TU Berlin, Institut für Mathematik Prof. Dr. Noemi Kurt
SoSe 19 András Tóbiás

Stochastik für Informatik, (9LP)
Klausur

19. Juli 2019

Musterlösung für die Einsicht

Aufgabe 1.1 (gemeinsame Verteilung) *10 Punkte*

Eine Urne enthält 4 rote und 6 weiße Kugeln. Es wird zwei Mal mit Zurücklegen gezogen. X bezeichne die Anzahl er roten gezogenen Kugeln, Y die Anzahl der weißen gezogenen Kugeln.

(a) Geben Sie die gemeinsame Verteilung von X und Y mit Randverteilung in einer Tabelle an.

(b) Berechnen Sie $\mathbb{E}[X], \mathbb{E}[Y]$ und $\operatorname{cov}(X, Y)$. Sind X und Y unabhängig?

(c) Berechnen Sie $\mathbb{E}[e^{2X+Y}]$.

Lösungsskizze zu Aufgabe 1.1.

(a)

X \ Y	0	1	2	p_X
0	0	0	0.36	0.36
1	0	0.48	0	0.48
2	0.16	0	0	0.16
p_Y	0.16	0.48	0.36	(1)

(b) $\mathbb{E}X = \sum_{k=0}^{2} k\mathbb{P}(X = k) = 0.48 + 2 \cdot 0.16 = 0.8$, $\mathbb{E}Y = \sum_{k=0}^{2} k\mathbb{P}(Y = k) = 0.48 + 2 \cdot 0.36 = 1.2$,
$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = 0,48 - 0,96 = -0,48$, denn $\mathbb{P}(XY = 1) = 1 - \mathbb{P}(XY = 0) = \mathbb{P}(X = Y = 1) = 0.48$ und damit $\mathbb{E}(XY) = 0.48$.
Da $\operatorname{cov}(X, Y) \neq 0$, sind X und Y nicht unabhängig.

(c) $\mathbb{P}(2X + Y) = 2 = \mathbb{P}(X = 0, Y = 2) = 0.36$, $\mathbb{P}(2X + Y = 3) = \mathbb{P}(X = Y = 1) = 0.48$, $\mathbb{P}(2X + Y = 4) = \mathbb{P}(X = 2, Y = 0) = 0.16$. Daraus folgt $\mathbb{E}[e^{2X+Y}] = \sum_{k=2}^{4} e^k \mathbb{P}(2X + Y = k) = e^2 \cdot 0.36 + e^3 \cdot 0.48 + e^4 \cdot 0.16 \approx 21.04$

Aufgabe 1.2 (Markov-Ketten) *10 Punkte*

Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette auf $S = \{1, 2, 3\}$ mit Übergangsmatrix

$$P = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \\ 0 & 3/4 & 1/4 \end{bmatrix}.$$

(a) Zeichnen Sie den Übergangsgraphen. Ist die Markov-Kette irreduzibel?

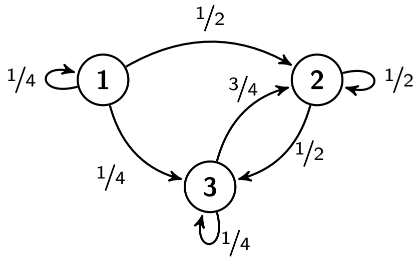
(b) Sei $Y = \min\{n \in \mathbb{N}_0 : X_n \neq 1\}$. Welche Verteilung hat Y, wenn $X_0 = 1$ ist? Berechnen Sie außerdem $\mathbb{P}(X_n = 1 \mid X_0 = 1), n \in \mathbb{N}$.

(c) Bestimmen Sie alle gegebenenfalls vorhandenen invarianten Verteilungen zu dieser Übergangsmatrix

(d) Berechnen Sie $\lim_{n\to\infty} \mathbb{P}(X_n = j \mid X_0 = 1), j \in \{1, 2, 3\}$.

Lösungsskizze zu Aufgabe 1.2.

(a)



Die Kette ist nicht irreduzibel, z.B. es gibt keinen Pfad von 2 nach 1 oder 3 nach 1.

(b) Wenn $X_0 = 1$, ist Y geometrisch verteilt mit Parameter $1 - \mathbb{P}(X = 1|Y = 1) = 3/4$. Falls $X_0 = 1$, dann für $n > 1$ gilt $X_n = 1$ genau dann, wenn $X_1 = X_2 = \ldots = X_{n-1} = 1$. Damit gilt

$$\mathbb{P}(X_n = 1|X_0 = 1) = \mathbb{P}(X_n = X_{n-1} = \ldots = X_1 = 1|X_0 = 1) = \mathbb{P}(Y > n|X_0 = 1) = (1-3/4)^n = \frac{1}{4^n}.$$

(c) Sei $\pi = (\pi_0, \pi_1, \pi_2)^T$ eine invariante Verteilung. $\pi^T P = \pi^T$ liefert $\pi_1 = 0$, $\pi_2/2 + 3\pi_3/4 = \pi_2$, also $\pi_3 = 2\pi_2/3$. Da $\pi_1 + \pi_2 + \pi_3 = 1$, daraus folgt, dass $\pi_2 + 2\pi_2/3 = 1$, also $\pi_2 = 3/5$ und $\pi_3 = 2/5$. Also ist $\pi = (0, 3/5, 2/5)$ die eindeutige invariante Verteilung.

(d) $\lim_{n\to\infty} \mathbb{P}(X_n = j|X_0 = 1) = \pi_j$, für $j = 1, 2, 3$, also 0 für $j = 1$, $3/5$ für $j = 2$ und $2/5$ für $j = 3$.

Aufgabe 1.3 *10 Punkte*

Sei X eine Zufallsvariable mit Verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & \text{falls } x < 0, \\ \frac{x^3}{8} & \text{falls } 0 \le x \le 2, \\ 1 & \text{falls } x > 2. \end{cases}$$

(a) Berechnen Sie den Wert der Dichte von X im Punkt $x = 0.5$.

(b) Berechnen Sie den Erwartungswert $\mathbb{E}[X]$ und die Varianz $\mathbb{V}[X]$.

(c) Es sei Y eine von X abhängige Zufallsvariable mit $\mathbb{V}[Y] = 1$ und $\mathbb{V}[X+Y] = \frac{1}{2}$. Berechnen Sie die Kovarianz sowie die Korrelation von X und Y.

Lösungsskizze zu Aufgabe 1.3.

(a)

$$f_X(0.5) = F_X'(0.5) = \frac{\mathrm{d}}{\mathrm{d}x}\Big|_{x=0.5} \frac{x^3}{8} = \Big(\frac{3}{8}x^2\Big)\Big|_{x=0.5} = \frac{3}{32}.$$

(b) Mit der letzten Teilaufgabe berechneten Dichte gilt

$$\mathbb{E}X = \int_0^2 x f_X(x)\mathrm{d}x = \int_0^2 x\frac{3x^2}{8}\mathrm{d}x = \Big[\frac{3x^4}{32}\Big]_0^2 = \frac{48}{32} = \frac{3}{2}.$$

$$\mathbb{E}[X^2] = \int_0^2 x^2 f_X(x)\mathrm{d}x = \int_0^2 \frac{3}{8}x^4\mathrm{d}x = \Big[\frac{3}{40}x^5\Big]_0^2 = \frac{96}{40} = \frac{12}{5},$$

und da $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}X)^2$, daraus folgt $\mathbb{V}[X] = \frac{12}{5} - \frac{9}{4} = 0.15$.

(c) Wir berechnen $\operatorname{cov}(X,Y)$ aus der Gleichung $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\operatorname{cov}(X,Y)$. Dann folgt

$$\operatorname{cov}(X,Y) = \frac{1}{2}\Big(\mathbb{V}(X+Y) - \mathbb{V}(X) - \mathbb{V}(Y)\Big) = \frac{1}{2}\Big(\frac{1}{2} - 1 - 0.15\Big) = -0.325.$$

Korrelation:

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}} = \frac{-0.325}{\sqrt{0.15}} \approx -0.839.$$

Aufgabe 1.4 (Verteilungen) *10 Punkte*

Ein stochastischer Algorithmus terminiert mit einer Wahrscheinlichkeit von $p = 0.2$.

(a) Der Algorithmus wird 150 Mal aufgerufen. Sei X die Anzahl Läufe, bei denen er terminiert. Welche Verteilung hat X? Berechnen Sie $\mathbb{E}[X]$ und $\mathbb{V}(X)$.

(b) Verwenden Sie eine geeignete Approximation, um in der Situation von (a) die Wahrscheinlichkeit für höchstens 20 Terminierungen zu berechnen.

(c) Y bezeichne die Anzahl Läufe, bis der Algorithmus zum ersten Mal terminiert. Welche Verteilung hat Y? Berechne $\mathbb{P}(Y \geq 10)$.

(d) Wie viele Male muss der Algorithmus aufgerufen werden, damit mit einer Wahrscheinlichkeit von mindestens 40% mindestens eine Terminierung stattfindet?

Lösungsskizze zu Aufgabe 1.4.

(a) $X \sim \operatorname{Bin}(n, p)$, wobei $n = 150$ und $p = 0.2$. Daraus folgt $\mathbb{E}X = np = 30$ und $\mathbb{V}X = np(1-p) = 24$.

(b)

$$\mathbb{P}(X \leq 20) = \mathbb{P}\Big(\frac{X-30}{\sqrt{24}} \leq \frac{20-30}{\sqrt{24}}\Big) \approx \mathbb{P}\Big(Y \leq \frac{-10}{\sqrt{24}}\Big) \approx \mathbb{P}(Y \leq -2.04) = \mathbb{P}(Y \geq 2.04) \approx 0.0207,$$

wobei Y standardnormalverteilt ist. [Alternativ könnte man $\mathbb{P}(X \leq 20) = \mathbb{P}(0 \leq X \leq 20)$ approximieren. Der Wert $\mathbb{P}(Y \geq 6.12)$ steht nicht in der Tabelle, aber ist offensichtlich kleiner als 0.0001 nach der Tabelle. Analog mit 1/2-Korrektur: Endergebnis $1 - \Phi(1.94) = 0.0268$.]

(c) Y ist geometrisch verteilt mit Parameter $p = 0.2$. Daraus folgt $\mathbb{P}(Y \geq 10) = \mathbb{P}(Y > 9) = (1-p)^9 = 0.8^9 \approx 0.13$.
[$\mathbb{P}(Y > n) = (1-p)^n$ muss man nicht auswendig kennen, man kann sie mit der geometrischen Reihe oder auch direkt mit dem Taschenrechner berechnen.]

(d) Gesucht ist das kleinste n, sodass $\mathbb{P}(Y \leq n) \geq 0.4$. Da $\mathbb{P}(Y \leq 1) = \mathbb{P}(Y = 1) = 0.2$, $\mathbb{P}(Y \leq 2) = \mathbb{P}(Y = 1) + 0.2 \cdot 0.8^2 = 0.36$, und $\mathbb{P}(Y \leq 3) = \mathbb{P}(Y \leq 2) + 0.2 * 0.8^2 = 0.488$, ist $n = 3$ die Lösung.

Aufgabe 1.5 *10 Punkte*

Die nächste Tabelle zeigt die Anzahl der Spätkauf-Läden und Dönerläden in den Berliner Bezirken. Ziel dieser Aufgabe ist es, die Anzahl Spätis und die Anzahl Dönerläden auf Unabhängigkeit zu untersuchen.

Bezirk	Anzahl Spätis (x)	Anzahl Dönerläden (y)
Charlottenburg-Wilmersdorf	47	19
Friedrichshain-Kreuzberg	123	47
Lichtenberg	22	30
Marzahn-Hellersdorf	19	27
Mitte	140	35
Neukölln	80	40
Pankow	64	32
Reinickendorf	35	35
Spandau	27	32
Steglitz-Zehlendorf	29	31
Tempelhof-Schöneberg	57	32
Treptow-Köpenick	48	29

Wir gruppieren die x-Daten in zwei Gruppen: (1) $x < 55$ und (2) $x \geq 55$ und die y-Daten in zwei Gruppen: (1) $y < 33$ und (2) $y \geq 33$.

(a) Berechnen Sie die empirischen Häufigkeiten $N_{i,j}$ sowie die Randhäufigkeiten $N_{i,*}$ und $N_{*,j}$ für $i, j = 1, 2$. Stellen Sie das Ergebnis in einer Tabelle dar.

(b) Berechnen Sie damit die theoretischen Häufigkeiten $F_{i,j}$ für $i, j = 1, 2$.

(c) Testen Sie zum Fehlerniveau $\alpha = 0.05$, ob die Anzahl der Spätis unabhängig von der Anzahl der Dönerladen sind.

(d) Wählen Sie die korrekte Antwort ohne Begründung aus: falls der χ^2-Test auf Unabhängigkeit zum beliebigen Fehlerniveau $\alpha \in (0, 1)$ eine Ablehnung der Unabhängigkeit liefert, ist der p-Wert immer kleiner als

(i) $\alpha/2$ (ii) α (iii) $1 - \alpha$ (iv) $1 - \alpha/2$.

Lösungsskizze zu Aufgabe 1.5.

(a)

X \ Y	1	2	Total
1	6	1	7
2	2	3	5
Total	8	4	(12)

(b) Die theoretischen Häufigkeiten sind $F_{i,j} = \frac{N_{i,*}N_{*,j}}{n}$, also

X \ Y	1	2	Total
1	$\frac{14}{3}$	$\frac{7}{3}$	7
2	$\frac{10}{3}$	$\frac{5}{3}$	5
Total	8	4	(12)

(c) Die Teststatistik ist

$$\chi^2 = \sum_{i,j} \frac{(F_{i,j} - N_{i,j})^2}{F_{i,j}} = \frac{16}{42} + \frac{16}{21} + \frac{8}{15} + \frac{16}{15} = 2.74.$$

Das muss man mit der $1 - \alpha = 0.95$-Quantil $\chi^2_{1-\alpha,f}$ der χ^2-Verteilung mit Freiheitsgrad $(k-1)(m-1) = 1$ vergleichen. In diesem Fall ist dieses Quantil $\chi^2_{0.95,1} = 3.84$. Da $\chi^2 \le \chi^2_{0.95,1}$, wird H_0 angenommen, also sind die x-Daten und die y-Daten nicht abhängig.

(d) (ii).

Tabelle: $\Phi_{0,1}(x) = \mathbb{P}(X \le x)$ für eine standardnormalverteilte Zufallsvariable X

z_α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998