Stochastik für Informatik(er) - Lösungsvorschlag für die Hauptklausur

01. August 2024

Aufgabe 1

(a) 2 Punkte - 1 Punkt pro $P(Y = \cdot)$ Da

$$1 = \mathbb{E}[Y] = 0 \cdot \mathbb{P}(Y = 0) + 1 \cdot \mathbb{P}(Y = 1) + 2 \cdot \mathbb{P}(Y = 2) = \frac{1}{2} + 2 \cdot \mathbb{P}(Y = 2)$$

erhalten wir $\mathbb{P}(Y = 2) = \frac{1}{4}$. Deshalb ist $\mathbb{P}(Y = 0) = 1 - \mathbb{P}(Y = 1) - \mathbb{P}(Y = 2) = \frac{1}{4}$.

(b) 2 Punkte - 0.5 Punkte Abzug bei einem Fehler, 1 Punkt Abzug bei 3 Fehlern, 1.5 Punkte Abzug bei 5 Fehlern Vollständige Tabelle:

X \ Y	0	1	2	$\mathbb{P}(X = x)$
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{16}$	$\frac{1}{16}$	0	$\frac{1}{8}$
2	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{3}{8}$
$\mathbb{P}(Y = y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

(c) 2 Punkte Wir berechnen

$$0.5P \quad \mathbb{E}[X] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{3}{8} = \frac{7}{8},$$

$$0.5P \quad \mathbb{E}[XY] = \sum_{x,y} xy\mathbb{P}(X = x, Y = y) = 1 \cdot \frac{1}{16} + 2 \cdot \frac{3}{16} + 4 \cdot \frac{1}{8} = \frac{15}{16}.$$

Damit folgt:

$$0.5P \quad \mathrm{Cov(X, Y)} = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{15}{16} - \frac{7}{8} = \frac{1}{16}.$$

0.5 P Der Korrelationskoeffizient hat dasselbe Vorzeichen wie die Kovarianz, also sind X und Y positiv korreliert.

(d) 2 Punkte - 1 für Antwort, 1 für Begründung Nein, weil z.B $\mathbb{P}(X = 2, Y = 1) = 0 \neq \mathbb{P}(X = 2)\mathbb{P}(Y = 1)$.

(e) 2 Punkte - 0.5 Abzug bei falscher Berechnung Man normaliziert die Zeile $Y = 0$:

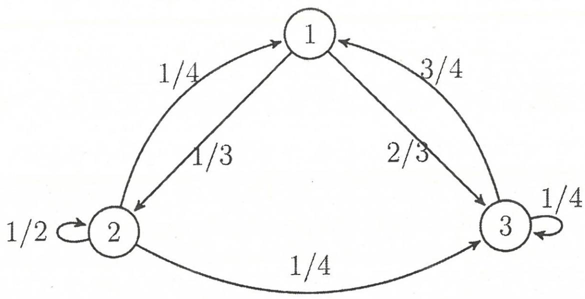
$$\mathbb{P}(X = 0|Y = 0) = \frac{\mathbb{P}(X = 0, Y = 0)}{\mathbb{P}(Y = 0)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}.$$

Analog für $X = 1$ und $X = 2$:

$$\mathbb{P}(X = 1|Y = 0) = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4}, \quad \mathbb{P}(X = 2|Y = 0) = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4}.$$

Aufgabe 2

(a) 1 Punkt Übergangsgraph:



(b) 3 Punkte - 1.5 P für irredutibel , 1.5 P für aperiodisch Die Markov-Kette ist irreduzibel, weil es von jedem Zustand zu jedem anderen Zustand einen Pfad gibt. Sie ist aperiodisch, da

$$\{n \geq 1 : p_{11}^{(n)} > 0\} = \{2, 3, 4, \dots\} \qquad \text{ggT} = 1,$$
$$\{n \geq 1 : p_{22}^{(n)} > 0\} = \{1, 2, 3, 4, \dots\} \qquad \text{ggT} = 1,$$
$$\{n \geq 1 : p_{33}^{(n)} > 0\} = \{1, 2, 3, 4, \dots\} \qquad \text{ggT} = 1.$$

(c) 4 Punkte Es handelt sich um eine irreduzible Kette auf einem endlichen Zustandsraum, daher hat sie eine 0.5 P eindeutige invariante Verteilung π. Die i.V. ist die einzige Lösung des Gleichungssystems 0.5 P

$$\pi(P - I) = 0,$$
$$\pi_1 + \pi_2 + \pi_3 = 1.$$

Wir erhalten:

$$-\pi_1 + \frac{1}{4}\pi_2 + \frac{3}{4}\pi_3 = 0,$$
$$\frac{1}{3}\pi_1 - \frac{1}{2}\pi_2 + 0\pi_3 = 0,$$
$$\frac{2}{3}\pi_1 + \frac{1}{4}\pi_2 - \frac{3}{4}\pi_3 = 0,$$
$$\pi_1 + \pi_2 + \pi_3 = 1.$$

Aus der 2. Gleichung folgt $\pi_1 = \frac{3}{2}\pi_2$. Setzen wir dies in die 3. Gleichung ein: $\frac{5}{4}\pi_2 = \frac{3}{4}\pi_3 \implies \pi_3 = \frac{5}{3}\pi_2$. Deshalb

$$1 = \pi_1 + \pi_2 + \pi_3 = \frac{3}{2}\pi_2 + \pi_2 + \frac{5}{3}\pi_2 = \frac{25}{6}\pi_2 \implies \pi_2 = \frac{6}{25}.$$

Dann $\pi_1 = \frac{3}{2}\pi_2 = \frac{9}{25}$ und $\pi_3 = \frac{5}{3}\pi_2 = \frac{2}{5}$. Lösung des LGS: 3 P (pro π_i 1P)

(d) 2 Punkte Die Startverteilung ist die invariante Verteilung,

$$\mathbb{P}_\nu(X_4 = 1) = \pi_1 = \frac{9}{25},$$
$$\mathbb{P}_\nu(X_4 = 2) = \pi_2 = \frac{6}{25},$$
$$\mathbb{P}_\nu(X_4 = 3) = \pi_3 = \frac{2}{5}.$$

Man kann auch die Matrixpotenz berechnen.

Aufgabe 3

(a) 2 Punkte Es ist bekannt, dass der 1 P Erwartungswert und die 1 P Varianz einer Poisson-verteilten Zufallsvariable gegeben sind durch

$$\mathbb{E}[X] = \lambda = 10, \quad \sigma(X) = \sqrt{\mathbb{V}[X]} = \sqrt{10}\,.$$

(b) 4 Punkte - 1 P für Definition der ZV S_{72}, 1 P für Identifikation von richtiger Standardisierung, 2 P für Berechnung Seien $X_1, \ldots, X_{72}$ die Anzahl an Fahrzeugen, die in jeder Stunde beobachtet wird. Mit dem Gesetz der großen Zahlen folgt

$$Z = \frac{S_{72} - 72 \times 10}{\sqrt{10 \times 72}} \sim \mathcal{N}(0, 1),$$

wobei

$$S_{72} = \sum_{i=1}^{72} X_i$$

die gemeinsame Anzahl Autos im Zeitraum ist. Deshalb

$$\mathbb{P}(S_{72} \geq 750) \approx 1 - \mathbb{P}\left(Z \leq \frac{750 - 720}{\sqrt{720}}\right) = 1 - \Phi(1.12) = 1 - 0.8686 = 0.1314.$$

(c) 4 Punkte - 2 P für Erkennen des Problems, 2 P für Berechnung Sei n die Anzahl Fahrzeuge eines solchen 3-tagigen Zeitraums. Sie erfüllt die Gleichung

$$\mathbb{P}(S_{72} \geq n) = 0.05$$

Mit einer ähnlichen Rechnung wie im obigen Teil:

$$0.05 = 1 - \Phi\left(\frac{n - 720}{\sqrt{720}}\right) \implies n = 720 + \sqrt{720}\Phi(0.95) \approx 764.$$

Aufgabe 4

(a) 2 Punkte - 0.5 P Abzug pro falsche Berechnung, wenn nicht FF Wir nutzen die Normalisierungseigenschaft aus und erhalten:

$$c\int_{-1}^{1}(1 - x^2)dx = c\left(\int_{-1}^{1} dx - \int_{-1}^{1} x^2 dx\right) = \frac{4}{3}c \overset{!}{=} 1 \Leftrightarrow c = \frac{3}{4}$$

(b) 2 Punkte - 0.5 P Abzug pro falsche Berechnung, wenn nicht FF Die Verteilungsfunktion ist rechtsstetig und monoton steigend. Damit erhalten wir: Für $x < -1$ ist $F_X(x) = 0$, für $x > 1$ ist $F_X(x) = 1$ und für $x \in [-1, 1]$ gilt:

$$\begin{aligned} F_X(x) &= c\int_{-1}^{x}(1 - t^2)dt = c\left(\int_{-1}^{x} dt - \int_{-1}^{x} t^2 dt\right) \\ &= c\left(-\frac{1}{3}x^3 + x + \frac{2}{3}\right) \overset{c=\frac{3}{4}}{=} -\frac{1}{4}x^3 + \frac{3}{4}x + \frac{1}{2} \end{aligned}$$

(c) 3 Punkte - 0.5 P Abzug pro falsche Berechnung, wenn nicht FF Man berechnet

$$\mathbb{P}(Z \leq z) = \mathbb{P}(\sqrt{Y} \leq z) = \mathbb{P}(Y \leq z^2) = 1 - e^{-z^2}.$$

Ableiten nach z ergibt die Dichte:

$$f_Z(z) = 2ze^{-z^2}.$$

(d) 3 Punkte - 0.5 P Abzug pro falsche Berechnung, wenn nicht FF Nach der Definition:

$$\mathbb{E}[Z] = \int_0^\infty 2z^2e^{-z^2}\,\mathrm{d}z = \int_{-\infty}^{\infty} z^2e^{-z^2}\,\mathrm{d}z = \frac{1}{2\sqrt{2}}\int_{-\infty}^{\infty} t^2e^{-t^2/2}\,\mathrm{d}t = \frac{1}{2\sqrt{2}} \times \sqrt{2\pi} = \frac{\sqrt{\pi}}{2}$$

Aufgabe 5 - 6LP Seien B_i, R_i die Ergebnisse der i-te Ziehung.

(a) 3 Punkte - 1 P für Identifikation was gesucht wird, 1 P für Nutzen der richtigen Formel, 1 P für richtige Berechnung Mit dem Gesetz der totalen Wahrscheinlichkeiten:

$$\mathbb{P}(B_2) = \mathbb{P}(B_2 \mid R_1)\mathbb{P}(R_1) + \mathbb{P}(B_2 \mid B_1)\mathbb{P}(B_1) = \frac{1}{3} \cdot \frac{4}{6} + \frac{2}{3} \cdot \frac{2}{6} = \frac{4}{9}.$$

(b) 5 Punkte - 1 P für Identifikation was gesucht wird, 2 P für Nutzen der richtigen Formel, 2 P für richtige Berechnung Aus der Bayes'sche Umkehrformel folgt:

$$\mathbb{P}(B_1 \mid B_2) = \frac{\mathbb{P}(B_2 \mid B_1)\mathbb{P}(B_1)}{\mathbb{P}(B_2)} = \frac{\frac{2}{3} \cdot \frac{1}{3}}{\frac{4}{9}} = \frac{1}{2}.$$

(c) 2 Punkte - 0.5 P für Identifikation was gesucht wird, 0.5 P für Nutzen der richtigen Formel, 1 P für richtige Berechnung Wir berechnen die gemeinsame Wahrscheinlichkeit

$$P(R_2 \text{ und } B_1) = P(R_2 \mid B_1) \cdot P(B_1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

Aufgabe 5 - 9LP

(a) 1 Punkt - 0.5 P pro $\bar{\mu}$ Leicht rechnet man:

$$\bar{\mu}_{10}(x) = \frac{6+6+6+2+2+5+4+5+1+4}{10} = 4.1$$
$$\bar{\mu}_{10}(y) = \frac{125+115+130+260+219+150+190+163+260+160}{10} = 177.2$$

(b) 2 Punkte Wir erhalten die empirische Varianz mittels

$$1P \quad \bar{\sigma}_n^2(x) = \frac{1}{n-1}\sum_i (x_i - \bar{\mu}_n(x))^2 = 3.43$$

Damit bestimmten wir den Korrelationskoeffizienten

$$1P \quad \bar{r}_{10}(x,y) = \frac{\bar{c}_n(x,y)}{\bar{\sigma}_n(x)\bar{\sigma}_n(y)} = -0.92$$

(c) 2 Punkte - 1 P pro Koeffizient Die Koeffizienten sind durch

$$\bar{a} = \bar{r}_n(x,y)\frac{\bar{\sigma}_n(y)}{\bar{\sigma}_n(x)} = -25.24$$
$$\bar{b} = \bar{\mu}_n(y) - \bar{a}\bar{\mu}_n(x) = 276.58$$

gegeben.

(d) 1 Punkt - 0.5 P richtiges Einsetzen, 0.5 P richtiges Ergebnis Schätzung durch die Gerade:

$$\bar{y} = \bar{a}\cdot 4 + \bar{b} = 175.62,$$

also 17, 562.00 Euro.

(e) 4 Punkte 0.5 P Aus der Behauptung folgt direkt, dass wir einen beidseitigen Hypothesentest durchführen möchten, da wir Gleichheit vorhanden haben. Da wir die empirische Standardabweichung vorliegen haben, wählen wir außerdem einen t-Test.

0.5 P Die Hypothesen definieren wir als $H_0 : \mu_y = 200$ vs. $\mu_y \neq 200$.

1 P Die Teststatistik ist gegeben als

$$T = \frac{\mu_y - \bar{\mu}_y}{\bar{\sigma}_y}\sqrt{n} \approx -1.42$$

1 P Der kritische Wert ist

$$t_{1-\frac{\alpha}{2},n-1} = t_{1-0.025,9} = 2.262$$

1 P Damit können wir schließen: Es gilt $|-1.42| < 2.626$. Wir nehmen H_0 an.