Løsningsforslag til eksamen i

STK1100 9. juni 2015

Oppgave 1

a) Den momentgenererende funksjonen til X er (for $t < 1/\beta$)

$$
\begin{aligned}
M_X(t) = E\left(e^{tX}\right) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx = \int_0^{\infty} e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)}\, x^{\alpha-1} e^{-x/\beta}\,dx \\
&= \int_0^{\infty} \frac{1}{\beta^\alpha \Gamma(\alpha)}\, x^{\alpha-1} e^{-(1-\beta t)x/\beta}\,dx \\
&= \int_0^{\infty} \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{u}{1-\beta t}\right)^{\alpha-1} e^{-u/\beta}\, \frac{du}{1-\beta t} \\
&= \frac{1}{(1-\beta t)^\alpha} \int_0^{\infty} \frac{1}{\beta^\alpha \Gamma(\alpha)}\, u^{\alpha-1} e^{-u/\beta}\,du \\
&= \frac{1}{(1-\beta t)^\alpha}.
\end{aligned}
$$

Her har vi substituert $u = (1-\beta t)x$ og brukt at integralet av Gamma(α, β)-tettheten er lik én.

b) Vi kan skrive $M_X(t) = (1-\beta t)^{-\alpha}$. Vi deriverer to ganger med hensyn på t og finner

$$
\begin{aligned}
M_X'(t) &= -\alpha(1-\beta t)^{-\alpha-1}(-\beta) = \alpha\beta(1-\beta t)^{-\alpha-1}, \\
M_X''(t) &= (-\alpha-1)\alpha\beta(1-\beta t)^{-\alpha-2}(-\beta) = (\alpha+1)\alpha\beta^2(1-\beta t)^{-\alpha-2}.
\end{aligned}
$$

Dermed er

$$E(X) = M_X'(0) = \alpha\beta,$$

og

$$V(X) = E(X^2) - (E(X))^2 = M_X''(0) - (\alpha\beta)^2 = (\alpha+1)\alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2.$$

c) Siden X_1 og X_2 er uavhengige, har vi at

$$M_{X_1+X_2}(t) = E\left(e^{t(X_1+X_2)}\right) = E\left(e^{tX_1} e^{tX_2}\right) = E\left(e^{tX_1}\right) E\left(e^{tX_2}\right) = M_{X_1}(t) M_{X_2}(t).$$

Nå har vi fra resultatet i punkt a at

$$M_{X_1}(t) = \frac{1}{(1-\beta t)^{\alpha_1}} \quad \text{og} \quad M_{X_2}(t) = \frac{1}{(1-\beta t)^{\alpha_2}}.$$

Den momentgenererende funksjonen til $X_1 + X_2$ blir dermed

$$M_{X_1+X_2}(t) = \frac{1}{(1-\beta t)^{\alpha_1}} \cdot \frac{1}{(1-\beta t)^{\alpha_2}} = \frac{1}{(1-\beta t)^{\alpha_1+\alpha_2}}.$$

Av resultatet i punkt a ser vi at dette er den momentgenererende funksjonen til gamma-fordelingen med formparameter $\alpha_1 + \alpha_2$ og skalaparameter β. Siden den momentgenererende funksjonen bestemmer fordelingen entydig, har vi at

$$X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta), \tag{A}$$

slik vi skulle vise.

Anta så at $X_1, X_2, \ldots, X_n$ er uavhengige og at $X_i \sim \text{Gamma}(\alpha_i, \beta)$ for $i = 1, 2, \ldots, n$. Vi vil vise at

$$\sum_{i=1}^{n} X_i \sim \text{Gamma}\left(\sum_{i=1}^{n} \alpha_i, \beta\right). \tag{B}$$

Av (A) følger det at (B) holder for $n = 2$. Anta så at (B) holder for $n = k-1$, altså at

$$\sum_{i=1}^{k-1} X_i \sim \text{Gamma}\left(\sum_{i=1}^{k-1} \alpha_i, \beta\right).$$

Sett $Y = \sum_{i=1}^{k-1} X_i$. Da har vi at $\sum_{i=1}^{k} X_i = Y + X_k$, der $Y \sim \text{Gamma}\left(\sum_{i=1}^{k-1} \alpha_i, \beta\right)$ og $X_k \sim \text{Gamma}(\alpha_k, \beta)$. Av (A) har vi da at

$$Y + X_k \sim \text{Gamma}\left(\sum_{i=1}^{k-1} \alpha_i + \alpha_k, \beta\right).$$

Dermed holder (B) også for $n = k$, og det følger ved induksjon at (B) holder for vilkårlig verdi av n.

Oppgave 2

a) Den kumulative fordelingsfunksjonen til T er gitt ved

$$F_T(t) = \int_{-\infty}^{t} f_T(u)du.$$

For $t \leq 0$ er $f_T(t) = 0$, og da er $F_T(t) = 0$. For $t > 0$ blir den kumulative fordelingsfunksjonen

$$F_T(t) = \int_0^t \lambda e^{-\lambda u} du = \left[-e^{-\lambda u}\right]_0^t = 1 - e^{-\lambda t}.$$

Altså har vi at

$$F_T(t) = \begin{cases} 0 & \text{for } t \leq 0, \\ 1 - e^{-\lambda t} & \text{for } t > 0. \end{cases}$$

b) Median tid til tilbakefall m er gitt ved

$$F_T(m) = P(T \leq m) = 1 - e^{-\lambda m} = \frac{1}{2}.$$

Det gir

$$\begin{aligned} e^{-\lambda m} &= \frac{1}{2} \\ -\lambda m &= -\ln(2) \\ m &= \frac{\ln(2)}{\lambda} \end{aligned}$$

Vi har at $P(T < m) = P(T \leq m) = 0.50$ og $P(T > m) = 1 - P(T \leq m) = 0.50$. Det er like sannsynlig at tid til tilbakefall er mindre enn medianen som at tid til tilbakefall er større enn medianen.

c) Eksponentialfordelingen er et spesialtilfelle av gamma-fordelingen der formparameteren er én og skalaparameteren er $1/\lambda$, dvs. $T_i \sim \text{Gamma}(1, 1/\lambda)$. Av resultatet i oppgave 1b har vi dermed at $E(T_i) = 1 \cdot (1/\lambda) = 1/\lambda$ og $V(T_i) = 1 \cdot (1/\lambda)^2 = 1/\lambda^2$.

Vi finner dermed at

$$E(\widehat{m}) = E\left(\frac{\ln(2)}{n}\sum_{i=1}^{n} T_i\right) = \frac{\ln(2)}{n}\sum_{i=1}^{n} E\left(T_i\right) = \frac{\ln(2)}{n} n \frac{1}{\lambda} = \frac{\ln(2)}{\lambda} = m,$$

så $\widehat{m}$ er forventningsrett. Videre har vi at

$$V(\widehat{m}) = V\left(\frac{\ln(2)}{n}\sum_{i=1}^{n} T_i\right) = \left(\frac{\ln(2)}{n}\right)^2 \sum_{i=1}^{n} V\left(T_i\right) = \left(\frac{\ln(2)}{n}\right)^2 n \frac{1}{\lambda^2} = \frac{(\ln(2))^2}{n\lambda^2},$$

så standardfeilen blir

$$\sigma_{\widehat{m}} = \sqrt{V(\widehat{m})} = \frac{\ln(2)}{\sqrt{n}\,\lambda}.$$

Av Chebyshevs ulikhet har vi at for enhver $\epsilon > 0$ er

$$P\left(|\widehat{m} - m| > \epsilon\right) = P\left(|\widehat{m} - E(\widehat{m})| > \epsilon\right) \leq \frac{V(\widehat{m})}{\epsilon^2} = \frac{(\ln(2))^2}{n\epsilon^2\lambda^2} \to 0 \quad \text{når} \quad n \to \infty.$$

Derfor er $\widehat{m}$ konsistent.

d) Vi setter $X_i = 2\lambda T_i$ for $i = 1, 2, \ldots, n$. For $x > 0$ blir den kumulative fordelingen til X_i-ene (jf. punkt a):

$$\begin{aligned} F_X(x) &= P(X_i \leq x) = P(2\lambda T_i \leq x) = P\left(T_i \leq \frac{x}{2\lambda}\right) \\ &= F_T\left(\frac{x}{2\lambda}\right) = 1 - e^{-\lambda\left(\frac{x}{2\lambda}\right)} = 1 - e^{-x/2}. \end{aligned}$$

For $x \leq 0$, er $F_X(x) = 0$. Det følger at sannsynlighetstettheten til X_i er

$$f_X(x) = F_X'(x) = \begin{cases} \frac{1}{2}e^{-x/2} & \text{for } x > 0, \\ 0 & \text{ellers.} \end{cases}$$

Hvis vi sammenholder denne tettheten med gamma-tettheten gitt i oppgave 1, og bruker at $\Gamma(1) = 1$, ser vi at X_i er gamma-fordelt med formparameter 1 og skalaparameter 2. Av resultatet (B) i oppgave 1, følger det da at $2\lambda \sum_{i=1}^{n} T_i = \sum_{i=1}^{n} X_i$ er gamma-fordelt med formparameter $1 + 1 + \cdots + 1 = n$ og skalaparameter 2.

e) Av resultatet i forrige punkt har vi at

$$P\left(a \leq 2\lambda \sum_{i=1}^{n} T_i \leq b\right) = 1 - \alpha.$$

Siden $m = \ln(2)/\lambda$ (punkt b) og $\widehat{m} = (\ln(2)/n)\sum_{i=1}^{n} T_i$ (punkt c), er $\lambda = \ln(2)/m$ og $\sum_{i=1}^{n} T_i = n\widehat{m}/\ln(2)$. Det følger at

$$P\left(a \leq 2n\left(\frac{\widehat{m}}{m}\right) \leq b\right) = 1 - \alpha.$$

Vi omformer ulikhetene og får at

$$P\left(\frac{2n}{b}\,\widehat{m} \leq m \leq \frac{2n}{a}\,\widehat{m}\right) = 1 - \alpha.$$

Et $100(1 - \alpha)\%$ konfidensinterval for medianen m er dermed gitt ved

$$\left[\frac{2n}{b}\,\widehat{m}\,,\ \frac{2n}{a}\,\widehat{m}\right].$$

f) Vi lar $t_1, t_2, \ldots, t_{21}$ være de observerte tidene til tilbakefall for de 21 kreftpasientene. Da har vi at $\sum_{i=1}^{21} t_i = 182$. Et estimat for median tid til tilbakefall er dermed (i uker):

$$\widehat{m} = \frac{\ln(2)}{n} \sum_{i=1}^{n} t_i = \frac{0.693}{21} 182 = 6.0\,.$$

Et 95% konfidensintervall for medianen er

$$\left[\frac{2 \cdot 21}{61.8} 6.0\,, \ \frac{2 \cdot 21}{26.0} 6.0\right],$$

dvs.

$$[4.1\,,\ 9.7]\,.$$

g) Hvis vi ikke antar at tilbakefallstidene er eksponentialfordelte, kan vi estimere medianen med den empiriske medianen $\widetilde{t}$, dvs. observasjon nummer 11 når tilbakefallstidene er skrevet i stigende rekkefølge. Vi finner at $\widetilde{t} = 8$.

For å bestemme standardfeilen til $\widetilde{t}$ går vi fram på følgende måte:

For $b = 1, 2, \ldots, B$ (for eksempel med $B = 1000$):

- Trekk 21 ganger *med tilbakelegging* fra mengden $\{t_1, t_2, \ldots, t_{21}\}$, og kall de uttrukne verdiene $t_1^*, t_2^*, \ldots, t_{21}^*$.
- Bestem den empiriske medianen av $t_1^*, t_2^*, \ldots, t_{21}^*$ og kall den $\widetilde{t}^{\,b*}$.

Estimer standardfeilen til $\widetilde{t}$ ved

$$\sigma_{\widetilde{t}} = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(\widetilde{t}^{\,b*} - \overline{\widetilde{t}^*}\right)^2},$$

der $\overline{\widetilde{t}^*} = \frac{1}{B} \sum_{b=1}^{B} \widetilde{t}^{b*}$.

Oppgave 3

a) Vi har at

$$\begin{aligned}
\mathrm{MSE}_X(\theta) &= E\{(\theta - X)^2\} = E\{[(\theta - \mu) - (X - \mu)]^2\} \\
&= E\{(\theta - \mu)^2 - 2(\theta - \mu)(X - \mu) + (X - \mu)^2\} \\
&= (\theta - \mu)^2 - 2(\theta - \mu)E(X - \mu) + E\{(X - \mu)^2\} \\
&= (\theta - \mu)^2 + V(X). \qquad \text{(C)}
\end{aligned}$$

I)(C) avhenger ikke $V(X)$ av θ, mens $(\theta-\mu)^2$ blir minst mulig når $\theta=\mu$. Dermed blir $\text{MSE}_X(\theta)$ minst mulig for $\theta=\mu$.

b) Vi har at

$$E(Y)=E\left(\rho X+\sqrt{1-\rho^2}\,Z\right)=\rho E(X)+\sqrt{1-\rho^2}\,E(Z)=0,$$

der den siste likeheten gjelder siden $E(X)=E(Z)=0$. Videre har vi at

$$V(Y)=V\left(\rho X+\sqrt{1-\rho^2}\,Z\right)=\rho^2V(X)+(1-\rho^2)\,V(Z)=\rho^2+(1-\rho^2)=1,$$

der den siste likeheten gjelder siden $V(X)=V(Z)=1$.

c) Vi har at

$$\begin{aligned}\text{Cov}(X,Y)&=\text{Cov}\left(X,\rho X+\sqrt{1-\rho^2}\,Z\right)\\&=\rho\,\text{Cov}(X,X)+\sqrt{1-\rho^2}\,\text{Cov}(X,Z)\\&=\rho\,V(X).\end{aligned}$$

Her har vi brukt at $\text{Cov}(X,X)=V(X)$ og at siden X og Z er uavhengige, er $\text{Cov}(X,Z)=0$. Nå er $V(X)=V(Z)=1$ og det følger at

$$\text{Corr}(X,Y)=\frac{\text{Cov}(X,Y)}{\sqrt{V(X)V(Z)}}=\frac{\rho\cdot 1}{\sqrt{1\cdot 1}}=\rho.$$

d) Siden X og Z er uavhengige og $E(Z)=0$, har vi at

$$\begin{aligned}E(Y\,|\,X=x_0)&=E\left(\rho X+\sqrt{1-\rho^2}\,Z\,|\,X=x_0\right)\\&=E\left(\rho x_0+\sqrt{1-\rho^2}\,Z\,|\,X=x_0\right)\\&=\rho x_0+\sqrt{1-\rho^2}\,E\left(Z\,|\,X=x_0\right)\\&=\rho x_0+\sqrt{1-\rho^2}\,E\left(Z\right)=\rho x_0.\end{aligned}$$

Tilsvarende finner vi siden $V(Z))=1$ at

$$\begin{aligned}V(Y\,|\,X=x_0)&=V\left(\rho X+\sqrt{1-\rho^2}\,Z\,|\,X=x_0\right)\\&=V\left(\rho x_0+\sqrt{1-\rho^2}\,Z\,|\,X=x_0\right)\\&=(1-\rho^2)\,V\left(Z\,|\,X=x_0\right)\\&=(1-\rho^2)\,V\left(Z\right)=1-\rho^2.\end{aligned}$$

e) På tilsvarende måte som i punkt a, har vi at

$$\mathrm{MSE}_{Y \mid X=x_0}(\theta) = \{\theta - E(Y \mid X = x_0)\}^2 + V(Y \mid X = x_0). \tag{D}$$

Her avhenger ikke $V(Y \mid X = x_0)$ av θ, mens $\{\theta - E(Y \mid X = x_0)\}^2$ blir minst mulig for $\theta = E(Y \mid X = x_0)$. Vårt beste gjett på Y hvis vi har observert $X = x_0$, er den betingede forventningen $E(Y \mid X = x_0) = \rho x_0$.

Når vi bruker anslaget $E(Y \mid X = x_0) = \rho x_0$, gir (D) at betinget forventet kvadratisk avvik er $V(Y \mid X = x_0) = 1 - \rho^2$. Hvis vi bruker forventningsverdien til Y, blir forventet kvadratiske avvik større. For da gir (C) at det er lik $V(Y) = 1$.