

Wahrscheinlichkeitstheorie und Statistik

Serie 9

MC 9-1. Seien $X_1, \dots, X_n$ unabhängig und identisch verteilte Zufallsvariablen mit unbekanntem Parameter $\theta \in \Theta$. Welche Aussagen über einen Schätzer $\hat{\theta} = g(X_1, \dots, X_n)$ sind immer korrekt? (Mehrere Antworten sind möglich.)

- (a) $\hat{\theta}$ muss eine messbare Funktion der Stichprobe $(X_1, \dots, X_n)$ sein.
- (b) $\hat{\theta}$ muss erwartungstreu für θ sein.
- (c) $\hat{\theta}$ ist selbst eine Zufallsvariable.
- (d) Wenn $\hat{\theta}$ eine konstante Funktion ist, ist $\hat{\theta}$ kein gültiger Schätzer.

MC 9-2. Seien $X_1, \dots, X_n$ unabhängig und identisch verteilte Zufallsvariablen mit unbekanntem Parameter $\theta \in \Theta$. Welche Ausdrücke sind formal gültige Schätzer für θ ? (Mehrere Antworten sind möglich.)

- (a) $\hat{\theta}(X_1, \dots, X_n) = 42$
- (b) $\hat{\theta}(X_1, \dots, X_n) = X_1 + \sin(X_2)$
- (c) $\hat{\theta}(X_1, \dots, X_n) = \theta + 1$
- (d) $\hat{\theta}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$

MC 9-3. Seien $X_1, \dots, X_n$ unabhängig und identisch verteilte Zufallsvariablen. Angenommen, weniger als $n/2$ Beobachtungen X_i werden kontaminiert, indem sie durch einen sehr großen Wert M ersetzt werden. Welche der folgenden Aussagen sind korrekt, wenn M gegen unendlich geht? (Mehrere Antworten sind möglich.)

- (a) Der Stichprobenmittelwert bleibt beschränkt.
- (b) Der Stichprobenmittelwert bleibt unverändert.
- (c) Der Stichprobenmedian bleibt beschränkt.
- (d) Der Stichprobenmedian bleibt unverändert.

Aufgabe 9-4. Seien $X_1, \dots, X_n$ unabhängige, Zufallsvariablen mit $X_i \sim \mathcal{N}(\theta\alpha_i, 1)$, wobei $\alpha_i \neq 0$ bekannte Parameter sind, aber $\theta \in \mathbb{R}$ unbekannt ist. Bestimmen Sie den Maximum-Likelihood-Schätzer für θ .

Aufgabe 9-5 Schreiben Sie eine Python-Implementierung der logistischen Regression. Die Aufgabe ist, eine Modellfunktion

$$p: \mathbb{R} \rightarrow \mathbb{R}, \quad p(x) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))}$$

mit Parametern β_0 und β_1 zu bestimmen, welche für gegebene Daten $(X_1, Y_1), \dots, (X_n, Y_n)$ die folgende logistische Log-Likelihood maximiert:

$$\text{LL}(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i \log p(X_i) + (1 - Y_i) \log (1 - p(X_i))).$$