Statistik I für Human- und Sozialwissenschaften
6. Übung
Lösungsvorschlag

Gruppenübung

G 18 a) Zur Feststellung der Anzahl N der in einem Revier lebenden Rothirsche wurden in einer Fangaktion insgesamt neun Tiere gefangen und gekennzeichnet. Anschließend wurden die gefangenen Tiere im gleichen Revier wieder freigelassen. Nach einer gewissen Zeit wurde eine weitere Fangaktion durchgeführt. Dabei wurden drei Rothirsche gefangen, und man stellte fest, dass unter diesen genau zwei Rothirsche gekennzeichnet waren. Es wird angenommen, dass zwischen beiden Fangaktionen keine Zu- oder Abgänge von Rothirschen im beobachteten Revier stattgefunden haben. Sei X die Zufallsvariable, welche die Anzahl der gefangenen und gekennzeichneten Rothirsche in der zweiten Fangaktion angibt.
Welche Verteilung besitzt X?

b) Beim Roulette tritt in einem Spiel eine der Zahlen $0, 1, 2, \ldots, 36$ auf. Ein abergläubischer Spieler beginnt erst mit dem Spiel, nachdem zum ersten Mal eine seiner Unglückszahlen 3, 13, 23 oder 33 aufgetreten ist. Die Zufallsgröße Y beschreibe die Anzahl von Runden, die dieser Spieler warten muss, bevor er mit seinem Spiel beginnen kann.
Welche Verteilung besitzt Y?

c) Bei einer Lotterie beträgt die Wahrscheinlichkeit für eine Niete bei jedem Zug 0.7. Die Zufallsvariable Z beschreibe die Anzahl an Nieten beim Ziehen von zehn Losen.
Welche Verteilung besitzt Z?

a) $X \sim H(3; N; 9)$

b) $Y \sim Geo(\frac{4}{37})$

c) $Z \sim B(10; 0.7)$

G 19 Sei X eine normalverteilte Zufallsvariable mit Erwartungswert μ und Varianz σ^2 mit $\sigma > 0$.

a) Zeigen Sie, dass für alle $a < b$ gilt

$$P\left(a < X \leq b\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist.

b) Es sei $\mu = -1$ und $\sigma^2 = 0.16$ gegeben. Berechnen Sie: $P(0 < X \leq 0.4)$, $P(X \leq -1.4)$ und $P(X > -\frac{37}{50})$.

a) Wenn $X \sim N(\mu, \sigma^2)$, dann ist die zugehörige standardisierte Zufallsvariable $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ und besitzt die Verteilungsfunktion Φ. Somit gilt:

$$\begin{aligned}
P\left(a < X \leq b\right) &= P\left(a - \mu < X - \mu \leq b - \mu\right) \\
&= P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= P\left(\frac{a-\mu}{\sigma} < Z \leq \frac{b-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).
\end{aligned}$$

b)

$$\begin{aligned}
P(0 < X \le 0.4) &= \Phi\left(\frac{0.4-(-1)}{\sqrt{0.16}}\right) - \Phi\left(\frac{0-(-1)}{\sqrt{0.16}}\right) = \Phi(3.5) - \Phi(2.5) \\
&\approx 0.999767 - 0.99379 = 0.005977 \\
P(X \le -1.4) &= \Phi\left(\frac{-1.4-(-1)}{\sqrt{0.16}}\right) = \Phi(-1) = 1 - \Phi(1) \\
&\approx 1 - 0.8413 = 0.1587 \\
P\left(X > -\frac{37}{50}\right) &= 1 - P\left(X \le -\frac{37}{50}\right) = 1 - \Phi\left(\frac{-\frac{37}{50}-(-1)}{\sqrt{0.16}}\right) = 1 - \Phi(0.65) \\
&\approx 1 - 0.7421 = 0.2579
\end{aligned}$$

G 20 Um die Präzision einer Waage zu überprüfen, wird n-mal das Gewicht eines Kilogramm-Prototyps gemessen. Die entstehende Messreihe soll als Realisierung von unabhängigen, identisch $N(0,\sigma^2)$-verteilten Zufallsvariablen $X_1,\ldots,X_n$ mit unbekannter Varianz σ^2 aufgefasst werden. Dabei wird $\sigma > 0$ angenommen.

Bestimmen Sie den Maximum-Likelihood-Schätzwert für den Parameter σ zur Messreihe
0.03; 0.01; −0.02; −0.05; 0.03; 0.1; −0.04; −0.07; 0.01; 0.02.

Die Zufallsvariable X_i besitzt die von dem Parameter σ abhängige Dichte

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad x \in \mathbb{R}.$$

Es ergibt sich für die Likelihood-Funktion bzgl. des Parameters σ

$$L(\sigma, x_1,\ldots,x_n) = \prod_{i=1}^n f_\sigma(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right).$$

Zur Bestimmung des Maximums gehen wir zur Log-Likelihood-Funktion über und formen mithilfe der Logarithmengesetze um.

$$\begin{aligned}
\ln(L(\sigma, x_1,\ldots,x_n)) &= \ln\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right)\right) \\
&= \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right)\right) \\
&= \sum_{i=1}^n \left(\ln\left(\frac{1}{\sqrt{2\pi}}\right) - \ln(\sigma) - \frac{x_i^2}{2\sigma^2}\right) \\
&= n\ln\left(\frac{1}{\sqrt{2\pi}}\right) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2
\end{aligned}$$

Zur Bestimmung der Extremstellen berechnen wir die Nullstellern der ersten Ableitung der Log-Likelihood-Funktion.

$$(\ln(L(\sigma, x_1,\ldots,x_n)))' = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n x_i^2$$

$$0 = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n} x_i^2$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3}\sum_{i=1}^{n} x_i^2$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\sigma_1 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \qquad \sigma_2 = -\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

Aus der Voraussetzung $\sigma > 0$ folgt, dass nur σ_1 möglich wäre. Wir überprüfen mit der zweiten Ableitung, ob σ_1 auch eine Maximalstelle ist.

$$(\ln(L(\sigma, x_1, \ldots, x_n)))'' = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n} x_i^2$$

$$\begin{aligned}(\ln(L(\sigma_1, x_1, \ldots, x_n)))'' &= \frac{n}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right)^2} - \frac{3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right)^4}\sum_{i=1}^{n} x_i^2 \\ &= \frac{n}{\frac{1}{n}\sum_{i=1}^{n} x_i^2} - \frac{3}{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^2}\sum_{i=1}^{n} x_i^2 \\ &= \frac{n^2}{\sum_{i=1}^{n} x_i^2} - \frac{3n^2}{\sum_{i=1}^{n} x_i^2} \\ &= -\frac{2n^2}{\sum_{i=1}^{n} x_i^2} < 0\end{aligned}$$

Somit ist σ_1 eine Maximalstelle der Likelihoodfunktion. Wir berechnen den Maximum-Likelihood-Schätzwert durch einsetzen der $n = 10$ Messdaten der Messreihe.

$$\sigma_1 = \sqrt{\frac{1}{10}\sum_{i=1}^{10} x_i^2} \approx 0.0467$$

Dass heisst, der Maximum-Likelihood-Schätzwert für den Parameter σ beträgt rund 0.0467.

G 21 Skizzieren Sie folgende Funktionen:

$f_1(x) := 2x$ $\quad$ $f_3(x) := x^2 - 1$ $\quad$ $f_5(x) := \ln(x), \quad$ für $x > 0$

$f_2(x) := -0.5x + 4$ $\quad$ $f_4(x) := e^x$ $\quad$ $f_6(x) := \frac{1}{\sqrt{8\pi}} \exp(-\frac{x^2}{8})$

Hausübung

H 11 a) Die Zufallsvariable X_1 sei $Poi(2)$-verteilt. Bestimmen Sie $E(X_1)$, $Var(X_1)$ und $P(\{X_1 = 3\})$.

b) Die Zufallsvariable X_2 sei $N(10; 25)$-verteilt. Bestimmen Sie $E(X_2)$, $Var(X_2)$ und $P(\{X_2 \leq 0\})$.

c) Die Zufallsvariable X_3 sei $Geo(\frac{4}{37})$-verteilt. Bestimmen Sie $E(X_3)$, $Var(X_3)$ und $P(\{2 \leq X_3 \leq 4\})$.

d) Die Zufallsvariable X_4 sei $H(6; 49; 6)$-verteilt. Bestimmen Sie $E(X_4)$, $Var(X_4)$ und $P(\{X_4 = 3\})$.

a) Aus $X \sim Poi(\lambda)$ folgt $E(X) = Var(X) = \lambda$ und $P(\{X = k\}) = \frac{\lambda^k}{k!}e^{-\lambda}$ für $k \in \mathbb{N}_0$. Somit gilt:

$$E(X_1) = Var(X_1) = 2 \qquad P(\{X_1 = 3\}) = \frac{2^3}{3!}e^{-2} \approx 0.18045$$

b) Aus $X \sim N(\mu, \sigma^2)$ folgt $E(X) = \mu$ und $Var(X) = \sigma^2$ Somit gilt:

$$E(X_2) = 10 \qquad Var(X_2) = 25$$

$$P(\{X_2 \leq 0\}) = P(\{\frac{X_2 - 10}{\sqrt{25}} \leq \frac{0 - 10}{\sqrt{25}}\}) = \Phi(-2) = 1 - \Phi(2) \approx 1 - 0.97725 = 0.02275$$

c) Aus $X \sim Geo(p)$ folgt $E(X) = \frac{1-p}{p}$, $Var(X) = \frac{1-p}{p^2}$ und $P(\{X = k\}) = (1-p)^{k-1}p$ für $k \in \mathbb{N}$. Somit gilt:

$$E(X_3) = \frac{1 - \frac{4}{37}}{\frac{4}{37}} = \frac{\frac{33}{37}}{\frac{4}{37}} = \frac{33}{4} = 8.25$$

$$Var(X_3) = \frac{1 - \frac{4}{37}}{\left(\frac{4}{37}\right)^2} = \frac{\frac{33}{37}}{\frac{4}{37}\,\frac{4}{37}} = \frac{33 \cdot 37}{16} = 76.3125$$

$$P(\{2 \leq X_3 \leq 4\}) = \left(\frac{33}{37}\right)\frac{4}{37} + \left(\frac{33}{37}\right)^2\frac{4}{37} + \left(\frac{33}{37}\right)^3\frac{4}{37} \approx 0.2591$$

d) Aus $X \sim H(n, N, M)$ folgt $E(X) = n\frac{M}{N}$, $Var(X) = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$ und $P(\{X = k\}) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$. Somit gilt:

$$E(X_4) = 6\frac{6}{49} = \frac{36}{49}$$

$$Var(X_4) = 6\frac{6}{49}\,\frac{49-6}{49}\,\frac{49-6}{49-1} = 0.5776$$

$$P(\{X_4 = 3\}) = \frac{\binom{6}{3}\binom{43}{3}}{\binom{49}{6}} \approx 0.01765$$

H 12 In einer Telefonzentrale wird an einem normalen Werktagvormittag die Anzahl der innerhalb von 5 Minuten ankommenden Telefongespräche ermittelt. Die entstehende Messreihe soll als Realisierung von unabhängigen, identisch $Poi(\lambda)$-verteilten Zufallsvariablen $X_1, \ldots, X_n$ mit unbekanntem Parameter $\lambda > 0$ aufgefasst werden.

Bestimmen Sie den Maximum-Likelihood-Schätzwert für den Parameter λ zur Messreihe
4; 9; 7; 3; 8; 12; 7; 9; 7; 10; 6; 8.

Die Zufallsvariable X_i besitzt die von dem Parameter λ abhängige Wahrscheinlichkeitsfunktion

$$f_\lambda(x) = \frac{\lambda^x}{x!}\exp(-\lambda) \qquad x \in \mathbb{N}_{\nvdash}.$$

Es ergibt sich für die Likelihood-Funktion bzgl. des Parameters λ

$$L(\lambda, x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\lambda(x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}\exp(-\lambda).$$

Zur Bestimmung des Maximums gehen wir zur Log-Likelihood-Funktion über und formen mithilfe der Logarithmengesetze um.

$$\begin{aligned}
\ln(L(\lambda, x_1, \ldots, x_n)) &= \ln\left(\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}\exp(-\lambda)\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{\lambda^{x_i}}{x_i!}\exp(-\lambda)\right) \\
&= \sum_{i=1}^{n} \left(x_i \ln(\lambda) - \ln(x_i!) - \lambda\right)
\end{aligned}$$

Zur Bestimmung der Extremstellen berechnen wir die Nullstellern der ersten Ableitung der Log-Likelihood-Funktion.

$$\begin{aligned}
(\ln(L(\lambda, x_1, \ldots, x_n)))' &= \sum_{i=1}^{n}\left(\frac{x_i}{\lambda} - 1\right) = \frac{1}{\lambda}\left(\sum_{i=1}^{n} x_i\right) - n \\
0 &= \frac{1}{\lambda}\left(\sum_{i=1}^{n} x_i\right) - n \\
\lambda &= \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}$$

Wir überprüfen mit der zweiten Ableitung, ob $\bar{x}$ auch eine Maximalstelle ist.

$$(\ln(L(\lambda, x_1, \ldots, x_n)))'' = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i < 0 \qquad \text{da die } x_i \geq 0.$$

Somit ist $\bar{x}$ eine Maximalstelle der Likelihoodfunktion. Wir berechnen den Maximum-Likelihood-Schätzwert durch einsetzen der $n = 12$ Messdaten der Messreihe.

$$\bar{x} = \frac{1}{12}\sum_{i=1}^{12} x_i = 7.5$$

Dass heisst, der Maximum-Likelihood-Schätzwert für den Parameter λ beträgt 7.5.