Prof. Josef Teichmann
David Martins

# Wahrscheinlichkeit & Statistik

## Serie 11

---

Die Übungen mit (*) markiert sind fakultativ.

**Aufgabe 11.1 [Schätzer I: gleichverteilte Zufallsvariablen]**
Seien $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen mit $X_i \sim \mathcal{U}([\theta - 1, \theta])$ unter $\mathbb{P}_\theta$, wobei $\theta \in \mathbb{R}$ ein unbekannter Parameter ist. Für $\theta$ bieten sich

$$T_1^{(n)} = \frac{1}{n} \sum_{i=1}^{n} (X_i + 1/2) \quad \text{und} \quad T_2^{(n)} = \max(X_1, \ldots, X_n)$$

als Schätzer an. Wir untersuchen in dieser Aufgabe, welche Eigenschaften diese beiden Schätzer besitzen.

(a) Untersuche, ob die Schätzer erwartungstreu sind.

(b) Berechne die Varianzen der Schätzer $\mathrm{Var}_\theta[T_1^{(n)}]$ und $\mathrm{Var}_\theta[T_2^{(n)}]$.

(c) Berechne die mittleren quadratischen Schätzfehler

$$\mathrm{MSE}_\theta[T_i^{(n)}] := E_\theta[(T_i^{(n)} - \theta)^2)]$$

für $i = 1, 2$.

**Lösung 11.1**

(a) Sei $\theta \in \mathbb{R}$ fixiert. Aus der Gleichverteilung folgt, dass

$$\mathrm{E}_\theta[T_1^{(n)}] = \left( \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_\theta[X_i] \right) + \frac{1}{2} = \mathrm{E}_\theta[X_1] + \frac{1}{2} = \theta - \frac{1}{2} + \frac{1}{2} = \theta.$$

Somit folgt, dass der Schätzer $T_1^{(n)}$ erwartungstreu ist.

Um die Rechnung für den anderen Schätzer zu vereinfachen, führen wir die Zufallsvariablen $Y_i = X_i - (\theta - 1)$ ein. Unter $\mathbb{P}_\theta$ sind $Y_1, \ldots, Y_n$ unabhängig und $\mathcal{U}([0,1])$-verteilt. Insbesondere ist

$$Y^{(n)} := \max\{Y_1, \ldots, Y_n\} = \max\{X_1, \ldots, X_n\} - (\theta - 1) = T_2^{(n)} - (\theta - 1).$$

Um den Erwartungswert von $Y^{(n)}$ unter $\mathbb{P}_\theta$ zu bestimmen, berechnen wir zunächst die Verteilungsfunktion und die Dichte von $Y^{(n)}$. Es gilt

$$\begin{aligned} F_{Y^{(n)}}(a) &= \mathbb{P}_\theta[Y^{(n)} \leq a] = \mathbb{P}_\theta[Y_1 \leq a, \ldots, Y_n \leq a] = \prod_{i=1}^{n} \mathbb{P}_\theta[Y_i \leq a] = \mathbb{P}_\theta[Y_1 \leq a]^n \\ &= \begin{cases} 0, & \text{für } a < 0, \\ a^n, & \text{für } a \in [0,1], \\ 1, & \text{für } a > 1, \end{cases} \end{aligned}$$

und somit

$$f_{Y^{(n)}}(a) = na^{n-1}\mathbb{1}_{a\in[0,1]}.$$

Es folgt, dass

$$\begin{aligned}\mathrm{E}_\theta[Y^{(n)}] &= \int_0^1 a \cdot na^{n-1}da = n\int_0^1 a^n da \\ &= n\left[\frac{1}{n+1}a^{n+1}\right]_{a=0}^{a=1} = \frac{n}{n+1} = 1 - \frac{1}{n+1}.\end{aligned}$$

Also ist

$$\mathrm{E}_\theta[T_2^{(n)}] = \mathrm{E}_\theta[Y^{(n)}] + (\theta-1) = 1 - \frac{1}{n+1} + (\theta-1) = \theta - \frac{1}{n+1}.$$

Der Schätzer $T_2^{(n)}$ ist somit nicht erwartungstreu.

(b) Wir erinnern zuerst daran, dass $\mathrm{Var}[Z+b] = \mathrm{Var}[Z]$ für eine beliebige Zufallsvariable $Z$ und $b \in \mathbb{R}$ gilt. Also folgt wegen der Unabhängigkeit von $X_1, \ldots, X_n$ und $Y_1 \sim \mathcal{U}(0,1)$, dass

$$\mathrm{Var}_\theta[T_1^{(n)}] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_\theta[X_i - (\theta-1)] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_\theta[Y_i] = \frac{1}{n}\mathrm{Var}_\theta[Y_1] = \frac{1}{12n}.$$

Um die Varianz des anderen Schätzers zu bestimmen, berechnen wir zuerst $\mathrm{Var}_\theta[Y^{(n)}]$. Auf ähnliche Weise wie in a) erhalten wir

$$\begin{aligned}\mathrm{E}_\theta[(Y(n))^2] &= n\int_0^1 a^2 a^{n-1}da \\ &= n\int_0^1 a^{n+1}da \\ &= n\left[\frac{1}{n+2}a^{n+2}\right]_{a=0}^{a=1} \\ &= \frac{n}{n+2}.\end{aligned}$$

Also ist

$$\begin{aligned}\mathrm{Var}_\theta[T_2^{(n)}] &= \mathrm{Var}_\theta[Y^{(n)} + (\theta-1)] = \mathrm{Var}_\theta[Y^{(n)}] \\ &= E_\theta[(Y(n))^2] - (E_\theta[Y^{(n)}])^2 = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{(n+1)^2(n+2)}.\end{aligned}$$

(c) Aus den Teilaufgaben (a) und (b) erhalten wir den mittleren quadratischen Schätzfehler der beiden Schätzer:

$$\begin{aligned}\mathrm{MSE}_\theta[T_1^{(n)}] &= \mathrm{Var}_\theta[T_1^{(n)} - \theta] + (\mathrm{E}_\theta[T_1^{(n)}] - \theta)^2 \\ &= \frac{1}{12n} + 0 = \frac{1}{12n}\end{aligned}$$

$$\begin{aligned}\mathrm{MSE}_\theta[T_2^{(n)}] &= \mathrm{Var}_\theta[T_2^{(n)} - \theta] + (\mathrm{E}_\theta[T_2^{(n)}] - \theta)^2 \\ &= \frac{n}{(n+1)^2(n+2)} + \left(\theta - \frac{1}{n+1} - \theta\right)^2 = \frac{2}{(n+1)(n+2)}\end{aligned}$$

**Aufgabe 11.2 [Schätzer II: Hochwasser im Zürichsee]**
Wir betrachten Pegelstände bei Hochwasser im Zürichsee. Hochwasser bedeute dabei, dass der Pegelstand die kritische Marke von 140 cm über Normalniveau überschreitet. Die Zufallsvariable $X$ messe die Wasserhöhe in cm über der kritischen Marke. 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 $\theta$ unter $\mathbb{P}_\theta$ u.i.v. sind mit Dichte $f_X(x;\theta)$. Als Schätzer für $\theta$ verwenden wir

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

(a) Berechne den Erwartungswert und die Varianz von $T^{(n)}$ im Modell $\mathbb{P}_\theta$ für jedes $\theta > 0$.
*Hinweis:* Benutze, dass $Y_i := \log(1+X_i) \sim \operatorname{Exp}(\frac{1}{\theta})$ ist, d.h. $Y_i$ hat unter $\mathbb{P}_\theta$ die Dichte $f_Y(y) = \frac{1}{\theta} e^{-\frac{1}{\theta}y} \mathbb{1}_{\{y \geq 0\}}$.

(b) Ist $T^{(n)}$ ein erwartungstreuer Schätzer für $\theta$?

(c) Berechne den mittleren quadratischen Schätzfehler

$$\operatorname{MSE}_\theta[T^{(n)}] := E_\theta[(T^{(n)} - \theta)^2)] = E_\theta[T^{(n)} - \theta]^2 + \operatorname{Var}_\theta[T^{(n)}]$$

im Modell $\mathbb{P}_\theta$ für jedes $\theta > 0$.

(d) Zeige, dass $T^{(n)}$ der Maximum-Likelihood-Schätzer für $\theta$ ist.

**Lösung 11.2**

(a) Die Linearität des Erwartungswertes liefert

$$\mathrm{E}_\theta[T^{(n)}] = \mathrm{E}_\theta\left[\frac{1}{n}\sum_{i=1}^{n}\log(1+X_i)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_\theta[\log(1+X_i)] = \mathrm{E}_\theta[\log(1+X_1)].$$

Nach dem Hinweis ist $Y_1 := \log(1+X_1) \sim \operatorname{Exp}(\frac{1}{\theta})$. Also ist $\mathrm{E}_\theta[Y_1] = \theta$ und $\operatorname{Var}_\theta[Y_1] = \theta^2$. Daher gilt $\mathrm{E}_\theta[T^{(n)}] = \mathrm{E}_\theta[Y_1] = \theta$.

Die Varianz rechnen wir auf eine ähnliche Weise aus:

$$\begin{aligned} \operatorname{Var}_\theta[T^{(n)}] &= \operatorname{Var}_\theta\left[\frac{1}{n}\sum_{i=1}^{n}\log(1+X_i)\right] = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}_\theta[\log(1+X_i)] \\ &= \frac{1}{n}\operatorname{Var}_\theta[\log(1+X_1)] = \frac{1}{n}\operatorname{Var}_\theta[Y_1] = \frac{1}{n}\theta^2. \end{aligned}$$

(b) Aus (a) folgt, dass $\mathrm{E}_\theta[T^{(n)}] = \theta$, d.h. $T^{(n)}$ ist erwartungstreu für $\theta$.

(c) Da der Schätzer erwartungstreu ist (siehe (b)), erhalten wir

$$\operatorname{MSE}_\theta[T^{(n)}] = \operatorname{Var}_\theta[T^{(n)}] = \frac{\theta^2}{n}.$$

(d) Die log-Likelihood-Funktion ist gegeben durch

$$\log L(x_1, \ldots, x_n; \theta) = \log\left(\frac{1}{\theta^n} \prod_{i=1}^{n} (1 + x_i)^{-(1+\frac{1}{\theta})}\right)$$
$$= -n \log\theta - (1 + \frac{1}{\theta}) \sum_{i=1}^{n} \log(1 + x_i).$$

Ableiten nach $\theta$ und Nullsetzen ergibt

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

für

$$\theta^* = \frac{1}{n} \sum_{i=1}^{n} \log(1 + x_i).$$

Die zweite Ableitung ist

$$\frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^{n} \log(1 + x_i),$$

also strikt kleiner als 0 an der Stelle $\theta^*$ und somit handelt es sich also um das Maximum. Der Maximum-Likelihood-Schätzer für $\theta$ ist also

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

**Aufgabe 11.3 [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 $\theta$ mit Hilfe eines Datensatzes schätzen.

(a) Sei $X_1, \ldots, X_n$ eine Stichprobe von unabhängigen Zufallsvariablen, welche alle die Dichte $f$ besitzen. Bestimme die Likelihood- und log-Likelihood-Funktion.

(b) Bestimme den zugehörigen Maximum-Likelihood-Schätzer für $\theta$. 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 |

**Lösung 11.3**

(a) Die Likelihood-Funktion ergibt sich aus dem Produkt der Dichten. Für $x_1, \ldots, x_n \geq 1$ erhalten wir

$$L(x_1, \ldots x_n; \theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{\theta}{x_i^{\theta+1}} = \theta^n \frac{1}{(\prod_{i=1}^{n} x_i)^{\theta+1}}\ .$$

Falls $x_i < 1$ für ein $1 \leq i \leq n$, ist die Likelihood-Funktion gleich 0, wir können uns also auf den Fall $x_1, \ldots, x_n \geq 1$ beschränken.

Die log-Likelihood-Funktion erhalten wir durch Logarithmieren obiger Formel als

$$\ell(x_1, \ldots x_n; \theta) = \log L(x_1, \ldots x_n; \theta) = n \log \theta - (\theta + 1) \sum_{i=1}^{n} \log x_i.$$

(b) Ableiten und Nullsetzen der log-Likelihood-Funktion ergibt

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

für $\theta^* = \frac{n}{\sum_{i=1}^{n} \log x_i}$. Für $\theta < \theta^*$ ist die Ableitug strikt positiv, für $\theta > \theta^*$ strikt negativ, es handelt sich also um das Maximum. Also ist der Maximum-Likelihood-Schätzer

$$T_{ML} = \frac{n}{\sum_{i=1}^{n} \log X_i} = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} \log X_i}.$$

Der realisierte Schätzwert für die gegebenen Daten ist dann

$$t_{ML} = \frac{5}{\sum_{i=1}^{5} \log x_i} = 0.4214.$$

*Hinweis: Wir verwenden den natürlichen Logarithmus.*