Wahrscheinlichkeit und Statistik

Serie 11

Version 2 (28. Mai 2024: kleine Änderungen), Version 1 (16. Mai)
Bitte stellt Fragen in den Übungen und/oder im Forum.
Bitte stell sicher, dass du die Webseite `https://kahoot.it/` in der Übung am **21. Mai** öffnen kannst.
Freiwillige Abgabe bis **23. Mai 8:00**. Nachher kann selbstständig mit der Lösung verglichen werden.

Aufgabe 11.1 [MLE III: Hochwasser im Zürichsee]
In Aufgabe 9.4 haben wir folgendes Modell betrachtet: Die Zufallsvariable X messe die Wasserhöhe in cm über der kritischen Marke von 140 cm über Normalniveau im Zürichsee. Zur Modellierung von X können wir eine sogenannte verallgemeinerte Pareto-Verteilung mit Dichte

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta}(1+x)^{-(1+\frac{1}{\theta})} & \text{falls } x > 0, \\ 0 & \text{falls } x \leq 0 \end{cases}$$

verwenden. Dabei ist $\theta > 0$ ein unbekannter Parameter, der auf Basis von Daten $x_1, \ldots, x_n > 0$ geschätzt werden soll; diese Daten werden wie üblich als Realisierungen von Zufallsvariablen $X_1, \ldots, X_n$ aufgefasst, die für jede Wahl des Parameters θ unter P_θ i.i.d. sind mit Dichte $f_X(x;\theta)$. Zeige, dass der Schätzer

$$T^{(n)} = \sum_{i=1}^{n} \frac{\log(1+X_i)}{n}$$

der Maximum-Likelihood-Schätzer für θ ist.

Aufgabe 11.2 [Mittlerer quadratischer Schätzfehler (MSE): Normalverteilung]
Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und je $\mathcal{N}(\mu, \sigma^2)$-verteilt unter P_θ, wobei $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ ein 2-dimensionaler unbekannter Parameter ist. Als Schätzer für σ^2 betrachte man

$$T^{(n)}(c) := c \sum_{i=1}^{n} (X_i - \overline{X}_n)^2,$$

wobei $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ und $c > 0$ ist.

(a) Für welches $c^* > 0$ wird der mittlere quadratische Schätzfehler $\mathbb{E}_\theta\left[\left(T^{(n)}(c^*) - \sigma^2\right)^2\right]$ minimiert?
Hinweis: Verwende, dass $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$-verteilt ist unter P_θ.

(b) Entspricht der in (a) gefundene Schätzer $T^{(n)}(c^*)$ dem Maximum-Likelihood-Schätzer?

Aufgabe 11.3 [Konfidenzintervall für $\mathcal{N}(\mu, \sigma^2)$ mit bekannter Standardabweichung]
In einem Sägewerk wird das Schnittholz einer bestimmten Sortierklasse einer Qualitätskontrolle unterzogen. An jedem Produktionstag wird ein Brett auf seine Steifigkeit getestet. $X_n, n \geq 1$, bezeichne das Ergebnis am n-ten Tag. Basierend auf Erfahrung kann angenommen werden, dass die Steifigkeit eines Brettes normalverteilt ist mit bekannter Standardabweichung $\sigma = 1430$ MPa[1]. Wir wählen also den Parameterraum $\Theta = \mathbb{R}$ und die Modellfamilie $(\mathrm{P}_\theta)_{\theta \in \Theta}$, wobei die $(X_n)_{n \geq 1}$ unter P_θ unabhängig und $\mathcal{N}(\theta, \sigma^2)$-verteilt sind.

[1] Das Pascal ist eine abgeleitete SI-Einheit des Drucks sowie der mechanischen Spannung. Sie ist nach Blaise Pascal benannt und folgendermassen definiert: 1 Pa = $1\text{kg}\times\text{m}^{-1}\times\text{s}^{-2} = 1\text{N}\times\text{m}^{-2}$. Ein Pascal ist also der Druck, den eine Kraft von einem Newton auf eine Fläche von einem Quadratmeter ausübt. 1 MPa = 10^6 Pa = 10 bar.

(a) Leite die Formel eines Konfidenzintervalls für θ mit Niveau 95% nach 150 Produktionstagen her.

(b) Berechne aus (a) das realisierte Konfidenzintervall für einen beobachteten Stichprobenmittelwert von $11'000$ MPa (nach 150 Produktionstagen).

(c) Wieviele Stichproben wären nötig, damit die Breite des Konfidenzintervalls kleiner als 200 MPa ist?

Aufgabe 11.4 [t-Verteilung mit m Freiheitsgraden]
Seien X und Y unabhängige Zufallsvariablen mit $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi^2_m$. Ziel dieser Aufgabe ist es zu zeigen, dass der Quotient

$$Z := \frac{X}{\sqrt{\frac{1}{m}Y}}$$

t-verteilt ist mit m Freiheitsgraden.

(a) Bestimme die Dichte von $Y' := \sqrt{\frac{1}{m}Y}$.

(b) Seien $\hat{X}$ und $\hat{Y}$ zwei unabhängige Zufallsvariablen mit Dichte $f_{\hat{X}}$ resp. $f_{\hat{Y}}$. Definiere $\hat{Z} := \frac{\hat{X}}{\hat{Y}}$ und bestimme die Dichte von $\hat{Z}$ (in Abhängigkeit von $f_{\hat{X}}$ und $f_{\hat{Y}}$).

Hinweis: Benutze in (a) und (b) die Charakterisierung der Dichte aus Satz 4.16 aus dem Skript.

(c) Zeige, dass Z t-verteilt ist mit m Freiheitsgraden.

Aufgabe 11.5 [Konfidenzintervall für $\mathcal{N}(\mu, \sigma^2)$ mit unbekannter Standardabweichung]
Wir betrachten erneut die Situation aus Aufgabe 11.3 und nehmen weiterhin an, dass die Steifigkeit eines Brettes normalverteilt ist. In diesem Fall gehen wir aber davon aus, dass die Standardabweichung σ unbekannt ist. Der unbekannte Parameter $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ ist in diesem Fall also 2-dimensional und wir wählen die Modellfamilie $(\mathrm{P}_\theta)_{\theta \in \Theta}$, wobei die $(X_n)_{n \geq 1}$ unter P_θ unabhängig und $\mathcal{N}(\mu, \sigma^2)$-verteilt sind. Zur Erinnerung: Aus der Vorlesung kennen wir bereits zwei Schätzer

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{und} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

für μ und σ^2.

(a) Zeige für jedes $\theta = (\mu, \sigma^2) \in \Theta$, dass die Zufallsvariable

$$\frac{\bar{X}_n - \mu}{\sqrt{S^2/n}}$$

t_{n-1}-verteilt ist unter P_θ.
Hinweis: Verwende $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$, $\frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$, die Unabhängigkeit von $\bar{X}_n$ und S^2 sowie das Resultat aus der vorherigen Aufgabe.

(b) Leite die Formel eines Konfidenzintervalls für μ mit Niveau 95% nach 10 Produktionstagen her.

Wenn du Feedback zur Serie hast, schreibe bitte in das Forum (oder eine Mail an Jakob Heiss).