SoSe 2023
16. Juni 2023

INSTITUT FÜR STOCHASTIK

Prof. Dr. Evgeny Spodarev
Philipp Rieder

Angewandte Stochastik – Übungsblatt 4

Dieses Übungsblatt wird in der Übung am 16. Juni 2023 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.

Hinweis: Für die folgenden Aufgaben sind die Vorlesungsvideos zu Statistik bis einschließlich Video 5 relevant

Aufgabe 1

Nehmt bitte an der Vorlesungsevaluation unter diesem Link teil.

Aufgabe 2 (2 + 2 + 3 = 7 Punkte)

Für welche $c \in \mathbb{R}$ handelt es sich bei folgenden Funktionen $f_i : \mathbb{R} \to \mathbb{R},\ i \in \{1, 2, 3\}$ um Wahrscheinlichkeitsdichten?

(a) $f_1(x) = \frac{c}{\sqrt{x-1}} \mathbb{1}_{(1,2]}(x),\ x \in \mathbb{R}$

(b) $f_2(x) = ce^{-|x|},\ x \in \mathbb{R}$

(c) Zeige außerdem, dass jeder Kerndichteschätzer

$$\hat{f}_n(x) = \sum_{i=1}^{n} \frac{1}{nh} K\left(\frac{x - x_i}{h}\right)$$

für jede Bandbreite $h > 0$, Kern K und Stichprobe $(x_1, \ldots, x_n)$ eine Dichte ist. Dabei ist ein Kern $K : \mathbb{R} \to \mathbb{R}$ eine nichtnegative messbare Funktion, sodass $\int_{\mathbb{R}} K(x)dx = 1$.

Aufgabe 3 (R-Aufgabe, 2 + 2 = 4 Punkte)

In dieser Aufgabe untersuchen wir die Wahl der Bandbreite h von Kerndichteschätzern.

(a) Schreibe eine Funktion in R, die den Funktionswert des Kerndichteschätzers für einen Vektor x an der Stelle t mit Bandbreite h mithilfe des Epanechnikov Kerns bestimmt.

(b) Teste deine Funktion für eine Realisierung von $n = 200$ standardnormalverteilten Zufallsvariablen und Bandbreiten $h \in \{0.1, 0.2, 0.5, 1, 2, 4, 5\}$. Plotte die Kerndichteschätzer inklusive der Dichte der Standardnormalverteilung in verschiedenen Farben in ein Koordinatensystem. Denke an eine sorgfältige Beschriftung der Achsen und an eine Legende.

Aufgabe 4 (R-Aufgabe, 2 + 2 = 3 Punkte)

Simuliere 500 Realisierungen von Exp(3)-verteilten Zufallsvariablen.

(a) Erstelle ein Histogramm der relativen Häufigkeiten der simulierten Daten und plotte in dieses Histogramm die Dichte der Exp(3)-Verteilung. Wähle hierbei eine Bandbreite, mit der das Histogramm gut zur Dichte passt. Es existiert keine eindeutige Lösung für die gesuchte Bandbreite.

(b) Plotte die empirische Verteilungsfunktion der simulierten Daten mit der in R enthaltenen Funktion `ecdf` oder `stat_ecdf` aus `ggplot2`. Plotte die theoretische Verteilungsfunktion in das selbe Schaubild.

Aufgabe 5 (2 +2 = 4 Punkte)

Es sei X eine absolutstetige Zufallsvariable mit streng monoton wachsender Verteilungsfunktion $F : \mathbb{R} \to [0, 1]$.

(a) Welche Verteilung besitzt die Zufallsvariable $F(X)$?

(b) Sei nun $Y \sim U(0, 1)$ eine auf dem Intervall $(0, 1)$ gleichverteilte Zufallsvariable und $Z = F^{-1}(Y)$. Welche Verteilung besitzt die Zufallsvariable Z?

Begründe in beiden Fällen deine Antwort.

Aufgabe 6 (4 Punkte)

Die Zufallsvariable $N \sim \text{Poi}(\lambda)$, $\lambda > 0$, beschreibe die Anzahl der Fehlermeldungen eines Systems an einem bestimmten Tag. Mit Wahrscheinlichkeit $p \in (0, 1)$ wird eine Fehlermeldung durch ein im System defektes Bauteil ausgelöst. Die Anzahl der Fehlermeldungen eines Tages, die auf ein defektes Bauteil zurückzuführen sind, sei durch die Zufallsvariable Z beschrieben. Zeige, dass $Z \sim \text{Poi}(\lambda p)$, falls davon ausgegangen wird, dass sowohl die Ursachen der einzelnen Fehlermeldungen unabhängig sind als auch deren Anzahl N unabhängig von den Fehlerursachen ist.

Aufgabe 7 (3 + 4 = 7 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 8 (5 Punkte)

Sei $X \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Für $a, b \in \mathbb{R}, a \neq 0$ sei die Zufallsvariable Y gegeben durch $Y = aX + b$. Zeige, dass $Y \sim N(a\mu + b, a^2\sigma^2)$.

Aufgabe 9 (2 + 2 + 2 + 2 = 8 Punkte)

Zu (seriösen) Forschungszwecken wurde ein mittelalterliches Katapult nachgebaut. Es steht 7,40 vom Ufer eines Baches entfernt, der 5,20m breit ist. Die Munition besteht aus Steinkugeln, die einen Durchmesser von 20cm aufweisen. Nach einigen Probeschüssen wurde die Theorie aufgestellt, dass die Schussweite normalverteilt ist, im Mittel 10m beträgt und eine Standardabweichung von 5m hat.

Hinweis: Skaliere die Normalverteilung zu $N(0, 1)$ und verwende die Quantiltabelle, die auf Moodle zur Verfügung gestellt wurde.
Für $X \sim N(0, 1)$ gilt $\mathbb{P}(X \leq 0.26) = \Phi(0.26) \overset{Tabelle}{=} 0,602568$, wobei $\Phi(x)$ die Verteilungsfunktion von X bezeichnet. Da f_X, die Dichte von X, achsensymmetrisch ist gilt außerdem $f_X(-x) = f_X(x)$ und $\Phi(-x) = 1 - \Phi(x)$.

(a) Wie groß ist die Wahrscheinlichkeit, dass eine geschossene Kugel in vollem Umfang im Bach landet?

(b) Wie groß ist die Wahrscheinlichkeit, dass eine Kugel über den Bach geschossen wird und (zumindest teilweise) am gegenüberliegenden Ufer aufschlägt?

(c) Wie groß ist die Wahrscheinlichkeit, dass ein Schuss nach hinten los geht (d.h. die Kugel fliegt nicht zum Bach, sondern in die entgegengesetzte Richtung)?

(d) Elf Meter hinter dem Katapult befindet sich ein geparkter Wagen. Ist es ausgeschlossen, dass eine Kugel den Wagen beschädigt? Begründe deine Antwort.