

Übungsblatt 10

Aufgabe P1 *Taxiproblem.*

In einer großen Stadt gibt es $N \in \mathbb{N}$ Taxis, die (von außen gut lesbar) die Nummern $1, \dots, N$ tragen. Ein Passant steht an einer viel befahrenen Straße und notiert sich alle Nummern der Taxis, die er sieht. Anschließend ordnet er die Taxinummern aufsteigend und löscht Wiederholungen, sodass er die n Beobachtungen x_1, \dots, x_n vorliegen hat mit $x_1 < x_2 < \dots < x_n$. Nun möchte er die Gesamtzahl N der Taxis schätzen.

In der Vorlesung wurde diese Situation als „Taxiproblem“ bezeichnet und Sie haben bereits gesehen, dass der Maximum-Likelihood-Schätzer

$$N_0(x_1, \dots, x_n) = x_n$$

ist. Also schlicht der größte beobachtete Wert.

Man könnte sich auch überlegen, dass es aus Symmetriegründen sinnvoll sein könnte anzunehmen, dass der größte beobachtete Wert von N genau so weit entfernt sein könnte, wie der kleinste beobachtete Wert von 1. Dies liefert den Schätzer

$$N_1(x_1, \dots, x_n) = x_n + x_1 - 1.$$

Eine etwas verfeinerte Variante dieser Strategie wäre es, die durchschnittliche Differenz von $x_j - j$ und $x_{j-1} - (j-1)$, $j = 1, \dots, n$, mit $x_0 = 0$, zu ermitteln und diese zur größten Beobachtung zu addieren. Diese Idee liefert den Schätzer

$$N_2(x_1, \dots, x_n) = x_n + \frac{x_n - n}{n}.$$

Untersuchen Sie, welche dieser drei Schätzer Erwartungstreu sind.

Hinweis: Sie dürfen ohne Beweis verwenden, dass $\mathbb{E}[x_i] = i \cdot \frac{N+1}{n+1}$ für $i \in \{1, n\}$.

Lösung: Ein Schätzer heißt erwartungstreu, falls der Erwartungswert des Schätzers der zu schätzende Parameter ist. Also in diesem Fall, falls $\mathbb{E}[N_i(x_1, \dots, x_n)] = N$ gilt. Mit dem Hinweis rechnet man folgendes:

$$\mathbb{E}[N_0(x_1, \dots, x_n)] = \mathbb{E}[x_n] = n \cdot \frac{N+1}{n+1} \neq N.$$

$$\mathbb{E}[N_1(x_1, \dots, x_n)] = \mathbb{E}[x_n + x_1 - 1] = n \cdot \frac{N+1}{n+1} + 1 \cdot \frac{N+1}{n+1} - 1 = (n+1) \cdot \frac{N+1}{n+1} - 1 = N.$$

$$\mathbb{E}[N_2(x_1, \dots, x_n)] = \mathbb{E}\left[x_n + \frac{x_n - n}{n}\right] = \mathbb{E}[x_n] + \frac{1}{n} \cdot \mathbb{E}[x_n] - 1 = \frac{n+1}{n} \cdot \mathbb{E}[x_n] - 1 = N.$$

Also ist der Maximum-Likelihood-Schätzer als einziger von den dreien nicht erwartungstreu.

Aufgabe P2 *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 H1 Maximum gleichverteilter ZVen.

Ein Zufallsgenerator erzeugt gleichverteilt Zahlen im Bereich $[0, a]$, wobei a unbekannt ist. Seien X_1, \dots, X_n unabhängig voneinander erzeugte Zahlen und sei T_n deren Maximum.

Berechnen Sie die Dichtefunktion von T_n . (Hinweis: Berechnen Sie zuerst die Verteilungsfunktion.) Zeigen Sie, dass T_n kein erwartungstreuer Schätzer ist, indem Sie den Erwartungswert von T_n unter Abhängigkeit von $a > 0$ explizit bestimmen.

Lösung: a) Für die Verteilungsfunktion gilt ($0 < t < a$):

$$\begin{aligned} F_{T_n}(t) &= \mathbb{P}(\max\{X_1, \dots, X_n\} \leq t) = \mathbb{P}(X_1 \leq t, \dots, X_n \leq t) = \mathbb{P}(X_1 \leq t) \cdots \mathbb{P}(X_n \leq t) \\ &= (F_{X_1}(t))^n = \left(\frac{t}{a}\right)^n. \quad [2 \text{ Pkt}] \end{aligned}$$

Damit ist die Dichte $f_{T_n} : \mathbb{R} \rightarrow \mathbb{R}_+$,

$$f_{T_n}(t) = \begin{cases} F'_{T_n}(t) = n \cdot \frac{t^{n-1}}{a^n}, & \text{falls } t \in (0, a) \\ 0, & \text{sonst.} \end{cases} \quad [1 \text{ Pkt}]$$

Da

$$\mathbb{E}_a[T_n] = \int_0^a t f_{T_n}(t) dt = n \int_0^a \frac{t^n}{a^n} dt = n \left[\frac{1}{n+1} \frac{t^{n+1}}{a^n} \right]_0^a = \frac{n}{n+1} \frac{a^{n+1}}{a^n} = \frac{n}{n+1} a \quad [3 \text{ Pkt}]$$

ist T_n kein erwartungstreuer Schätzer.

Aufgabe H2 Schätzer.

- a) Seien X_1, \dots, X_n unabhängige mathematische Stichproben vom Umfang n der Grundgesamtheit X . Zeigen Sie, dass $Y := \sum_{i=1}^n \lambda_i X_i$ für beliebige Werte $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ mit $\lambda_1 + \dots + \lambda_n = 1$ einen erwartungstreuen Schätzer für $\mathbb{E}[X]$ darstellt.
- b) Zeigen Sie, dass unter den möglichen Schätzern Y derjenige mit $\lambda_1 = \dots = \lambda_n = 1/n$, also das Stichprobenmittel, höchstmögliche Effizienz besitzt. Verwenden Sie ohne Beweis, dass $\sum_{i=1}^n \lambda_i^2 \geq 1/n$ gilt.

Lösung: a) Wegen der Linearität des Erwartungswerts gilt

$$\mathbb{E}[Y] = (\lambda_1 + \dots + \lambda_n) \cdot \mathbb{E}[X] = \mathbb{E}[X].$$

b) Da Y erwartungstreu ist, ist der mittlere quadratische Fehler die Varianz $\text{Var}(Y)$. Für diese gilt wegen der Unabhängigkeit der X_i

$$\text{Var}(Y) = \sum_{i=1}^n \text{Var}(\lambda_i X_i) = \sum_{i=1}^n \lambda_i^2 \text{Var}(X_i) \geq \frac{\text{Var}(X)}{n}.$$

Für $\lambda_1 = \dots = \lambda_n = 1/n$ gilt Gleichheit und folglich minimiert diese Wahl der λ_i die Varianz von Y .

Aufgabe H3 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.

- a) Bestimmen Sie für unabhängige Stichproben x_1, \dots, x_n den Maximum-Likelihood-Schätzwert für θ .

b) 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: 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

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

mit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Man sieht leicht, dass es sich um ein Maximum handelt. [1 Pkt]

b) Für diese gegebenen Werte gilt

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

Entsprechend gilt nach a)

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

Abgabe der Hausübungen (Aufgaben H1 bis H3): Mittwoch, 18. Dezember, 16:00 Uhr

Viel Erfolg! :)