$ inneholder alle åpne bokser, er $B_i \in \mathcal{E}$ for alle i. $\mathcal{E}$ er også lukket under tellbart snitt, som betyr at $\cap_{i=1}^{\infty} B_i = \{\omega\} \in \mathcal{E}$. Derfor er $\{\omega\}$ en hendelse.

3.12

Tettheten for en persons alder er

$$f(t) = kt^2(100-t)^2 = k(t^4 - 200t^3 + 100^2t^2), \quad 0 \leq t \leq 100.$$

Vi finner den kumulative sannsynlighetsfordelingen

$$F(t) = \int_0^t f(x) = \int_0^t k(x^4 - 200x^3 + 100^2x^2)dx = k(\frac{1}{5}t^5 - 50t^4 + \frac{100^2}{3}t^3).$$

Vi vet at integralet over tettheten fra 0 til 100 må være lik 1, så da har vi

$$F(100) = 1 = k(\frac{1}{5} \cdot 100^5 - 50 \cdot 100^4 + \frac{1}{3} \cdot 100^5) = k \cdot 100^5 \left(\frac{1}{5} - \frac{1}{2} + \frac{1}{3}\right) = k \cdot 100^5/30.$$

Dette gir oss at $k = 30/100^5$. Vi kan nå finne P(personen lever mer enn 60 år).

$$P(T > 60) = 1 - P(T \leq 60) = 1 - \frac{30}{100^5}(\frac{1}{5} \cdot 60^5 - 50 \cdot 60^4 + \frac{100^2}{3} \cdot 60^3) \approx 0.3174.$$

4.5

Hvorvidt du bør vedde imot eller ikke er veldig avhengig av hvor risikoavers du er. Vi kan derimot se på sannsynlighetene. Sannsynligheten for å trekke kortet med to røde sider gitt at du har trukket et kort med en rød side opp er lik

$$P(\text{begge sider er røde}|\text{siden opp er rød}) = \frac{P(\text{begge sider er røde})}{P(\text{siden opp er rød})} = \frac{1/3}{1/2} = \frac{2}{3},$$

siden vi antar at det er lik sannsynlighet for å trekke de tre kortene. De tre kortene har totalt seks sider, hvorav tre av dem er røde, som gir at sannsynligheten for at siden opp er rød er lik 1/2. Det er altså 2/3 sjanse for å tape veddemålet, som vil si at man ikke bør vedde imot med mindre man har en veldig god grunn.

5.2

Det er ingen av de tre parentesene som inneholder den samme variabelen. Da vil tre parenteser med henholdsvis 3, 3 og 4 ledd gi oss totalt $3 \cdot 3 \cdot 4 = 36$ ledd. Vi kan se at leddene aeu og cdx finnes i ekspansjonen. For de to andre leddene er det derimot flere variabler med fra samme parentes. Disse vil aldri bli ganget sammen, som betyr at det ikke er mulig å få disse leddene fra denne ekspansjonen.

6.10

En tilfeldig variabel er en observator som har utfallsrom $\mathbb{R}$. At noe er en observator betyr at det er en funksjon $T : \Omega \to \Omega_T$ slik at $(T \in A)$ er en hendelse i Ω for alle hendelser A i Ω_T. At F_X er fordelingsfunksjonen til X betyr at $F_X(t) = P(X \in (-\infty, t])$ for alle $t \in \mathbb{R}$.

7.10

Vi har at $X_1, \dots, X_n$ er uavhengige og kontinuerlige tilfeldige variabler. Da har alle de tilfeldige variablene en sannsynlighetstetthet $f_{X_i}(x)$, $i = 1, \dots, n$. Siden variablene er uavhengige kan vi skrive den simultane sannsynlighetstettheten som

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

Vi finner forventningsverdien

$$\begin{aligned}
E[X_1 \cdots X_n] &= \int \cdots \int x_1 \cdots x_n f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n \\
&= \int x_1 f_{X_1}(x_1) \left[\int x_2 f_{X_2}(x_2) \left[\cdots \left[\int x_n f_{X_n}(x_n) dx_n \right] \cdots \right] dx_2 \right] dx_1 \\
&= E[X_1] E[X_2] \cdots E[X_n].
\end{aligned}$$

Vi antar at $\phi(X_1, X_2)$ og $\psi(X_3)$ er tilfeldige variabler. For at disse skal være uavhengige må vi ha at $P(\phi(X_1, X_2) \in A \cap \psi(X_3) \in B) = P(\phi(X_1, X_2) \in A)P(\psi(X_3) \in B)$ for alle hendelser A og B. Vi bruker at forventningsverdien $E[I_A]$ til en indikatorfunksjon for hendelsen A er lik $E[I_A] = P(A)$. Da får vi, på grunn av variabelskifteteoremet

$$
\begin{aligned}
&P(\phi(X_1, X_2) \in A \cap \psi(X_3) \in B)\\
=&E[I(\phi(X_1, X_2) \in A \cap \psi(X_3) \in B)]\\
=&\int\int\int I(\phi(x_1, x_2) \in A \cap \psi(x_3) \in B) f_{X_1,X_2,X_3}(x_1, x_2, x_3) dx_1 dx_2 dx_3\\
=&\int\int\int I(\phi(x_1, x_2) \in A) \cdot I(\psi(x_3) \in B) f_{X_1,X_2}(x_1, x_2) f_{X_3}(x_3) dx_1 dx_2 dx_3\\
&\text{(på grunn av uavhengigheten mellom } X_1,\ X_2 \text{ og } X_3)\\
=&\int\int I(\phi(x_1, x_2) \in A) f_{X_1,X_2}(x_1, x_2) \left[\int I(\psi(x_3) \in B) f_{X_3}(x_3) dx_3\right] dx_1 dx_2\\
=&\int\int I(\phi(x_1, x_2) \in A) f_{X_1,X_2}(x_1, x_2) \cdot P(\psi(X_3) \in B) dx_2 dx_3\\
=&P(\phi(X_1, X_2) \in A)P(\psi(X_3) \in B).
\end{aligned}
$$

Altså er $\phi(X_1, X_2)$ og $\psi(X_3)$ uavhengige tilfeldige variabler.

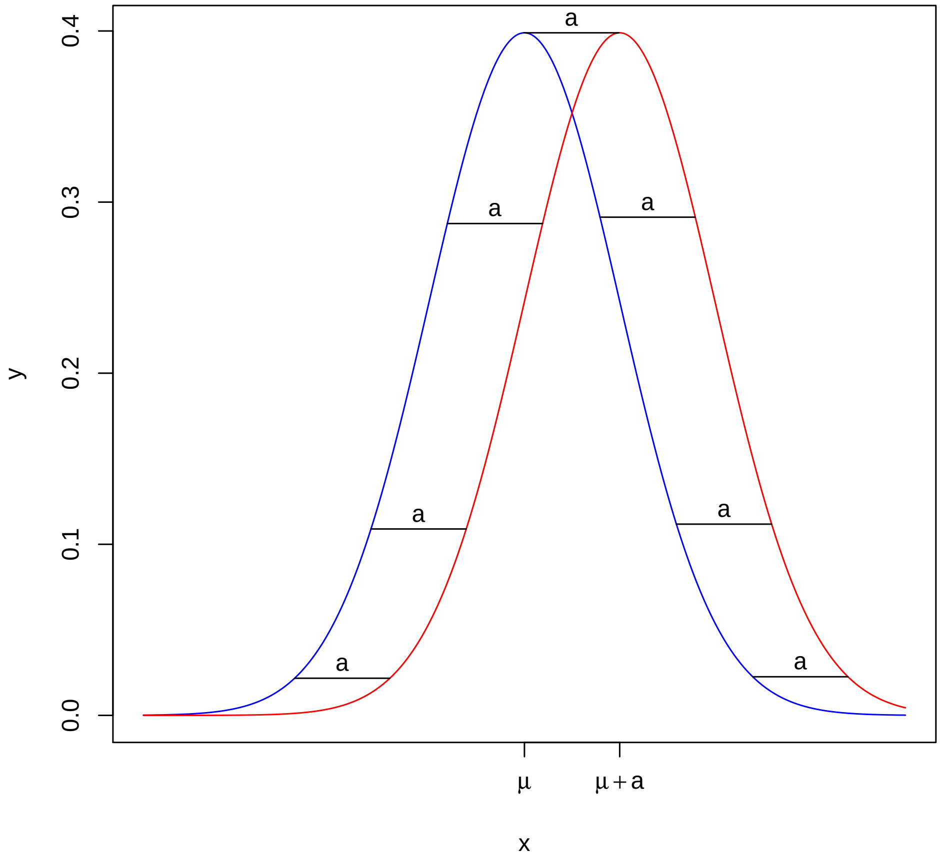
8.10

Vi har en tilfeldig variabel X med tetthet f. Vi vil se på sammenhengen mellom tettheten f og tettheten g til den tilfeldige variabelen $Y = X + a$. I dette tilfellet er Y bare en flytting av den tilfeldige variabelen X. Vi antar derfor at tettheten g vil se helt lik ut som tettheten f, men at den vil være flyttet en lengde a unna grafen til f. Vi kan illustrere i tilfellet hvor X er normalfordelt med forventning μ og standardavvik σ. Da vil også Y være normalfordelt, med forventning $\mu + a$ og standardavvik σ.

```
a = 1
mu = 0
sigma = 1
xvec = seq(-4, 4, length.out = 500)
y1 = dnorm(xvec, mean = mu, sd = sigma)
y2 = dnorm(xvec, mean = mu + a, sd = sigma)

plot(xvec, y1, "l", col = "blue", xlab = "x",
ylab = "y", xaxt = "n")
lines(xvec, y2, col = "red")
for (i in seq(100, 400, by = 50)) {
  lines(c(xvec[i], xvec[i] + 1), rep(y1[i], 2))
  text(x = xvec[i] + a / 2, y = .008 + y1[i], labels = "a")
}
xticks = c(mu, mu + a)
xlabs = expression(mu, mu + a)
```

```
axis(side = 1, at = xticks, labels = xlabs)
```



9.16

At tettheten til gammafordelingen er normalisert betyr at integralet over tettheten er lik 1. Vi ønsker å vise dette. Tettheten til en gammafordeling med parametre α og β er lik

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0.$$

Vi vil vise at

$$\int_0^\infty f(x) = \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx = 1.$$

Dette kan se ut som et vanskelig integral å løse. Vi kan derimot bruke et triks, nemlig at vi kjenner definisjonen av gammafunksjonen Γ. Fra side 33 i statistikk-heftet ser vi at

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-t}dx.$$

Dette er nesten det samme som integralet vi ønsker å løse. Vi bruker derfor substitusjonen $y = x/\beta$ som gir

$$\begin{aligned}
\int_0^\infty f(x) &= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-x/\beta}dx \\
&= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty (\beta y)^{\alpha-1}e^{-y}\beta dy \\
&= \frac{\beta^\alpha}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty y^{\alpha-1}e^{-y}dy \\
&= \frac{\beta^\alpha\Gamma(\alpha)}{\beta^\alpha\Gamma(\alpha)} \\
&= 1.
\end{aligned}$$

10.4

Vi betegner diameteren til en produsert styrestang med X. Sannsynligheten for at fabrikken produserer en ubrukelig styrestang er lik

$$P(\text{ubrukelig}) = 1 - P(1.480\text{cm} \leq X \leq 1.500\text{cm}).$$

Vi ønsker å finne denne sannsynligheten. X er normalfordelt med forventning $\mu = 1.495$cm og standardavvik $\sigma = 0.005$cm. Vi standardiserer X slik at vi kan finne sannsynligheten fra en tabell. Dette gir

$$\begin{aligned}
P(\text{ubrukelig}) &= 1 - P(1.480\text{cm} \leq X \leq 1.500\text{cm}) \\
&= 1 - P\left(\frac{1.480\text{cm} - 1.495\text{cm}}{0.005\text{cm}} \leq \frac{X-\mu}{\sigma} = Z \leq \frac{1.500\text{cm} - 1.495\text{cm}}{0.005\text{cm}}\right) \\
&= 1 - P(-3 \leq Z \leq 1) \\
&= 1 - (\Phi(1) - \Phi(-3)) \\
&\approx 1 - (0.8413 - 0.0013) \\
&= 0.1600,
\end{aligned}$$

som betyr at ca. 16% av de produserte styrestengene blir ubrukelige.

11.19

Vi har fire uniformfordelte variabler $Y_1, \ldots, Y_4 \sim U(0, \theta)$, og en estimator $W = Y_{\max}$. Fordelingsfunksjonen for Y er da gitt ved $F_Y(t) = t/\theta,\ t \leq \theta$. W er maksimumsverdien av de fire variablene, som betyr at fordelingsfunksjonen for W er gitt ved

$$F_W(t) = P(W \leq t) = P(Y_1 \leq t, \ldots, Y_4 \leq t) = (F_Y(t))^4 = t^4/\theta^4.$$

Vi ønsker å finne sannsynligheten for at W ikke er lenger enn $1/8$ unna θ, altså at $|\theta - W| \leq 1/8$. Siden Y_i er uniformfordelt med parameter θ er det ikke mulig at Y_i er større enn $1/8$, som igjen vil si at det ikke er mulig at W er større enn θ. Da får vi

$$\begin{aligned}
P(|\theta - W| \leq 1/8) &= P(\theta - W \leq 1/8) \\
&= P(W \geq \theta - 1/8) \\
&= 1 - P(W < \theta - 1/8) \\
&= 1 - F_W(\theta - 1/8) \\
&= 1 - \left(\frac{\theta - 1/8}{\theta}\right)^4.
\end{aligned}$$

I tilfellet når $\theta = 1$ blir denne sannsynligheten lik

$$P(|\theta - W| \leq 1/8) = 1 - \left(\frac{\theta - 1/8}{\theta}\right)^4 = 1 - \left(\frac{1 - 1/8}{1}\right)^4 = 1 - \left(\frac{7}{8}\right)^4 \approx 0.4138.$$