26. April 2024

Prof. Dr. Evgeny Spodarev
Sabrina Weber

Angewandte Stochastik – Übungsblatt 1

Dieses Übungsblatt wird in der Übung am 26. April 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 (1 + 6 = 7 Punkte)

Wir betrachten einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ um das gleichzeitige Würfeln mit zwei sechsseitigen Würfeln zu beschreiben. Die Grundmenge Ω wählen wir wie in Beispiel 2.1.2 im Skript. Ein Elementarereignis (also ein Element $\omega \in \Omega$ der Grundmenge) gibt also an, welche Zahlen die beiden Würfel zeigen. Beachte, dass die Würfel hier unterscheidbar sind, wir haben also z.B. einen roten Würfel, dem der erste Eintrag von ω entspricht und einen blauen Würfel, dessen Augenzahl im zweiten Eintrag von ω steht. Das Elementarereignis $(4, 2)$ bedeutet dann, dass der rote Würfel eine 4 zeigt, der blaue eine 2; das Elementarereignis $(2, 4)$ aber dass der rote Würfel eine 2 zeigt, der blaue eine 4.

Ereignisse, deren Wahrscheinlichkeit wir berechnen wollen können unter anderem sein:

- A = “Der rote Würfel zeigt eine Zahl kleiner als 4”
- B = “Der rote und der blaue Würfel zeigen die gleiche Zahl”
- C = “Der blaue Würfel zeigt eine größere Zahl als der rote Würfel”

Die σ-Algebra $\mathcal{F}$ repräsentiert die Menge aller Ereignisse, deren Wahrscheinlichkeit wir später berechnen können. Aus der Intuition des Gegenereignisses (also z.B. “der rote Würfel zeigt *nicht* eine Zahl kleiner als 4” bzw. “der rote Würfel zeigt mindestens 4 Augen”) wird klar, wieso es wichtig ist, dass $\mathcal{F}$ auch das Komplement $\overline{A}$ (also das Gegenereignis von A) enthält, wenn A selbst enthalten ist. Mit der zweiten Forderung der (σ-)Vereinigungsstabilität wird ein Mengensystem konstruiert, in dem im Wesentlichen alle intuitiv definierbaren Ereignisse enthalten sind, also auch Vereinigungen von Ereignissen (A oder B tritt ein), Schnitte (A und B tritt ein) und auch kompliziertere wiederholte Anwendung dieser Operationen (z.B. A und B tritt ein, aber nicht C).

(a) Gib eine geeignete σ-Algebra $\mathcal{F}$ an, die alle Ereignisse aus der Liste oben enthält sowie einfache Abwandlungen davon (Ersetze die 4 durch eine beliebige Zahl; Ersetze "rot" durch "blau" und umgekehrt; Ersetze "kleiner" durch "größer"; ...).

(b) Gehe davon aus, dass die beiden Würfel fair sind, also jede Seite mit der gleichen Wahrscheinlichkeit zeigen. Gib das Wahrscheinlichkeitsmaß $P\colon \mathcal{F} \to [0,1]$ an und gib die Wahrscheinlichkeiten für die drei Ereignisse aus der Liste oben an. Es ist hilfreich, wenn du ein Ereignis als Menge von Elementarereignissen betrachtest.

Aufgabe 2 (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 3 Punkte)

Sei $\Omega = \{5, 6, 7, 8, 9, 16, 17, 18, 19, 20\}$ mit Teilmengen $A = \{6, 8, 16, 18, 20\}$, $B = \{5, 7, 9, 17, 19\}$, $C = \{16, 17, 18, 19\}$ und $D = \{6, 9, 20\}$. Schreibe folgende Mengen als Aufzählung ihrer Elemente:

(a) $A \cup C$

(b) $A \cap B$

(c) C^c

(d) $C^c \cap D$

(e) $(\Omega \setminus C)^c$

(f) $A \setminus C$

Aufgabe 3 (3 Punkte)

Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{F}$. Zeige, dass

$$P(A \triangle B) = P(A) + P(B) - 2P(A \cap B) \tag{1}$$

Aufgabe 4 (2 + 3 + 3 = 8 Punkte)

Betrachte folgende Ereignisse bzgl. der Wertentwicklung des Bitcoinkurses nach einem Jahr:

- A = "Der Kurs steigt um weniger als 5%, fällt aber nicht."
- B = "Der Kurs steigt um 5 - 10%."
- C = "Der Kurs steigt über 10%."

Betrachte die folgenden Ereignisse

(a) $A \cup B$

(b) $(A \cup B \cup C)^c$

(c) $B^c \cap C^c$

Beschreibe die Ereignisse in (a)-(c) jeweils zunächst in Worten und bestimme ihre Wahrscheinlichkeit, falls bekannt ist, dass $P(A) = 0.4$ und $P(B) = P(C) = 0.05$

Aufgabe 5 (2 + 5 = 7 Punkte)

Begründe, ob es sich bei den Ereignissystemen $\mathcal{F}_1, \mathcal{F}_2$ um σ-Algebren handelt.

(a) $\Omega = \{23, U, \varepsilon, ☺, \square, ☼\}$ und $\mathcal{F}_1 = \{\emptyset,\ \Omega,\ \{23, U, \varepsilon, ☺\},\ \{23, \square, ☼\},\ \{U, \varepsilon, ☺\},\ \{23\}\}$

(b) $\Omega = \mathbb{R}$ und $\mathcal{F}_2 = \{A \subseteq \mathbb{R} \mid A \subseteq [0, 1] \text{ oder } A^c \subseteq [0, 1]\}$

Aufgabe 6 (3 + 4 = 7 Punkte)

Im Folgenden ist $(\Omega, \mathcal{F})$ stets ein Maßraum und $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ eine Funktion. In welchen Fällen ist $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum?

(a) $\Omega = \mathbb{N} \cup \{0\},\ \mathcal{F} = \mathcal{P}(\Omega)$ und $\mathbb{P}(\{k\}) = (1/2)^k,\ k \in \Omega$

(b) Sei $n \in \mathbb{N}$ fest gewählt, $\Omega = \{1, 2, \ldots, n\},\ \mathcal{F} = \mathcal{P}(\Omega)$ und

$$\mathbb{P}(A) = \sum_{j \in A} \frac{2j}{n(n+1)},\ A \in \mathcal{F}$$

10. Mai 2024

Prof. Dr. Evgeny Spodarev
Sabrina Weber

Angewandte Stochastik – Übungsblatt 2

Dieses Übungsblatt wird in der Übung am 10. Mai 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 (3 Punkte)

Bei der Multiple-Choice Aufgabe in der Klausur zu "Angewandte Stochastik" kennt Marco auf 20 der insgesamt 25 Fragen die richtige Antwort. Da er einige Vorlesungen geschwänzt hat, muss er bei den restlichen 5 Fragen raten. Nimm an, dass bei jeder Frage 4 Antwortmöglichkeiten zur Auswahl stehen, von denen jeweils genau eine richtig ist. Wie groß ist die Wahrscheinlichkeit, dass Marco genau 3 der 5 Fragen richtig beantwortet?

Aufgabe 2 (2 + 2 + 2 + 2 + 2 = 10 Punkte)

In einer Urne seien drei rote, zwei schwarze und fünf weiße Kugeln. Wir ziehen zufällig vier Kugeln nacheinander mit Zurücklegen. Bestimme die Wahrscheinlichkeit folgender Ereignisse:

(a) Die erste gezogene Kugel ist eine weiße, die letzte eine rote.

(b) Alle drei Farben werden gezogen.

(c) Alle vier Kugeln sind (i) weiß, (ii) rot, (iii) schwarz.

(d) Genau zwei Farben werden gezogen.

(e) Genau eine weiße und eine rote Kugel werden gezogen.

Aufgabe 3 (2 + 3 = 5 Punkte)

In einer Studie von Vanderpump et al. (Clinical Endocrinology 1995, 43, 55–69) wurden Todesereignisse innerhalb von 10 Jahren bei RaucherInnen (139 von 582) und NichtraucherInnen (230 von 732) erhoben.

(a) Berechne die bedingten Wahrscheinlichkeiten für Tod nach 10 Jahren im Laplaceschen Wahrscheinlichkeitsraum, bedingt auf RaucherInnenstatus.

(a) Welche Aussagen treffen zu:

(i) Die bedingten Wahrscheinlichkeit für Tod nach 10 Jahren ist höher für RaucherInnen als für NichtraucherInnen.

(ii) Die Ereignisse *Tod nach 10 Jahren* und *RaucherIn* sind stochastisch unabhängig.

(iii) Ist damit gezeigt, dass das Rauchverhalten einen kausalen Einfluss auf die Sterbewahrscheinlichkeit hat.

Begründe deine Antwort.

Aufgabe 4 (2+2+2 = 6Punkte)

Beim Scrabble werden vor Spielstart aus einem Säckchen mit 102 Buchstaben (35 Vokabeln, 3 Umlaute, 62 Konsonanten und 2 Joker/Blanks) 7 zufällig herausgezogen. Berechne die Wahrscheinlichkeit dafür, dass

(a) 5 Konsonanten und 2 Vokale

(a) mindestens 1 Joker

(a) Keine Vokale

gezogen werden.

Aufgabe 5 (4 Punkte)

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \mathcal{F}$. Zeige die Bonferroni-Ungleichung:

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) \geq \sum_{i=1}^{n} \mathbb{P}(A_i) - (n-1) = 1 - \sum_{i=1}^{n} \mathbb{P}(A_i^c).$$

Aufgabe 6 (7 Punkte)

Führe die Induktion aus dem Beweis der Siebformel (vgl. Folgerung 2.17, Teil 3) bis zum Ende durch.

24. Mai 2024

Prof. Dr. Evgeny Spodarev
Sabrina Weber

Angewandte Stochastik – Übungsblatt 3

Dieses Übungsblatt wird in der Übung am 24. Mai 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 (1 + 2 + 3 = 6 Punkte)

Betrachte das zweimalige Würfeln eines fairen Würfels. Die Zufallsvariable $X : \Omega \to \mathbb{R}$ beschreibe die Größere der beiden geworfenen Augenzahlen (d.h. wurde z.B. eine 3 und eine 2 gewürfelt, so hat X den Wert 3).

(a) Gib einen geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ an.

(b) Bestimme die kleinste Menge C, so dass $\mathbb{P}(X \in C) = 1$.

(c) Bestimme die Wahrscheinlichkeitsfunktion $\{p_k\}$ von X mit der Menge C aus Teil (b), d.h. gib für alle $x_k \in C$ jeweils $p_k = \mathbb{P}(X = x_k)$ an.

Aufgabe 2 (2 + 3 = 5 Punkte)

Gegeben sei der Wahrscheinlichkeitsraum $([0,1], \mathcal{B}([0,1]), \lambda)$. Hierbei sind $\mathcal{B}([0,1])$ die Borel'sche σ-Algebra auf dem Intervall $[0,1]$ (d.h. alle Teilmengen von $\mathcal{B}(\mathbb{R})$, die in $[0,1]$ liegen) sowie λ das Lebesguemaß auf $[0,1]$, d.h. λ ordnet jedem Teilintervall von $[0,1]$ seine Länge zu (insbesondere also $\lambda([0,1]) = 1$. Im Folgenden sei stets $p \in (0,1)$ beliebig aber fest vorgegeben. Konstruiere Zufallsvariablen X, Y auf dem Wahrscheinlichkeitsraum $([0,1], \mathcal{B}([0,1]), \lambda)$ mit

(a) $X \sim \mathrm{Bin}(1, p)$, sowie

(b) $Y \sim \mathrm{Bin}(2, p)$.

Begründe auch, warum es sich bei deiner Wahl um Zufallsvariablen handelt.

Aufgabe 3 (3 + 1 = 4 Punkte)

Beim Lotto "6 aus 49" betrage der aktuelle Jackpot 15 Millinonen Euro. Es sei bekannt, dass an der nächsten Ziehung 5 Millionen Spieler teilnehmen, von denen jeder jeweils genau einen Tipp abgegeben hat. Den Jackpot gewinnt man, wenn man die "6 Richtigen" und zusätzlich die auf dem Tippschein bereits vorher aufgedruckte Superzahl (eine der Zahlen $0, \ldots, 9$) hat.

(a) Wie groß ist die Wahrscheinlichkeit dafür, dass genau ein Spieler den Jackpot gewinnt (der Gewinner ihn also mit keinem anderen Spieler teilen muss), wenn davon ausgegangen wird, dass sowohl die Tipps der Spieler als auch die Superzahl zufällig ist (also ohne System getippt wurde)?

(b) Bestimme die Wahrscheinlichkeit in (a) auch approximativ mit dem "Gesetz der seltenen Ereignisse". Gib das Ergebnis auf neun Nachkommastellen gerundet an.

Aufgabe 4 (3 + 2 + 2 = 7 Punkte)

An der Haltestelle "Lehrer Tal" fahren abwechselnd Busse[1] der Linien 2 und 5 zur Universität ab. Teilweise sind die Busse jedoch so voll, dass nicht alle Busse tatsächlich halten. Es sei bekannt, dass die Wahrscheinlichkeit, dass ein Bus der Linie 2 an der Haltestelle "Lehrer Tal" hält $1/3$ ist, während die Wahrscheinlichkeit bei der Linie 5 eine Konstante $p \in (0, 1)$ ist.
Du kommst an der Haltestelle "Lehrer Tal" an, und beschließt, den nächsten Bus, der an der Haltestelle anhält, zur Uni zu nehmen. Laut Fahrplan (von dem wir ausnahmsweise mal davon ausgehen, dass er korrekt ist) ist der nächste Bus ein Bus der Linie 2.

(a) Die Zufallsvariable X sei wie folgt definiert: Für $k \in \mathbb{N}$ beschreibe $\{X = k\}$ das Ereignis "Der k-te ankommende Bus ist der erste Bus, der anhält". Bestimme die Zähldichte von X. *Hinweis: Unterscheide zwischen k gerade und k ungerade.*

(b) Die Zufallsvariable Y sei die Nummer der Buslinie mit der du an der Uni ankommen wirst, d.h. $Y \in \{2, 5\}$. Bestimme die Zähldichte von Y.

(c) Für welchen Wert von p gilt $\mathbb{P}(Y = 3) = \mathbb{P}(Y = 5)$?

Aufgabe 5 (2 + 2 = 4 Punkte)

Gegeben sei eine Zufallsvariable $X \sim Geo(p)$ für einen Parameter $p \in (0, 1)$, die der geometrischen Verteilung folgt.

(a) Berechne die Verteilungsfunktion der Zufallsvariable X.

(b) Zeige, dass die geometrische Verteilung gedächtnislos ist, d.h., es gilt für $n, k = 1, 2, \ldots$:

$$P(X = n + k | X > k) = P(X = n) .$$

[1] O.B.d.A. Straßenbahnen sind Busse.

Statistik

Für die folgende Aufgabe sind die Vorlesungsvideos zu Statisik 1-3 relevant.

Aufgabe 6 (1 + 3 + 2 + 3 = 9 Punkte)

Die folgende Stichprobe beschriebt das Gewicht von 10 Bananen (in Gramm).

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
179	172	150	206	198	230	247	205	182	179

(a) Gib die Ordnungsstatistiken an.

(b) Berechne folgende Größen

- (i) Arithmetischer Mittelwert
- (ii) geometrischer Mittelwert
- (iii) harmonischer Mittelwert
- (iv) Median
- (v) 25%-Quantil
- (vi) 75%-Quantil

Programmieraufgaben

Lade auf Moodle den kompletten Datensatz *bananas.csv* herunter. Achte bei den folgenden Aufgaben auf eine korrekte Beschriftung der Achsen.

(c) Plotte in R die empirische Verteilungsfunktion der Nettogewichte aller Bananen.

(d) Zeige drei Boxplots, jeweils für das Gewicht der Schalen, das Nettogewicht sowie das Gesamtgewicht der Bananen. Plotte die drei Boxplots in einen Plot.

07. Juni 2024

Prof. Dr. Evgeny Spodarev
Sabrina Weber

Angewandte Stochastik – Übungsblatt 4

Dieses Übungsblatt wird in der Übung am 07. 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.

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

Bitte nehmt an der Evaluation für die Übung unter diesem Link und an der Evaluation für die Vorlesung unter disem Link teil.

Aufgabe 1 (2 + 2 + 2 = 6 Punkte)

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

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

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 2 (R-Aufgabe, Statistik, 2 + 2 + 2 = 6 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.

(c) Für eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$, wobei $X_1, \ldots, X_n$ unabhängig und identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F_X : X_i \overset{d}{=} X \sim F_X$ und zugehöriger Wahrschienlichkeitsdichte f_X sind, ist die optimale Bandbreite eines Kerndichteschätzers gegeben durch

$$h^* = \frac{c_1^{-2/5} c_2^{1/5} c_3^{-1/5}}{n^{1/5}},$$

wobei $c_1 = \int_{\mathbb{R}} x^2 K(x) dx$, $c_2 = \int_{\mathbb{R}} K(x)^2 dx$ und $c_3 = \int_{\mathbb{R}} (f_X''(x))^2 dx$. Bestimme diese optimale Bandbreite für die Stichprobe aus Teilaufgabe b) und ergänze den zugehörigen Kerndichteschätzer im Plot von Teilaufgabe b). Erweitere die Legende entsprechend.
Hinweis:

- $f_X''(x) = \frac{(x^2-1)\exp(-\frac{x^2}{2})}{\sqrt{2\pi}}$
- Benutze in R die Funktion *integrate*

Aufgabe 3 (2 + 2 = 4 Punkte)

Gegeben seien die Funktionen

$$f(x) = \begin{cases} cx^2 & \text{falls } x \in [0, 1] \\ 0 & \text{sonst} \end{cases}$$

und

$$g(x) = \begin{cases} 1/\sqrt{x} & \text{falls } x \in (0, d] \\ 0 & \text{sonst.} \end{cases}$$

(a) Bestimme c und d so, dass f und g tatsächlich Dichten sind.

(b) Bestimme für eine Zufallsvariable X mit Dichte f die Wahrscheinlichkeiten

$$P\left(X \leq \tfrac{1}{2}\right), \quad P\left(X = \tfrac{1}{2}\right), \quad P\left(X \in \left[\tfrac{1}{3}, \tfrac{2}{3}\right]\right), \quad P\left(X \in \left(\tfrac{1}{3}, 2\right)\right).$$

Aufgabe 4 (2 +1 = 3 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 5 (4 Punkte)

Die Zufallsvariable $N \sim \mathrm{Poi}(\lambda)$, $\lambda > 0$, 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 $Z \sim \mathrm{Poi}(\lambda p)$, falls davon ausgegangen wird, dass sowohl die Ursachen der einzelnen Fehlermeldungen unabhängig sind als auch deren Anzahl N unabhängig von den Fehlerursachen ist.

Aufgabe 6 (4 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 7 (2 + 2 + 2 + 2 = 8 Punkte)

Zu (seriösen) Forschungszwecken wurde ein mittelalterliches Katapult nachgebaut. Es steht $7,40$m vom Ufer eines Baches entfernt, der $5,20$m 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.

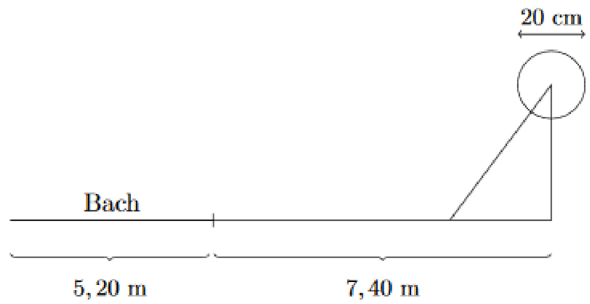
Hinweis: Skaliere die Normalverteilung zu $N(0,1)$ und verwende die Quantiltabelle, die auf Moodle zur Verfügung gestellt wurde.
Für $X \sim N(0,1)$ gilt $\mathbb{P}(X \leq 0.26) = \Phi(0.26) \overset{Tabelle}{=} 0,602568$, wobei $\Phi(x)$ die Verteilungsfunktion von X bezeichnet. Da f_X, die Dichte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



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 $i = 1, \ldots, n$ 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 $X_1, X_2 \sim \mathrm{Exp}(\lambda)$ 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.

05. Juli 2024

Prof. Dr. Evgeny Spodarev
Sabrina Weber

Angewandte Stochastik – Übungsblatt 6

Dieses Übungsblatt wird in der Übung am 05. Juli 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 (3 + 3 + 3 = 9 Punkte)

Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit Wertebereich $X(\Omega)$. Sei weiterhin $Y = g(X)$. Bestimme in den folgenden Fällen $\mathbb{E}[Y^r]$, $r \in \mathbb{N}$, und $\mathrm{Var}(Y)$:

(a) $X(\Omega) = \{0, \pi/2, \pi, 3\pi/2, 2\pi, \dots, 4\pi\} = \{k\pi/2 \mid k = 0, \dots, 8\}$ und X ist gleichverteilt[1], $g(x) = \sin(x)$.

(b) $X(\Omega) = \{0, 1, \dots, n\}$, $n \in \mathbb{N}$ und $X \sim \mathrm{Bin}(n, p)$, $p \in (0, 1)$, $g(x) = 2^x$.

(c) $X(\Omega) = \mathbb{N}_0$ und $X \sim \mathrm{Poi}(\lambda)$, $\lambda > 0$, $g(x) = e^{2x}$.

Aufgabe 2 (2 + 1 = 3 Punkte)

Es sei X eine beliebige Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$[2] und $A \in \mathcal{B}(\mathbb{R})$ eine Teilmenge von $\mathbb{R}$. Die Zufallsvariable Y sei definiert durch $Y = \mathbb{1}_A(X)$. Zeige:

(a) $\mathbb{E}[Y^r] = P(X \in A)$, $r \in \mathbb{N}$

(b) $\mathrm{Var}(Y) = P(X \in A) - P(X \in A)^2$

[1] d.h. X nimmt alle Werte in $X(\Omega)$ mit gleicher Wahrscheinlichkeit an

[2] Grundsätzlich ist es irrelevant, welchen Wahrscheinlichkeitsraum man wählt. Zur Anschauung kannst du einen diskreten Wahrscheinlichkeitsraum verwenden.

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

Für die Aufgabenteile (b), (c) und (d) ist das Vorlesungsvideo 7 relevant.
Es sei eine Zufallsstichprobe von stochastisch unabhängigen identisch verteilten Zufallsvariablen $X_1, \dots, X_n \sim \mathrm{Exp}(\lambda)$ gegeben, wobei $\lambda > 0$ unbekannt sei.

(a) Zeige: $\min\limits_{i=1,\dots,n} X_i \sim \mathrm{Exp}(n\lambda)$.

(b) Zeige, dass $n \cdot \min\limits_{i=1,\dots,n} X_i$ ein erwartungstreuer Schätzer für $1/\lambda$ ist.

(c) Berechne die mittlere quadratische Abweichung des Schätzers aus b) für $1/\lambda$.

(d) Zeige, dass der Schätzer aus b) weder stark noch schwach konsistent für $1/\lambda$ ist.

Aufgabe 4 (3 + 3 + 3 = 9 Punkte)

Es seien X und Y Zufallsvariablen mit $0 < \mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$ und $\mathbb{E}[X] = \mathbb{E}[Y] = 0$.

(a) Für welchen Wert von $a \in \mathbb{R}$ wird $\mathbb{E}[(X - aY)^2]$ minimal?

(b) Zeige, dass die Zufallsvariablen $X - Y$ und $X + Y$ genau dann unkorreliert sind, wenn $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$.

(c) Gib ein Beispiel für X und Y so, dass $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$, aber $X - Y$ und $X + Y$ nicht stochastisch unabhängig sind.

Aufgabe 5 (2 + 1 = 3 Punkte)

Es sei X eine gleichverteilte Zufallsvariable mit Wertebereich $X(\Omega) = \{-2, -1, 0, 1, 2\}$ und $Y = X^2$.

(a) Bestimme die Kovarianz von X und Y

(b) Sind X und Y stochastisch unabhängig?

Aufgabe 6 (je 1 Punkt = 6 Bonuspunkte)

Sei (X, Y) ein zweidimensionaler Zufallsvektor mit gemeinsamer Dichtefunktion $f(x, y)$. Die Dichte der Randverteilungen seinen $f_X(x)$ und $f_Y(y)$. Gib an, welche Aussagen korrekt sind:

Jede richtige Antwort gibt einen Punkt, für jede falsche Antwort wird ein Punkt abgezogen. Es können minimal 0 Punkte erreicht werden.

(a) X und Y sind abhängige Zufallsvariablen, da sie eine gemeinsame Dichtefunktion besitzen.

(b) Sind X und Y unkorreliert, so gilt $f(x, y) = f_X(x) \cdot f_Y(y)$.

(c) Falls $(X, Y) \sim N(\mu, K)$ und $f(x, y) = f_X(x) \cdot f_Y(y)$, so gilt X und Y unkorreliert.

(d) Falls $(X, Y) \sim N(\mu, K)$ und die Nichtdiagonalelemente von K gleich Null sind, so ist der Korrelationskoeffizient von X und Y gleich Null.

(e) Falls $f(x, y) \neq f_X(x) \cdot f_Y(y)$, so ist der Korrelationskoeffizient von X und Y negativ.

(f) Für stochastisch abhängige Zufallsvariablen X und Y sind die standardisierten Zufallsvariablen stochastisch unabhängig.