Stochastik für Informatiker, (6LP)
Musterlösung
02. Oktober 2018

Aufgabe 1 *10 Punkte*

In einem Krankenhaus werden Patienten von drei verschiedenen Ärzten A, B und C behandelt. Arzt A behandelt 50%, Arzt B behandelt 20% und Arzt C 30% der Patienten. Unter den Diagnosen von Arzt A sind 3% falsch, von Arzt B sind 2% der Diagnosen falsch, und von Arzt C sind 5% der Diagnosen falsch.

(a) Definieren Sie sinnvolle Ereignisse zur Bearbeitung der Aufgabe und formulieren Sie die gegebenen Wahrscheinlichkeiten mit Hilfe dieser Ereignisse.

(b) Mit welcher Wahrscheinlichkeit ist eine in diesem Krankenhaus gestellte Diagnose falsch?

(c) Gegeben dass eine falsche Diagnose gestellt wird, mit welcher Wahrscheinlichkeit kommt sie von Arzt A, B oder C?

(d) Um die Diagnostik zu verbessern, kontrolliert neuerdings Arzt B die Diagnosen seiner Kollegen A und C, und entdeckt dabei 80% der falschen Diagnosen. Die Diagnosen von Arzt B werden nicht überprüft. Mit welcher Wahrscheinlichkeit wird nach diesem neuen Verfahren eine falsche Diagnose gestellt?

Lösungsskizze zu Aufgabe 1.

(a) Wir definieren die folgenden Ereignisse:

$$\begin{aligned}
A &:= \text{Patient wird von Arzt A behandelt.}\\
B &:= \text{Patient wird von Arzt B behandelt.}\\
C &:= \text{Patient wird von Arzt C behandelt.}\\
D &:= \text{Die gestellt Diagnose ist falsch.}
\end{aligned}$$

Die Wahrscheinlichkeiten ergeben sich als

$$\begin{aligned}
&\mathbb{P}(A) = \frac{1}{2} && \mathbb{P}(B) = \frac{1}{5} && \mathbb{P}(C) = \frac{3}{10}\\
&\mathbb{P}(D|A) = 0.03 && \mathbb{P}(D|B) = 0.02 && \mathbb{P}(D|C) = 0.05
\end{aligned}$$

(b) Die Wahrscheinlichkeit für eine falsche Diagnose ergibt sich aus dem Gesetz der totalen Wahrscheinlichkeit als

$$\begin{aligned}
\mathbb{P}(D) &= \mathbb{P}(D|A)\mathbb{P}(A) + \mathbb{P}(D|B)\mathbb{P}(B) + \mathbb{P}(D|C)\mathbb{P}(C)\\
&= 0.03 * \frac{1}{2} + 0.02 * \frac{1}{5} + 0.05 * \frac{3}{10} = 0.034
\end{aligned}$$

(c) Die Wahrscheinlichkeiten ergeben sich aus dem Satz von Bayes als

$$\begin{aligned}
\mathbb{P}(A|D) &= \frac{\mathbb{P}(D|A)}{\mathbb{P}(D)}\mathbb{P}(A) = \frac{0.03}{0.034}\frac{1}{2} \approx 0.44\\
\mathbb{P}(B|D) &= \frac{\mathbb{P}(D|B)}{\mathbb{P}(D)}\mathbb{P}(B) = \frac{0.02}{0.034}\frac{1}{5} \approx 0.12\\
\mathbb{P}(C|D) &= \frac{\mathbb{P}(D|C)}{\mathbb{P}(D)}\mathbb{P}(C) = \frac{0.05}{0.034}\frac{3}{10} \approx 0.44
\end{aligned}$$

(d) Die gesuchte Wahrscheinlichkeit ergibt sich als

$$\begin{aligned}\mathbb{P}(D) &= \mathbb{P}(D|A)\mathbb{P}(A) * 0.2 + \mathbb{P}(D|B)\mathbb{P}(B) + \mathbb{P}(D|C)\mathbb{P}(C) * 0.2 \\ &= 0.03 * \frac{1}{2} * 0.2 + 0.02 * \frac{1}{5} + 0.05 * \frac{3}{10} * 0.2 = 0.01\end{aligned}$$

Aufgabe 2 *10 Punkte*

Es sei $(X_n)_{n\in\mathbb{N}}$ eine Markov-Kette auf dem Zustandsraum $\{1,2,3,4\}$ mit Übergangsmatrix

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}.$$

(a) Skizzieren Sie den Übergangsgraphen. Ist die Kette irreduzibel?
(b) Bestimmen Sie die Verteilung von X_2, falls $X_0 = 2$ ist.
(c) Bestimmen Sie alle gegebenenfalls vorhandenen invarianten Verteilungen der Kette.
(d) Was bedeutet das Ergebnis von (c) für das Langzeitverthalten der Kette, d.h. für $n \to \infty$? Erläutern Sie kurz (2-3 Sätze).

Lösungsskizze zu Aufgabe 2.

(a) *Skizze.* Nicht irreduzibel, da man bspw. nicht von Zustand 1 zu Zustand 2 kommen kann.

(b) Die Verteilung von X_2 ergibt sich aus der folgenden Vektor-Matrix-Multiplikation:

$$(0,1,0,0) * P^2 = (1/2, 0, 1/4, 1/4).$$

(c) Sei $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$. Die invarianten Verteilungen ergeben sich aus dem Gleichungssystem $\pi P = \pi$. Die Lösungsmenge ist gegeben durch

$$\Pi = \left\{ \pi \in \mathbb{R}^4 \middle| \pi_1 \in [0,1], \pi_2 = 0, \pi_3 = \pi_4 = \frac{1-\pi_1}{2} \right\}.$$

(d) Das Langzeitverhalten hängt von der Startverteilung ab. Ein Teil der Masse wird von Zustand 1 absorbiert, der Rest verteilt sich gleichmäßig auf die Zustände 3 und 4.

Aufgabe 3 *10 Punkte*

Es sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$, welche das Ergebnis beim einmaligen Werfen eines fairen Würfels beschreibt. Es sei Y die Zufallsvariable auf demselben Wahrscheinlichkeitsraum, welche gegeben ist durch

$$Y(\omega) = \begin{cases} X(\omega) & \text{falls } X(\omega) \text{ gerade ist,} \\ 1 & \text{sonst.} \end{cases}$$

(a) Bestimmen Sie die Verteilung von Y.
(b) Sei nun $Z := 1_{\{X \leq 4\}}$. Geben Sie in einer Tabelle die gemeinsame Verteilung von Y und Z an.
(c) Bestimmen Sie $\mathbb{E}[Y]$ und $\mathbb{E}[Z]$.
(d) Bestimmen Sie $\mathrm{cov}(Y, Z)$. Sind Y und Z unabhängig? Begründen Sie.

Lösungsskizze zu Aufgabe 3.

(a) Der Wertebereich von Y ist gegeben durch $Y(\Omega) = \{1, 2, 4, 6\}$. Die Verteilung ist

$$\begin{aligned}
\mathbb{P}(Y = 1) &= \mathbb{P}(X = 1) + \mathbb{P}(X = 3) + \mathbb{P}(X = 5) = \frac{1}{2} \\
\mathbb{P}(Y = 2) &= \mathbb{P}(X = 2) = \frac{1}{6} \\
\mathbb{P}(Y = 4) &= \mathbb{P}(X = 4) = \frac{1}{6} \\
\mathbb{P}(Y = 6) &= \mathbb{P}(X = 6) = \frac{1}{6}.
\end{aligned}$$

(b) Die gemeinsame Verteilung von Y und Z ist in folgender Tabelle gegeben:

Z \ Y	1	2	4	6
0	$1/6$	0	0	$1/6$
1	$1/3$	$1/6$	$1/6$	0

(c) Die Erwartungswerte ergeben sich als

$$\begin{aligned}
\mathbb{E}[Y] &= 1 * \mathbb{P}(Y = 1) + 2 * \mathbb{P}(Y = 2) + 4 * \mathbb{P}(Y = 4) + 6 * \mathbb{P}(Y = 6) \\
&= 1 * \frac{1}{2} + 2 * \frac{1}{6} + 4 * \frac{1}{6} + 6 * \frac{1}{6} = \frac{5}{2} \\
\mathbb{E}[Z] &= 0 * \mathbb{P}(Z = 0) + 1 * \mathbb{P}(Z = 1) \\
&= \mathbb{P}(Z = 1) = \frac{2}{3}.
\end{aligned}$$

(d) Die Kovarianz ergibt sich als

$$\begin{aligned}
\mathrm{cov}(Y, Z) &= \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z] = 1 * \mathbb{P}(YZ = 1) + 2 * \mathbb{P}(YZ = 2) + 4 * \mathbb{P}(YZ = 4) - \frac{2}{3} * \frac{5}{2} \\
&= 1 * \frac{1}{3} + 2 * \frac{1}{6} + 4 * \frac{1}{6} - \frac{5}{3} = -\frac{1}{3}.
\end{aligned}$$

Daraus folgt, dass Y und Z korreliert sind und somit nicht unabhängig.

Aufgabe 4 *10 Punkte*

Gegeben sei die Funktion $f : \mathbb{R} \to \mathbb{R}$ mit Parametern $a, b \in \mathbb{R}$ und

$$f(x) = \begin{cases} ax^2 & 0 \leq x \leq 1, \\ b & -1 \leq x < 0, \\ 0 & \text{sonst.} \end{cases}$$

(a) Bestimmen Sie alle Paare $a, b \in \mathbb{R}$, so dass es sich bei f um eine Dichte handelt.
(b) Bestimmen Sie, in Abhängigkeit von a und b, den Erwartungswert $\mathbb{E}[X]$ der zugehörigen Zufallsvariablen X.
(c) Bestimmen Sie, in Abhängigkeit von a und b, die Verteilungsfunktion F_X.
(d) Es sei nun $b = 0$. Skizzieren Sie f und markieren Sie in der Skizze die Wahrscheinlichkeit $\mathbb{P}(X \geq 1/2)$.

Lösungsskizze zu Aufgabe 4.

(a) Es muss gelten $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ und $f(x) \geq 0$ für alle $x \in \mathbb{R}$. Also

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-1}^{0} b\mathrm{d}x + \int_{0}^{1} ax^2\mathrm{d}x = b + \frac{a}{3}x^3\Big|_0^1 = b + \frac{a}{3} = 1,$$

also $a = 3(1 - b)$. Mit der Bedingung $f(x) \geq 0$, für alle $x \in \mathbb{R}$, folgt als Lösungsmenge

$$\mathbb{L} = \{(a, b) \in \mathbb{R} | b \in [0, 1], a = 3(1 - b)\}.$$

(b) Der Erwartungswert ergibt sich als

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x = \int_{-1}^{0} bx \mathrm{d}x + \int_{0}^{1} ax^3 \mathrm{d}x = -\frac{b}{2} + \frac{a}{4}.$$

(c) Zur Bestimmung der Verteilungsfunktion machen wir eine Fallunterscheidung:
1. Fall $t \leq -1$:

$$F_X(t) = \int_{-\infty}^{t} f(x) \mathrm{d}x = 0.$$

2. Fall $t \in (-1, 0]$:

$$F_X(t) = \int_{-\infty}^{-1} f(x) \mathrm{d}x + \int_{-1}^{t} f(x) \mathrm{d}x = b(t+1).$$

3. Fall $t \in (0, 1]$:

$$F_X(t) = \int_{-\infty}^{-1} f(x) \mathrm{d}x + \int_{-1}^{0} f(x) \mathrm{d}x + \int_{0}^{t} f(x) \mathrm{d}x = b + \frac{a}{3} t^3.$$

4. Fall $t > 1$:

$$F_X(t) = \int_{-\infty}^{-1} f(x) \mathrm{d}x + \int_{-1}^{0} f(x) \mathrm{d}x + \int_{0}^{1} f(x) \mathrm{d}x + \int_{1}^{t} f(x) \mathrm{d}x = 1.$$

(d) *Graph.* $\mathbb{P}(X \geq 1/2)$ ist die Fläche unterhalb der Dichte ab $x = 1/2$.

Aufgabe 5 *10 Punkte*

Ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p = 0.2$ wird wiederholt unter gleichbleibenden Bedingungen ausgeführt. Es sei X die Anzahl Wiederholungen bis zum ersten Erfolg, und Y die Anzahl Erfolge bei $n = 100$ Wiederholungen.

(a) Welche Verteilungen haben die Zufallsvariablen X bzw. Y?

(b) Beweisen Sie: $\mathbb{P}(X \geq n) = 0.8^{n-1}$ für $n \in \mathbb{N}$.

(c) Berechnen Sie $\mathbb{P}(15 \leq Y \leq 25)$ approximativ mit Hilfe der Normalverteilung. Warum darf diese Approximation verwendet werden?

(d) Es sei Z Poisson-verteilt mit Parameter λ. Wie müssen Sie λ wählen, damit $\mathbb{P}(Y = k) \approx \mathbb{P}(Z = k)$, $k \in Y(\Omega)$, gilt? Welcher Satz liegt hier zu Grunde?

Lösungsskizze zu Aufgabe 5.

(a) X ist geometrisch verteilt mit Parameter $p = 0.2$, Y ist binomialverteilt mit Parametern $p = 0.2$ und $n = 100$.

(b) Da $X \sim$Geo(0.2) ist, gilt mit der endlichen geometrischen Reihe

$$\begin{aligned}\mathbb{P}(X \geq n) = 1 - \mathbb{P}(X < n) = 1 - \mathbb{P}(X \leq n-1) &= 1 - \sum_{k=0}^{n-2} 0.2(0.8)^k \\ &= 1 - 0.2 \frac{1 - 0.8^{n-1}}{1 - 0.8} = 1 - (1 - 0.8^{n-1}) = 0.8^{n-1}.\end{aligned}$$

Alternative Lösung: $\mathbb{P}(X \geq n)$ bedeutet, dass man erst $n-1$ Misserfolge hatte. Was danach geschieht ist egal. Da die Wahrscheinlichkeit für Misserfolg $1 - p = 0.8$ ist, folgt die Behauptung.

(c) Da Y ~Bin(100,0.2) ist, ist $\mathbb{E}[Y] = np = 20$, $\text{var}(Y) = np(1-p) = 16$ folgt (nach dem zentralen Grenzwertsatz bzw. wie in Beispiel 7.6) für $Z \sim \mathcal{N}(0,1)$

$$\begin{aligned}\mathbb{P}(15 \leq Y \leq 25) =&\mathbb{P}\left(\frac{15-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{25-\mu}{\sigma}\right) \approx \mathbb{P}\left(\frac{15-20}{4} \leq Z \leq \frac{25-20}{4}\right)\\ =&\mathbb{P}(-1.25 \leq Z \leq 1.25) = 2\Phi_{0,1}(1.25) - 1 = 2 \cdot 0.8944 - 1 = 0.7888.\end{aligned}$$

Die Approximation darf wegen des zentralen Grenzwertsatzes verwendet werden, und weil die Faustregel $np \geq 5$ und $n(1-p) \geq 5$ erfüllt ist.

Alternative Rechnung unter der Verwendung der 1/2-Korrektur (Satz 7.7 im Skript): $14.4 \leq Y \leq 25.5$, also

$$\begin{aligned}\mathbb{P}(14.4 \leq Y \leq 25.5) =&\mathbb{P}\left(\frac{14.5-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{25.5-\mu}{\sigma}\right) \approx \mathbb{P}\left(\frac{14.5-20}{4} \leq Z \leq \frac{25.5-20}{4}\right)\\ =&\mathbb{P}(-1.375 \leq Z \leq 1.375) \approx 2\Phi_{0,1}(1.38) - 1 = 2 \cdot 0.9162 - 1 = 0.8324.\end{aligned}$$

(d) $\lambda = np = 100 \cdot 0.2 = 20$. Poisson-Grenzwertsatz.