Oppg 3

X inntekten til en tilfeldig lønnsmottaker i en spesifikk befolkningsgruppe

Det er da vanlig å anta at X er Pareto-fordelt.

Med sannsynlighetstetthet.

$$f(x) = \begin{cases} \theta K^{\theta}\left(\frac{1}{x}\right)^{\theta+1} & \text{hvis } x > K \\ 0 & \text{ellers} \end{cases}$$

K: minsteinntekten i gruppen. $\theta > 1$ parameter på lønnsforskjeller

a) Vis den kumulative sannsynlighetsfordelingen til X

$$F(x) = \begin{cases} \left(1 - \left(\frac{K}{x}\right)^{\theta}\right) & \text{Hvis } x > K \\ 0 & \text{ellers} \end{cases}$$

Integrerer $f(u)$

$$F(x) = \int_K^x f(u)\,du = \int_K^x \theta k^{\theta}\left(\frac{1}{u}\right)^{\theta+1} du$$

$$= \theta k^{\theta} \int_K^x u^{-\theta-1}\, du = \theta k \left[-\theta^{-1} u^{-\theta}\right]_k^x$$

$$= \theta k (-\theta)^{-1} \left[x^{-\theta} - k^{-\theta}\right] = -\frac{k}{x} + 1$$

Videre for $x \leq k$

$$F(x) = \int_k^x 0 \, du = \underline{0}$$

Så vi har da

$$F(x) = \begin{cases} 1 - \frac{k}{x} & \text{for } x > k \\ 0 & \text{ellers} \end{cases}$$

$\#$

Bruk dette til å vise at medianen er $\tilde{\mu} = 2^{1/\theta} k$

Vet at medianen er $F(\tilde{\mu}) = \frac{1}{2}$, dermed har vi

$$1 - \left(\frac{k}{\tilde{\mu}}\right)^{\theta} \stackrel{!}{=} \frac{1}{2} \quad / -1 \,, \cdot -1$$

$$\left(\frac{k}{\tilde{\mu}}\right)^{\theta} = \frac{1}{2}$$

$$\sqrt[\theta]{\tilde{\mu}^{\theta}} = \sqrt[\theta]{2 \cdot k^{\theta}}$$

$$\tilde{\mu} = \underline{\underline{2^{1/\theta} k}} = \text{Medianen}$$

$\boxed{\#}$

b) Vis at forventet innfekt er $E(X) = \theta k / (\theta - 1)$

Definisjonen av forventning er $E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$

$x\left(\frac{1}{x}\right)^{\theta+1} = \left(\frac{1}{x}\right)^{\theta+1}$

Det blir da

$$E(x) = \int_k^\infty x\, \theta k^\theta \left(\frac{1}{x}\right)^{\theta+1} dx = \theta k^\theta \int_k^\infty x \left(\frac{1}{x}\right)^{\theta+1} dx$$

$$= \theta k^\theta \int_k^\infty x^{-\theta} dx = \theta k^\theta (-\theta + 1)^{-1} \left[x^{-\theta+1} \right]_k^\infty$$

$$= \frac{\theta k^\theta}{1-\theta} \left(\lim_{x \to \infty} \overset{0}{\overset{\shortparallel}{\frac{1}{x^{\theta+1}}}} - k^{-\theta+1} \right) = \frac{\theta k^\theta k^{-\theta+1}}{1-\theta} = \frac{\theta k}{1-\theta} \quad \#$$

c) Vis at $Y = 2\theta[\ln(x) - \ln(k)]$ er kjikvadrat-fordelt med 2 frihetsgrader

Må først finne kumulative fordelingen

$F(Y) = P(2\theta[\ln(X) - \ln(k)] \leq y)$

Løser ulikheten for X

$$2\theta[\ln(X) - \ln(k)] \leq y \quad \Big| \cdot \frac{1}{2\theta}$$

$$\ln(X) - \ln(k) \leq \frac{y}{2\theta} \quad , \quad \ln(X) - \ln(Y) = \ln\left(\frac{X}{Y}\right)$$

$$\ln\left(X/k\right) \leq \frac{y}{2\theta}$$; opphøyer i e, $\cdot k$

$$X \leq k e^{y/2\theta}$$

Da har vi

$$F(y) = P(X \le k e^{y/2\theta})$$

Setter,

$$F(y) = 1 - \left(\frac{\cancel{k}}{\cancel{k} e^{y/2\theta}}\right)^{\theta} = 1 - e^{(-y/2\theta)^{\theta}}$$

$$e^{(-y/2\theta)\cdot\theta} = e^{-y/2}$$

$$F(y) = 1 - e^{-y/2} =$$

Siden $F(0) = 0$ så får vi sannsynlighetsfordelingen

$$G(y) = \begin{cases} 1 - e^{-y/2} & \text{for } y > 0 \\ 0 & \text{ellers} \end{cases}$$

Sjekker om fordelingen har 2 frihetsgrader

Finner derfor sannsynlighetstettheten

$$g(y) = G'(y)$$

$$g(y) = \begin{cases} \frac{1}{2} e^{-y/2} & \text{for } y \geq 0 \\ 0 & \text{ellers} \end{cases}$$

Ser dermed at det er 2 frihetsgrader $\square$

d) Bestem momentestimatoren for θ,

$$\bar{X} = E(X) = \frac{\theta k}{\theta - 1}$$, fra punkt b

Løser for $\bar{\theta} \approx \theta$

$$\bar{X} = \frac{\bar{\theta} k}{\bar{\theta} - 1}$$

$$\bar{x}(\bar{\theta}-1) = \bar{\theta}k$$

$$\bar{x}\bar{\theta} - \bar{x} = \bar{\theta}k$$

$$\bar{x}\bar{\theta} - \bar{\theta}k = \bar{x}$$

$$\bar{\theta}(\bar{x}-k) = \bar{x} \quad / : (\bar{x}-k)$$

$$\bar{\theta} = \frac{\bar{x}}{\bar{x}-k}$$ = momentestimatoren for θ

e) Sett opp likelihooden og vis at maksimum estimatoren for θ blir $\hat{\theta}$

Bruker den kumulative sannsynlighetsfordelingen til x og setter opp log-likelihooden

$$L(x;\theta) = \prod_{i=1}^{n} \theta k^{\theta}\left(\frac{1}{x_i}\right)^{\theta+1}$$, setter θ og k^{θ} utenfor

$$= \theta^n k^n \prod_{i=1}^{n} x_i^{-\theta-1}$$

Finner log-L

$$l(x;\theta) = n\ln(\theta) + n\ln(k) - (\theta+1)\ln\left(\prod x_i\right)$$

$$= n\ln(\theta) + n\ln(k) - (\theta+1)\sum_{i=1}^{n} \ln x_i$$ "

$\ln(x^2) = \ln(x) + \ln(x)$

Optimerer m.h.p θ.

$$l'(x;\theta) = \frac{n}{\theta} + n\ln k - \sum \ln(x_i) = 0 \quad / -\frac{n}{\theta}, \; \cdot -1$$

$$= \hat{\theta} = \frac{n}{\sum(\ln x_i) - n\ln(k)}$$

f) Vis at $2n\theta/\hat{\theta}$ er kji^2-fordelt med $2n$ frihetsgrader

Bruker tipset å skriver

$$\frac{2n\theta}{\hat{\theta}} = 2\theta \sum_{i=1}^{n} \ln(x_i) - n\ln(k)$$

$$= \sum_{i=1}^{n} 2\theta\left(\ln(x_i) - n\ln(k)\right) \quad (*)$$

Det er en sum av n kji^2-fordelinger med hver en frihetsgrad på 2 og siden de er antatt uavhengige så er $(*)$ en kji^2-fordeling med $2n$ frihetsgrader #

g) Hvis V er en stokastisk variabel som er kji^2-fordelt med v frihetsgrader, så er

$$E(V^k) = \left(\frac{2^k \Gamma\left(\frac{v}{2}+k\right)}{\Gamma\frac{v}{2}}\right) \quad \text{så sant } k > -v/2$$

g) Fin forventningen og Variansen til MLEen

$$E\left[\left(\frac{2n\theta}{\hat{\theta}}\right)^{-1}\right] = \frac{2^{-1}\Gamma\left(\frac{\cancel{2}n}{\cancel{2}}-1\right)}{\Gamma\left(\frac{\cancel{2}n}{\cancel{2}}\right)}$$

$$= \frac{\Gamma(n-1)}{2\Gamma(n)} = \frac{\Gamma(n-1)}{2(n-1)\Gamma(n-1)} = \frac{1}{2(n-1)} \quad 1$$

Forventningen:

$$E(\hat{\theta}) = E\left[2n\theta\left(\frac{2n\theta}{\hat{\theta}}\right)^{-1}\right] = 2n\theta\, E\left[\left(\frac{2n\theta}{\hat{\theta}}\right)^{-1}\right]$$

$$= 2n\theta \frac{1}{2(n-1)} = \frac{n\theta}{n-1}$$

For å finne variansen trenger vi

$E(\hat{\theta}^2)$, der med trenger vi $E(V^{-2})$

$$E\left(\frac{2n\theta}{\hat{\theta}}\right)^{-2} = \frac{\Gamma(n-2)}{4\Gamma(n)} = \frac{\cancel{\Gamma(n-2)}}{4(n-1)(n-2)\cancel{\Gamma(n-2)}}$$

$$= \frac{1}{4(n-2)(n-1)} - 2$$

Finner så $E(\hat{\theta}^2)$

$$= E\left[4n^2\theta^2\left(\frac{2n\theta}{\hat{\theta}}\right)^{-2}\right] = 4n^2\theta^2\, E\left[\left(\frac{2n\theta}{\hat{\theta}}\right)^{-2}\right]$$

2

$$= \cancel{4}n^2\theta^2 \frac{1}{\cancel{4}(n-2)(n-1)} = \frac{n^2\theta^2}{(n-2)(n-1)}$$

Dermed kan vi finne variansen

$$V(\hat{\theta}) = E(\hat{\theta}^2) - (E(\hat{\theta}))^2 = \frac{n^2\theta^2}{(n-2)(n-1)} - \left(\frac{n\theta}{n-1}\right)^2$$

$$= \frac{n^2\theta^2}{(n-1)^2(n-2)}$$

Dermed har vi

Forventningen $\frac{n-\theta}{n-1}$

og

Variansen $\frac{n^2\theta^2}{(n-1)^2(n-2)}$

Oppgave 1. (Utregning ligger i et vedlegg mot slutten)

a) Estimert alder til bergart:

$\mu = 276{,}89$ millioner år

90% CI:

t CI = [266, 288]

b) Estimert måleusikkerhet σ

$\sigma = 27{,}08$ millioner år

konfidensintervall

chi CI = [21,4, 37,5]

c) Det forutsetter at begge CI'ene er normalfordelte og uavhengige. Jeg sjekker om verdiene er tilnærmet normalfordelt i Plot 1. Der vises det at dataen er tilnærmet normal-fordelt.

Oppg 2

a) Ta utgangspunkt i ① og utled et $100(1-\alpha)\%$ CI for μ

$$P\left(-t_{\alpha/2, n-1} < \frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{\alpha/2, n-1}\right) = 1-\alpha$$

Løser ulikheten for μ, og setter $t_{\alpha/2, n-1} = t$ for å slippe rot

$$P\left(-t < \frac{\bar{x}-\mu}{s\sqrt{n}} < t\right) = 1-\alpha$$

$$P(-t s\sqrt{n} < \bar{x}-\mu < t s\sqrt{n}) = 1-\alpha$$

$$P(-t s\sqrt{n} - \bar{x} < -\mu < t s\sqrt{n} - \bar{x}) = 1-\alpha$$

$$P(t s\sqrt{n} + \bar{x} > \mu > -t s\sqrt{n} + \bar{x}) \quad 1-\alpha$$

Dermed er $100(1-\alpha)\%$ CI

$$\underline{(\bar{x} - t s\sqrt{n}, \ \bar{x} + t s\sqrt{n})} \quad \#$$

b) Ta utgangspunkt i (2) og utled $100(1-\alpha)\%$ CI

Setter $\chi^2_{1-\alpha/2, n-1} = X_1$ og

$\chi^2_{\alpha/2, n-1} = X_2$, løser for σ

$$P\left(X_1 < \frac{(n-1)S^2}{\sigma^2} < X_2\right) = 1-\alpha$$

$$P\left(\frac{1}{X_1} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{X_2}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{X_1} < \sigma^2 < \frac{(n-1)S^2}{X_2}\right) = 1-\alpha$$

$$P\left(S\sqrt{\frac{n-1}{X_1}} < \sigma < S\sqrt{\frac{n-1}{X_2}}\right) = 1-\alpha$$

Det gir oss $100(1-\alpha)$ CI

$$\left(S\sqrt{\frac{n-1}{X_1}}, S\sqrt{\frac{n-1}{X_2}}\right)$$

#

c) Utregnet i vedlegg

antall verdier som var innenfor i

$\mu = 940$ $\sigma = 952$

d) Utregnet i vedlegg

antall verdier som var innenfor når $n=30$

$\mu = 946$ $\sigma = 942$

(946, 942)

(960, 955)

f) Vis at $2n\theta/\hat{\theta}$ er kji^2-fordelt med $2n$ frihetsgrader

Bruker tipset å skriver

$$\frac{2n\theta}{\hat{\theta}} = 2\theta \sum_{i=1}^{n} \ln(x_i) - n\ln(k)$$

$$= \sum_{i=1}^{n} 2\theta\left(\ln(x_i) - n\ln(k)\right) \quad (*)$$

Det er en sum av $\overset{n}{\cdot}$ kji^2-fordelinger med hver en frihets grad på 2 og siden de er antatt uavhengige så er $(*)$ en kji^2-fordeling med $2n$ frihets grader #

g) Hvis V er en stokastisk variabel som er kji^2-fordelt med v frihetsgrader, så er

$$E(V^k) = \left(\frac{2^k\,\Gamma\frac{v}{2}+k}{\Gamma\frac{v}{2}}\right)$$ så lenge $k > -v/2$

g) Fin forventningen og Variansen til MLEen

$$E\left[\left(\frac{2n\theta}{\hat{\theta}}\right)^{-1}\right] = \frac{2^{-1}\,\Gamma\left(\frac{\cancel{2}n}{\cancel{2}} - 1\right)}{\Gamma\left(\frac{\cancel{2}n}{\cancel{2}}\right)}$$

$$= \frac{\Gamma(n-1)}{2\,\Gamma(n)} = \frac{\Gamma(n-1)}{2(n-1)\,\Gamma(n-1)} = \frac{1}{2(n-1)} \quad 1$$

Forventningen:

$$E(\hat{\theta}) = E\left[2n\theta\left(\frac{2n\theta}{\hat{\theta}}\right)^{-1}\right] = 2n\theta\, E\left[\left(\frac{2n\theta}{\hat{\theta}}\right)^{-1}\right]$$

$$= 2n\theta \frac{1}{2(n-1)} = \frac{n\theta}{n-1}$$

For å finne variansen trenger vi

$E(\hat{\theta}^2)$, dermed trenger vi $E(V^{-2})$

$$E\left(\frac{2n\theta}{\hat{\theta}}\right)^{-2} = \frac{\Gamma(n-2)}{4\Gamma(n)} = \frac{\cancel{\Gamma(n-2)}}{4(n-1)(n-2)\cancel{\Gamma(n-2)}}$$

$$= \frac{1}{4(n-2)(n-1)} - (2)$$

Finner så $E(\hat{\theta}^2)$

$$= E\left[4n^2\theta^2\left(\frac{2n\theta}{\hat{\theta}}\right)^{-2}\right] = 4n^2\theta^2\, E\left[\left(\frac{2n\theta}{\hat{\theta}}\right)^{-2}\right]$$

$$= \cancel{4}n^2\theta^2 \frac{1}{\cancel{4}(n-2)(n-1)} = \frac{n^2\theta^2}{(n-2)(n-1)}$$

Dermed kan vi finne variansen

$$V(\hat{\theta}) = E(\hat{\theta}^2) - (E(\hat{\theta}))^2 = \frac{n^2\theta^2}{(n-2)(n-1)} - \left(\frac{n\theta}{n-1}\right)^2$$

$$= \frac{n^2\theta^2}{(n-1)^2(n-2)}$$

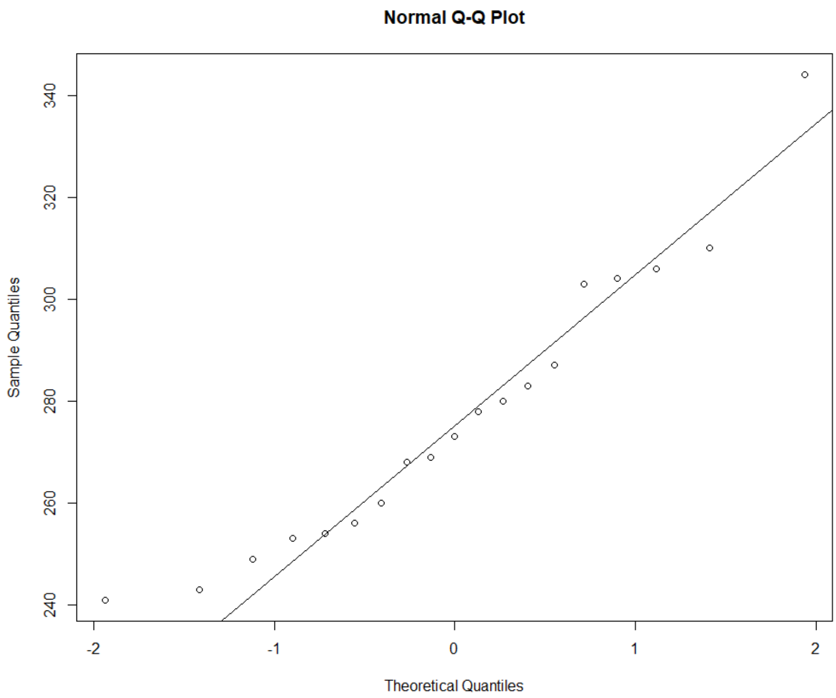
Kode oppg1 (a-c)

```r
init = c(249,254, 243, 268, 253, 269, 287, 241, 273, 306, 303, 280, 260, 256, 278, 344, 304, 283, 310)
#a
#finner estimert forventing, til de uavhengige og stokastiske veridene
n = length((init))
mu = 1/n * sum(init); mu #mean(init)

#finner forventingsrett estimert standardavvik.
S = sdinit = sqrt(1/(n-1)*sum((init - mu)^2));S#sd(init)

#finner 90% CI for mu
tCI = mu +c(1,-1)*qt(0.05,df=n-1)*(S/sqrt(n)); tCI

#b
#sigma estimeres til S
sigma = S

#finner 90% CI for sigma
chiCI = S*sqrt(18/qchisq(c(0.95,1-0.95), n-1));chiCI

#c
#plotter dataen for ? se om det er tilnemet en normallinjes
qqnorm(init)
qqline(init)
```

Values	
chiCI	num [1:2] 21.4 37.5
init	num [1:19] 249 254 243 268 253 269 287 241 273 306 ...
mu	276.894736842105
n	19L
S	27.0819635281124
sdinit	27.0819635281124
sigma	27.0819635281124
tCI	num [1:2] 266 288

chiCI: 90% konfedensintervall for sigma, init: inndaten, mu: Gjennomsnitt, n: anttall verdier i init, S: estimert standardavvik, sdinit = S, sigma = S = sdinit, 90% konfedensintervall for my.

Plot 1



Oppg2

```r
set.seed(11)
mu = 1
sigma = 1
n = 8
ant_setts = 1000
Clvl = 0.975 #konfidens nivaa

simulation <- function(mu, sigma, n, ant_sets, Clvl){
  simres=numeric(0)
  for (i in 1:ant_setts) {
    x=rnorm(n,mu,sigma)
    muhat=mean(x)
    sdx=sd(x)

    muCIt = muhat +c(-1,1)*qt(Clvl,df=n-1)*sdx/sqrt(n) #CI for  my
    sdxCIKJI = sdx*sqrt((n-1)/qchisq(c(Clvl,1-Clvl), n-1))#CI for sigma

    nyres = c(muhat,sdx, muCIt,sdxCIKJI)
    simres=rbind(simres,nyres)
  }

  muCItinInterval= sum((simres[,3]<1)*(1<simres[,4]));
  sdxCIKJIinIntervall = sum((simres[,5]<1)*(1<simres[,6]));
  c(muCItinInterval, sdxCIKJIinIntervall)
}
print(simulation(mu, sigma, n, ant_sets, Clvl))
#d

n = c(30,200)
for (i in n){
  print(simulation(mu, sigma, i, ant_sets, Clvl))
}
#e
n = 8
my = 1
sigma= 1
exponent <- function(mu, sigma, n, ant_sets, Clvl){
  simres=numeric(0)
  for (i in 1:ant_setts) {
    x=rexp(n,rate=1)
    muhat=mean(x)
    sdx=sd(x)

    muCIt = muhat +c(-1,1)*qt(Clvl,df=n-1)*sdx/sqrt(n) #CI for  my
    sdxCIKJI = sdx*sqrt((n-1)/qchisq(c(Clvl,1-Clvl), n-1))#CI for sigma

    nyres = c(muhat,sdx, muCIt,sdxCIKJI)
    simres=rbind(simres,nyres)
  }

  muCItinInterval= sum((simres[,3]<1)*(1<simres[,4]));
  sdxCIKJIinIntervall = sum((simres[,5]<1)*(1<simres[,6]));
  c(muCItinInterval, sdxCIKJIinIntervall)
}
exponent(mu, sigma, n, ant_sets, Clvl)
#f
n  = c(30,200)
for (i in n){
  print(exponent(mu, sigma, i, ant_sets, Clvl))
}
```

```
ant_setts                   1000
Clvl                        0.975
i                           200
mu                          1
my                          1
n                           num [1:2] 30 200
sigma                       1
Functions
exponent                    function (mu, sigma, n, ant_sets, Clvl)
simulation                  function (mu, sigma, n, ant_sets, Clvl)
```

Kjøreeksempel

```
> set.seed(11)
> mu = 1
> sigma = 1
> n = 8
> ant_setts = 1000
> Clvl = 0.975 #konfidens nivaa
>
> simulation <- function(mu, sigma, n, ant_sets, Clvl){
+   simres=numeric(0)
+   for (i in 1:ant_setts) {
+     x=rnorm(n,mu,sigma)
+     muhat=mean(x)
+     sdx=sd(x)
+
+     muCIt = muhat +c(-1,1)*qt(Clvl,df=n-1)*sdx/sqrt(n) #CI for  my
+     sdxCIKJI = sdx*sqrt((n-1)/qchisq(c(Clvl,1-Clvl), n-1))#CI for sigma
+
+     nyres = c(muhat,sdx, muCIt,sdxCIKJI)
+     simres=rbind(simres,nyres)
+   }
+
+   muCItinInterval= sum((simres[,3]<1)*(1<simres[,4]));
+   sdxCIKJIinIntervall = sum((simres[,5]<1)*(1<simres[,6]));
+   c(muCItinInterval, sdxCIKJIinIntervall)
+ }
> print(simulation(mu, sigma, n, ant_sets, Clvl))
[1] 940 952
> #d
>
> n = c(30,200)
> for (i in n){
+   print(simulation(mu, sigma, i, ant_sets, Clvl))
+ }
[1] 946 942
[1] 960 955
> #e
> n = 8
> my = 1
> sigma= 1
> exponent <- function(mu, sigma, n, ant_sets, Clvl){
```

```
+   simres=numeric(0)
+   for (i in 1:ant_setts) {
+     x=rexp(n,rate=1)
+     muhat=mean(x)
+     sdx=sd(x)
+
+     muCIt = muhat +c(-1,1)*qt(Clvl,df=n-1)*sdx/sqrt(n) #CI for  my
+     sdxCIKJI = sdx*sqrt((n-1)/qchisq(c(Clvl,1-Clvl), n-1))#CI for sigma
+
+     nyres = c(muhat,sdx, muCIt,sdxCIKJI)
+     simres=rbind(simres,nyres)
+   }
+
+   muCItinInterval= sum((simres[,3]<1)*(1<simres[,4]));
+   sdxCIKJIinIntervall = sum((simres[,5]<1)*(1<simres[,6]));
+   c(muCItinInterval, sdxCIKJIinIntervall)
+ }
> exponent(mu, sigma, n, ant_sets, Clvl)
[1] 899 787
> #f
> n  = c(30,200)
> for (i in n){
+   print(exponent(mu, sigma, i, ant_sets, Clvl))
+ }
[1] 932 714
[1] 950 686
> View(exponent)
> View(exponent)
```