

Übungsblatt 10

Aufgabe P1 *Fischpopulation.*

In einem Teich befinden sich n Fische, wobei n (die Populationsgröße) unbekannt sei. Um die Populationsgröße n zu schätzen, kann man wie folgt vorgehen. Im ersten Schritt werden aus dem Teich n_1 (eine bekannte Zahl, kleinergleich n) verschiedene Fische gefangen und markiert. Danach werden die n_1 Fische wieder in den Teich zurückgeworfen. Im zweiten Schritt werden x (eine bekannte Zahl, kleinergleich n) verschiedene Fische gefangen. Unter diesen x Fischen ist eine Anzahl an Fischen markiert, diese Anzahl nennen wir $x_1 \in \{0, \dots, x\}$. Entsprechend sind als $x - x_1$ der im zweiten Schritt gefangenen Fische nicht markiert.

Bestimmen Sie den Maximum-Likelihood-Schätzwert für die Populationsgröße n .

Lösung: Sei die Zufallsgröße X die Anzahl der markierten Fische unter den $x \leq n$ gezogenen Fischen. Damit ist für $x_1 \in \{0, \dots, x\}$

$$L(x_1, n) = \mathbb{P}_n(X = x_1) = \frac{\binom{n_1}{x_1} \binom{n-n_1}{x-x_1}}{\binom{n}{x}}.$$

Wir müssen nun $n \in \mathbb{N}$ (!) so bestimmen, dass dies maximal wird. Betrachte

$$D(n) = \frac{L(x_1, n)}{L(x_1, n-1)} = \frac{(n-x)(n-n_1)}{n(n-n_1-x+x_1)}.$$

Sei $\hat{n} = n_1 x / x_1$. Eine einfache Rechnung zeigt:

- $D(n) > 1$, falls $n < \hat{n}$.
- $D(n) < 1$, falls $n > \hat{n}$.
- $D(n) = 1$, falls $n = \hat{n}$.

Folglich wächst die Likelihoodfunktion L für $n < \hat{n}$ und fällt für $n > \hat{n}$. Ist $\hat{n} \notin \mathbb{N}$, so ist $L(x_1, n)$ bei $n = \lfloor \hat{n} \rfloor$ maximal (hier ist $\lfloor \cdot \rfloor : \mathbb{R} \rightarrow \mathbb{Z}$, $\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \leq x\}$). Ist \hat{n} hingegen ganzzahlig, so gibt es zwei Maxima bei $n = \hat{n}$ und bei $n = \hat{n} - 1$, denn wegen $D(\hat{n}) = 1$ ist $L(x_1, n) = L(x_1, n-1)$. Somit lautet der MLS für n

$$\hat{n} = \begin{cases} \lfloor \frac{n_1 x}{x_1} \rfloor, & \text{falls } \frac{n_1 x}{x_1} \notin \mathbb{N} \\ \frac{n_1 x}{x_1} \text{ oder } \frac{n_1 x}{x_1} - 1, & \text{sonst.} \end{cases}$$

Aufgabe P2 *Körpergröße.*

Die Körpergröße von Menschen kann als normalverteilt angenommen werden. Sie vermessen 20 Männer und erhalten die folgenden Körpergrößen in cm

185, 176, 152, 178, 181, 194, 163, 168, 171, 179,
181, 175, 164, 202, 176, 181, 180, 174, 178, 182.

Bestimmen Sie Konfidenzintervalle für den Erwartungswert und die Varianz der Körpergröße von Männern, die die gesuchten Parameter mit einer Wahrscheinlichkeit von 95% enthalten.

Hinweis: Verwenden Sie ohne Beweis

$$(n-1)S^2 = \left(\sum_{i=1}^n X_i^2 \right) - n\bar{X}^2$$

wobei $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Lösung: *Erstellt von Meier-Hans 13.01.2016*

Aus dem Text erhalten wir $n = 20$ und $\alpha = 0,05$. Aus den Daten berechnet man $\bar{X} = 177$ und $S^2 \approx 115,1579$.

Laut Vorlesung konstruiert man bei unbekannter Varianz ein Konfidenzintervall I_μ für den Erwartungswert einer Normalverteilung gemäß

$$I_\mu = \left(\bar{X} - t_{1-\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}; \bar{X} + t_{1-\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \right)$$

Aus der Tabelle der t-Verteilung ermittelt man $t_{1-\alpha/2, n-1} = t_{0,975;19} = 2,0930$. Somit erhält man

$$I_\mu = (171,85; 182,15).$$

Ein entsprechendes Intervall I_{σ^2} erhält man nach Vorlesung als

$$\left(\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}; \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2} \right).$$

Aus der Tabelle der χ^2 -Verteilung ermittelt man $\chi_{0,975,19}^2 = 32,85$ und $\chi_{0,025,19}^2 = 8,907$. Somit erhält man

$$I_{\sigma^2} = (66,61; 245,65).$$

Aufgabe H1 *Konfidenzintervall für Exponentialverteilung.*

Seien X_1, \dots, X_n , $n \in \mathbb{N}$, unabhängig und identisch verteilt mit $X_1 \sim \text{Exp}(\lambda)$ und $\lambda > 0$ sei der unbekannte zu schätzende Parameter. Das heißt, X_1 hat die Dichte $f_\lambda(x) = \lambda e^{-\lambda x} 1_{\{x>0\}}$. Weiterhin seien $X_{(1)} := \min\{X_1, \dots, X_n\}$ das Minimum der Daten und $\alpha \in (0, 1)$. Zeigen Sie, dass

$$K := \left[-\frac{\ln(1 - \frac{\alpha}{2})}{nX_{(1)}}, -\frac{\ln(\frac{\alpha}{2})}{nX_{(1)}} \right]$$

ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha$ für λ ist.

Lösung: Zunächst rechnet man nach, dass

$$\begin{aligned} \mathbb{P}_\lambda(\lambda \in K) &= \mathbb{P}_\lambda \left(-\frac{\ln(1 - \frac{\alpha}{2})}{n\lambda} \leq X_{(1)} \leq -\frac{\ln(\frac{\alpha}{2})}{n\lambda} \right) \\ &= \mathbb{P}_\lambda \left(X_{(1)} \leq -\frac{\ln(\frac{\alpha}{2})}{n\lambda} \right) - \mathbb{P}_\lambda \left(X_{(1)} \leq -\frac{\ln(1 - \frac{\alpha}{2})}{n\lambda} \right). \end{aligned} \quad (1)$$

Unter Beachtung, dass die Verteilungsfunktion von X_i gegeben ist durch $F_\lambda(x) = (1 - e^{-\lambda x}) 1_{\{x>0\}}$ und dass die X_i i.i.d. sind, rechnen wir nach, dass für alle $x > 0$ gilt

$$\begin{aligned} \mathbb{P}_\lambda(X_{(1)} \leq x) &= 1 - \mathbb{P}_\lambda(X_{(1)} \geq x) = 1 - \mathbb{P}_\lambda(X_1 \geq x, \dots, X_n \geq x) \\ &= 1 - \prod_{i=1}^n \mathbb{P}_\lambda(X_i \geq x) = 1 - (1 - \mathbb{P}_\lambda(X_1 \leq x))^n = 1 - e^{-n\lambda x}. \end{aligned}$$

Wenden wir dies in (1) mit $x = -\frac{\ln(\frac{\alpha}{2})}{n\lambda}$ bzw. $x = -\frac{\ln(1 - \frac{\alpha}{2})}{n\lambda}$ an, erhalten wir, dass Gleichung (1) gleich ist zu

$$1 - \frac{\alpha}{2} - (1 - 1 + \frac{\alpha}{2}) = 1 - \alpha.$$

Aufgabe H2 *Wahlbeteiligung.*

Die Fachschaft möchte vor den Hochschulwahlen die voraussichtliche Wahlbeteiligung schätzen. Auf dem Weg zur der Mensa befragt jeder der Mitarbeiter solange Studenten, bis er jemanden trifft, der angibt, nicht zur Wahl zu gehen.

X bezeichne die Anzahl der Studenten, die überprüft wurden, bis ein potentieller Nichtwähler gefunden wurde. Eine sinnvolle Verteilung für X ist beispielsweise die modifizierte geometrische Verteilung mit dem Parameter $1 - \theta$, das heißt

$$\mathbb{P}_\theta(X = i) = \theta^{i-1} \cdot (1 - \theta), \quad i \in \mathbb{N}.$$

Die unbekannte Größe $\theta \in (0, 1)$ kann dabei als die gesuchte prozentuale Wahlbeteiligung interpretiert werden.

- Bestimmen Sie für unabhängige Stichproben x_1, \dots, x_n den Maximum-Likelihood-Schätzwert für θ .
- Nun haben Sie folgende konkrete Stichprobenwerte gegeben

3, 1, 2, 3, 2, 1, 2, 1, 5, 3, 2, 3.

Berechnen Sie für diese Werte den Maximum-Likelihood-Schätzwert für θ .

Lösung: Insgesamt fünf Punkte.

a) Die Likelihood-Funktion ist

$$L(x, \theta) = \prod_{i=1}^n \mathbb{P}(X = x_i) = \prod_{i=1}^n \theta^{x_i-1} \cdot (1 - \theta) = (1 - \theta)^n \cdot \theta^{-n} \cdot \prod_{i=1}^n \theta^{x_i}. \quad [1 \text{ Pkt}]$$

Die Log-Likelihood-Funktion lautet

$$\ln L(x, \theta) = n \cdot \ln(1 - \theta) - n \ln(\theta) + \sum_{i=1}^n x_i \ln(\theta) \quad [1 \text{ Pkt}]$$

und ihre Ableitung

$$\frac{\partial \ln L(x, \theta)}{\partial \theta} = -\frac{n}{1 - \theta} - \frac{n}{\theta} + \sum_{i=1}^n \frac{x_i}{\theta} \quad [1 \text{ Pkt}].$$

Nullsetzen und auflösen nach θ liefert als Schätzer

$$\hat{\theta} = 1 - \frac{1}{\bar{x}}.$$

mit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Die zweite Ableitung ist wegen

$$\frac{\partial^2}{\partial \theta^2} L(x, \theta) = -\frac{n}{(1 - \theta)^2} + \frac{n}{\theta^2} - \sum_{i=1}^n \frac{x_i}{\theta^2} = -\frac{n}{(1 - \theta)^2} - \sum_{i=1}^n \frac{x_i - 1}{\theta^2}$$

überall negativ, denn $x_i \geq 1$. Damit muss $\hat{\theta}$ ein Maximum sein [1 Pkt]. (Alternativ kann man auch begründen, dass die erste Ableitung auf ähnliche Weise als überall fallend erkannt werden kann.)

b) Für diese gegebenen Werte gilt

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{28}{12} = \frac{7}{3}. \quad [0.5 \text{ Pkt}]$$

Entsprechend gilt nach a)

$$\hat{\theta} = 1 - \frac{1}{\frac{7}{3}} = \frac{4}{7}. \quad [0.5 \text{ Pkt}]$$

Aufgabe H3 Konfidenzintervalle.

In einem Online-Shop wird festgestellt, dass von $n = 100$ überprüften Benutzern genau $H = 80$ Benutzer einen Webbrowser verwenden, der nicht über alle von den Webdesignern des Shops gewünschten Fähigkeiten verfügt. Es soll ein Konfidenzintervall für den relativen Anteil der Benutzer mit veraltetem Browser bestimmt werden, dass höchstens mit Wahrscheinlichkeit $\alpha = 0,05$ nicht eingehalten wird.

- Bestimmen Sie das Konfidenzintervall approximativ unter Verwendung der Quantile der Normalverteilung.
- Es stellt sich heraus, dass die Stichprobenvarianz 0,16 beträgt. Bestimmen Sie das Konfidenzintervall unter Verwendung der Quantile der t -Verteilung.

Lösung: Insgesamt fünf Punkte.

Erstellt von Meier-Hans 17.12.2014, modifiziert von Alexander Klump 01.2022

Dass ein Benutzer einen veralteten Browser benutzt ist Bernoulli-verteilt mit unbekanntem Parameter p , aber wir kennen den geschätzten Anteil $h_n = H/n = 0,8$ und es ist $\alpha = 0,05$, $n = 100$. [0.5 Pkt]

a)

Will man die Quantile der Normalverteilung nutzen, kennt aber die Varianz nicht, muss die Varianz geschätzt werden. Laut Vorlesung gilt für das Schätzen eines Konfidenzintervalles für p bei unbekannter Varianz approximativ

$$K_a = \left(h_n - z_{(1-\frac{\alpha}{2})} \sqrt{\frac{h_n(1-h_n)}{n}}; h_n + z_{(1-\frac{\alpha}{2})} \sqrt{\frac{h_n(1-h_n)}{n}} \right) \quad [1 \text{ Pkt}]$$

Mit $z_{(1-\frac{\alpha}{2})} = z_{0,975} = 1,96$ gilt

$$z_{(1-\frac{\alpha}{2})} \sqrt{\frac{h_n(1-h_n)}{n}} \approx 0,0784 \quad [0.5 \text{ Pkt}]$$

Somit gilt $K_a = (0,7216; 0,8784)$. [0.5 Pkt]

b)

Auf analoge Weise wie in der Vorlesung für Teil a) nehmen wir an dass

$$\tilde{Z} = \frac{h_n - p}{S} \cdot \sqrt{n}$$

approximativ t -verteilt ist [0.5 Pkt].

Die Stichprobenvarianz sei als $s^2 = 0,16$ gegeben. Ebenfalls nach Vorlesung findet man ein Konfidenzintervall für p als

$$K_b = \left(h_n - t_{(n-1; 1-\frac{\alpha}{2})} \sqrt{\frac{s^2}{n}}; h_n + t_{(n-1; 1-\frac{\alpha}{2})} \sqrt{\frac{s^2}{n}} \right) \quad [1 \text{ Pkt}]$$

Mit $t_{(n-1; 1-\frac{\alpha}{2})} = t_{(99; 0,975)} = 1,9842$ gilt

$$t_{(n-1; 1-\frac{\alpha}{2})} \sqrt{\frac{s^2}{n}} \approx 0,0794 \quad [0.5 \text{ Pkt}]$$

Somit gilt $K_b = (0,7206; 0,8794)$. [0.5 Pkt]

Abgabe der Hausübungen (Aufgaben H1 bis H3): Mittwoch, 12. Januar, 13:00 Uhr

Viel Erfolg! :)