Aufgabe 1

Die Zufallsvariablen $X_1, \cdots, X_n$ seien unabhängig und jeweils $\mathcal{N}(0;\theta)$-verteilt, dabei ist $\theta > 0$ unbekannt. Die Dichte von X_1 ist also gegeben durch

$$f(x,\theta) = \frac{1}{\sqrt{2\pi\theta}} \cdot \mathrm{e}^{-\frac{x^2}{2\theta}}, \quad x \in \mathbb{R}.$$

a) Bestimmen Sie den Maximum-Likelihood-Schätzer für den Parameter θ.

b) Welcher der beiden Schätzer

i. $S_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i^2$

ii. $T_n = \frac{1}{n-1} \cdot \sum_{i=1}^{n} X_i^2$

ist erwartungstreu, welcher ist asymptotisch erwartungstreu
(d.h. Betrachtung des Grenzwert vom Erwartungswert der Schätzfunktion)?

Lösung 1a

Likelihoodfunktion zur Realisierung $(x_1, \ldots, x_n)$:

$$L(\mu{=}0, \sigma^2{=}\theta; x_1, \ldots, x_n) = \frac{1}{\left(\sqrt{2\pi\theta}\right)^n} \exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right)$$

Log-Likelihoodfunktion (logarithmierte Likelihoodfunktion):

$$\begin{aligned} l(\mu{=}0, \sigma^2{=}\theta; x_1, \ldots, x_n) &= \ln L(\mu{=}0, \sigma^2{=}\theta; x_1, \ldots, x_n) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\theta - \frac{1}{2\theta}\sum_{i=1}^{n} x_i^2 \end{aligned}$$

Berechnung der ersten partiellen Ableitung:

$$\frac{\partial}{\partial\theta} l(0, \theta; x_1, \ldots, x_n) = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2$$

Berechnung der Nullstellen:

$$\begin{aligned} & -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 &= 0 \\ \Leftrightarrow \quad & \frac{1}{n}\sum_{i=1}^{n} x_i^2 &= \theta \end{aligned}$$

Der Punkt $(0,\theta) = (0, \frac{1}{n}\sum_{i=1}^{n} x_i^2)$ ist eine Maximalstelle, da die zweite partielle Ableitung der Likelihoodfunktion l'' nach θ

$$\frac{\partial^2}{\partial\theta\partial\theta} l(0,\theta;x_1,\ldots,x_n) = \frac{n}{2\theta^2} - \frac{2}{2\theta^3}\sum_{i=1}^{n} x_i^2$$

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

$$\begin{aligned}
& \frac{n}{2\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^2} - \frac{2}{2\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^3} \cdot \sum_{i=1}^{n} x_i^2 \\
= & \frac{n}{2\frac{1}{n^2}\left(\sum_{i=1}^{n} x_i^2\right)^2} - \frac{2}{2\frac{1}{n^3}\left(\sum_{i=1}^{n} x_i^2\right)^2} \\
= & \frac{n^3}{2\left(\sum_{i=1}^{n} x_i^2\right)^2} - \frac{2n^3}{2\left(\sum_{i=1}^{n} x_i^2\right)^2} \\
= & \frac{n^3 - 2n^3}{2\left(\sum_{i=1}^{n} x_i^2\right)^2} \\
= & \frac{-n^3}{2\left(\sum_{i=1}^{n} x_i^2\right)^2} < 0 \quad \checkmark
\end{aligned}$$

ist.
Die Maximum-Likelihood-Schätzung für den Parameter θ einer $\mathcal{N}(0;\theta)$-Verteilung lautet also

$$\hat{\theta}_{\text{ML}} = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

Lösung 1b

Ein Schätzer ist erwartungstreu, wenn $E(S_n) = \theta$.

$$\begin{aligned}
E(S_n) &= E\left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i^2\right) \\
&= \frac{1}{n}\cdot\sum_{i=1}^{n} E\left(X_i^2\right) \\
&= \frac{1}{n}\cdot n \cdot E\left(X^2\right) \\
&= \theta \quad \checkmark
\end{aligned}$$

Da $E(T_n) = \frac{n}{n-1}\theta$ prüfen wir auf asymptotische Erwartungstreue:

$$\begin{aligned}
\lim_{n\to\infty} E(T_n) &= \lim_{n\to\infty} \frac{n}{n-1}\theta \\
&= \lim_{n\to\infty} \frac{1}{1-\frac{1}{n}}\theta \\
&= \theta \quad \checkmark
\end{aligned}$$

Aufgabe 2

Die durchschnittliche Länge von Metallstiften soll geschätzt werden. Eine Stichprobe vom Umfang 36 liefert eine mittlere Länge von $\overline{X} = 38{,}5$ mm. Aus früheren Untersuchungen sei bekannt, dass die Länge der Metallstifte normalverteilt ist und die produzierende Maschine mit einer Standardabweichung von $\sigma = 1{,}8$ mm arbeitet.

a) Geben Sie ein Konfidenzintervall zum Niveau 0,95 für die erwartete Metallstiftlänge an.

b) Welchen Umfang muss eine Stichprobe haben, damit das Konfidenzintervall zum Niveau 0,95 für die mittlere Stiftlänge halb so breit ist, wie das unter a) berechnete?

Lösung 2

Sei $X = \{\text{Länge der Metallstifte}\} \sim \mathcal{N}(\mu;\ 1{,}8^2)$. Gesucht ist das Konfidenzintervall zum Niveau 0,95 bei bekannter Varianz und unbekanntem Mittelwert. Eine Stichprobe vom Umfang $n = 36$ mit dem empirischen Mittelwert $\overline{X} = 38{,}5$ ist gegeben.

$$\begin{aligned} & \overline{X} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\ \Leftrightarrow\ & \overline{X} - u_{0{,}975} \cdot \frac{1{,}8}{6} \leq \mu \leq \overline{X} + u_{0{,}975} \cdot \frac{1{,}8}{6} \\ \Leftrightarrow\ & 38{,}5 - 0{,}588 \leq \mu \leq 38{,}5 + 0{,}588 \\ \Leftrightarrow\ & 37{,}912 \leq \mu \leq 39{,}088 \end{aligned}$$

Damit das Konfidenzintervall zum Niveau 0,95 für die mittlere Stiftlänge halb so breit ist, muss $n \geq 144$ sein.

Aufgabe 3

Das Umweltreferat einer Großstadt will Aufschluss darüber gewinnen, wie viele Asbestfasern pro Kubikmeter Luft im Freien in ca. einem Meter Abstand von asbestzementhaltigen Gebäudeteilen zu erwarten sind. Bei $n = 14$ diesbezüglichen Messungen traten die Werte

980	1340	610	750	880	1250	2410
1100	470	1040	910	1860	730	820

auf, die als Ergebnisse unabhängiger normalverteilter Stichprobenvariablen angesehen werden.

a) Führen Sie für den Erwartungswert μ der Anzahl X der unter den obigen Bedingungen vorhandenen Asbestfasern eine Intervallschätzung zum Konfidenzniveau 0,95 durch.

b) Wie müsste das Konfindezniveau gewählt sein, damit die Länge des entstehenden Schätzintervalls gleich 500 ist?

Lösung 3

Um eine Intervallschätzung für den Erwartungswert μ der Anzahl X der Asbestfasern durchzuführen, berechnen wir zunächst den Mittelwert $\bar{x}$ und die Standardabweichung der Stichprobe. Der Mittelwert beträgt $\bar{x} \approx 1082{,}142857$ Fasern/m^3 und die Standardabweichung $s \approx 514{,}903704$ Fasern/m^3.
Sodann bestimmen wir den kritischen t-Wert für das gegebene Konfidenzniveau von $1 - \frac{\alpha}{2} = 0{,}975$ und $n - 1 = 13$ Freiheitsgraden und wenden die Formel für das Konfidenzintervall an:

$$\bar{x} \pm t_{\text{kritisch}} \cdot \left(\frac{s}{\sqrt{n}}\right)$$

Mit $t_{\text{kritisch}} \approx 2{,}1603687$ erhalten wir also ein Konfidenzintervall von 784,8463 bis 1379,4394 Fasern/m^3.
Das bedeutet, dass mit einer Sicherheit von 95% der wahre Mittelwert der Anzahl der Asbestfasern pro Kubikmeter Luft innerhalb dieses Intervalls liegt.
Um herauszufinden, wie das Konfidenzintervall gewählt werden muss, damit die Länge des Schätzintervalls genau 500 beträgt, suchen wir den t-Wert, der diese Gleichung erfüllt:

$$\begin{aligned}
& 2 \cdot t_{\text{benötigt}} \cdot \left(\frac{s}{\sqrt{n}}\right) &&= 500 \\
\Leftrightarrow \quad & t_{\text{benötigt}} &&= 250 \cdot \frac{\sqrt{n}}{s} \\
\Leftrightarrow \quad & t_{\text{benötigt}} &&= 250 \cdot \frac{\sqrt{14}}{514{,}903704} \\
\Leftrightarrow \quad & t_{\text{benötigt}} &&= 1{,}816678
\end{aligned}$$

Dies entspricht einem benötigten Kofidenzintervall $\approx 9{,}23848\%$.
Das bedeutet, dass, um ein Konfidenzintervall mit einer Länge von 500 zu erhalten, das Konfidenzniveau auf etwa 9,24% gesetzt werden muss.

Aufgabe 4

Bei der Anlieferung von Bauteilen mit einem Drehgewinde werden einige Teile zufällig ausgewählt und deren Gewindedurchmesser vermessen. Die Abweichung (in µm) von der untersten zulässigen Durchmessergrenze, das so genannte Spiel, werden wie folgt notiert:

$$1 \quad 2 \quad 3 \quad 4 \quad 4 \quad 4 \quad 6 \quad 7 \quad 9 \quad 10$$

Das Spiel eines Drehgewindes kann durch eine normalverteilte Zufallsgröße beschrieben werden.

a) Bestimmen Sie ein zweiseitiges Konfidenzintervall für die Varianz des Spiels zum 80% Niveau.

b) Bestimmen Sie ein einseitig nach oben begrenztes Konfidenzintervall für die Varianz des Spiels zum 95% Niveau.

Lösung 4

Sei das Spiel eines Drehgewindes normalverteilt. Wir berechnen zunächst die Stichprobenvarianz, die als Schätzer für die wahre Varianz μ der Grundgesamtheit dient.
Der Mittelwert $\bar{x}$ der Messwerte beträgt 5 µm². Die Stichprobenvarianz s^2 berechnet sich aus der Summe der quadrierten Abweichungen der Messwerte vom Mittelwert, geteilt durch $n-1$. Sie beträgt $s^2 = \frac{26}{3} \approx 8{,}67$ µm².

Für das zweiseitige Konfidenzintervall (80% Niveau) verwenden wir die Chi-Quadrat-Verteilung. Das Intervall wird durch die Formel

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}};\ \frac{(n-1)s^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right]$$

definiert, wobei $\alpha = 1 - 0{,}80 = 0{,}20$ und $n-1$ die Freiheitsgrade sind. Es liegt demnach zwischen 5,312028 und 18,713298 µm².

Für das einseitige Konfidenzintervall (95% Niveau), das nach oben begrenzt ist, nutzen wir ebenfalls die Chi-Quadrat-Verteilung. Das Intervall wird durch

$$\left[0;\ \frac{(n-1)s^2}{\chi^2_{n-1;\alpha}}\right]$$

definiert, wobei $\alpha = 1 - 0{,}95 = 0{,}05$. Es liegt demnach zwischen 0 und 23,457850 µm².

Diese Intervalle geben den Bereich an, in dem wir mit einer bestimmten Wahrscheinlichkeit erwarten können, dass die wahre Varianz des Spiels liegt.

Hinweis: Für die Berechnung der genauen Werte wurde Python und nicht die Tabelle verwendet, sodass es zu kleineren Abweichungen kommen kann.

Aufgabe 5

Messungen des systolischen Blutdrucks bei $n = 10$ Personen ergaben folgende Werte in mmHg:

$$124, 145, 112, 124, 136, 129, 125, 131, 142, 114$$

Unter der Annahme, dass der Blutdruck normalverteilt ist, bestimmen Sie jeweils zum Niveau 90%

a) ein einseitig nach oben begrenztes Konfidenzintervall für den Erwartungswert μ.

b) ein zweiseitiges Konfidenzintervall für die Standardabweichung σ.

Lösung 5

Da die Standardabweichung der Grundgesamtheit nicht bekannt ist und die Stichprobengröße $n = 10$ relativ klein ist, verwenden wir das t-Verteilungsbasierte Konfidenzintervall. Die Formel für ein einseitiges Konfidenzintervall ist:

$$\mu \leq \overline{x} + t_{n-1,\ \alpha} \cdot \frac{s}{\sqrt{n}}$$

Das arithmetisches Mittel der Stichprobe ist $\overline{x} = 128{,}2$ mmHg und die Standardabweichung der Stichprobe ist $s \approx 10{,}809$ mmHg. Der t-Wert aus der t-Verteilung für das Konfidenzniveau 90 % und $n - 1 = 9$ Freiheitsgraden ist $t_{9,\ 0{,}9} \approx 1{,}3830287$.
Wir erhalten somit das einseitige Konfidenzintervall $\mu \leq 132{,}92754$. Das bedeutet, wir können mit einer Sicherheit von 90% sagen, dass der wahre Erwartungswert des systolischen Blutdrucks maximal 132,93 mmHg beträgt.

Bei dem 90%-Konfidenzintervall für die Standardabweichung σ bestimmen wir die zwei Chi-Quadrat-Wert für $n - 1$ Freiheitsgrade. Zum einen $\chi^2_{n-1;\ \alpha/2}$ für das $\alpha/2$-Quantil, und $\chi^2_{n-1;\ 1-\alpha/2}$ für das obere $(1 - \alpha/2)$-Quantil.
Das Konfidenzintervall für σ wird berechnet, indem die Wurzel aus der skalierten Stichprobenvarianz genommen wird, wobei die Skalierung durch die Chi-Quadrat-Werte bestimmt wird. Das Intervall ist dann

$$\left[\sqrt{(n-1) \cdot \frac{s^2}{\chi^2_{n-1;\ \alpha/2}}};\ \sqrt{(n-1) \cdot \frac{s^2}{\chi^2_{n-1;\ 1-\alpha/2}}}\right]$$

also ungefähr

$$7{,}8838478 \leq \sigma \leq 17{,}7836988.$$