21. Juni 2024

Prof. Dr. Evgeny Spodarev
Sabrina Weber

Angewandte Stochastik – Übungsblatt 5

Dieses Übungsblatt wird in der Übung am 21. Juni 2024 um 10:15 Uhr diskutiert. Bitte gebt eure Lösungen vor Beginn der Übung über Moodle ab. Gebt eure Lösungen als *exakt zwei Dateien* ab: Eine PDF-Datei für die theoretischen Aufgaben und eine R- oder R-Markdown-Datei (Dateiendung *.R* bzw. *.RMD*) für die Programmieraufgaben. Andere Dateiformate oder mehrere Dateien sind nicht erlaubt. Bearbeitet das Übungsblatt in *Gruppen von 3- 4 Personen* und gebt bitte euren jeweiligen Studiengang an. Nur eine Person aus der Gruppe soll das Übungsblatt abgeben. Beachtet außerdem die Hinweise zu den Übungen in Moodle.

Aufgabe 1 (2 + 3 = 5 Punkte)

Für $i = 1, 2$ sei im Folgenden $(X_i, Y_i) : \Omega \to \mathbb{R}^2$ jeweils ein Zufallsvektor mit Verteilungsfunktion $F_i : \mathbb{R}^2 \to [0, 1]$. Bestimme die Randverteilungen von X_i und $Y_i,\ i = 1, 2$, falls

(a) F_1 die Zähldichte

Y_1	$X_1 = 1$	2	3
2	0.1	0.05	0.2
4	0.05	0.15	0.1
5	0	0.2	0.15

besitzt.

(b) F_2 die Wahrscheinlichkeitsdichte

$$f_2(x, y) = \frac{1}{\pi\sqrt{3}} e^{-\frac{2}{3}(x^2 - xy + y^2)},\ (x, y) \in \mathbb{R}^2$$

besitzt.

Aufgabe 2 (7 Punkte)

Ein Zufallsvektor $X = (X_1, X_2)$ heißt zweidimensional normalverteilt, falls er die Dichte

$$f(x) = \frac{1}{2\pi\sqrt{\det(K)}} \exp\Big(-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)\Big),\ x = (x_1, x_2)^T \in \mathbb{R}^2$$

besitzt, wobei $\mu = (\mu_1, \mu_2)^T \in \mathbb{R}^2$ ein Vektor und

$$K = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \in \mathbb{R}^{2\times 2},\ \rho \in (-1, 1),\ \sigma_1, \sigma_2 > 0$$

eine symmetrische und positiv definite (d.h. zusätzlich gilt $\det(K) > 0$) Matrix sei. Zeige: Die Komponenten X_1 und X_2 von X sind genau dann stochastisch unabhängig, wenn K eine Diagonalmatrix ist, d.h. wenn $\rho = 0$ gilt.[1]

Aufgabe 3 (3 + 6 + 5 = 12 Punkte)

(a) Seien $X_1, X_2 \sim U(0, 1)$ stochastisch unabhängig. Bestimme die Dichte von X_1/X_2.

(b) Für $i = 1, \dots, n$ seien $X_i \sim \mathrm{Poi}(\lambda_i)$, $\lambda_i > 0$, stochastisch unabhängige poissonverteilte Zufallsvariablen. Zeige, dass dann $\sum_{i=1}^n X_i$ ebenfalls poissonverteilt ist und bestimme in diesem Fall auch den Parameter.[2]

(c) Seien X_1, X_2 $\sim$Exp(λ) stochastisch unabhängig. Bestimme die Dichte der Zufallsvariablen X_1/X_2 und $X_1 + X_2$.

Hinweis: Für die folgende Aufgabe sind die Vorlesungsvideos zu Statistik bis einschließlich Video 6 relevant

Aufgabe 4 (3 + 2 + 2 + 2 + 2 = 11 Punkte)

Für diese Aufgabe sollen die Teilaufgaben (a)-(d) ohne die Verwendung von **R** bearbeitet werden. Die folgende Tabelle enthält Stichprobendaten für die Anzahl von Stunden, die 8 Studierende zum Lernen für die Klausur in 'Angewandter Stochastik' aufgewendet haben, sowie die Ergebnisse, die sie in der Prüfung erzielt haben.

Lernzeit in Stunden (x)	20	16	34	23	27	32	18	22
Punkte in der Prüfung (y)	64	61	84	70	88	92	72	77

Es wird angenommen, dass sich die erzielten Punkte als lineare Funktion der Lernzeit darstellen lassen.

(a) Bestimme und interpretiere die Regressionsparameter $\hat{\alpha}$ und $\hat{\beta}$ und zeichnen Sie die Regressionsgerade zusammen mit den Daten in ein Streudiagramm (scatter plot).

(b) Wie ändert sich das Prüfungsergebnis eines Studierenden (laut unserem Modell), der sich entschließt 9 Stunden mehr zu lernen als ursprünglich geplant? Prognostiziere das Prüfungsergebnis eines Studierenden, der 30 Stunden für die Prüfung gelernt hat bzw. der überhaupt nicht für die Prüfung gelernt hat.

[1] Es darf ohne Beweis verwendet werden, dass $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ gilt.

[2] Betrachte zunächst den Fall $n = 2$ und stelle dann eine Hypothese für allgemeines $n \in \mathbb{N}$ auf. Beweise diese mittels vollständiger Induktion.

(c) Ein Studierender möchte die Klausur ohne Abzug, d.h. mit vollen 100 Punkten bestehen. Wie viele Stunden müsste er dafür lernen, wenn man die Störterme außer acht lässt?

(d) Bestimme und interpretiere das Bestimmtheitsmaß R^2.

(e) Bestimme nun die Regressionsparameter α und β mit Hilfe von **R** und interpretiere die Ausgabe in **R** des linearen Models.