

**Løsningsforslag til eksamen i
STK1100 31. mai 2019**

Oppgave 1

a) Antall fødsler X i løpet av én time er Poissonfordelt med parameter $\lambda = t/2$. For $x = 0, 1, 2, \dots$ er da

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} = \frac{(t/2)^x}{x!} e^{-t/2}.$$

Med $t = 1$ finner vi at

$$\begin{aligned} &P(\text{minst ett barn i løpet av én time}) \\ &= 1 - P(\text{ingen barn i løpet av én time}) \\ &= 1 - P(X = 0) = 1 - \frac{(1/2)^0}{0!} e^{-1/2} = 1 - e^{-1/2} = 1 - 0.607 = 0.393. \end{aligned}$$

Videre finner vi når $t = 2$ at

$$\begin{aligned} &P(\text{minst to barn i løpet av to timer}) \\ &= 1 - P(\text{høyst ett barn i løpet av to timer}) \\ &= 1 - P(X = 0) - P(X = 1) = 1 - \frac{(2/2)^0}{0!} e^{-2/2} - \frac{(2/2)^1}{1!} e^{-2/2} \\ &= 1 - 2e^{-1} = 1 - 0.736 = 0.264. \end{aligned}$$

b) For $t > 0$ har vi at $T > t$ hvis og bare hvis $X = 0$. Derfor er

$$P(T > t) = P(X = 0) = \frac{(t/2)^0}{0!} e^{-t/2} = e^{-t/2}.$$

For $t > 0$ er dermed den kumulative fordelingen til T gitt ved

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-t/2},$$

mens $F(t) = 0$ for $t \leq 0$. Tettheten til T blir dermed

$$f(t) = F'(t) = \begin{cases} \frac{1}{2}e^{-t/2} & \text{for } t > 0, \\ 0 & \text{ellers.} \end{cases}$$

Forventningsverdien til T er (substituerer $u = t/2$)

$$E(T) = \int_{-\infty}^{\infty} tf(t)dt = \int_0^{\infty} t \frac{1}{2} e^{-t/2} dt = \frac{1}{2} \int_0^{\infty} (2u) e^{-u} 2du = 2 \int_0^{\infty} u e^{-u} du = 2\Gamma(1) = 2.$$

(Det er her også helt fint å bruke formelen for forventningsverdien i eksponentialfordelingen.)

Forventet tid for første fødsel etter midnatt er kl 02.00.

c) Vi har at

$$\begin{aligned} P(\text{ingen barn blir født de to første timene etter midnatt}) \\ = P(T > 2) = e^{-2/2} = e^{-1} = 0.368 \end{aligned}$$

Hvis vi må vente minst to timer etter kl. 02.00, har klokka blitt 04.00. Vi har derfor at

$$\begin{aligned} P(\text{vente minst to timer til} | \text{ingen barn blir født de to første timene etter midnatt}) \\ = P(T > 4 | T > 2) = \frac{P((T > 4) \cap (T > 2))}{P(T > 2)} = \frac{P(T > 4)}{P(T > 2)} = \frac{e^{-4/2}}{e^{-2/2}} = e^{-1} = 0.368 \end{aligned}$$

Oppgave 2

a) For $x > 0$ er den kumulative fordelingen til X gitt ved $F(x) = 1 - e^{-x^2/(2\theta^2)}$. Medianen m i fordelingen er gitt ved at $F(m) = 1/2$. Vi løser denne likningen:

$$\begin{aligned} F(m) &= \frac{1}{2} \\ 1 - e^{-m^2/(2\theta^2)} &= \frac{1}{2} \\ e^{-m^2/(2\theta^2)} &= \frac{1}{2} \\ \frac{m^2}{2\theta^2} &= \ln(2) \\ m^2 &= 2\theta^2 \ln(2) \\ m &= \theta\sqrt{2 \ln(2)} \end{aligned}$$

For $x > 0$ er sannsynlighetstettheten til X er gitt ved

$$f(x) = F'(x) = -e^{-x^2/(2\theta^2)} \left(-\frac{2x}{2\theta^2} \right) = \frac{1}{\theta^2} x e^{-x^2/(2\theta^2)},$$

mens $f(x) = 0$ for $x \leq 0$. Oppsummert har vi at

$$f(x) = \begin{cases} \frac{1}{\theta^2} x e^{-x^2/(2\theta^2)} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

b) Vi har at [substituerer $u = x^2/(2\theta^2)$]

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \frac{1}{\theta^2} x e^{-x^2/(2\theta^2)} dx \\ &= \frac{1}{\theta^2} \int_0^{\infty} x^2 e^{-x^2/(2\theta^2)} dx = \frac{1}{\theta^2} \int_0^{\infty} (\theta\sqrt{2u})^2 e^{-u} \frac{\theta du}{\sqrt{2u}} \\ &= \theta\sqrt{2} \int_0^{\infty} u^{(3/2)-1} e^{-u} du = \theta\sqrt{2} \Gamma(3/2) = \theta\sqrt{2} (1/2) \Gamma(1/2) \\ &= \theta \frac{1}{\sqrt{2}} \sqrt{\pi} = \theta \sqrt{\frac{\pi}{2}} \end{aligned}$$

Tilsvarende har vi at

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^{\infty} x^2 \frac{1}{\theta^2} x e^{-x^2/(2\theta^2)} dx \\ &= \frac{1}{\theta^2} \int_0^{\infty} x^3 e^{-x^2/(2\theta^2)} dx = \frac{1}{\theta^2} \int_0^{\infty} (\theta\sqrt{2u})^3 e^{-u} \frac{\theta du}{\sqrt{2u}} \\ &= 2\theta^2 \int_0^{\infty} u^{2-1} e^{-u} du = 2\theta^2 \Gamma(2) = 2\theta^2 \cdot 1 \Gamma(1) = 2\theta^2 \end{aligned}$$

Variansen blir dermed

$$V(X) = E(X^2) - [E(X)]^2 = 2\theta^2 - \theta^2 \frac{\pi}{2} = \frac{4-\pi}{2} \theta^2$$

c) For $y > 0$ er den kumulative fordelingen til Y gitt ved

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P((X/\theta)^2 \leq y) = P(X/\theta \leq \sqrt{y}) = P(X \leq \theta\sqrt{y}) \\ &= 1 - e^{-(\theta\sqrt{y})^2/(2\theta^2)} = 1 - e^{-y/2}, \end{aligned}$$

mens $F_Y(y) = 0$ for $y \leq 0$. Tettheten til Y blir dermed

$$f_Y(y) = F'_Y(y) = \begin{cases} \frac{1}{2} e^{-y/2} & \text{for } y > 0, \\ 0 & \text{ellers.} \end{cases}$$

Dette er gamma-tettheten med formparameter $\alpha = 1$ og skalaparameter $\beta = 2$, så Y er gamma $(1, 2)$ -fordelt.

Oppgave 3

a) Ved resultatene i spørsmål b i oppgave 2, har vi at

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n \theta \sqrt{\frac{\pi}{2}} = \theta \sqrt{\frac{\pi}{2}},$$

og

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n V(X_i) = \frac{1}{n^2} n \frac{4-\pi}{2} \theta^2 = \frac{4-\pi}{2n} \theta^2.$$

For estimatoren $\hat{\theta} = \bar{X} \sqrt{\frac{2}{\pi}}$ får vi dermed at

$$E(\hat{\theta}) = E\left(\bar{X} \sqrt{\frac{2}{\pi}}\right) = \sqrt{\frac{2}{\pi}} E(\bar{X}) = \sqrt{\frac{2}{\pi}} \theta \sqrt{\frac{\pi}{2}} = \theta,$$

så estimatoren er forventningsrett. Variansen til estimatoren er gitt ved

$$V(\hat{\theta}) = V\left(\bar{X} \sqrt{\frac{2}{\pi}}\right) = \left(\sqrt{\frac{2}{\pi}}\right)^2 V(\bar{X}) = \frac{2}{\pi} \cdot \frac{4-\pi}{2n} \theta^2 = \frac{4-\pi}{\pi n} \theta^2.$$

b) Sett $Y_i = (X_i/\theta)^2$. Ved resultatet i spørsmål c i oppgave 2 har vi at Y_i -ene er gamma $(1, 2)$ -fordelte. Videre vet vi at en sum av uavhengige gammafordelte variabler som alle har samme skalaparameter, selv er gammafordelt. Derfor er $\sum_{i=1}^n Y_i = \sum_{i=1}^n (X_i/\theta)^2$ gamma $(n, 2)$ -fordelt, dvs. gammafordelt med formparameter $\alpha = n$ og skalaparameter $\beta = 2$.

Ved å bruke formelen for forventningen til en gammafordelt variabel, finner vi nå at

$$E\left(\sum_{i=1}^n (X_i/\theta)^2\right) = \alpha\beta = n \cdot 2 = 2n.$$

Derfor er

$$E\left(\frac{1}{2n} \sum_{i=1}^n X_i^2\right) = \frac{\theta^2}{2n} E\left(\sum_{i=1}^n (X_i/\theta)^2\right) = \frac{\theta^2}{2n} 2n = \theta^2.$$

c) Vi har estimatoren $\tilde{\theta} = \sqrt{\frac{1}{2n} \sum_{i=1}^n X_i^2}$. Vi merker oss at $\sum_{i=1}^n X_i^2 = 2n\tilde{\theta}^2$. Dermed har vi at $2n\tilde{\theta}^2/\theta^2 = \sum_{i=1}^n (X_i/\theta)^2$ er gamma $(n, 2)$ -fordelt ved resultatet i spørsmål b. Hvis a og b er henholdsvis 2.5% og 97.5% persentilene i gammafordelingen med formparameter $\alpha = n$ og skalaparameter $\beta = 2$, har vi dermed at

$$P\left(a \leq \frac{2n\tilde{\theta}^2}{\theta^2} \leq b\right) = 0.95.$$

Vi omformer ulikhetene og får dermed at

$$P\left(\frac{2n\tilde{\theta}^2}{b} \leq \theta^2 \leq \frac{2n\tilde{\theta}^2}{a}\right) = 0.95,$$

som gir

$$P\left(\sqrt{\frac{2n}{b}}\tilde{\theta} \leq \theta \leq \sqrt{\frac{2n}{a}}\tilde{\theta}\right) = 0.95,$$

Et 95% konfidensinterval for θ er dermed gitt ved

$$\left[\sqrt{\frac{2n}{b}}\tilde{\theta}, \sqrt{\frac{2n}{a}}\tilde{\theta}\right]$$

d) Vi finner estimatene

$$\hat{\theta} = \bar{x}\sqrt{\frac{2}{\pi}} = 7.19\sqrt{\frac{2}{\pi}} = 5.74,$$

og

$$\tilde{\theta} = \sqrt{\frac{1}{40} \sum_{i=1}^n x_i^2} = \sqrt{30.51} = 5.52.$$

Et 95% konfidensintervall for θ er gitt ved

$$\left[\sqrt{\frac{40}{49.3}} 5.52, \sqrt{\frac{40}{24.4}} 5.52 \right],$$

dvs.

$$[4.97, 7.07].$$

e) For å bestemme standardfeilen med parametrisk bootstrap går vi fram på følgende måte:

For $b = 1, 2, \dots, B$ (for eksempel med $B = 10000$):

- Vi generer 20 uavhengige observasjoner fra Rayleigh fordelingen med parameter θ lik den estimerte verdien $\theta = 5.52$. Kall de genererte verdiene $x_1^*, x_2^*, \dots, x_{20}^*$. (For å trekke fra Rayleigh fordelingen kan vi bruke inversjonsmetoden.)
- Beregn bootstrap estimatene $\tilde{\theta}_b^* = \sqrt{\frac{1}{40} \sum_{i=1}^{20} (x_i^*)^2}$ og $\hat{\theta}_b^* = \bar{x}_b^* \sqrt{\frac{2}{\pi}}$, der $\bar{x}_b^* = \frac{1}{20} \sum_{i=1}^{20} x_i^*$.

Vi finner så de estimerte standardfeilene ved

$$s_{\tilde{\theta}} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\tilde{\theta}_b^* - \bar{\tilde{\theta}}^*)^2}$$

$$s_{\hat{\theta}} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2}$$

der $\bar{\tilde{\theta}}^* = \frac{1}{B} \sum_{b=1}^B \tilde{\theta}_b^*$ og $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*$.

Oppgave 4

a) For $x > 0$ er den marginale tettheten til X gitt ved

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_0^{\infty} xy e^{-(x+y)} dy = x e^{-x} \int_0^{\infty} y e^{-y} dy = x e^{-x} \Gamma(2) = x e^{-x}.$$

For $x \leq 0$ er $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = 0$. Altså har vi at

$$f_X(x) = \begin{cases} x e^{-x} & \text{for } x > 0, \\ 0 & \text{ellers.} \end{cases}$$

Den simultane tettheten $f_{XY}(x, y)$ er symmetrisk i x og y . Den marginale tettheten til Y er derfor

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

For $x > 0$ og $y > 0$ har vi at

$$f_{XY}(x, y) = xye^{-(x+y)} = xe^{-x}ye^{-y} = f_X(x)f_Y(y).$$

Hvis $x \leq 0$ og/eller $y \leq 0$ er

$$f_{XY}(x, y) = 0 = f_X(x)f_Y(y).$$

Altså er $f_{XY}(x, y) = f_X(x)f_Y(y)$ for alle x, y , så X og Y er uavhengige.

b) Vi har $U = X + Y$ og $V = X/(X + Y)$. Den simultane tettheten $f_{XY}(x, y) > 0$ for $x > 0$ og $y > 0$. Det følger at den simultane tettheten $f_{UV}(u, v)$ til U og V er positiv for $u > 0$ og $0 < v < 1$.

Den inverse transformasjonen er gitt ved $X = UV$ og $Y = U(1 - V)$, og Jacobi-determinanten er

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 1 - v & -u \end{vmatrix} = -uv - (1 - v)u = -u.$$

For $u > 0$ og $0 < v < 1$ har vi dermed at simultantettheten til U og V er gitt ved

$$f_{UV}(u, v) = f_{XY}(uv, u(1 - v))|-u| = uv \cdot u(1 - v)e^{-(uv+u(1-v))}u = u^3v(1 - v)e^{-u},$$

mens $f_{UV}(u, v) = 0$ ellers.