

Übungsblatt 9

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 wird diese Situation als „Taxiproblem“ bezeichnet. Sie werden sehen, dass ein möglicher 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 Erwartungstreue sind.

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

Aufgabe P2 *Acceptance-Rejection-Algorithmus.*

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $B \in \mathcal{A}$ ein Ereignis mit $\mathbb{P}(B) > 0$. Wir betrachten folgenden Algorithmus:

1. Generiere $x \sim \mathbb{P}$ gemäß Verteilung \mathbb{P} , d.h. $\mathbb{P}(x \in A) = \mathbb{P}(A)$ für alle $A \in \mathcal{A}$.
2. Falls $x \in B$, so akzeptiere $y := x$. Sonst lehne ab: Kehre zu Schritt 1 zurück.

Von einem eher mathematischen Standpunkt her gesehen, bedeutet das, dass die Zufallsvariablen $X_1, X_2, \dots \sim \mathbb{P}$ unabhängig sind und für das Ergebnis $Y := X_T$ mit $T := \min\{n \in \mathbb{N} \mid X_n \in B\}$ gilt.

Zeigen Sie:

- a) $T \sim \text{Geo}(\mathbb{P}(B))$ und somit $\mathbb{E}[T] = 1/\mathbb{P}(B)$.
- b) $Y \sim \mathbb{P}(\cdot \mid B)$.

Hinweis: Für $n \in \mathbb{N}$ ist $\{T = n\} = \{X_1 \notin B, \dots, X_{n-1} \notin B, X_n \in B\}$ und es gilt $\mathbb{P}(Y \in A) = \mathbb{P}(X_T \in A) = \sum_{n=1}^{\infty} \mathbb{P}(X_n \in A, T = n)$.

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.

Ist T_n asymptotisch erwartungstreu?

Lösung: Insgesamt sechs Punkte.

Es gilt $F_{X_i}(t) = \frac{t}{a} 1_{[0, a]}(t)$ (0,5P).

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

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

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

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

Da

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

ist T_n kein erwartungstreuer Schätzer. (1,5P)

Es gilt jedoch $\mathbb{E}_a[T_n] = \frac{n}{n+1} a \rightarrow a$, wenn $n \rightarrow \infty$. Also ist T_n asymptotisch erwartungstreu. (1P)

Aufgabe H2 Schätzer.

- Seien X_1, \dots, X_n unabhängige mathematische Stichproben vom Umfang n der Grundgesamtheit X . Zeigen Sie, dass $Y_n := \sum_{i=1}^n \lambda_i X_i$ 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.
- Zeigen Sie, dass unter den möglichen Schätzern Y_n 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.
- Angenommen es gilt $\sum_{i=1}^n \lambda_i^2 \rightarrow 0$, falls $n \rightarrow \infty$. Zeigen Sie unter dieser Voraussetzung, dass Y_n dann ein konsistenter Schätzer für den Parameter $\mathbb{E}[X]$ ist.

Lösung: Insgesamt vier Punkte

a) Wegen der Linearität des Erwartungswerts gilt

$$\mathbb{E}[Y] = (\lambda_1 + \dots + \lambda_n) \cdot \mathbb{E}[X] = \mathbb{E}[X]. \quad (1P)$$

b) Da Y erwartungstreu ist, ist der mittlere quadratische Fehler die Varianz $\text{Var}(Y)$ (0,5P). Für diese gilt wegen der Unabhängigkeit der X_i

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

Für $\lambda_1 = \dots = \lambda_n = 1/n$ gilt Gleichheit und folglich minimiert diese Wahl der λ_i die Varianz von Y . (0,5P)

c) Mit der Chebychev-Ungleichung und der Unabhängigkeit der X_i folgt

$$\mathbb{P}(|Y_n - \mathbb{E}[X]| > \varepsilon) \leq \frac{\text{Var}(\sum_{i=1}^n \lambda_i X_i)}{\varepsilon^2} = \frac{\text{Var}(X) \sum_{i=1}^n \lambda_i^2}{\varepsilon^2} \rightarrow 0,$$

für $n \rightarrow \infty$. (1P)

Aufgabe H3 Effizienz bei steigender Datenmenge.

Gegeben sei eine Zufallsvariable X , deren Erwartungswert und Varianz existieren mit $\text{Var}[X] \neq 0$. Für $n \in \mathbb{N}$ seien X_1, \dots, X_n unabhängige Zufallsvariablen, welche alle dieselbe Verteilung wie X besitzen. In der Vorlesung haben Sie bereits gesehen, dass

$$\bar{X}(n) := \frac{1}{n} \sum_{i=1}^n X_i$$

für alle $n \in \mathbb{N}$ ein erwartungstreuer Schätzer für $\mathbb{E}[X]$ ist. Zeigen Sie, dass $\bar{X}(n)$ effizienter ist als $\bar{X}(n-1)$ für alle $n \geq 2$. Dabei heißt unter zwei erwartungstreuen Schätzern derjenige mit der geringeren Varianz effizienter.

Lösung: Wir berechnen $\text{Var}[\bar{X}(n)]$.

$$\text{Var}[\bar{X}(n)] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \cdot n \cdot \text{Var}[X] = \frac{1}{n} \text{Var}[X].$$

Dies ist offenbar fallend in n . Somit ist $\bar{X}(n)$ effizienter als $\bar{X}(n-1)$.

Abgabe der Hausübungen (Aufgaben H1 bis H3): **Mittwoch, 22. Dezember, 13:00 Uhr**

Viel Erfolg! :)