Prof. Vincent Tassion
Laurin Köhler-Schindler

Wahrscheinlichkeit & Statistik

Serie 9

Abgabe bis Mittwoch (04.05.2022) um 10:15 Uhr

Diese Serie beschäftigt sich mit dem Konzept von Schätzern und mit der χ^2-Verteilung.

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

Aufgabe 9.1 [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 θ 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 $\mathrm{MSE}_\theta[T_1^{(n)}]$ und $\mathrm{MSE}_\theta[T_2^{(n)}]$.

Aufgabe 9.2 [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 $x_1, \ldots, x_n$ 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 $\mathbb{P}_\theta$ u.i.v. sind mit Dichte $f_X(x;\theta)$. Als Schätzer für θ 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 $Y_i := \log(1+X_i) \sim \mathrm{Exp}(\frac{1}{\theta})$ 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 θ?

(c) Berechne den mittleren quadratischen Schätzfehler $\mathrm{MSE}_\theta[T^{(n)}]$ im Modell $\mathbb{P}_\theta$ für jedes $\theta > 0$.

Aufgabe 9.3 [χ^2**-Verteilung**]
Sei Y eine χ^2_n-verteilte Zufallsvariable mit $n \in \mathbb{N}$, das heisst

$$Y = \sum_{i=1}^{n} X_i^2,$$

wobei $X_1, \ldots, X_n$ unabhängig und $\mathcal{N}(0,1)$-verteilt sind.

(a) Zeige, dass
$$\mathbb{E}[Y] = n \quad \text{und} \quad \sigma_Y^2 = 2n$$
gilt.

(b) Gebe mit Hilfe der Chebyshev-Ungleichung eine untere Schranke für die Wahrscheinlichkeit
$$\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \leq 0.75\right].$$
Berechne die Schranke für $n = 12$.

(c) Berechne für $n = 12$ eine Annäherung für die obige Wahrscheinlichkeit mit Hilfe des zentralen Grenzwertsatzes.

Lösung 9.1

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

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

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 $Y_1, \dots, Y_n$ unabhängig und $\mathcal{U}([0,1])$-verteilt. Insbesondere ist

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

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

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

und somit

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

Es folgt, dass

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

Also ist

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

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 $X_1, \dots, X_n$ 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

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

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:

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

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

Lösung 9.2

(a) Die Linearität des Erwartungswertes liefert

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

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

Die Varianz rechnen wir auf eine ähnliche Weise aus:

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

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

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

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

Lösung 9.3

(a) Aus der Definition von Y folgt

$$\mathbb{E}[Y] = \sum_{i=1}^n \mathbb{E}[X_i^2] = \sum_{i=1}^n 1 = n.$$

Andererseits ist wegen Unabhängigkeit und identischer Verteilung

$$\mathbb{E}[Y^2] = \mathbb{E}[(X_1^2 + X_2^2 + \cdots + X_n^2)^2] = n\mathbb{E}[X_1^4] + n(n-1)(\mathbb{E}[X_1^2])^2.$$

Mit $\mathbb{E}[X_i^2] = 1$ und

$$\begin{aligned}\mathbb{E}[X_i^4] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-x^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \left[-x^3 e^{-x^2/2}\right]_{-\infty}^{\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx = 3\end{aligned}$$

erhalten wir schliesslich $\mathbb{E}[Y^2] = 3n + n(n-1) = n^2 + 2n$, und daraus

$$\mathrm{Var}[Y] = (n^2 + 2n) - n^2 = 2n.$$

Alternativ: Mit $X_1, \ldots, X_n$ sind auch $X_1^2, \ldots, X_n^2$ unabhängig und damit ist

$$\mathrm{Var}[Y] = \sum_{i=1}^{n} \mathrm{Var}[X_i^2] = n\mathrm{Var}[X_1^2].$$

Wie oben ist

$$\mathrm{Var}[X_1^2] = \mathbb{E}[X_1^4] - (\mathbb{E}[X_1^2])^2 = 3 - 1 = 2,$$

also $\mathrm{Var}[Y] = 2n$.

(b) Mit der Chebyshev-Ungleichung erhalten wir

$$\begin{aligned}\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \leq 0.75\right] &= \mathbb{P}\left[\left|\frac{Y-n}{n}\right| \leq \frac{3}{4}\right] = 1 - \mathbb{P}\left[|Y - \mathbb{E}[Y]| > \frac{3n}{4}\right] \\ &\geq 1 - \frac{\mathrm{Var}[Y]}{9n^2/16} = 1 - \frac{32}{9n} \quad \left(= \frac{19}{27} \approx 0.7037 \text{ für } n = 12\right).\end{aligned}$$

(c) Seien $Z_i := X_i^2$ für $i = 1, \ldots, n$. Dann gilt $Z_i \sim \chi_1^2$, also $\mathbb{E}[Z_i] = 1$ und $\mathrm{Var}[Z_i] = 2$. Mit dem zentralen Grenzwertsatz folgt, dass

$$\begin{aligned}\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \leq 0.75\right] &= \mathbb{P}\left[\left|\frac{Y-n}{n}\right| \leq \frac{3}{4}\right] \\ &= \mathbb{P}\left[\left|\frac{Y-n}{\sqrt{2n}}\right| \leq \frac{3}{4}\sqrt{\frac{n}{2}}\right] \\ &\approx \Phi\left(\frac{3}{4}\sqrt{\frac{n}{2}}\right) - \Phi\left(-\frac{3}{4}\sqrt{\frac{n}{2}}\right) \\ &= 2\Phi\left(\frac{3}{4}\sqrt{\frac{n}{2}}\right) - 1.\end{aligned}$$

Wir schreiben hier $\approx$ für die Annäherung mit Hilfe des zentralen Grenzwertsatzes, die für $n \to \infty$ gilt. Für $n = 12$ ist

$$\mathbb{P}\left[\left|\frac{Y}{n} - 1\right| \leq 0.75\right] \approx 2\Phi\left(\frac{3}{4}\sqrt{6}\right) - 1 = 2\Phi(1.837) - 1 \approx 0.9338.$$