Løsningsforslag øving 9

9.1

En indikatorvariabel er den tilfeldige variabelen

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A. \end{cases}$$

Da er $P(I_A = 1) = P(A) = p$ og $P(I_A = 0) = P(A^c) = 1 - p$. Dette gir

$$P(I_A = x) = p^x(1-p)^{(1-x)}, \quad x = 0, 1,$$

som vil si at indikatorvariabelen er Bernoullifordelt.

9.3

Et forsøk som gjentas helt til man oppnår suksess kan beskrives ved den geometriske fordelingen. Vi antar at sannsynligheten for å føde en jente er like stor som sannsynligheten for å føde en gutt. Vi ser også bort ifra muligheten for tvillinger, trillinger, etc. Da er fordelingen til den tilfeldige variabelen $N = \#$ barn som blir født

$$P(N = n) = p(1-p)^{n-1}, \text{ med } p = 1/2.$$

Vi finner forventningsverdien til N fra Tabell 9.1, og ser at den er lik $1/p$. Da er den forventede familiestørrelsen $EN = 2$.

9.7

Antall ganger vi feiler, X før vi oppnår r suksesser kan beskrives ved den negative binomiske fordelingen. I dette tilfellet er vi interessert i å finne når de defekte bildekkene produseres, så vi ser på et defekt bildekk som en "suksess", og et riktig bildekk som en "feil". Da er antall riktige bildekk som produseres før maskinen blir stoppet negativt binomisk fordelt med $p = 0.15$ og $r = 3$. Sannsynligheten for at det produseres 5 eller flere bildekk før maskinen stoppes er den samme som at det produseres 2 eller flere fungerende bildekk før maskinen stoppes, siden defekte bildekk også telles som bildekk. Da får vi

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \binom{r-1}{0} p^r - \binom{r}{1} p^r (1-p) \approx 0.9880$$

I gjennomsnitt vil maskinen produsere $EX + 3$ dekk før den blir stoppet. Fra Tabell 9.1 får vi $EX + 3 = r(1-p)/p + 3 = 17 + 3 = 20$.

9.9

Denne oppgaven ble også gitt i forrige øving, så her var dere heldige. For kompletthetens skyld kommer beviset en gang til.

$$\begin{aligned} M_X(t) = E[e^{Xt}] &= \sum_{x=0}^{\infty} e^{xt} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \sum_{x=0}^{\infty} e^{-\lambda} \frac{(e^t \lambda)^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!}. \end{aligned}$$

Dette kjenner vi igjen som den algebraiske kjernen til en Poisson-fordeling med parameter $\tilde{\lambda} = e^t \lambda$. Vi legger derfor til den rette konstanten for å få tettheten til en Poisson-fordeling på innsiden av summetegnet.

$$\begin{aligned} M_X(t) &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} \\ &= e^{-\lambda} e^{e^t \lambda} \sum_{x=0}^{\infty} e^{-e^t \lambda} \frac{(e^t \lambda)^x}{x!} \\ &= e^{\lambda(e^t - 1)}, \end{aligned}$$

fordi summen over sannsynlighetstettethen summeres til 1.
Vi kan nå benytte den momentgenererende funksjonen til å finne første- og andremomentet til X.

$$M_X'(t) = e^{\lambda(e^t-1)} \cdot \lambda e^t.$$
$$M_X''(t) = e^{\lambda(e^t-1)} \cdot (\lambda e^t)^2 + e^{\lambda(e^t-1)} \cdot \lambda e^t.$$

Ved å sette inn $t = 0$ finner vi

$$E[X] = e^{\lambda(1-1)} \cdot \lambda e^0 = \lambda.$$
$$E[X^2] = e^0 \cdot (\lambda e^0)^2 + e^0 \cdot \lambda e^0 = \lambda^2 + \lambda,$$

som gir

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

9.11

Betegn antall telefonsamtaler i morgen som X. Vi antar at X er Poisson-fordelt med intensitet $\lambda = 4$. Da vet vi fra Tabell 9.1 at $EX = \lambda = 4$ og $\mathrm{Var}X = \lambda = 4 = 2^2$.. standardavviket pluss gjennomsnittet til X er lik $\sigma + \mu = \lambda + \sqrt{\lambda} = 6$. Vi finner $P(X > 6) = 1 - P(X \leq 6)$ enten ved å slå opp i "Tabeller og formler i statistikk" eller ved hjelp av et regneverktøy. Sannsynligheten er lik 0.1107.

```r
p <- 1 - ppois(q = 6, lambda = 4)
p
```

```
[1] 0.110674
```

9.12

Vi har $X \sim \text{Poisson}(\lambda)$ og $[Y|X=x] \sim B(x,p)$. Siden $[Y|X=x] \sim B(x,p)$ har vi $P(Y=y|X=x) = 0$ for $y > x$. Vi finner simultanfordelingen til X og Y.

$$P(X=x, Y=y) = P(Y=y|X=x)P(X=x) = \binom{x}{y} p^y (1-p)^{x-y} \cdot \frac{\lambda^x}{x!} e^{-\lambda}$$

Vi skriver ut uttrykket for $\binom{x}{y}$ og samler x'er og y'er for seg selv, og får

$$P(X=x, Y=y) = \frac{1}{y!(x-y)!} \left(\frac{p}{1-p}\right)^y ((1-p)\lambda)^x \, e^{-\lambda}, \; x = 0, 1, \ldots, y = 0, 1, \ldots, x.$$

For å finne marginalfordelingen til Y må vi summere ut X.

$$\begin{aligned}
P(Y=y) &= \sum_{x=0}^{\infty} P(X=x, Y=y) = \sum_{x=y}^{\infty} P(X=x, Y=y) \\
&= \frac{1}{y!} \left(\frac{p}{1-p}\right)^y e^{-\lambda} \sum_{x=y}^{\infty} \frac{1}{(x-y)!} ((1-p)\lambda)^x \\
&= C(y) \sum_{x=y}^{\infty} \frac{1}{(x-y)!} ((1-p)\lambda)^x \\
&\overset{z=x-y}{=} C(y) \sum_{0}^{\infty} \frac{1}{z!} ((1-p)\lambda)^{z+y} \\
&= C(y)((1-p)\lambda)^y \sum_{0}^{\infty} \frac{1}{z!} ((1-p)\lambda)^z.
\end{aligned}$$

Vi kjenner igjen uttrykket inne i summen som den algebraiske kjernen til sannsynlighetstettheten til en Poisson-fordeling med intensitet $(1-p)\lambda$. Da får vi

$$\begin{aligned}P(Y=y) &= C(y)((1-p)\lambda)^y \sum_0^\infty \frac{1}{z!}((1-p)\lambda)^z \\ &= C(y)((1-p)\lambda)^y e^{(1-p)\lambda} \sum_0^\infty \frac{1}{z!}((1-p)\lambda)^z e^{-(1-p)\lambda} \\ &= C(y)((1-p)\lambda)^y e^{(1-p)\lambda} \cdot 1 \\ &= \frac{1}{y!}\left(\frac{p}{1-p}\right)^y e^{-\lambda}((1-p)\lambda)^y e^{(1-p)\lambda} \\ &= \frac{(p\lambda)^y}{y!} e^{-p\lambda}.\end{aligned}$$

Dette er sannsynlighetstettheten til en Poisson-fordeling med intensitet λp, så vi får at $Y \sim \text{Poisson}(\lambda p)$.

9.13

Betegn antall flyulykker i det visse landet med X. Vi antar at forekomsten av flyulykker følger en Poissonfordeling med intensitet $\lambda = 2.5$. Da er $P(X \geq 4) = 1 - P(X \leq 3)$. Vi finner sannsynligheten fra tabeller og formler i statistikk, ved å regne ut uttrykket for hånd eller ved hjelp av et regneverktøy. Vi får `p <- 1 - ppois(3, lambda = 2.5); print(p) [1] 0.2424239`. Tidsrommet mellom to flyulykker i en Poissonfordeling er eksponentialfordelt med parameter $1/\lambda$. Dette betyr at sannsynligheten for at tidsrommet mellom to flyulykker er mindre enn tre måneder er

$$\begin{aligned}P(T < \frac{3}{12} = \frac{1}{4}) &= \int_0^{\frac{1}{4}} \lambda e^{-\lambda t} dt \\ &= -e^{-\lambda t}\Big|_0^{\frac{1}{4}} = -e^{-\frac{\lambda}{4}} + e^0 \\ &= 1 - e^{-\frac{2.5}{4}} \\ &\approx 0.4647.\end{aligned}$$

9.17

Gammafordelingen har sannsynlighetstetthet $f(x) = \lambda^r \Gamma^{-1}(x) x^{r-1} e^{-\lambda x}$ for $x > 0$. Da får vi

$$\begin{aligned}M(t) &= \int_0^\infty e^{xt} \frac{\lambda^r}{\Gamma(x)} x^{r-1} e^{-\lambda x} \\ &= \int_0^\infty \frac{\lambda^r}{\Gamma(x)} x^{r-1} e^{-(\lambda - t)x}.\end{aligned}$$

Vi kjenner igjen $x^{r-1}\exp\{-(\lambda-t)x\}$ som den algebraiske kjernen til en gammafordeling Gamma$(r,(\lambda-t)^{-1})$. Her må vi ha $\lambda > t$. Hvis ikke vil integralet i likningen over være lik ∞. Vi trekker enkelte konstanter utenfor integralet så vi får en sannsynlighetstetthet som integreres til 1 på innsiden av integralet.

$$\begin{aligned} M(t) &= \int_0^\infty \frac{\lambda^r}{\Gamma(x)} x^{r-1} e^{-(\lambda-t)x} \\ &= \left(\frac{\lambda}{\lambda-t}\right)^r \int_0^\infty \frac{(\lambda-t)^r}{\Gamma(x)} x^{r-1} e^{-(\lambda-t)x} \\ &= \left(\frac{\lambda}{\lambda-t}\right)^r \\ &= \left(\frac{1}{1-t/\lambda}\right)^r. \end{aligned}$$

Den momentgenererende funksjonen til gammafordelingen er altså $M(t) = (1-t/\lambda)^r,\ t<\lambda$.

9.18

Dette kan vi med en gang se intuitivt i tilfellet hvor gammafordelingen beskriver ventetiden til den r'te suksessen. Tiden X_1 til vi har fått r_1 suksesser er gammafordelt med parametre r_1 og β, og tiden X_2 til r_2 suksesser er gammafordelt med parametre r_2 og β. Da er det klart at tiden $X = X_1 + X_2$ til vi har fått $r = r_1 + r_2$ suksesser er gammafordelt med parametre r og β. Dette kan enkelt generaliseres til å gi $\sum_i X_i \sim \text{Gamma}(\sum_i r_i, \beta)$.
For et mer rigorøst bevis kan vi bruke momentgenererende funksjoner. Vi vet at hvis to momentgenererende funksjoner er like, så er også de tilhørende fordelingsfunksjonene like. Vi har

$$M_{\sum_i X_i}(t) = E[e^{\sum_i X_i t}] \overset{\text{uavhengighet}}{=} E[e^{X_1 t}]E[e^{X_2 t}]\dots E[e^{X_n t}] = \prod_i M_{X_i}(t).$$

Fra Tabell 9.1 får vi oppgitt at den momentgenererende funksjonen til en gammafordeling er $M_X(t) = \left(\frac{\lambda}{\lambda-t}\right)^r$. Dette gir at den momentgenererende funksjonen til en sum av gammafordelte variabler er

$$M_{\sum_i X_i}(t) = \prod_i^n \left(\frac{\lambda}{\lambda-t}\right)^{r_i} = \left(\frac{\lambda}{\lambda-t}\right)^{\sum_i^n r_i}.$$

Vi kjenner igjen dette som den momentgenererende funksjonen til en gammafordeling med parametre $\sum_i^n r_i$ og β. Derfor kan vi konkludere at hvis $X_i \sim \text{Gamma}(r_i,\beta)$ så er $\sum_i X_i \sim \text{Gamma}(\sum_i r_i, \beta)$.
Merk: Her kunne vi også brukt konvolusjonsformelen til å finne fordelingen til summen av n gammafordelinger.

9.19

Fra tabell 9.1 kan vi se at eksponensialfordelingen er et spesialtilfelle av gammafordelingen, Gamma$(1,\lambda)$ = Exp(λ). Siden det bare er en av de tre vindmålerne som er i drift av gangen får vi at den totale tiden før systemet er nede er lik $T = T_1 + T_2 + T_3$, hvor T_i er levetiden til vindmåler nummer i. Siden eksponensialfordelingen er et spesialtilfelle av gammafordelingen har vi fra oppgave 9.18 at $T \sim \text{Gamma}(3, \lambda)$, hvor $\lambda = 1000$ timer. Den forventede levetiden til systemet er da $E[T] = 3\lambda = 3000$ timer.