Eksamen i STK1100 våren 2023 - løsningsforslag

Oppgave 1

Definerer hendelsene O: 'tilfeldig valgt bakterie overlever' og M: 'tilfeldig valgt bakterie har mutasjonen'. Fra teksten finner vi følgende sannsynligheter:

$$P(M) = 0.1 \qquad P(O|M) = 0.2 \qquad P(O|M') = 0.01$$

og det følger videere at $P(M') = 1 - P(M) = 0.9$.

a

Bruker setningen om total sannsynlighet:

$$P(O) = P(O|M) \cdot P(M) + P(O|M') \cdot P(M') = 0.2 \cdot 0.1 + 0.01 \cdot 0.9 = 0.029$$

b

Her kan vi bruke Bayes formel:

$$P(M|O) = \frac{P(O|M) \cdot P(M)}{P(O)} = \frac{0.2 \cdot 0.1}{0.029} = 0.69$$

c

Vi antar at vi har som før

$$P(O|M) = 0.2 \qquad P(O|M') = 0.01$$

Men i den nye situasjonen er forekomsten av mutasjoner nå (fra forrige punkt) $P(M) = 0.69$ og dermed $P(M') = 1 - P(M) = 0.31$. Da får vi en overlevelsessannsynlighet på

$$P(O) = P(O|M) \cdot P(M) + P(O|M') \cdot P(M') = 0.2 \cdot 0.69 + 0.01 \cdot 0.31 = 0.141$$

og sannsynligheten for at en av disse har mutasjon, har økt til

$$P(M|O) = \frac{P(O|M) \cdot P(M)}{P(O)} = \frac{0.2 \cdot 0.69}{0.141} = 0.98.$$

Oppgave 2

a

Vi finner først marginaltettheten til X. Vi har

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{\infty} \frac{\alpha}{\mu} y^{\alpha-1} e^{-\frac{x}{\mu} - y^\alpha} dy = \left[-\frac{1}{\mu} e^{-\frac{x}{\mu} - y^\alpha} \right]_0^{\infty} = \frac{1}{\mu} e^{-\frac{x}{\mu}}.$$

Altså er marginaltettheten til X gitt ved

$$f_X(x) = \begin{cases} \frac{1}{\mu}e^{-\frac{x}{\mu}}, & x > 0 \\ 0, & ellers, \end{cases}$$

som er en eksponentialfordeling med parameter $\lambda = \frac{1}{\mu}$. Videre har vi:

$$f_Y(y) = \int_0^{\infty} \frac{\alpha}{\mu} y^{\alpha-1} e^{-\frac{x}{\mu}-y^{\alpha}} dx = \left[-\alpha y^{\alpha-1} e^{-\frac{x}{\mu}-y^{\alpha}} \right]_0^{\infty} = \alpha y^{\alpha-1} e^{-y^{\alpha}}.$$

Altså er marginaltettheten til Y gitt ved

$$f_Y(y) = \begin{cases} \alpha y^{\alpha-1} e^{-y^{\alpha}}, & y > 0 \\ 0, & ellers \end{cases}.$$

b

Når både $x > 0$ og $y > 0$, har vi

$$f_X(x) \cdot f_Y(y) = \frac{\alpha}{\mu} y^{\alpha-1} e^{-\frac{x}{\mu}-y^{\alpha}} = f(x, y).$$

Og dersom enten $x \leq 0$ eller $y \leq 0$ eller begge, så er

$$f_X(x) \cdot f_Y(y) = 0 = f(x, y).$$

Altså er $f(x, y) = f_X(x) \cdot f_Y(y)$ for alle mulige par (x, y), hvilket betyr at X og Y er uavhengige.

c

Momentgenererende funksjon for X finner vi ved:

$$\begin{aligned} M_X(t) = \mathrm{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx &= \int_0^{\infty} e^{tx} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx \\ &= \int_0^{\infty} \frac{1}{\mu} e^{-x\left(\frac{1-t\mu}{\mu}\right)} dx \\ &= \left[-\frac{1}{1-t\mu} e^{-x\left(\frac{1-t\mu}{\mu}\right)} \right]_0^{\infty} \\ &= \frac{1}{1-t\mu}. \end{aligned}$$

Da øvre gense av integralet bare er denlig dersom $(1-t\mu)/\mu > 0$, er mgf bare definert dersom $1 - t\mu > 0$, dvs.dersom $t < 1/\mu$.

d

Vi vet at $\mathrm{E}(X) = M_X'(0)$ og $\mathrm{E}(X^2) = M_X''(0)$. Vi har:

$$M_X'(t) = -\frac{1}{(1-t\mu)^2} \cdot (-\mu) = \frac{\mu}{(1-t\mu)^2}$$

og

$$M_X''(t) = -2\frac{\mu}{(1-t\mu)^3} \cdot (-\mu) = \frac{2\mu^2}{(1-t\mu)^3}$$

Det gir

$$\mathrm{E}(X) = M_X'(0) = \frac{\mu}{(1-0\cdot\mu)^2} = \mu$$

og

$$\mathrm{E}(X^2) = M_X''(0) = \frac{2\mu^2}{(1-0\cdot\mu)^3} = 2\mu^2,$$

slik at $\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 2\mu^2 - \mu^2 = \mu^2$.

Oppgave 3

a

Vi får oppgitt at X_i er eksponensialfordelt. Hvis vi skriver opp tettheten til eksponensialfordelingen med parameter λ, der $\lambda = 1/\mathrm{E}(X_i)$, og reparameteriserer med $\mu = 1/\lambda = \mathrm{E}(X_i)$, får vi nettopp

$$f_{X_i}(x) = \begin{cases} \frac{1}{\mu}e^{-\frac{x}{\mu}}, & x > 0 \\ 0, & ellers. \end{cases}$$

Likelihood-funksjonen finner vi ved

$$L(\mu) = f(x_1, x_2, \cdots, x_n; \mu) = \prod_{i=1}^{n} \frac{1}{\mu}e^{-\frac{x_i}{\mu}} = \frac{1}{\mu^n}e^{-\frac{\sum_{i=1}^{n} x_i}{\mu}},$$

der $f(x_1, x_2, \cdots, x_n; \mu)$ er simultantettheten til $X_1, \ldots, X_n$ (her har vi brukt antakelsen om uavhengighet), og log-likelihood-funksjonen blir dermed

$$l(\mu) = \log L(\mu) = -n\log\mu - \frac{1}{\mu}\sum_{i=1}^{n} x_i.$$

b

Vi deriverer $l(\mu)$ m.h.p. μ og setter lik 0:

$$l'(\mu) = -\frac{n}{\mu} + \frac{\sum x_i}{\mu^2} = 0$$

Løser med hensyn på μ og får ML-estimatoren

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

Hvis vi setter inn tallene fra oppgaven, får vi estimatet

$$\hat{\mu} = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{1}{30}\cdot 1200 = 40 \text{ sekunder.}$$

Momentestimator finner vi ved å sette første moment $\mathrm{E}(X)$ lik gjennomsnittet, som her gir momentestimatoren $\hat{\hat{\mu}} = \bar{X}$, som er det samme som ML-estimatoren $\hat{\mu}$.

c

Fra 2d) vet vi at $\mathrm{E}(X_i) = \mu$ og $\mathrm{V}(X_i) = \mu^2$. Da får vi

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}(\bar{X}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu,$$

$$\mathrm{V}(\hat{\mu}) = \mathrm{V}(\bar{X}) = \mathrm{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mu^2 = \frac{1}{n}\mu^2,$$

der vi har brukt antakelsen om uavhengighet i beregningen av $\mathrm{V}(\hat{\mu})$. Sentralgrenseteoremet sier at når n er stor, vil gjennomsnittet $\bar{X} = 1/n\sum_{i=1}^{n} X_i$ av et tilfeldig utvalg på størrelse n fra en fordeling med forventning μ og standardavvik σ være asymptotisk normalfordelt med forventning μ og standardavvik $\sigma/\sqrt{n}$. Her er $\sigma = \mu$, og resultatet følger.

Standardfeilen til $\hat{\mu}$ er $\sqrt{\mathrm{V}(\hat{\mu})} = \mu/\sqrt{n}$. Her kan vi sette inn estimatet for μ, og får da estimert standardfeil

$$\hat{SE}_{\hat{\mu}} = \hat{\mu}/\sqrt{n} = 40/\sqrt{30} = 7.3 \text{ sekunder.}$$

d

Vi vet fra 2c) at $M_{X_i}(t) = 1/(1 - t\mu)$. Vi vil finne MGF for $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$, som en-entydig identifiserer fordelingen til $\hat{\mu}$. Bruker først regelen for MGF til en sum av uavhengige variable,

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \frac{1}{(1-t\mu)} = \frac{1}{(1-t\mu)^n}.$$

Bruker deretter regelen for MGF når vi multipliserer med en konstant, som gir oss

$$M_{\frac{1}{n}\sum_{i=1}^n X_i}(t) = M_{\sum_{i=1}^n X_i}(t/n) = \frac{1}{(1-t\mu/n)^n}.$$

Fra formelsamlingen finner vi at gamma$(n, \mu/n)$ har MGF som ovenfor. Derfor kan vi konkludere med at $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i \sim$ gamma$(n, \mu/n)$.

Oppgave 4

a

Vi begynner med å finne den kumulative fordelingsfunksjonen til Y:

$$F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}\left(\frac{2}{\mu}X \le y\right) = \mathrm{P}\left(X \le \frac{\mu y}{2}\right) = F_X\left(\frac{\mu y}{2}\right).$$

der $F_X(x)$ er den kumulative fordelingsfunksjonen til eksponentialfordelingen med parameter $\lambda = 1/\mu$, dvs. $F_X(x) = 1 - e^{-\frac{x}{\mu}}$, og vi får

$$F_Y(y) = F_X\left(\frac{\mu y}{2}\right) = 1 - e^{-\frac{1}{\mu}\cdot\frac{\mu y}{2}} = 1 - e^{-\frac{y}{2}}.$$

Tettheten til Y finner vi så ved å derivere $F_Y(y)$ m.h.p. y:

$$f_Y(y) = F_Y'(y) = \frac{1}{2}e^{-\frac{y}{2}}.$$

Dette er tettheten for $y > 0$, mens den er 0 for $y \le 0$, som er tettheten til *gamma* $(1, 2)$-fordelingen.

Da Y er en monoton funksjon av X, kunne en her brukt transformasjonsformelen til å finne tettheten til Y direkte. En annen mulighet er å bruke mgf for X.

b

Vi har:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\sum_{i=1}^n \frac{\mu}{2}Y_i = \frac{\mu}{2n}\sum_{i=1}^n Y_i,$$

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

c

Fra Oppgave b) vet vi at $\frac{2n\hat{\mu}}{\mu} \sim \chi^2_{2n}$. Dermed har vi

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

Et 95% konfidensintervall for μ er dermed gitt ved

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

Med $\chi^2_{60,0.975} = 40.48$, $\chi^2_{60,0.025} = 83.30$ og $\hat{\mu} = \bar{x} = 40$ får vi følgende intervall:

$$
\left(\frac{2 \cdot 30 \cdot 40}{83.30}, \frac{2 \cdot 30 \cdot 40}{40.48}\right) = (28.8, 59.3).
$$

d

Når en har et stort utvalg, kan en benytte seg av at $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, der $\sigma = \sqrt{\mathrm{V}(X)} = \mu$, er tilnærmet standard normalfordelt i henhold til sentralgrenseteoremet. Dermed er

$$
Z = \frac{\bar{X} - \mu}{\mu/\sqrt{n}}
$$

tilnærmet standard normalfordelt. Vi får:

$$\begin{aligned}
&\mathrm{P}\left(-1.96 \leq Z \leq 1.96\right) \approx 0.95\\
\rightarrow &\mathrm{P}\left(-1.96 \leq \frac{\bar{X}-\mu}{\mu/\sqrt{n}} \leq 1.96\right) \approx 0.95\\
\rightarrow &\mathrm{P}\left(-\frac{1.96}{\sqrt{n}} \leq \frac{\bar{X}}{\mu} - 1 \leq \frac{1.96}{\sqrt{n}}\right) \approx 0.95\\
\rightarrow &\mathrm{P}\left(\frac{1-1.96/\sqrt{n}}{\bar{X}} \leq \frac{1}{\mu} \leq \frac{1+1.96/\sqrt{n}}{\bar{X}}\right) \approx 0.95\\
\rightarrow &\mathrm{P}\left(\frac{\bar{X}}{1-1.96/\sqrt{n}} \geq \mu \geq \frac{\bar{X}}{1+1.96/\sqrt{n}}\right) \approx 0.95.
\end{aligned}$$

Et tilnærmet 95% konfidensintervall for μ er dermed gitt ved $\left(\frac{\bar{x}}{1+1.96/\sqrt{n}}, \frac{\bar{x}}{1-1.96/\sqrt{n}}\right)$. Alternativt kan en benytte seg av at for store n vil også

$$Z = \frac{\bar{X}-\mu}{S/\sqrt{n}},$$

med $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ også er tilnærmet standard normalfordelt. Da oppnår en i stedet det tilnærmede konfidensintervallet $\left(\bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}}\right)$.