STK1110

Obligatorisk oppgave 1

Jonas Semprini Næss

20. oktober 2021

NB! Kode på oppgave 1 f.) er skrevet i samarbeid med “erlek@math.uio.no”

Oppgave 0:

a.)
Jeg tar STK1110 fordi jeg synes statistikk er viktig og det er et obligatorisk emne for den masteren jeg ønsker å søke meg inn på. I tillegg skal jeg videre ta STK2100 og da er det greit å ha tatt STK1110 fra før av.
b.)
Jeg forventer at STK1110 gir oss en god innføring i håndtering og modellering av data samt gir oss et godt teoretisk grunnlag til hvorfor vi kan praktiske statistikk som vi gjør. Men viktigst av alt så forventer jeg at kurset gir en interesse for faget så både fler og de som allerede tar kurset får sterke og nyttige verktøy å bruke i hverdagen samt fortsetter å se viktighet og anvendeligheten i statistikk.

Oppgave 1:

a.) *Vis at* $\mathrm{E}(\mathrm{X_i}) = \mathrm{e}^{\mu+\frac{\sigma^2}{2}}$ *og at* $\mathrm{E}(\mathrm{X_i^2}) = \mathrm{e}^{2\mu+2\sigma^2}$ *(Hint: du kan få bruk for momentgenererende funksjon for normalfordelingen* $\mathrm{N}(\mu, \sigma^2)$*, som er* $\mathrm{M_Y(t)} = \mathrm{e}^{\mu \mathrm{t}+\frac{1}{2}\sigma^2\mathrm{t}^2}$*).*

Løsning:

Siden Y er en normalfordelt variabel med forventning μ og varians σ^2 lønner det seg å analysere den momentgenererende funksjonen $\mathrm{M_Y(t)}$ for å finne forventningen av den lognormalfordelte variablen X.
Ved nærmere observasjon kan en således se at

$$\begin{aligned}\mathrm{M_{Y_i}(1)} &= \mathrm{e}^{\mu(1)+\frac{1}{2}\sigma^2(1)^2} \\ &= \mathrm{e}^{\mu+\frac{1}{2}\sigma^2} \qquad \blacksquare\end{aligned}$$

hvilket svarer til forventningen av $\mathrm{X_i}$. For å finne forventningen av $\mathrm{X_i^2}$ kan vi på lik måte analysere nte momente til $\mathrm{Y_i}$.

$$\begin{aligned}\mathrm{E(X_i) = M_{Y_i}(1)} \implies \mathrm{E(X_i^n)} &= \mathrm{E((e^{Y_i})^n)} \\ &= \mathrm{M_{Y_i}(n)} \\ &= \mathrm{e}^{\mathrm{n}\mu+\mathrm{n}^2\frac{\sigma^2}{2}}\end{aligned}$$

Som gir følgelig

$$\begin{aligned}\mathrm{E(X_i^2)} &= \mathrm{E((e^{Y_i})^2)} \\ &= \mathrm{e}^{2\mu+4\frac{\sigma^2}{2}} \\ &= \mathrm{e}^{2\mu+2\sigma^2} \qquad \blacksquare\end{aligned}$$

b.) *Finn momentestimatorene $\hat{\mu}_{\mathrm{mom}}$ og $\hat{\sigma^2}_{\mathrm{mom}}$ for μ og σ^2, og beregn de tilsvarende estimatene for bilforsikringskravene.*

Løsning:

Momentestimatorene til respektive parametere i en fordeling kan utrykkes ved å se på momentene til fordelingen. Ved å benytte oss av resultatene i a.) har vi

$$\mathrm{E(X_i)} = \mathrm{e}^{\mu+\frac{\sigma^2}{2}}.$$

For å finne momentestimatoren til μ sammenligner vi dermed det teoretiske momente med det empiriske momente

$$\begin{aligned}\mathrm{E(X_i)} &= \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{X_i} \\ \mathrm{e}^{\mu+\frac{\sigma^2}{2}} &= \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{X_i}\end{aligned}$$

tar så log på begge sider

$$\mu + \frac{\hat{\sigma^2}}{2} = \ln\left(\sum_{i=0}^{n} X_i\right) - \ln(n)$$

og løser for û

$$\hat{\mu} = \ln\left(\sum_{i=0}^{n} X_i\right) - \ln(n) - \frac{\hat{\sigma^2}}{2}. \tag{1}$$

Videre ser vi på andre momente av X_i for å finne $\hat{\sigma^2}$.

$$\begin{aligned} E(X_i^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ e^{2\hat{\mu}+2\hat{\sigma^2}} &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ 2\hat{\mu} + 2\hat{\sigma^2} &= \ln\left(\sum_{i=1}^{n} X_i^2\right) - \ln(n) \end{aligned}$$

Løser igjen for $\hat{\mu}$

$$\hat{\mu} = \frac{\ln\left(\sum_{i=1}^{n} X_i^2\right) - \ln(n)}{2} - \hat{\sigma^2} \tag{2}$$

og setter (1) = (2)

$$\ln\left(\sum_{i=0}^{n} X_i\right) - \ln(n) - \frac{\hat{\sigma^2}}{2} = \frac{\ln\left(\sum_{i=1}^{n} X_i^2\right) - \ln(n)}{2} - \hat{\sigma^2}$$

hvor vi sluttvis løser for $\hat{\sigma^2}$

$$\begin{aligned} \frac{\hat{\sigma^2}}{2} + \ln\left(\sum_{i=0}^{n} X_i\right) &= \frac{\ln\left(\sum_{i=1}^{n} X_i^2\right)}{2} + \ln(n) \\ \hat{\sigma^2} &= \ln\left(\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\sum_{i=0}^{n} X_i\right) + \ln(n). \qquad \blacksquare \end{aligned}$$

Herfra kan vi tilslutt løse for $\hat{\mu}$

$$
\begin{aligned}
\hat{\mu} &= \ln\left(\sum_{i=0}^{n} X_i\right) - \ln(n) - \frac{\hat{\sigma^2}}{2}. \\
&= \ln\left(\sum_{i=0}^{n} X_i\right) - \ln(n) - \frac{\ln\left(\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\sum_{i=0}^{n} X_i\right) + \ln(n)}{2} \\
&= \ln\left(\sum_{i=0}^{n} X_i\right) + \ln\left(\sum_{i=0}^{n} X_i\right) - \frac{3}{2}\ln(n) - \frac{\ln\left(\sum_{i=1}^{n} X_i^2\right)}{2} \\
&= 2\ln\left(\sum_{i=0}^{n} X_i\right) - \frac{3}{2}\ln(n) - \frac{\ln\left(\sum_{i=1}^{n} X_i^2\right)}{2}. \qquad \blacksquare
\end{aligned}
$$

For å beregne estimatorene våre regner vi på datasettet i R.

```r
forsikring.data = read.table("https://www.uio.no/studier/emner/matnat/math/STK1110/data/forsikringskrav.txt", header=T)

forsikring = forsikring.data$X7.708

mu_moment = round(2*log(sum(forsikring)) - (3/2)*(log(length(forsikring))) - (1/2)*(log(sum(forsikring^2))), 4)

sigma_moment = log(sum(forsikring^2)) - 2*log(forsikring) + log(length(forsikring))

```

Som gir verdiene for $\hat{\mu}$ og $\hat{\sigma^2}$

```
> mu_moment
[1] 2.7389
> sigma_moment
[1] 0.8901
```

c.) *Finn maksimum likelihood-estimatorene* $\hat{\mu}_{\text{mle}}$ *og* $\hat{\sigma^2}_{\text{mle}}$ *for* μ *og* σ^2 *på vanlig måte*

Løsning:

Ettersom verdiene i fordeling er uavhengig identisk fordelte vil vi kunne bruke det faktum at likelihoodfunksjonen er definert som produktet av sannsynlighetstetthetene til alle $\mathrm{X_i}$. Siden $\mathrm{X_i}$ er lognormalfordelt har den sannsynlighetsttethetsfunksjon

$$\mathrm{f}(\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}; \mu, \sigma^2) = \frac{1}{\mathrm{x_i}\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln(\mathrm{x_i})-\mu)^2}{2\sigma^2}}.$$

likelihoodfunksjonen for $\mathrm{X_i}$ er dermed definert som

$$\begin{aligned}
\mathrm{L}(\mu, \sigma^2 \mid \mathrm{X_i}) &= \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{f}(\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}; \mu, \sigma^2) \\
&= \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{\mathrm{X_i}\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln(\mathrm{X_i})-\mu)^2}{2\sigma^2}} \\
&= (2\pi\sigma)^{-\frac{\mathrm{n}}{2}} \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{\mathrm{X_i}} \mathrm{e}^{-\frac{(\ln(\mathrm{X_i})-\mu)^2}{2\sigma^2}}
\end{aligned}$$

Så tar vi logaritmen av denne som gir

$$\begin{aligned}
\mathcal{L}(\mu, \sigma^2 \mid \mathrm{X_i}) &= \ln\left((2\pi\sigma^2)^{-\frac{\mathrm{n}}{2}} \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{1}{\mathrm{X_i}} \mathrm{e}^{-\frac{(\ln(\mathrm{X_i})-\mu)^2}{2\sigma^2}}\right) \\
&= -\frac{\mathrm{n}}{2}\ln(2\pi\sigma^2) - \ln\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}\right) - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\ln(\mathrm{X_i}) - \mu)^2}{2\sigma^2} \\
&= -\frac{\mathrm{n}}{2}\ln(2\pi\sigma^2) - \ln\left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}\right) - \frac{\sum_{\mathrm{i}}^{\mathrm{n}} \ln(\mathrm{X_i})^2}{2\sigma^2} + \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})\mu}{\sigma^2} - \frac{\mathrm{n}\mu^2}{2\sigma^2}.
\end{aligned}$$

For å finne $\hat{\mu}_{\mathrm{mle}}$ og $\hat{\sigma^2}_{\mathrm{mle}}$ må vi partiellderivere $\mathcal{L}(\mu, \sigma^2 \mid \mathrm{X_i})$ med hensyn på de respektive parametrene og løse

$$\frac{\partial \mathcal{L}}{\partial \mu} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = 0$$

som gir

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})}{\sigma^2} - \frac{\mathrm{n}\mu}{\sigma^2} = 0$$

$$\frac{\mathrm{n}\mu}{\sigma^2} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})}{\sigma^2}$$

$$\hat{\mu}_{\mathrm{mle}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})}{\mathrm{n}}$$

for μ og

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = \frac{-\mathrm{n}}{2\sigma^2} - \left(-\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\ln(\mathrm{X_i}) - \mu)^2}{2(\sigma^2)^2}\right) = 0$$

$$\frac{\mathrm{n}}{2\sigma^2} = \left(\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\ln(\mathrm{X_i}) - \mu)^2}{2(\sigma^2)^2}\right)$$

$$\hat{\sigma^2}_{\mathrm{mle}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\ln(\mathrm{X_i}) - \hat{\mu}_{\mathrm{mle}})^2}{\mathrm{n}}$$

setter vi inn for $\hat{\mu}_{\mathrm{mle}}$ har vi sluttvis at

$$\hat{\sigma^2}_{\mathrm{mle}} = \frac{1}{\mathrm{n}}\left(\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\ln(\mathrm{X_i}) - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})}{\mathrm{n}}\right)^2\right)$$

.

Beregner vi så maksimumestimatorene i R har vi at

```
forsikring.data = read.table("https://www.uio.no/studier/emner/matnat/math/STK1110/data/forsikringskrav.txt", header=T)

forsikring = forsikring.data$X7.708

u_mle = round((sum(log(forsikring)))/length(forsikring),4)

sigma = sum((log(forsikring) - sum(log(forsikring))/length(forsikring))^2)

sigma_mle = round(sigma/length(forsikring), 4)
```

```
> u_mle
[1] 2.7824
 > sigma_mle
[1] 0.7658
```

hvor vi fra oppgave b ser at estimatorene er rimelig like, men med noen differanser.

d.) *Husk at* $\mathrm{Y_i} = \log(\mathrm{X_i})\tilde{\mathrm{N}}(\mu, \sigma^2), \mathrm{i} = 1, \ldots, \mathrm{n}$. *Bruk dette, samt maksimum likelihood-estimatorene for parameterne i normalfordelingen (disse behøver du ikke å utlede) til å finne* $\hat{\mu}_{\mathrm{mle}}$ *og* $\hat{\sigma^2}_{\mathrm{mle}}$ *på en alternativ (og mye enklere) måte.*

Løsning:

Vi vet at $\mathrm{Y_i} = \ln(\mathrm{X_i})$ er normalfordelt så vi kan skrive

$$\mu = \bar{\mathrm{Y}} = \ln(\bar{\mathrm{X}}_\mathrm{i})$$

hvor en kan skrive

$$\bar{\mathrm{Y}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Y_i} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})$$

hvilket hvis observerer fra c.) er det samme som

$$\hat{\mu}_{\mathrm{mle}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i}). \qquad \blacksquare$$

Likledes har vi for σ^2 at

$$\sigma^2 = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{Y_i} - \bar{\mathrm{Y}})^2$$

hvilket hvis vi setter inn for $\mathrm{Y_i}$ og $\bar{\mathrm{Y}}$ gir

$$\hat{\sigma^2} = \frac{1}{\mathrm{n}} \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\ln(\mathrm{X_i}) - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})}{\mathrm{n}} \right)^2 \right) \qquad \blacksquare$$

som er det samme som $\hat{\sigma^2}_{\mathrm{mle}}$.

e.) *Vis at informasjonsmatrisa for èn observasjon er gitt ved*

$$\mathrm{I}(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Løsning:

Informasjonselemente i Fishermatrisen for én observasjon er gitt ved Hessematrisen

$$\mathrm{I}(\mu,\sigma^2) = \begin{pmatrix} -\mathrm{E}\left(\frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial^2\mu}\right) & -\mathrm{E}\left(\frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial\mu\partial\sigma^2}\right) \\ -\mathrm{E}\left(\frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial\sigma^2\partial \mathrm{u}}\right) & -\mathrm{E}\left(\frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial(\sigma^2)^2}\right) \end{pmatrix}$$

Starter vi med å ta den andre ordens partiellderiverte av μ gir det

$$\begin{aligned} \frac{\partial \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial\mu} &= \frac{\partial}{\partial\mu} - \left(\frac{\ln(\mathrm{X}_\mathrm{i})^2 - 2\ln(\mathrm{X}_\mathrm{i})\mu + \mu^2}{2\sigma^2}\right) - \ln(\sqrt{2\pi}\mathrm{x}\sigma) \\ &= -\frac{1}{\sigma^2}\left(-\ln(\mathrm{X}_\mathrm{i}) + \mu\right) \\ \frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial^2\mu} &= -\frac{1}{\sigma^2} \end{aligned}$$

og sluttvis har vi

$$\begin{aligned} -\mathrm{E}\left(\frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial^2\mu}\right) &= -\mathrm{E}(-\frac{1}{\sigma^2}) \\ &= \frac{1}{\sigma^2}. \end{aligned}$$

Følger vi samme regime for σ^2 har vi

$$\begin{aligned} \frac{\partial \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial \mathrm{u}} &= \frac{1}{2}\left(\frac{(\ln(\mathrm{X}_\mathrm{i}) - \mu)^2}{\mathrm{u}^2}\right) + \frac{1}{2\sqrt{\mathrm{u}}}\frac{1}{\sqrt{\mathrm{u}}} \qquad ; \qquad \boxed{\mathrm{u} = \sigma^2} \\ &= \frac{1}{2}\left(\frac{(\ln(\mathrm{X}_\mathrm{i}) - \mu)^2}{\mathrm{u}}\right) - \frac{1}{2\mathrm{u}} \\ \frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial^2 \mathrm{u}} &= -\frac{1}{4}\left(\frac{(\ln(\mathrm{X}_\mathrm{i}) - \mu)^2}{\mathrm{u}^3}\right) - \frac{1}{2\mathrm{u}^2} \\ \frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n};\mu,\sigma^2))}{\partial(\sigma^2)^2} &= -\frac{1}{4}\left(\frac{(\ln(\mathrm{X}_\mathrm{i}) - \mu)^2}{\sigma^6}\right) - \frac{1}{2\sigma^4} \end{aligned}$$

som gir

$$
\begin{aligned}
-\mathrm{E}\left(-\frac{1}{4}\left(\frac{(\ln(\mathrm{X_i})-\mu)^2}{\sigma^6}\right)-\frac{1}{2\sigma^4}\right) &= -\mathrm{E}\left(-\frac{1}{4}\left(\frac{(\ln(\mathrm{X_i})-\mu)^2}{\sigma^6}\right)\right)-\mathrm{E}\left(-\frac{1}{2\sigma^4}\right)\\
&= 0+\frac{1}{2\sigma^4}\\
&= \frac{1}{2\sigma^4}.
\end{aligned}
$$

Og sluttvis har vi for de førsteordens partiellderiverte respektive parametrene

$$
\begin{aligned}
\frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x_n};\mu,\sigma^2))}{\partial\mu\partial\sigma^2} &= \frac{\ln(\mathrm{X_i})}{\sigma^2}-\frac{\mu}{\sigma^2}\\
&= \frac{\ln(\mathrm{X_i})-\mu}{\sigma^2}
\end{aligned}
$$

hvor den negative forventningen gir

$$
\begin{aligned}
-\mathrm{E}\left(\frac{\partial^2 \ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x_n};\mu,\sigma^2))}{\partial\mu\partial\sigma^2}\right) &= -\mathrm{E}\left(\frac{\ln(\mathrm{X_i})-\mu}{\sigma^2}\right)\\
&= \frac{\mu}{\sigma^2}-\mathrm{E}\left(\frac{\ln(\mathrm{X_i})}{\sigma^2}\right)\\
&= \frac{\mu}{\sigma^2}-\frac{\mu}{\sigma^2}\\
&= 0
\end{aligned}
$$

og setter vi sammen resultatene har vi at

$$
\mathrm{I}(\mu,\sigma^2)=\begin{pmatrix}-\mathrm{E}\left(\frac{\partial^2\ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x_n};\mu,\sigma^2))}{\partial^2\mu}\right) & -\mathrm{E}\left(\frac{\partial^2\ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x_n};\mu,\sigma^2))}{\partial\mu\partial\sigma^2}\right)\\ -\mathrm{E}\left(\frac{\partial^2\ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x_n};\mu,\sigma^2))}{\partial\sigma^2\partial\mathrm{u}}\right) & -\mathrm{E}\left(\frac{\partial^2\ln(\mathrm{f}(\mathrm{x}_1,\ldots,\mathrm{x_n};\mu,\sigma^2))}{\partial(\sigma^2)^2}\right)\end{pmatrix}=\begin{pmatrix}\frac{1}{\sigma^2} & 0\\ 0 & \frac{1}{2\sigma^4}\end{pmatrix} \qquad \blacksquare
$$

som var det vi skulle vise. Standardfeilen til informasjonsmatrisen er gitt ved å ta kvadratroten til diagonalelementene i den inverse Fishermatrisen. Da har vi

$$
\begin{aligned}
\mathrm{I}(\mu,\sigma^2)^{-1} &= \frac{1}{\frac{1}{2\sigma^6}}\begin{pmatrix}\frac{1}{\sigma^2} & 0\\ 0 & \frac{1}{2\sigma^4}\end{pmatrix}\\
&= \begin{pmatrix}\sigma^2 & 0\\ 0 & 2\sigma^4\end{pmatrix}
\end{aligned}
$$

og $\text{SD}(\mu, \sigma^2) = \sigma, \sqrt{2}\sigma^2$.

f.) *Estimer nå standardfeilen til $\hat{\mu}_{\text{mle}}$ og σ^2_{mle} for foriskringsdataene ved hjelp av ikke-parametrisk bootstrapping. Sammenlign med estimatene du fikk i e) og kommenter.*

Løsning:

Vi løser oppgaven ved hjelp av følgende R-script

```
bilforsikring.krav = read.table("https://www.uio.no/studier/emner/matnat/math/STK1110/data/forsikringskrav.txt")
n = length(bilforsikring.krav$V1)

iterations = 1000
arr_mu_mle = rep(0,iterations)
arr_sigma_2_mle = rep(0,iterations)

for (i in 1:iterations){
 temp = sample(bilforsikring.krav$V1, replace=T)
 mu_mle = sum(log(temp))/n
 sigma_2_mle = sum((log(temp) - mu_mle)^2)/n

 arr_mu_mle[i] = mu_mle
 arr_sigma_2_mle[i] = sigma_2_mle
}

sd_mu_mle_boot = sd(arr_mu_mle)*sqrt(n)
sd_sigma_2_mle_boot = sd(arr_sigma_2_mle)*sqrt(n)

sigma = sd(log(bilforsikring.krav$V1))

sd_mu_mle = sigma^2
sd_sigma_2_mle = 2*sigma^4

abs(sd_mu_mle - sd_mu_mle_boot)
abs(sd_sigma_2_mle - sd_sigma_2_mle_boot)
```

som gir oss verdiene

```
> sd_mu_mle_boot
[1] 0.9104013

> sd_sigma_2_mle_boot
[1] 1.1427
```

```
> sigma
[1] 0.875151

> sd_mu_mle
[1] 0.7658893

> sd_sigma_2_mle
[1] 1.173173

> abs(sd_mu_mle - sd_mu_mle_boot)
[1] 0.144512

> abs(sd_sigma_2_mle - sd_sigma_2_mle_boot)
[1] 0.03047338
```

hvor vi observerer at estimatorene våre fra maksimumlikelihoodfunksjonen er ganske gode sammenlignet med standardavviket fra informasjonsmatrisen.

g.) *Fra a) vet vi at* $\phi = \mathrm{E(Xi)} = \mathrm{e}^{\mu+\frac{\sigma^2}{2}} = \phi(\mu, \sigma^2)$. *Hva er maksimum likelihood-estimatoren* $\hat{\phi}_{\mathrm{mle}}$ *for* ϕ*? Beregn det estimatet for bilforsikringskravene. En alternativ estimator for* ϕ *er den forventningsrette* $\bar{\mathrm{X}} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}$*. Beregn estimatet* $\bar{\mathrm{x}}$ *for forsikringsdataene og sammenlign med maksimum likelihood-estimatet.*

Løsning:

Vi vet at $\mathrm{E(Xi)} = \mathrm{e}^{\mu+\frac{\sigma^2}{2}}$ og ønsker vi å finne maksimumlikelihoodestimatoren av $\mathrm{E(X_i)}$ kan vi simpelthen sette inn estimatore våre fra tidligere

$$\begin{aligned}
\mathrm{E}(\hat{\mathrm{X}}_{\mathrm{i}})_{\mathrm{mle}} &= \mathrm{e}^{\hat{\mu}_{\mathrm{mle}}+\frac{\hat{\sigma}_{\mathrm{mle}}}{2}} \\
&= \mathrm{e}^{\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})}{\mathrm{n}} + \frac{1}{2\mathrm{n}}\left(\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\ln(\mathrm{X_i}) - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \ln(\mathrm{X_i})}{\mathrm{n}}\right)^2\right)}
\end{aligned}$$

og sammenligne med den forventingsrette estimatoren $\bar{\mathrm{X}}$. Dette kan løses ved følgende R-script

```r
forsikring.data = read.table("https://www.uio.no/studier/emner/matnat/math/STK1110/data/forsikringskrav.txt", header=T)
```

```

forsikring = forsikring.data$X7.708

u_mle = round((sum(log(forsikring)))/length(forsikring),4)

sigma = sum((log(forsikring) - sum(log(forsikring))/length(forsikring))^2)

sigma_mle = round(sigma/length(forsikring), 4)

forventing_mle = round(exp(u_mle + (1/2)*sigma_mle),4)

X_strek = round(sum(forsikring)/length(forsikring),4)
```

som gir følgende verdier

```
> forventing_mle
[1] 23.6959

> X_strek
[1] 24.141
```

hvilket viser at likelihoodestimatoren gir en ganske god estimering.

Oppgave 2:

a.) $\mathrm{E}(\mathrm{X_i})$ *og* $\mathrm{V}(\mathrm{X_i})$

Løsning:

For en kontinuerlig sannsynlighetstetthetfunksjon har vi at

$$\mathrm{E(X)} = \int_{-\infty}^{\infty} \mathrm{xf(x)\ dx}$$

bruker vi dette for den uniforme sannsynlighetsfordelingen gir det

$$
\begin{aligned}
\mathrm{E}(\mathrm{X_i}) &= \int_{-\infty}^{\infty} \mathrm{x}f(\mathrm{x})\ \mathrm{dx} \\
&= \underbrace{\int_{-\infty}^{0} \mathrm{x}f(\mathrm{x})\ \mathrm{dx}}_{0} + \int_{0}^{\theta} \mathrm{x}f(\mathrm{x})\ \mathrm{dx} \\
&= \int_{0}^{\theta} \mathrm{x}\frac{1}{\theta}\ \mathrm{dx} \\
&= \frac{1}{\theta}\left[\frac{\mathrm{x}^2}{2}\right]_0^{\theta} \\
&= \frac{\theta}{2}.
\end{aligned}
$$

Videre husker vi at $\mathrm{Var}(\mathrm{X_i}) = \mathrm{E}(\mathrm{X_i^2}) - (\mathrm{E}(\mathrm{X_i}))^2$ og for å kunne gå videre må vi beregne $\mathrm{E}(\mathrm{X_i^2})$.

$$
\begin{aligned}
\mathrm{E}(\mathrm{X_i^2}) &= \int_{0}^{\theta} \mathrm{x}^2\frac{1}{\theta}\ \mathrm{dx} \\
&= \frac{1}{\theta}\left[\frac{\mathrm{x}^3}{3}\right]_0^{\theta} \\
&= \frac{\theta^2}{3}.
\end{aligned}
$$

Da har vi at

$$
\begin{aligned}
\mathrm{Var}(\mathrm{X_i}) &= \mathrm{E}(\mathrm{X_i^2}) - (\mathrm{E}(\mathrm{X_i}))^2 \\
&= \frac{\theta^2}{3} - \frac{\theta^2}{4} \\
&= \frac{\theta^2}{12}.
\end{aligned}
$$

b.) *Vis at momentestimatoren for θ er $\hat{\theta}_{\mathrm{mom}} = 2\bar{\mathrm{X}}$. Er estimatoren forventningsrett?*

Løsning:

Vi vet at $\bar{\mathrm{X}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}$ og $\mathrm{E}(\mathrm{X_i}) = \frac{\theta}{2}$. Bruker vi denne informasjonen har vi

$$\mathrm{E}(\mathrm{X_i}) = \bar{\mathrm{X}}$$

$$\frac{\theta}{2} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}$$

$$\hat{\theta}_{\mathrm{mom}} = 2\left(\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}\right)$$

$$\hat{\theta}_{\mathrm{mom}} = 2\bar{\mathrm{X}}$$

hvor vi ser at

$$\mathrm{E}(\hat{\theta}_{\mathrm{mom}}) = 2\mathrm{E}(\bar{\mathrm{X}}) = 2\frac{\theta}{2} = \theta$$

$\hat{\theta}_{\mathrm{mom}}$ er forventningsrett.

c.) *Bestem standardfeilen til* $\hat{\theta}_{\mathrm{mom}}$.

Løsning:

Standardfeilen til estimatoren $\hat{\theta}_{\mathrm{mom}}$ er det samme som standardavviket til estimatoren. Det betyr at

$$\begin{aligned}
\mathrm{SD}(\hat{\theta}_{\mathrm{mom}}) &= \sqrt{\mathrm{Var}(\hat{\theta}_{\mathrm{mom}})} \\
&= \sqrt{\mathrm{E}(\hat{\theta}^2_{\mathrm{mom}}) - (\mathrm{E}(\hat{\theta}_{\mathrm{mom}}))^2} \\
&= \sqrt{\mathrm{E}(4\bar{\mathrm{X}}) - \theta^2} \\
&= \sqrt{\frac{4\theta^2}{3} - \frac{3\theta^2}{3}} \\
&= \sqrt{\frac{\theta^2}{3}} \\
&= \frac{\sqrt{3}\theta}{3}.
\end{aligned}$$

d.) *Vis at likelihooden er gitt ved*

$$\mathrm{f}(\mathrm{x}_1, \ldots, \mathrm{x_n}; \theta) = \begin{cases} \frac{1}{\theta^{\mathrm{n}}}, & 0 \leq \mathrm{x} \leq \theta \\ 0, & \text{ellers} \end{cases}$$

og forklar at maksimum likelihood-estimatoren for θ er $\hat{\theta}_{\mathrm{mle}} = \max(X_i)$

Løsning:

Vi husker fra oppgave 1 c.) at likelihoodfunksjonen er gitt ved

$$
\begin{aligned}
L(\theta|X_i) &= \prod_{i=1}^{n} f(x_1, \ldots, x_n; \theta) \\
&= \prod_{i=1}^{n} \frac{1}{\theta} \\
&= \frac{1}{\theta^n} \qquad \blacksquare
\end{aligned}
$$

Når vi skal finne maksimumlikelihoodestimatoren til θ hjelper det å se på hvordan likelihoodfunksjonen utvikler seg. Selve funksjonen $\frac{1}{\theta^n}$ er strengt avtakende for $\theta \geq x_n$ og maksimert hvis $x_n = \theta$ jmf. $[0 \leq x \leq \theta]$. For $x_n \geq \theta$ er funksjonen udefinert. Dermed må maksimumlikelihoodestimatoren være $\max(X_i)$.

e.) *Det kan vises at forventningen og andremomentet til $\hat{\theta}_{\mathrm{mle}}$ er henholdsvis*

$$E(\hat{\theta}_{\mathrm{mle}}) = \frac{n\theta}{n+1} \text{ og } E(\hat{\theta}^2_{\mathrm{mle}}) = \frac{n\theta^2}{n+2}$$

hvilket betyr at $\hat{\theta}_{\mathrm{mle}}$ ikke er forventningsrett for θ. Bruk resultatene over til å vise at estimatoren $\tilde{\theta} = \frac{n+1}{n}\hat{\theta}_{\mathrm{mle}}$ er forventningsrett og til å finne standardfeilen til $\tilde{\theta}$.

Løsning:

Vi har at

$$
\begin{aligned}
E(\tilde{\theta}) &= E\left(\frac{n+1}{n}\hat{\theta}_{\mathrm{mle}}\right) \\
&= \frac{n+1}{n} E(\hat{\theta}) \\
&= \frac{n+1}{n} \frac{n\theta}{n+1} \\
&= \theta \qquad \blacksquare.
\end{aligned}
$$

Standardavviket beregnes på samme måte som tidligere

$$\begin{aligned}
\mathrm{Var}(\tilde{\theta}) &= \mathrm{E}(\tilde{\theta}^2) - (\mathrm{E}(\tilde{\theta}))^2 \\
&= \mathrm{E}\left(\left(\frac{\mathrm{n}+1}{\mathrm{n}}\right)^2 \hat{\theta}^2\right) - \theta^2 \\
&= \frac{(\mathrm{n}+1)^2}{\mathrm{n}^2} \frac{\mathrm{n}\theta^2}{\mathrm{n}+2} - \theta^2 \\
&= \frac{\theta^2}{\mathrm{n}(\mathrm{n}+2)} \\
\mathrm{SD}(\tilde{\theta}) &= \frac{\theta}{\sqrt{\mathrm{n}(\mathrm{n}+2)}}
\end{aligned}$$

f.) *Sammenlign momentestimatoren* $\hat{\theta}_{\mathrm{mom}}$ *fra b) med* $\tilde{\theta}$*. Hvilken estimator vel- ger du? Svaret skal begrunnes.*

Løsning:

For å sammenligne estimatorene ser vi på standarfeilen til begge to, nemlig

$$\begin{aligned}
\mathrm{SD}(\tilde{\theta}) &= \frac{\theta}{\sqrt{\mathrm{n}(\mathrm{n}+2)}} \\
\mathrm{SD}(\hat{\theta}_{\mathrm{mom}}) &= \frac{\sqrt{3}\theta}{3}
\end{aligned}$$

Siden standardfeilen til $\tilde{\theta}$ varierer med n er det verdt å sjekke når denne vil bli mindre enn standarfeilien til $\hat{\theta}_{\mathrm{mom}}$.

$$\begin{aligned}
\mathrm{n}(\mathrm{n}+2) &> 3 \\
\mathrm{n}^2 + 2\mathrm{n} &> 3 \\
\mathrm{n}^2 + 2\mathrm{n} - 3 &> 0 \\
(\mathrm{n}+3)(\mathrm{n}-1) &> 0
\end{aligned}$$

Siden n ikke kan være negativ (ville gitt et komplekst tall) må $\mathrm{n} > 1$ for at $\tilde{\theta}$ skal være en bedre estimator. Derfor ville jeg heller valgt $\tilde{\theta}$ enn $\hat{\theta}_{\mathrm{mom}}$ siden standardfeilen er konstant og veldig mye dårligere for større n. g.) *Sjekk hvordan teorien virker i praksis ved å simulere* $\mathrm{n} = 20$ *observasjoner* *fra uniform fordeling på* $[0, 1]$ *(du kan bruke data = runif(20) i R). I dette tilfellet er den sanne verdien for* θ *lik 1. Hvor nær kommer du med de to estimatorene du sammenlignet i forrige punkt? Gjenta prosedyren over mange ganger (f.eks. 1000 til sammen) for å illustrere poenget i f), og presenter resultatene grafisk (f.eks. i et boksplott eller et histogram).*

Løsning:

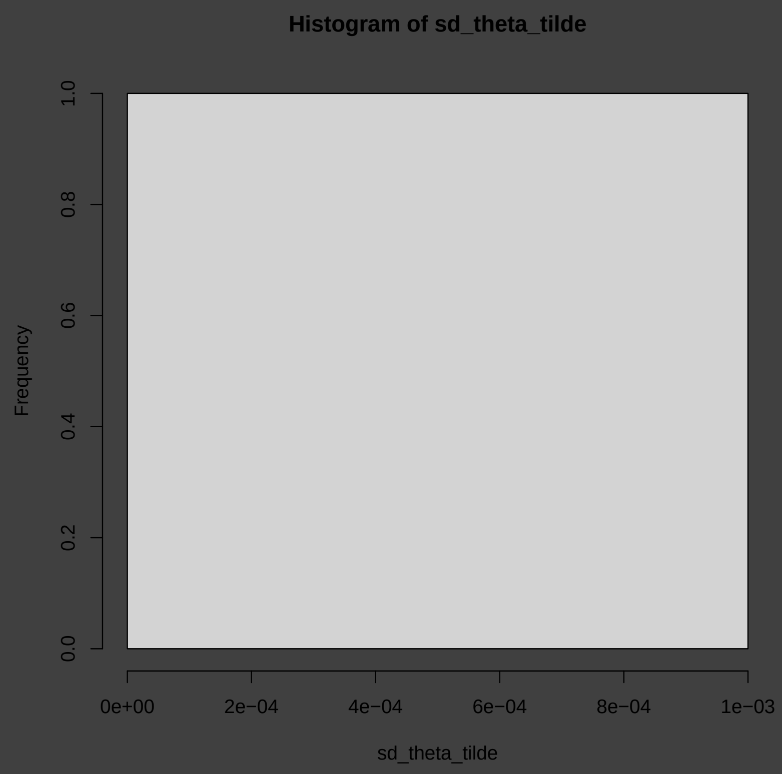
Kan løse oppgaven ved følgende R-script

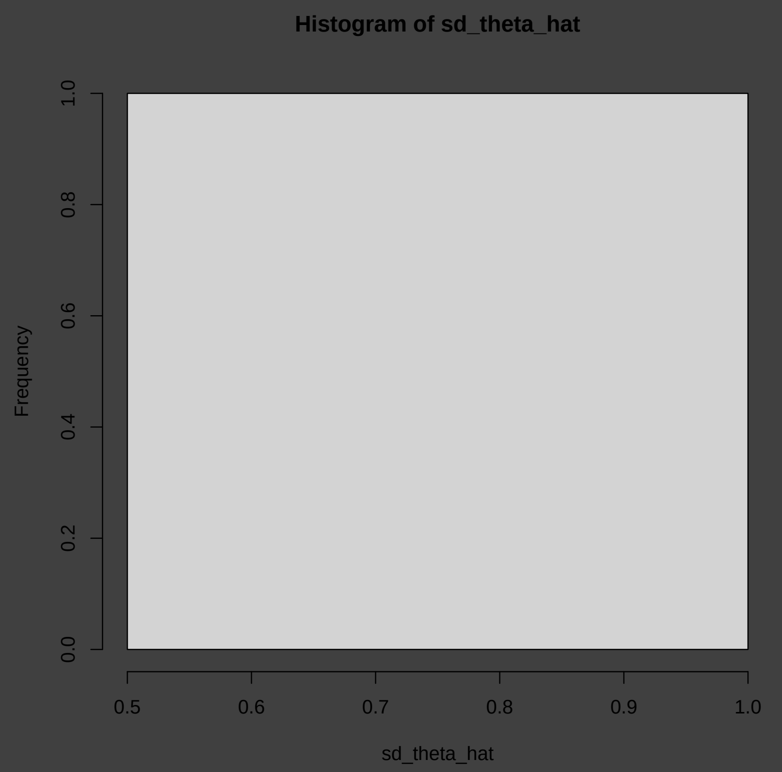
```
n = 1000

uni = runif(n)

sd_theta_tilde = 1/(sqrt(n*(n+2)))

sd_theta_hat = (sqrt(3)*1)/3

pdf(file="estimatorer.pdf")

hist(sd_theta_tilde, breaks = "Sturges")

hist(sd_theta_hat, breaks = "Sturges")

dev.off()
```

hvor tisvarende verdier er

```
> sd_theta_hat
[1] 0.5773503

> sd_theta_tilde
[1] 0.0009990015
```

hvor vi ut i fra teorien ser at det stemmer ganske godt hvilken estimator som er ideel. Ett mulig plot av standardavvikene kan vises i dette histogrammet.


Histogram of sd_theta_tilde
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0e+00
2e−04
4e−04
6e−04
8e−04
1e−03
sd_theta_tilde


Histogram of sd_theta_hat
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0.5
0.6
0.7
0.8
0.9
1.0
sd_theta_hat