STK1110 - Statistiske metoder og dataanalyse høsten 2019
Løsningsforslag til oppgaver i læreboka for uke 36

Vinnie Ko & Ørnulf Borgan

September 6, 2019

Lærebok: *Modern Mathematical Statistics with Applications (2nd ed.)* - Devore & Berk

Exercise 7.23

a) Ved definisjonen av forventing har vi

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x(\theta+1)x^\theta\,dx = (\theta+1)\int_0^1 x^{\theta+1}\,dx = \frac{\theta+1}{\theta+2}$$

Momentestimatoren er derfor gitt som løsningen av likningen:

$$\frac{\widehat{\theta}+1}{\widehat{\theta}+2} = \overline{X}$$

Så momentestimatoren blir

$$\widehat{\theta} = \frac{2\overline{X}-1}{1-\overline{X}} = \left(1-\overline{X}\right)^{-1} - 2.$$

Ved å bruke dataene gitt i oppgaven, finner vi estimatet $\widehat{\theta} = 3$.

```
> # Lager en vektor med dataene:
> time.data = c(0.92, 0.79, 0.90, 0.65, 0.86, 0.47, 0.73, 0.97, 0.94, 0.77)
>
> # Beregner estimatet:
> theta.hat = (1 - mean(time.data))^(-1) - 2
> print(theta.hat)
[1] 3
```

b) Likelihood funksjonen er

$$\mathcal{L}(\theta) = \prod_{i=1}^{n}(\theta+1)X_i^\theta,$$

og log-likelihood funksjonen er

$$\ell(\theta) = n\ln(\theta+1) + \theta\sum_{i=1}^{n}\ln X_i.$$

Vi vil finne θ som maksimerer $\ell(\theta)$. Vi løser da likningen:

$$\frac{\partial\ell(\theta)}{\partial\theta} = \frac{n}{\theta+1} + \sum_{i=1}^{n}\ln X_i = 0.$$

Vi finner at maximum likelihood estimatoren er gitt ved

$$\widehat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln X_i} - 1.$$

Ved å bruke dataene gitt i oppgaven, finner vi estimatet $\widehat{\theta} = 3.12$.

```
> # Beregner estimatet.
> theta.hat = -1/mean(log(time.data)) - 1
> print(theta.hat)
[1] 3.116068
```

Exercise 7.27

a) Vi har at $\mathrm{V}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$.

Siden det er gitt i oppgaven at $E[X] = 0$, har vi $\sigma^2 = \mathrm{V}(X) = \mathrm{E}[X^2]$.

Momentmetoden gir estimatoren $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

b) Likelihood funksjonen for n observasjoner fra normalfordelingen med $\mu = 0$ er

$$\mathcal{L}(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{X_i^2}{2\sigma^2}\right].$$

Log-likelihood funksjonen er

$$\ell(\sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} X_i^2.$$

Vi vil finne σ^2 som maksimerer $\ell(\sigma^2)$. Vi løser da likningen

$$\frac{\partial \ell(\sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n} X_i^2 = 0.$$

Ved å løse denne likningen, finner vi at maximum likelihood estimatoren er

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

dvs. den samme estimatoren som vi fikk ved momentmetoden.

Exercise 7.29

a) Likelihood funksjonen for Rayleighfordeling er gitt ved

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{X_i}{\theta} \exp\left[-\frac{X_i^2}{2\theta}\right].$$

Log-likelihood funksjonen er

$$\ell(\theta) = -n\ln\theta + \sum_{i=1}^{n} \ln X_i - \frac{1}{2\theta}\sum_{i=1}^{n} X_i^2.$$

Vi vil finne θ som maksimerer $\ell(\theta)$. Vi løser da likningen:

$$\frac{\partial \ell(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n} X_i^2 = 0.$$

Vi finner at maximum likelihood estimatoren er gitt ved

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

som er identisk med den forventningsrette estimatoren fra Exercise 15.

Vi bruker `R` til å bestemme $\widehat{\theta}$ ut fra dataene i Exercise 15.

```
> # Lager en vektor med dataene:
> kons = c(16.88, 10.23, 4.59, 6.66, 13.68, 14.23, 19.87, 9.40, 6.51, 10.95)
>
> # Beregner estimatet:
> theta.hat = mean(kons^2)/2
> print(theta.hat)
[1] 74.50529
```

b) Den kumulative Rayleigh fordelingen er gitt ved

$$F(x;\theta) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x} \frac{y}{\theta}\exp\left(-\frac{y^2}{2\theta}\right)\,dy = \left[\exp\left(-\frac{y^2}{2\theta}\right)\right]_0^x = 1 - \exp\left(-\frac{x^2}{2\theta}\right).$$

Siden medianen m er gitt ved at $F(m;\theta) = 0.5$, har vi at m er løsningen av likningen

$$1 - \exp\left(-\frac{m^2}{2\theta}\right) = 0.5.$$

Det gir $m = \sqrt{2\theta \ln 2}$.

Ved å bruke *invariansprinsippet* for maximum likelihood estimatoren (jf. side 357 læreboka), får vi at maximum likelihood estimatoren for medianen er

$$\widehat{m} = \sqrt{2\widehat{\theta}\ln 2},$$

der $\widehat{\theta}$ er maximum likelihood estimatoren fra punkt **a)**.

Vi bruker `R` til å bestemme $\widehat{m}$ ut fra dataene i Exercise 15.

```
> m.hat = sqrt(2*theta.hat*log(2))
> print(m.hat)
[1] 10.16298
```

Exercise 7.31

Vi har $Y \sim \text{Bin}(n,p)$. Likelihooden er

$$\mathcal{L}(p) = f(y,n,p) = \binom{n}{y} p^y (1-p)^{n-y},$$

og log-likelihooden er

$$\ell(p) = \ln\binom{n}{y} + y\ln p + (n-y)\ln(1-p).$$

Vi vil finne p som maksimerer $\ell(p)$. Vi løser da likningen

$$\frac{\partial \ell(p)}{\partial p} = \frac{y}{p} - \frac{n-y}{1-p} = 0.$$

Vi finner at maximum likelihood estimatet for p er $\widehat{p} = \dfrac{y}{n}$ og maximum likelihood estimatoren er $\widehat{p} = \dfrac{Y}{n}$.

Siden levetiden for hver komponent (X_i) er eksponentialt fordelt, har vi at

$$p = P(X_i > 24) = 1 - P(X_i \leq 24) = 1 - F(24) = e^{-24\lambda}.$$

Ved å bruke *invariansprinsippet* for maximum likelihood estimatoren (jf. side 357 i lærebokaboka), får vi at maximum likelihood estimatoren for λ er

$$\widehat{\lambda} = -\frac{1}{24}\ln(\widehat{p}).$$

Ved å bruke opplysningen i oppgaven om at $y = 15$ og $n = 20$, får vi maximum likelihood estimatene $\widehat{p} = \dfrac{15}{20} = 0.75$ og $\widehat{\lambda} = -\dfrac{1}{24}\ln(0.75) = 0.0120$.