## Oppgave 1: Feil på mobiltelefoner

### a)

Sannsynlighetene i oppgaven blir

$$\begin{aligned} P(F_1 \cup F_2) &= P(F_1) + P(F_2) - P(F_1 \cap F_2) \\ &= P(F_1) + 1 - P(F_2^C) - P(F_1 \cap F_2) \\ &= 0.080 + 0.075 - 0.006 \ = \ \underline{\underline{0.149}} \end{aligned}$$

$$P(F_1 \mid F_2) \ = \ \frac{P(F_1 \cap F_2)}{P(F_2)} \ = \ \frac{P(F_1 \cap F_2)}{1 - P(F_2^C)} \ = \ \frac{0.006}{1 - 0.925} \ = \ \underline{\underline{0.080}}$$

$$P(F_1^C \cap F_2^C) \ = \ P((F_1 \cup F_2)^C) \ = \ 1 - P(F_1 \cup F_2) \ = \ 1 - 0.149 \ = \ \underline{\underline{0.851}}$$

Siden $P(F_1 \mid F_2) = P(F_1)$ er $F_1$ og $F_2$ uavhengige. Vi har også at $P(F_1) \cdot P(F_2) = P(F_1) \cdot (1 - P(F_2^C)) = 0.080 * (1 - 0.925) = 0.006 \ = \ P(F_1 \cap F_2)$, som betyr at $F_1$ og $F_2$ er uavhengige.

To hendelser er disjunkte dersom snittet er den tomme mengde $\emptyset$, som oppfyller $P(\emptyset) = 0$. Siden $P(F_1 \cap F_2) > 0$, må $F_1 \cap F_2 \neq \emptyset$, og $F_1$ og $F_2$ er ikke disjunkte.

### b)

Sannsynligheten for at en mottatt klage gjelder feil av type 2:

Ved å bruke Bayes regel får vi at

$$\begin{aligned} P(F_2 \mid R) &= \frac{P(F_2 \cap R)}{P(R)} \\ &= \frac{P(R \mid F_2) P(F_2)}{P(R)} \\ &= \frac{P(R \mid F_2)(1 - P(F_2^C)}{P(R)} \\ &= \frac{0.70 \cdot (1 - 0.925)}{0.15} \ = \ \underline{\underline{0.35}} \end{aligned}$$

Sannsynligheten for at en mottatt klage gjelder feil av type 1 eller 2:

$$
\begin{aligned}
P(F_1 \cup F_2 \mid R) &= \frac{P((F_1 \cup F_2) \cap R)}{P(R)} \\
&= \frac{P((F_1 \cap R) \cup (F_2 \cap R))}{P(R)} \\
&= \frac{P(F_1 \cap R) + P(F_2 \cap R) - P(F_1 \cap F_2 \cap R)}{P(R)} \\
&= \frac{P(R \mid F_1)P(F_1) + P(R \mid F_2)P(F_2) - P(R \mid F_1 \cap F_2)P(F_1 \cap F_2)}{P(R)} \\
&= \frac{0.90 \cdot 0.080 + 0.70 \cdot (1 - 0.925) - 0.95 \cdot 0.006}{0.15} = \frac{0.1188}{0.15} = \underline{\underline{0.792}}
\end{aligned}
$$

Alternativ metode: Bruker addisjonssetninga, og deretter Bayes regel på hvert ledd i summen.

$$
\begin{aligned}
P(F_1 \cup F_2 \mid R) &= P(F_1 \mid R) + P(F_2 \mid R) - P(F_1 \cap F_2 \mid R) \\
&= \frac{P(R \mid F_1)P(F_1)}{P(R)} + \frac{P(R \mid F_2)P(F_2)}{P(R)} - \frac{P(R \mid F_1 \cap F_2)P(F_1 \cap F_2)}{P(R)} \\
&= \frac{P(R \mid F_1)P(F_1) + P(R \mid F_2)P(F_2) - P(R \mid F_1 \cap F_2)P(F_1 \cap F_2)}{P(R)} \\
&= \frac{0.90 \cdot 0.080 + 0.70 \cdot (1 - 0.925) - 0.95 \cdot 0.006}{0.15} = \frac{0.1188}{0.15} = \underline{\underline{0.792}}
\end{aligned}
$$

## Oppgave 2 - Trafikkmåling

### a)

Vi finner sannsynlighetene ved å slå opp i tabell:

$$
\begin{aligned}
&P(X > 20) = P(X \leq 20) = 1 - 0.9170 = \underline{\underline{0.0830}} \\
&P(10 \leq X < 20) = P(X \leq 19) - P(X \leq 9) = 0.8750 - 0.0699 = \underline{\underline{0.8051}}.
\end{aligned}
$$

Utleder $E(X)$ ved å benytte definisjonen av forventningsverdi

$$
\begin{aligned}
E(X) &= \sum_x x \cdot f(x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x}{x!} = e^{-\lambda} \cdot 0 + e^{-\lambda} \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x}{x!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda\, e^{-\lambda}\, e^{\lambda} = \underline{\underline{\lambda}}.
\end{aligned}
$$

Som stemmer med resultatet i formelsamlingen. I nest siste overgangen gjenkjenner vi taylorrekka til $e^{\lambda}$.

## b)

Ved bruk av sentralgrenseteoremet har vi at

$$\frac{\hat{\lambda} - E(\hat{\lambda})}{\sqrt{Var(\hat{\lambda})}} \approx n(0,1),$$

siden $\hat{\lambda}$ er gjennomsnittet av uavhengige og identisk fordelte stokasiske variable. Finner $E(\hat{\lambda})$ og $Var(\hat{\lambda})$,

$$E(\hat{\lambda}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E(X_i)}_{=\lambda} = \lambda$$
$$Var(\hat{\lambda}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{Var(X_i)}_{=\lambda} = \lambda/n.$$

Dette betyr at

$$P\left(-z_{\alpha/2} < \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} < z_{\alpha/2}\right) \approx 1 - \alpha,$$

hvor $z_{\alpha/2}$ er $\alpha/2$ kvantilen i standard normalfordelingen. Det er vanskelig å løse denne ulikheten mht på $\lambda$ så vi velger derfor å bytte ut $\lambda$ med $\hat{\lambda}$ i nevner

$$P\left(-z_{\alpha/2} < \frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}/n}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

Løser så mhp $\lambda$

$$P\left(\hat{\lambda} - z_{\alpha/2}\sqrt{\hat{\lambda}/n} < \lambda < \hat{\lambda} + z_{\alpha/2}\sqrt{\hat{\lambda}/n}\right) \approx 1 - \alpha. \tag{1}$$

Vi vet da fra (1) at invervallet $\underline{\underline{\left[\hat{\lambda} - z_{\alpha/2}\sqrt{\hat{\lambda}/n}, \hat{\lambda} + z_{\alpha/2}\sqrt{\hat{\lambda}/n}\right]}}$ dekker den ukjente $\lambda$ med tilnærmet sannsynlighet $1 - \alpha$ (når $n$ er tilstrekkelig stor).

Numerisk: $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = 359/30 = 11.9667$. Setter inn i konfidensintervallet og får

$$[11.9667 - 2.576\sqrt{11.9667/30}, 11.9667 + 2.576\sqrt{11.9667/30}] = \underline{\underline{[10.02, 13.91]}}$$

## c)

Likelihoodfunksjonen er lik simultanfordelingen til dataene,

$$L(\lambda; x_1, \ldots, x_n, y_1, \ldots, y_m) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \prod_{j=1}^{m} \frac{(\lambda/2)^{y_j}}{y_j!} e^{-\lambda/2} = \frac{\lambda^{(\sum_{i=1}^{n} x_i)}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda} \frac{(\lambda/2)^{(\sum_{j=1}^{m} y_j)}}{\prod_{j=1}^{m} y_j!} e^{-m\lambda}.$$

Tar som vanlig logaritmen av likelihoodfunksjonen og får

$$\begin{aligned} & l(\lambda; x_1, \ldots, x_n, y_1, \ldots, y_m) \\ & \quad = \ln(\lambda) \sum_{i=1}^{n} x_i - \ln(\prod_{i=1}^{n} x_i!) - n\lambda + \ln(\lambda/2) \sum_{j=1}^{m} y_j - \ln(\prod_{j=1}^{m} y_j!) - m\lambda/2. \end{aligned} \tag{2}$$

Ønsker nå å finne den verdien av $\lambda$ som maksimerer likelihoodfunksjonen. Finner dette ved å deriverere (2), sette $= 0$ og løse mht på $\lambda$

$$\begin{aligned} \frac{\mathrm{d}l}{\mathrm{d}\lambda} &= \frac{\sum_{i=1}^{n} x_i}{\lambda} - n + \frac{\sum_{j=1}^{m} y_j}{\lambda} - m/2 = 0 \quad | \cdot -\lambda \\ \lambda(n + m/2) &= \sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j. \end{aligned}$$

Dermed får vi sannsynlighetsmaksimeringsestimatoren

$$\hat{\lambda} = \underline{\underline{\frac{\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} y_j}{n + m/2}}}.$$

## Oppgave 3 - Tomat-produksjon

Vi gjennkjenner dette som en regresjonsmodel:

$$Y = 100 + \beta(x - x_r) + E,$$

hvor $E \sim n(e; 0, 15)$.

### a)

La $x = x_r$ som gir at $Y_r = 100 + E \sim n(y_r; 100, 15)$

$$\begin{aligned} P(Y_r > 110) &= 1 - P(Y_r \leq 110) \\ &= 1 - P(\frac{Y_r - 100}{15} < \frac{110 - 100}{15}) \\ &= 1 - \Phi(0.667) = \underline{\underline{0.2514}} \end{aligned}$$

$$\begin{aligned} P(90 < Y_r < 110) &= 1 - P(\frac{90 - 100}{15} < \frac{Y_r - 100}{15} < \frac{110 - 100}{15}) \\ &= \Phi(0.667) - \Phi(-0.667) \\ &= 0.7486 - 0.2514 = \underline{\underline{0.4972}} \end{aligned}$$

$$P(\text{En mer enn dobbelt så tung som den andre}) = 2P(Y_r^1 > 2Y_r^2) = 2P(Y_r^1 - 2Y_r^2 > 0),$$

hvor $Y_r^1 - 2Y_r^2 \sim n(-100, \sqrt{(15^2 + 4 \cdot 15^2)})$. Dette gir videre

$$\begin{aligned} 2 * P(Y_r^1 > 2Y_r^2) &= 2P(\frac{Y_r^1 - 2Y_r^2 - (-100)}{\sqrt{5} \cdot 15} > \frac{100}{\sqrt{5} \cdot 15}) \\ &= 2(1 - \Phi(2.98)) = 2(1 - 0.9986) = \underline{\underline{0.0028}} \end{aligned}$$

**b)**

Drivhus lysintensitet: $x_1, \ldots, x_5$
Observasjoner: $Y_{ij}, i = 1, \ldots, 5, j = 1, \ldots, 3$
Velger å finne en estimator ved å bruke maksimum likelihood metoden. (Det finnes klart andre måter å finne estimatorer på som feks minste kvadraters metode eller mer på ren intuisjon.)

$$L(\beta) = \prod_{j=1}^{3} \prod_{i=1}^{5} \text{const} \cdot \exp\left\{ -\frac{1}{2 \cdot 15^2} \left[ y_{ij} - 100 - \beta(x_i - x_r) \right]^2 \right\}$$

Tar som vanlig logaritmen

$$l(\beta) = \sum_{j=1}^{3} \sum_{i=1}^{5} \left\{ \text{const} - \frac{1}{2 \cdot 15^2} \left[ y_{ij} - 100 - \beta(x_i - x_r) \right]^2 \right\}$$

Deriviverer så $l(\beta)$ mht $\beta$ og setter $= 0$. Løser så mht $\beta$. Dette vil da være maksimum likelihoodestimatoren

$$\frac{\mathrm{d}l(\beta)}{\mathrm{d}\beta} = \sum_{j=1}^{3} \sum_{i=1}^{5} \left[ y_{ij} - 100 - \beta(x_i - x_r) \right] \cdot (-(x_i - x_r)) = 0$$

som gir estimatoren

$$\hat{\beta} = \underline{\frac{\sum_{j=1}^{3} \sum_{i=1}^{5} (Y_{ij} - 100)(x_i - x_r)}{3 \sum_{i=1}^{5} (x_i - x_r)^2}}.$$

Vi sjekker nå forventnigen til estimatoren vi har funnet. Hvis estimatoren er forventningsrett, er alt vel og bra, hvis ikke må vi gjøre en justering for å få den forventningsrett, dette er da typisk å multiplisere med en passende konstant.

$$\begin{aligned} E(\hat{\beta}) &= E\left[ \frac{\sum_{j=1}^{3} \sum_{i=1}^{5} (Y_{ij} - 100)(x_i - x_r)}{3 \sum_{i=1}^{5} (x_i - x_r)^2} \right] \\ &= \frac{\sum_{j=1}^{3} \sum_{i=1}^{5} E[(Y_{ij} - 100)](x_i - x_r)}{3 \sum_{i=1}^{5} (x_i - x_r)^2} \\ &= \frac{\sum_{j=1}^{3} \sum_{i=1}^{5} \beta(x_i - x_r)(x_i - x_r)}{3 \sum_{i=1}^{5} (x_i - x_r)^2} \\ &= \underline{\underline{\beta}} \end{aligned}$$

Altså er estimatoren vi har funnet forventningsrett! Videre skal vi nå finne variansen til estimatoren

$$
\begin{aligned}
Var(\hat{\beta}) &= Var\left[\frac{\sum_{j=1}^{3}\sum_{i=1}^{5}(Y_{ij}-100)(x_i-x_r)}{3\sum_{i=1}^{5}(x_i-x_r)^2}\right] \\
&= \frac{\sum_{j=1}^{3}\sum_{i=1}^{5}(x_i-x_r)^2 Var((Y_{ij}-100))}{[3\sum_{i=1}^{5}(x_i-x_r)^2]^2} \\
&= \frac{15^2\cdot 3\sum_{i=1}^{5}(x_i-x_r)^2}{[3\sum_{i=1}^{5}(x_i-x_r)^2]^2} = \underline{\underline{\frac{15^2}{3\sum_{i=1}^{5}(x_i-x_r)^2}}}
\end{aligned}
$$

## c)

Prediktoren er $\hat{Y_0} = 100 + \hat{\beta}(x_0 - x_r)$. Fra oppgaveteksten har vi at $\hat{\beta} = 2.0$ og at $x_0 - x_r = 5$ som gir at $\hat{y_0} = 110$. Vi undersøker prediksjonsfeil ved å ta utgangspunkt i fordelingen til

$$Y_0 - \hat{Y_0}. \tag{3}$$

Først observerer vi at (3) er normalfordelt fordi det kun er en lineærkombinasjon av normalfordelte stokastiske variable. Det neste er da å finne forventning og varians i denne normalfordelingen,

$$
\begin{aligned}
&E(Y_0-\hat{Y_0}) = E(Y_0) - E(\hat{Y_0}) = 100 + \beta(x_0-x_r) - (100 + E(\hat{\beta})(x_0-x_r)) = \underline{0} \\
&Var(Y_0-\hat{Y_0}) = Var(Y_0) + Var(\hat{Y_0}) = 15^2 + (x_0-x_r)^2\cdot Var(\hat{\beta}) = 15^2 + 5^2\cdot 1.2 = \underline{255}.
\end{aligned}
$$

I den første overgangen i utregning av varians kan vi gjøre fordi $Y_0$ og $\hat{Y_0}$ er uavhengige. Dette er de fordi $\hat{Y_0}$ kun er en funksjon av observesjonene $Y_{ij}, i = 1,\ldots,5, j = 1,\ldots,3$ og altså ikke $Y_0$ (alle $Y$'ene er uavhengige pr definisjon i regresjonsmodellen). Dermed har vi at

$$\frac{Y_0 - \hat{Y_0} - 0}{\sqrt{255}} \sim n(0,1),$$

som gir at

$$P\left(z_{0.025} < \frac{Y_0 - \hat{Y_0} - 0}{\sqrt{255}} < z_{0.025}\right) = 0.95.$$

Løser så med hensyn på det "ukjente" ($Y_0$) og får at $Y_0$ ligger i intervallet

$$\left[\hat{y_0} - z_{0.025}\cdot\sqrt{255}, \hat{y_0} + z_{0.025}\cdot\sqrt{255}\right] = \underline{\underline{[78.7, 141.3]}}$$

med 95% sannsynlighet.