

Wahrscheinlichkeit & Statistik

Serie 10

Abgabe bis Mittwoch (11.05.2022) um 10:15 Uhr

Diese Serie beschäftigt sich mit der Maximum-Likelihood-Methode.

Weitere Informationen und Instruktionen zur Abgabe unter
<https://metaphor.ethz.ch/x/2022/fs/401-0614-00L/>

Aufgabe 10.1 [MLE I: Stetige Verteilung]

Wir betrachten eine stetige Verteilung mit Dichte

$$f(x) = \begin{cases} \frac{\theta}{x^{\theta+1}} & x \geq 1, \\ 0 & x < 1, \end{cases}$$

wobei $\theta > 0$ ein unbekannter Parameter ist. Wir wollen den Parameter θ mit Hilfe eines Datensatzes schätzen.

- Sei X_1, \dots, X_n eine Stichprobe von unabhängigen Zufallsvariablen, welche alle die Dichte f besitzen. Bestimme die Likelihood- und log-Likelihood-Funktion.
- Bestimme den zugehörigen Maximum-Likelihood-Schätzer für θ . Schreibe zuerst die allgemeine Formel für n Beobachtungen hin und berechne dann den Schätzwert für die folgende konkrete Stichprobe:

x_1	x_2	x_3	x_4	x_5
12.0	4.0	6.9	27.9	15.4

Aufgabe 10.2 [MLE II: Hochwasser im Zürichsee]

In Aufgabe 2 von Serie 9 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, \dots, x_n geschätzt werden soll; diese Daten werden wie üblich als Realisierungen von Zufallsvariablen X_1, \dots, X_n aufgefasst, die für jede Wahl des Parameters θ unter \mathbb{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 10.3 [Mittlerer quadratischer Schätzfehler: Normalverteilung]

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und je $\mathcal{N}(\mu, \sigma^2)$ -verteilt unter \mathbb{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

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

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

- (a) Für welches $c^* > 0$ wird der mittlere quadratische Schätzfehler $\mathbb{E}_\theta [(T^{(n)}(c^*) - \sigma^2)^2]$ minimiert?

Hinweis: Verwende, dass $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2$ -verteilt ist unter \mathbb{P}_θ .

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

Aufgabe 10.4 [MLE III: Exponentialverteilung]

Eine Tankstelle veranschlagt für einen Ölwechsel mindestens $\theta_1 > 0$ Minuten. Die tatsächlich benötigte Zeit X variiert natürlich im Bereich $X \geq \theta_1$ und ist je Kunde/-in verschieden. Man kann jedoch annehmen, dass die zusätzliche Zeit durch eine exponentialverteilte Zufallsvariable gut modelliert wird.

- (a) Sei $\theta_2 > 0$ und Z eine $\text{Exp}(\theta_2)$ -verteilte Zufallsvariable, Zeige, dass die Zufallsvariable $X = \theta_1 + Z$ die folgende Dichte hat:

$$f_{\theta_1, \theta_2}(x) = \begin{cases} \theta_2 e^{\theta_1 \theta_2 - \theta_2 x} & \text{falls } x \geq \theta_1, \\ 0 & \text{sonst.} \end{cases}$$

Wir betrachten nun den Parameterraum $\Theta = \mathbb{R}_+ \times \mathbb{R}_+$ mit $\theta = (\theta_1, \theta_2)$ und die Modellfamilie $(\mathbb{P}_\theta)_{\theta \in \Theta}$, wobei die Zufallsvariablen X_1, \dots, X_n unter \mathbb{P}_θ unabhängig und identisch verteilt sind mit Dichte f_{θ_1, θ_2} .

- (b) Bestimme den Maximum-Likelihood-Schätzer T_{ML} für $\theta = (\theta_1, \theta_2)$.
- (c) Die benötigte Arbeitszeit in Minuten wurde für 10 zufällig ausgewählte Kunden/-innen notiert:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
4.2	3.1	3.6	4.5	5.1	7.6	4.4	3.5	3.8	4.3

Der Stichprobenumfang ist hier also $n = 10$. Welcher Schätzwert ergibt sich basierend auf dem Maximum-Likelihood-Schätzer aus (b)?