Wahrscheinlichkeitstheorie und Statistik

Serie 6 - Lösungen

MC 6-1. Sei X eine Zufallsvariable mit Werten in $\{0, 1, 2, \dots\}$. Welche der folgenden Aussagen sind korrekt? (Mehrere richtige Antworten sind möglich.)

(a) $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X = k)$

(b) $\mathbb{E}(X) = \sum_{k=0}^{\infty} k \cdot \mathbb{P}(X = k)$

(c) $\mathbb{E}(X) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$

(d) $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X \geq k)$

Lösung: Nur (b) und (c) sind korrekt.

(a) Falsch. Gegenbeispiel: Die konstante Zufallsvariable $X = 0$.

(b) Wahr, denn $k \mapsto \mathbb{P}(X = k)$ ist die Gewichtsfunktion von X.

(c) Wahr, denn

$$\sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \mathbb{1}_{k \leq \ell} \mathbb{P}(X = \ell) = \sum_{\ell=1}^{\infty} \mathbb{P}(X = \ell) \sum_{k=1}^{\infty} \mathbb{1}_{k \leq \ell} = \sum_{\ell=1}^{\infty} \ell \, \mathbb{P}(X = \ell) = \mathbb{E}(X).$$

(d) Falsch. Gegenbeispiel: Die konstante Zufallsvariable $X = 0$.

MC 6-2. Sei X Bernoulli-verteilt mit Parameter $p \in [0, 1]$ und $Y = (2X - 1)^2$. Was ist die Varianz von Y? (Nur eine Antwort ist richtig.)

(a) 1

(b) p

(c) p^2

(d) 0

Lösung: Nur (d) ist richtig. Die Zufallsvariable Y ist konstant gleich 1. Daher gilt $\mathbb{E}(Y) = 1$ und $\mathbb{V}\mathrm{ar}(Y) = 0$.

Aufgabe 6-3. Sie $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $X_1, \dots, X_n : \Omega \to \mathbb{R}$ Zufallsvariablen mit Verteilung $P : \mathcal{B}(\mathbb{R}) \to [0, 1]$. Die empirische Verteilung von $X_1, \dots, X_n$ ist

$$P_n : \Omega \times \mathcal{B}(\mathbb{R}) \to [0, 1], \qquad P_n(\omega, A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A \left(X_i(\omega) \right)).$$

Für jedes Ereignis $A \in \mathcal{B}(\mathbb{R})$ ist also $P_n(A) : \Omega \to \mathbb{R}$ eine Zufallsvariable. Zeigen Sie, dass $P_n(A)$ im folgenden Sinn eine gute Näherung für $P(A)$ ist:

(a) $\mathbb{E}\left(P_n(A)\right) = P(A)$.

(b) $\mathbb{V}\mathrm{ar}\left(P_n(A)\right) \leq 1/n$, wenn $X_1, \ldots, X_n$ unabhängig sind.

Hier sind $\mathbb{E}$ bzw. $\mathbb{V}\mathrm{ar}$ der Erwartungswert bzw. die Varianz unter $\mathbb{P}$.

Lösung: Wir schreiben $\mathbb{E}$ bzw. $\mathbb{V}\mathrm{ar}$ für den Erwartungswert und die Varianz unter $\mathbb{P}$. Analog schreiben wir E bzw. Var für den Erwartungswert bzw. die Varianz unter P.

(a) Berechnung des Erwartungswertes:

$$\begin{aligned}\mathbb{E}(P_n(A)) &= \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_A(X_i)\Big) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_A(X_i)) \\ &= \frac{1}{n}\sum_{i=1}^{n} P(A) = P(A).\end{aligned}$$

(b) Berechnung der Varianz: Da die X_i unabhängig sind, gilt:

$$\begin{aligned}\mathbb{V}\mathrm{ar}(P_n(A)) &= \mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_A(X_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(\mathbb{1}_A(X_i)) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} P(A)(1-P(A)) \\ &= \frac{P(A)(1-P(A))}{n} \leq \frac{1}{n}.\end{aligned}$$

Aufgabe 6-4. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und P_n die empirische Verteilung von Zufallsvariablen $X_1, \ldots, X_n : \Omega \to \mathbb{R}$. Berechnen Sie den Erwartungswert und die Varianz des Wahrscheinlichkeitsmasses $P_n(\omega) : \mathcal{B}(\mathbb{R}) \to [0,1]$, für eine beliebige aber fixe Realisierung $\omega \in \Omega$.

Hilfestellung: Es geht also um die Berechnung von $E_n(X)$ bzw. $\mathrm{Var}_n(X)$, wobei $X : \mathbb{R} \to \mathbb{R}$ die Identitätsabildung ist und E_n bzw. Var_n die Erwartung bzw. Varianz von P_n bezeichnen, jeweils für fixes $\omega \in \Omega$.

Lösung: Das Wahrscheinlichkeitsmass $P_n(\omega)$ ordnet jedem der beobachteten Werte $X_1(\omega), \ldots, X_n(\omega)$ die gleiche Wahrscheinlichkeit zu:

$$P_n(\omega)(\{x\}) = \frac{1}{n}, \quad \text{für} \quad x \in \{X_1(\omega), \ldots, X_n(\omega)\}.$$

Zur Vereinfachung unterdrücken wir die fixe Realisierung $\omega \in \Omega$ in der Notation und definieren $X : \mathbb{R} \to \mathbb{R}$ als Identitätsabbildung. Dann gilt:

$$P_n(X = x) = \frac{1}{n}, \quad \text{für} \quad x \in \{X_1, \ldots, X_n\}.$$

Wir schreiben E_n bzw. Var_n für den Erwartungswert bzw. die Varianz unter P_n, wiederum für fixes

$\omega \in \Omega$. Dann gilt:

$$E_n(X) = \sum_{i=1}^{n} X_i \cdot P_n(X = X_i) = \sum_{i=1}^{n} X_i \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\mathrm{Var}_n(X) = E_n\left((X - E_n(X))^2\right) = \sum_{i=1}^{n} (X_i - E_n(X))^2 \cdot P_n(X = X_i) = \frac{1}{n} \sum_{i=1}^{n} (X_i - E_n(X))^2 .$$

Aufgabe 6-5. Schreiben Sie ein Python Programm, welches die lineare Regression Z einer Zufallsvariable Y auf eine Zufallsvariable X berechnet. Dabei soll die unbekannte Verteilung von (X, Y) näherungsweise durch die empirischen Verteilung von $(X_1, Y_1), \ldots, (X_n, Y_n)$ ersetzt werden. Wenn wir E_n, Var_n und Cov_n für die Erwartung, Varianz und Kovarianz unter der empirischen Verteilung P_n schreiben, geht es also um die Berechnung von

$$Z = \beta_0 + \beta_1 X, \qquad \beta_0 = E_n(Y) - \frac{\mathrm{Cov}_n(X, Y)}{\mathrm{Var}_n(X)} E_n(X), \qquad \beta_1 = \frac{\mathrm{Cov}_n(X, Y)}{\mathrm{Var}_n(X)}$$

für einen gegebenen Datensatz $(X_1, Y_1), \ldots, (X_n, Y_n)$ von Paaren reeller Zahlen.

Lösung:

```
import numpy as np
import numpy.typing as npt
import matplotlib.pyplot as plt

def linear_regression(x: npt.NDArray, y: npt.NDArray) -> tuple[float, float]:
    n = len(x)
    mean_x = sum(x) / n
    mean_y = sum(y) / n
    cov_xy = sum((x-mean_x) * (y-mean_y)) / n
    var_x = sum((x-mean_x) ** 2) / n
    beta_1 = cov_xy / var_x
    beta_0 = mean_y - beta_1 * mean_x
    return beta_0, beta_1

x = np.linspace(0, 10, 50)
y = 3 * x + 7 + np.random.normal(0, 2, 50)
beta_0, beta_1 = linear_regression(x, y)
y_pred = beta_0 + beta_1 * x
plt.scatter(x, y, color='blue')
plt.plot(x, y_pred, color='red')
plt.xlabel('X')
plt.ylabel('Y')
plt.show()
```