Extra Opgaven over Hoofdstuk 21

21.1. Gegeven zijn de volgende drie observaties: 3, 7, 2. We veronderstellen deze data een realisatie te zijn van een steekproef uit een Poissonverdeling met (onbekende) parameter λ. De Maximum Likelihood schatter voor λ gebaseerd op deze gegevens is

a. 1/12 **b.** 1/4 **c.** 1/3 **d.** 3 **e.** 4 **f.** 12

21.2. De Pareto-verdeling is een kansverdeling waarbij de cumulatieve verdelingsfunctie wordt gegeven door $F(x) = 1 - x^{-\alpha}$ voor $x \geq 1$ en $F(x) = 0$ als $x < 1$. Hier is $\alpha > 0$ een (onbekende) parameter. Voor een dataset $x_1, x_2, \ldots, x_n$ veronderstellen we een Pareto verdeling. De likelihoodfunctie voor α wordt gegeven door

a. $L(\alpha) = \sum_{i=1}^{n} \log(1 - x_i^{-\alpha})$

b. $L(\alpha) = \sum_{i=1}^{n} \log(x_i^{-\alpha})$

c. $L(\alpha) = n \log \alpha - \alpha \sum_{i=1}^{n} \log x_i$

d. $L(\alpha) = \left(1 - \frac{1}{x_1^\alpha}\right)\left(1 - \frac{1}{x_2^\alpha}\right) \cdots \left(1 - \frac{1}{x_n^\alpha}\right)$

e. $L(\alpha) = \frac{\alpha^n}{x_1^{\alpha+1} x_2^{\alpha+1} \cdots x_n^{\alpha+1}}$

f. $L(\alpha) = \frac{\alpha^n}{x_1^\alpha x_2^\alpha \cdots x_n^\alpha}$

21.3. Men schat een onbekende parameter $0 \leq \theta \leq 1$ met de methode van maximum likelihood. Gegeven is dat de likelihood $L(\theta)$ voor de data gelijk is aan:

$$L(\theta) = 32\theta^3 - 54\theta^2 + 21\theta + 1$$

Dan is de maximum likelihood schatting voor θ gelijk aan

a. 0 **b.** $\frac{1}{8}$ **c.** $\frac{1}{4}$ **d.** $\frac{1}{2}$ **e.** $\frac{7}{8}$ **f.** 1

21.4. Bekijk de volgende dataset

$$5 \quad 5 \quad 3 \quad 1$$

die we beschouwen als realisaties uit een exponentiele verdeling met parameter λ. Gebaseerd op deze data is de Maximum Likelihood Schatter voor λ gelijk aan

a. $\lambda = \frac{1}{2}$ **b.** $\lambda = \frac{1}{14}$ **c.** $\lambda = -\frac{2}{7}$ **d.** $\lambda = \frac{2}{7}$

e. $\lambda = -\frac{1}{2}$ **f.** $\lambda = -\frac{1}{14}$

21.5. In een laboratorium worden op de ochtend van dag 1 van een experiment 100 ratten ingespoten met een virus. Aan het begin van dag 2, 3 en 4 wordt gekeken hoeveel ratten er de vorige dag overleden zijn. Dit levert de volgende gegevens:

dag 1	dag 2	dag 3	overlevend na drie dagen
31	16	18	35

We werken met het volgende model: elke rat heeft kans p om op een dag te sterven, onafhankelijk van de andere ratten en onafhankelijk van hoeveel dagen hij/zij al overleefd heeft. We willen de kans p schatten op grond van de gegeven data.

a. Beargumenteer dat de likelihood $L(p)$ gegeven wordt door

$$L(p) = Cp^{65}(1-p)^{157},$$

waarbij C een constante is die niet van p afhangt.

b. Bepaal de maximum likelihood schatting voor p.

Extra Opgaven over Hoofdstuk 22

22.1. Gegeven is een bivariate dataset $(x_1, y_1), (x_2, y_2), \ldots, (x_{100}, y_{100})$, waarvan bekend is dat de kleinste kwadraten schatting voor het intercept van de regressielijn gelijk is aan $\hat{\alpha} = 2$ en dat de kleinste kwadraten schatting voor de helling van de regressielijn $\hat{\beta}$ gelijk is aan 2. Zij verder gegeven dat het gemiddelde $\bar{y}_{100}$ van de y_i's gelijk is aan 8. Bereken $\bar{x}_{100}$.

22.2. Gegeven is een bivariate dataset $(x_1, y_1), (x_2, y_2), \ldots, (x_{50}, y_{50})$, waarvan bekend is dat

$$\sum_{i=1}^{50} x_i = 150 \quad \text{en} \quad \sum_{i=1}^{50} y_i = 400.$$

De kleinste kwadraten schatting voor de helling van de regressielijn is $\hat{\beta} = 2$. Bereken de kleinste kwadraten schatting $\hat{\alpha}$ voor het intercept (startgetal).

22.3. Bekijk de volgende data

x	1	1	4	9
y	-3	-2	-1	2

Bereken de kleinste-kwadratenschatters $\hat{\alpha}$ en $\hat{\beta}$.

22.4. Voor een bivariate dataverzameling $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, beschouwen we een lineair regressie model zonder intercept: $Y_i = \beta x_i + U_i$ voor $i = 1, 2, \ldots, n$, waarbij $U_1, U_2, \ldots, U_n$ onafhankelijk zijn met verwachting 0 en variantie σ^2. De kleinste-kwadraten-schatting voor β wordt gegeven door

a. $\dfrac{\sum y_i}{\sum x_i}$ **b.** $\dfrac{1}{n}\sum \dfrac{y_i}{x_i}$ **c.** $\dfrac{\sum x_i y_i}{\sum x_i^2}$ **d.** $\dfrac{\sum x_i^2 y_i^2}{\sum x_i^4}$ **e.** $\dfrac{\sqrt{\sum y_i^2}}{\sqrt{\sum x_i^2}}$ **f.** $\sqrt{\dfrac{1}{n}\sum \dfrac{y_i^2}{x_i^2}}$

Extra Opgaven over Hoofdstukken 23 en 24

23.1. Gegeven is een dataverzameling die een realisatie is van een steekproef van omvang 16 uit een $N(\mu, \sigma^2)$ verdeling, met μ en $\sigma^2 > 0$ onbekend. Verder is gegeven dat het gemiddelde van de data 15 is, en de steekproefvariantie gelijk is aan 4. Het 90% betrouwbaarheidsinterval voor μ wordt dan gegeven door:

a. (13.934, 16.066) **b.** (13.940, 16.006) **c.** (14.020, 15.980)
d. (14.124, 15.877) **e.** (14.127, 15.873) **f.** (14.177, 15.822)

23.2. Een onderzoeker maakt op basis van 15 metingen (die beschouwd worden als realisaties van onafhankelijke trekkingen uit dezelfde $\mathcal{N}(\mu, \sigma^2)$ verdeling met μ en σ onbekend) een 90 % betrouwbaarheidsinterval voor μ. Als hij in plaats van 90 procent betrouwbaarheid 98 procent betrouwbaarheid neemt, wat gebeurt er dan met het betrouwbaarheidsinterval?

a. Dat wordt ongeveer 5 keer zo groot **b.** Dat wordt ongeveer 5 keer zo klein
c. Dat wordt ongeveer 1.5 keer zo klein **d.** Dat wordt ongeveer 1.5 keer zo groot
e. Dat wordt ongeveer $\sqrt{5}$ keer zo klein **f.** Dat hangt van de waarden van de metingen af

23.3. Een lijmfabrikant heeft een gestandaardiseerde test voor het bepalen van de uithardingstijd van lijm. Onder de testcondities zijn de uithardingstijden bij benadering normaal verdeeld en onafhankelijk. Er worden 17 tests gedaan voor lijmtype *White Rabbit*, met als uitkomsten een gemiddelde uithardingstijd van 54.0 seconden en een standaarddeviatie van 20.4. Het 95% betrouwbaarheidsinterval voor de verwachte uithardingstijd is:

a. $43.5 < \mu < 64.5$ **b.** $44.3 < \mu < 63.7$ **c.** $43.6 < \mu < 64.4$
d. $45.9 < \mu < 62.1$ **e.** $45.4 < \mu < 62.6$ **f.** $51.5 < \mu < 56.5$

23.4. Stel dat $X_1, X_2, \ldots, X_n$ een steekproef is uit een normale verdeling met verwachting μ en variantie σ^2. Beschouw de volgende beweringen:

A Het 95%-betrouwbaarheidsinterval voor μ wordt smaller naarmate n groter wordt.

B Het 95%-betrouwbaarheidsinterval voor μ is smaller dan het 99%-betrouwbaarheidsinterval voor μ

Welke van deze beweringen zijn waar?

a. geen **b.** A **c.** B **d.** A en B

23.5. Om het vermogen tot het schatten van lengte te onderzoeken, is aan 30 willekeurige personen gevraagd de lengte te schatten van een in een rechte lijn liggend stuk touw. De volgende schattingen, in meters en in volgorde van grootte, werden gedaan:

2.2	2.5	2.5	2.7	3.0	3.0
3.0	3.0	3.0	3.2	3.5	3.6
3.6	3.6	3.7	3.7	4.0	4.0
4.0	4.1	4.1	4.2	4.2	4.2
4.3	4.3	4.4	4.5	4.6	4.8

Voor deze dataset geldt $\bar{x}_{30} = 3.65$ en $s_{30} = 0.696$. De werkelijke lengte was 4.0 meter. Neem aan dat de dataset opgevat kan worden als een realisatie van een steekproef $X_1, \ldots, X_{30}$ uit een normale verdeling $N(\mu, \sigma^2)$. Een 90%-betrouwbaarheidsinterval voor μ wordt gegeven door:

a. (3.43, 3.87)
b. $(3.45, 3.85)$
c. $(3.41, 3.89)$
d. $(3.40, 3.90)$
e. $(3.38, 3.92)$
f. Kan niet bepaald worden op grond van deze gegevens.

23.6. Stel dat we data $x_1, \ldots, x_n$ hebben die we opvatten als een realisatie van een steekproef $X_1, \ldots, X_n$ uit een normale verdeling met onbekende verwachting μ en bekende variantie σ^2.

a. Stel dat $\sigma = 2$ en $\alpha = 0.05$. Hoe groot moet n zijn opdat de lengte van het betrouwbaarheidsinterval maximaal 1 is?

b. Stel dat we berekend hebben dat het betrouwbaarheidsinterval voor μ gelijk is aan $\left(1\frac{1}{2}, 3\right)$. Geef een betrouwbaarheidsinterval voor $1 - \mu$ (bij dezelfde data en dezelfde betrouwbaarheid).

Extra Opgaven over Hoofdstukken 25 en 26

25.1. Voor een zekere statistische toets wordt de toestingsgrootheid gegeven door T. Het significantie-niveau is gelijk aan 0.1. Het kritieke gebied behorende bij deze toets noteren we met K. De kans dat we de nulhypothese verwerpen als deze niet waar is, is gelijk aan 0.6. De kans op een fout van de tweede soort is gelijk aan:

a. 0.1 **b.** 0.9 **c.** 0.4
d. 0.6 **e.** 0.5 **f.** 0.7

25.2. Gegeven data $x_1, x_2, \ldots, x_{15}$ die we opvatten als een steekproef uit een verdeling met kansdichtheid

$$f(x) = \begin{cases} 0 & x < \theta \\ e^{-(x-\theta)} & x \geq \theta \end{cases}.$$

We toetsen de nulhypothese $\theta = 0$ tegen de alternatieve hypothese $\theta > 0$. Als toetsingsgrootheid nemen we $T = \min\{X_1, X_2, \ldots, X_{15}\}$. Grote waarden van T duiden op de alternatieve hypothese. Stel dat de geobserveerde waarde van T gelijk is aan $t = 0.1$. De p-waarde bij deze toets (afgerond op 2 decimalen) is gelijk aan:

a. 0.14 **b.** 0.90 **c.** 0.22 **d.** 0.06 **e.** 0.12 **f.** 0.43

Hint: Als $X_1, X_2, \ldots, X_n$ een steekproef uit een Exp(λ) verdeling is, dan heeft

$$\min\{X_1, X_2, \ldots, X_n\}$$

een Exp($n\lambda$) verdeling.

25.3. Stel dat X een uniforme verdeling heeft op het interval $[0, \theta]$, waarbij $\theta > 0$ een onbekende parameter is. We toetsen $H_0 : \theta = 3$ versus $H_1 : \theta > 3$. Als toetsingsgrootheid nemen we X zelf.

a. Bereken het kritieke gebied bij significantieniveau 0.1.

b. Stel dat we $x = 1$ observeren. Bereken de bijbehorende p-waarde.

25.4. Uit een verzameling objecten (bijv. tanks) genummerd 1 tot en met K worden *met* teruglegging 20 objecten getrokken. We willen toetsen $H_0 : K = 100000$ tegen $H_1 : K < 100000$, met het hoogste rangnummer M van onze steekproef als toetsingsgrootheid. We vinden als realisatie voor M de waarde 81115. De p-waarde van deze uitkomst is:

a. 0.013 **b.** 0.041 **c.** 0.015 **d.** 0.520 **e.** 0.154 **f.** 0.852

25.5. Stel een meting X van de snelheid van een passerende auto op de snelweg wordt gemodelleerd als een normale variable met parameters $\mu = v$ (in km/u) en $\sigma = 5$, waarbij v de exacte snelheid is van de auto. Om te toetsen of een automobilist harder rijdt dan de toegestane 120 km/u wordt getoetst H_0: $v = 120$ tegen H_1: $v > 120$.
Stel verder dat de nulhypothese wordt verworpen (en dus: de automobilist krijgt een boete) als de gemeten snelheid 128 km/u of hoger is. Wat geldt dan voor P_{II}, de fout van de tweede soort, voor een automobilist die 133 km/u rijdt?

a. $P_{II} = 0.08$ **b.** $P_{II} = 0.16$ **c.** $P_{II} = 0.32$ **d.** $P_{II} = 0.01$
e. $P_{II} = 0.12$ **f.** $P_{II} = 0.02$

25.6. Stel X heeft een uniforme verdeling op $[0, \mu]$ waarbij μ onbekend is. De nulhypothese is dat $\mu = 2.5$, en de alternatieve hypothese dat $\mu > 2.5$. Iemand gaat een toets uitvoeren door twee trekkingen X_1 en X_2 te doen en neemt als toetsingsgrootheid T het maximum van X_1 en X_2. Stel dat hij besluit H_0 te verwerpen ten gunste van H_1 als $T \geq 2$. Als de werkelijke waarde van μ gelijk is aan 3 wat is dan (op 2 decimalen nauwkeurig) de kans op een fout van de tweede soort?

a. 0.12 **b.** 0.44 **c.** 0.28 **d.** 0.33 **e.** 0.53 **f.** 0.21

Antwoorden:

21.1. e.

21.2. e.

21.3. c.

21.4. **d.**

21.5.b $p \approx 0.293$

22.1. 3

22.2. 2

22.3. $\hat{\alpha}$ =-3.105, $\hat{\beta} = 0.561$.

22.4. **c.**

23.1. d.

23.2. d.

23.3. a.

23.4. d.

23.5. a.

23.6.a $n \geq 62$.

23.6.b $(-2, -\frac{1}{2})$

25.1. c.

25.2. c.

25.3.a $[2.7, \infty)$

25.3.b $\frac{2}{3}$

25.4. c.

25.5. b.

25.6. b.