Klausur SS2020 Fragen und Antworten

Diskrete Wahrscheinlichkeitstheorie (IN0018) (Technische Universität München)

Lehrstuhl für Algorithmen und Komplexität
Fakultät für Informatik
Technische Universität München

Esolution
Sticker mit SRID hier einkleben

Hinweise zur Personalisierung:

- Ihre Prüfung wird bei der Anwesenheitskontrolle durch Aufkleben eines Codes personalisiert.
- Dieser enthält lediglich eine fortlaufende Nummer, welche auch auf der Anwesenheitsliste neben dem Unterschriftenfeld vermerkt ist.
- Diese wird als Pseudonym verwendet, um eine eindeutige Zuordnung Ihrer Prüfung zu ermöglichen.

Diskrete Wahrscheinlichkeitstheorie

Klausur:	IN0018 / Endterm	**Datum:**	Mittwoch, 5. August 2020
Prüfer:	Prof. Dr. Susanne Albers	**Uhrzeit:**	14:15 – 16:15

Bearbeitungshinweise

- Diese Klausur umfasst **18 Seiten** mit insgesamt **6 Aufgaben**.
  Bitte kontrollieren Sie jetzt, dass Sie eine vollständige Angabe erhalten haben.
- Die Gesamtpunktzahl in dieser Prüfung beträgt 40 Punkte.
- Das Heraustrennen von Seiten aus der Prüfung ist untersagt.
- Als Hilfsmittel sind zugelassen:
  - Vorlesungsfolien und Übungsblätter, inkl. selbst geschriebener Zusammenfassungen und Notizen in beliebigem Umfang
  - ein **analoges Wörterbuch** Deutsch $\leftrightarrow$ Muttersprache **ohne Anmerkungen**
- Unterschreiben Sie in dem obigen Unterschriftenfeld. Damit versichern Sie, dass Sie
  - alle Antworten selbstständig und ohne Austausch mit Dritten angefertigt haben,
  - keine anderen als die erlaubten Hilfsmittel benutzt haben und
  - unter Ihrem eigenen Namen abgeben.
- Ich erkläre mich mit einer **Videoüberwachung** während der elektronischen Übungsleistung einverstanden.

  [ ] ja [ ] nein

  Wenn ich das Einverständnis verweigere, wird eine **mündliche Nachprüfung** stattfinden, ob die Prüfungsleistung eigenständig von mir erbracht wurde.
- Es werden nur solche Ergebnisse gewertet, bei denen der **Lösungsweg erkennbar ist.** Begründen Sie alle Antworten, solange es in der jeweiligen Teilaufgabe nicht ausdrücklich anders vermerkt ist.
- Schreiben Sie weder mit roter / grüner Farbe noch mit Bleistift.
- Beachten Sie, dass Sie in der Klausur in jeder Aufgabe eine von mehreren Aufgabenvarianten erhalten haben. Aus diesem Lösungsvorschlag können die Lösungen aller Varianten entnommen werden.

Hörsaal verlassen von __________ bis __________ / Vorzeitige Abgabe um __________

Aufgabe 1 Diskreter Wahrscheinlichkeitsraum (6 Punkte)

Gegeben seien zwei Behälter mit Gegenständen, die durch ihre Form spezifiziert sind:

Variante 1 Behälter 1: △, △, □, ◯, □ Behälter 2: ◯, □, △, ◯, ◯

Variante 2 Behälter 1: □, ◯, ◯, △, ◯ Behälter 2: ◯, △, □, △, □

Variante 3 Behälter 1: △, △, □, ◯, △ Behälter 2: ◯, □, △, ◯, ◯

Variante 4 Behälter 1: ◯, △, □, ◯, ◯ Behälter 2: □, △, △, △, ◯

Nun wird zunächst eine (nicht faire) Münze geworfen, die Wahrscheinlichkeit für Kopf sei $p \in (0, 1)$. Bei Kopf wird Behälter 1, bei Zahl Behälter 2 gewählt. Aus dem gewählten Behälter wird dann ein Gegenstand gezogen, wobei jeder Gegenstand mit der gleichen Wahrscheinlichkeit gewählt wird.

Seien B_1 und B_2 die Ereignisse, dass Behälter 1 bzw. Behälter 2 gewählt wurde. Weiterhin sei E das Ereignis, dass die in (a) gesuchte Form gezogen wurde.

0
1
2

a) **Varianten 1-2:** Mit welcher Wahrscheinlichkeit ist der gezogene Gegenstand ein Dreieck?
Varianten 3-4: Mit welcher Wahrscheinlichkeit ist der gezogene Gegenstand ein Kreis?

Bedingen wir auf einen Behälter, entspricht die Wahrscheinlichkeit für eine Form gerade der relativen Häufigkeit dieser Formen in diesem Behälter. Mit dem Satz der totalen Wahrscheinlichkeit erhalten wir so (exemplarisch für Variante 1):

$$\Pr[E] = \Pr[E \mid B_1] \cdot \Pr[B_1] + \Pr[E \mid B_2] \cdot \Pr[B_2] = \frac{2}{5} \cdot p + \frac{1}{5} \cdot (1-p) = \frac{1+p}{5}\,.$$

Ergebnisse der Varianten (sei E_i das Ereignis E in Variante i):

$$\Pr[E_1] = \frac{2}{5} \cdot p + \frac{1}{5} \cdot (1-p) = \frac{1+p}{5}$$
$$\Pr[E_2] = \frac{1}{5} \cdot p + \frac{2}{5} \cdot (1-p) = \frac{2-p}{5}$$
$$\Pr[E_3] = \frac{1}{5} \cdot p + \frac{3}{5} \cdot (1-p) = \frac{3-2p}{5}$$
$$\Pr[E_4] = \frac{3}{5} \cdot p + \frac{1}{5} \cdot (1-p) = \frac{1+2p}{5}$$

0
1

b) **Varianten 1-2:** Angenommen es wurde ein Dreieck gezogen. Mit welcher Wahrscheinlichkeit kommt es aus Behälter 2?
Varianten 3-4: Angenommen es wurde ein Kreis gezogen. Mit welcher Wahrscheinlichkeit kommt er aus Behälter 2?

Gesucht ist $\Pr[B_2 \mid E]$. Durch Umformen ergibt sich (wiederum exemplarisch für Variante 1)

$$\Pr[B_2 \mid E] = \frac{\Pr[B_2 \cap E]}{\Pr[E]} = \frac{\Pr[E \mid B_2] \cdot \Pr[B_2]}{\Pr[E]} = \frac{1/5 \cdot (1-p)}{(1+p)/5} = \frac{1-p}{1+p}\,.$$

Ergebnisse der Varianten: $\Pr[B_2 \mid E_1] = \frac{1-p}{1+p}$, $\Pr[B_2 \mid E_2] = \frac{2-2p}{2-p}$, $\Pr[B_2 \mid E_3] = \frac{3-3p}{3-2p}$, $\Pr[B_2 \mid E_4] = \frac{1-p}{1+2p}$.

c) Ermitteln Sie je einen Gegenstand aus jedem Behälter, sodass durch Austauschen dieser das Ereignis
Varianten 1-2: „Der gezogene Gegenstand ist ein Kreis"
Varianten 3-4: „Der gezogene Gegenstand ist ein Dreieck"
unabhängig vom Ereignis „Die Münze zeigt Zahl" wird. Dabei sollten die Wahrscheinlichkeiten aus (a) und (b) unverändert bleiben.

0 1 2 3

Varianten 1-2: Sei K das Ereignis, dass ein Kreis gezogen wird. Ist die relative Häufigkeit von Kreisen in beiden Behältern gleich, so ist das Ereignis K unabhängig von der Wahl des Behälters, d.h. der Seite der Münze. Weiterhin sollte die relative Häufigkeit der Dreiecke in beiden Behältern unverändert bleiben.
Mit dieser Überlegung entscheiden wir uns dazu, ein Quadrat aus Behälter 1 und einen Kreis aus Behälter 2 miteinander zu vertauschen (Variante 1) bzw. einen Kreis aus Behälter 1 und einen Quadrat aus Behälter 2 miteinander zu vertauschen (Variante 2). Dann gilt

$$\Pr[K] = \Pr[K \mid B_1] \cdot \Pr[B_1] + \Pr[K \mid B_2] \cdot \Pr[B_2] = \frac{2}{5} \cdot p + \frac{2}{5} \cdot (1-p) = \frac{2}{5}$$

und andererseits

$$\Pr[K \cap B_2] = \Pr[K|B_2] \cdot \Pr[B_2] = \frac{2}{5} \cdot (1-p)\,.$$

Insbesondere folgt somit $\Pr[K] \cdot \Pr[B_2] = \Pr[K \cap B_2]$, womit die Unabhängigkeit beider Ereignisse gezeigt ist. Durch den Tausch wurde die relative Häufigkeit der Dreiecke in beiden Behältern nicht verändert. Die in (a) und (b) gezeigten Gleichungen gelten daher weiterhin.

Varianten 3-4: Sei D das Ereignis, dass ein Dreieck gezogen wird. Ist die relative Häufigkeit von Dreiecken in beiden Behältern gleich, so ist das Ereignis D unabhängig von der Wahl des Behälters, d.h. der Seite der Münze. Weiterhin sollte die relative Häufigkeit der Kreise in beiden Behältern unverändert bleiben. Mit dieser Überlegung entscheiden wir uns dazu, ein Dreieck aus Behälter 1 und ein Quadrat aus Behälter 2 miteinander zu vertauschen (Variante 3) bzw. ein Quadrat aus Behälter 1 und ein Dreieck aus Behälter 2 miteinander zu vertauschen (Variante 4). Dann gilt

$$\Pr[D] = \Pr[D \mid B_1] \cdot \Pr[B_1] + \Pr[D \mid B_2] \cdot \Pr[B_2] = \frac{2}{5} \cdot p + \frac{2}{5} \cdot (1-p) = \frac{2}{5}$$

und andererseits

$$\Pr[D \cap B_2] = \Pr[D|B_2] \cdot \Pr[B_2] = \frac{2}{5} \cdot (1-p)\,.$$

Insbesondere folgt somit $\Pr[D] \cdot \Pr[B_2] = \Pr[D \cap B_2]$, womit die Unabhängigkeit beider Ereignisse gezeigt ist. Durch den Tausch wurde die relative Häufigkeit der Kreise in beiden Behältern nicht verändert. Die in (a) und (b) gezeigten Gleichungen gelten daher weiterhin.

Aufgabe 2 **Diskrete Zufallsvariablen (7 Punkte)**

Zwei diskrete Zufallsvariablen X und Y haben die gemeinsame Dichte

$$f_{X,Y}(x,y) = \begin{cases} c \cdot \frac{1}{k^{x+y}} & x \in \mathbb{N},\ y \in \mathbb{N}_0,\ |x-y| = 1 \\ 0 & \text{sonst.} \end{cases}$$

Sei $k \geq 2$ die Basis im Nenner der Dichtefunktion. In Variante 1,2,3,4 gilt $k = 2$, $k = 3$ $k = 4$, $k = 5$.

0 1 2

a) Zeigen Sie, dass
$c = 6/5$ (Var. 1) $c = 12/5$ (Var. 2) $c = 60/17$ (Var. 3) $c = 60/13$ (Var. 4)
die einzige zulässige Konstante ist.

Vorüberlegung: Unter der Annahme $X = x$ für $x \in \mathbb{N}$ nimmt die Zufallsvariable Y nur die Werte $x-1$ und $x+1$ mit positiver Wahrscheinlichkeit an. Daher gilt

$$\begin{aligned} \sum_{x \in W_X} \sum_{y \in W_Y} f_{X,Y}(x,y) &= \sum_{x \in \mathbb{N}} \left(f_{X,Y}(x, x-1) + f_{X,Y}(x, x+1) \right) \\ &= c \cdot \sum_{x \in \mathbb{N}} \left(\frac{1}{k^{x+x-1}} + \frac{1}{k^{x+x+1}} \right) \\ &= c \cdot \sum_{x \in \mathbb{N}} \frac{k + 1/k}{k^{2x}} \\ &= c \cdot \frac{k^2+1}{k} \cdot \sum_{x \in \mathbb{N}} \left(\frac{1}{k^2} \right)^x \end{aligned}$$

Da $|1/k^2| < 1$, nimmt die geometrische Reihe den Wert

$$\sum_{x \in \mathbb{N}} \left(\frac{1}{k^2} \right)^x = \left(\sum_{x=0}^{\infty} \left(\frac{1}{k^2} \right)^x \right) - 1 = \frac{1}{1 - 1/k^2} - 1 = \frac{1}{k^2 - 1}$$

an. Aus der Bedingung $\sum_{x \in W_X} \sum_{y \in W_Y} f_{X,Y}(x,y) = 1$ folgt somit $c = \frac{k \cdot (k^2-1)}{k^2+1}$.
Für $k = (2,3,4,5)$ ergibt sich $c = (\frac{6}{5}, \frac{12}{5}, \frac{60}{17}, \frac{60}{13})$.

b) Bestimmen Sie die Randdichte von X.

0 1 2

Für $x \notin \mathbb{N}$ gibt es keine Ereignisse mit positiver Wahrscheinlichkeit, daher $f_X(x) = 0$. Für $x \in \mathbb{N}$ folgt analog zum vorherigen Aufgabenteil

$$f_X(x) = \sum_{y \in W_Y} f_{X,Y}(x, y) = f_{X,Y}(x, x-1) + f_{X,Y}(x, x+1) = c \cdot \frac{k^2+1}{k} \cdot \left(\frac{1}{k^2}\right)^x .$$

Mit $c = \frac{k\cdot(k^2-1)}{k^2+1}$ vereinfacht sich der letzte Ausdruck weiter zu $(k^2-1) \cdot \left(\frac{1}{k^2}\right)^x$, sodass insgesamt gilt

$$f_X(x) = \begin{cases} (k^2-1) \cdot \left(\frac{1}{k^2}\right)^x & \text{falls } x \in \mathbb{N} \\ 0 & \text{sonst.} \end{cases}$$

Für $k = (2, 3, 4, 5)$ gilt also $f_X(x) = \left(\frac{3}{4^x}, \frac{8}{9^x}, \frac{15}{16^x}, \frac{24}{25^x}\right)$, falls $x \in \mathbb{N}$.

0
1
2
3

c) Bestimmen Sie die Dichte von $Z = X + Y$.

Man beachte, dass die Faltungsformel hier nicht anwendbar ist, da es sich um abhängige Zufallsvariablen handelt (betrachte etwa $X = 1$ und $Y = 0$). Stattdessen berechnen wir die Dichte von Z wie folgt:

Für das Ereignis $Z = z$ müssen zwei Zahlen $x \in \mathbb{N}$ und $y \in \mathbb{N}_0$ mit $x + y = z$ existieren. Die gemeinsame Dichtefunktion sagt allerdings aus, dass nur solche Zahlenpaare (x, y) mit positiver Wahrscheinlichkeit auftreten, für die $|x - y| = 1$ gilt. Insbesondere ist genau eine der beiden Zahlen ungerade.

Ist z gerade, so lässt sich die Zahl nicht als Summe einer geraden und einer ungeraden Zahl darstellen. Sei daher $z \in \mathbb{N}$ ungerade. Für $z = 1$ folgt $(x, y) = (1, 0)$ und somit $f_Z(1) = c \cdot \frac{1}{k^{1+0}} = c/k$. Für $z \geq 3$ muss $(x, y) = (\frac{z+1}{2}, \frac{z+1}{2} - 1)$ oder $(x, y) = (\frac{z+1}{2} - 1, \frac{z+1}{2})$ gelten. Insgesamt tritt dieser Fall mit Wahrscheinlichkeit

$$f_{X,Y}\left(\frac{z+1}{2}, \frac{z+1}{2} - 1\right) + f_{X,Y}\left(\frac{z+1}{2} - 1, \frac{z+1}{2}\right) = 2c \cdot \frac{1}{k^z}$$

auf. Insgesamt erhalten wir die Dichtefunktion

$$f_Z(z) = \begin{cases} c/k & z = 1 \\ 2c/k^z & z \in \mathbb{N} \text{ ungerade}, z \geq 3 \\ 0 & \text{sonst}. \end{cases}$$

Für $k = (2, 3, 4, 5)$ ergibt sich $f_Z(z) = \left(\frac{3}{5}, \frac{4}{5}, \frac{15}{17}, \frac{12}{13}\right)$ falls $z = 1$ sowie $f_Z(z) = \left(\frac{12/5}{2^z}, \frac{24/5}{3^z}, \frac{120/17}{4^z}, \frac{120/13}{5^z}\right)$ falls $z \geq 3$ ungerade.

Aufgabe 3 Kontinuierliche Zufallsvariablen (6 Punkte)

Biologin Alison stellt fest, dass die Halslänge (in Metern) der westafrikanischen Giraffe durch eine kontinuierliche Zufallsvariable X mit Dichtefunktion

$$f_X(x) = \begin{cases} 1/(lx) & \text{falls } 1 \le x \le e^l \\ 0 & \text{sonst} \end{cases}$$

beschrieben werden kann.

Sei $l \in \mathbb{N}_{\ge 2}$ die Konstante in der Dichtefunktion. In Variante 1,2,3,4 gilt $l = 2$, $l = 3$, $l = 4$, $l = 5$.

a) Bestimmen Sie die zugehörige Verteilungsfunktion.

0 1 2

Für $x < 1$ gilt $F_X(x) = 0$. Gelte daher $x \ge 1$ und sei $u = \min\{x, e^l\}$, dann gilt

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{1}^{u} \frac{1}{t \cdot l}\, dt = \frac{1}{l} \cdot [\ln(t)] = \frac{\ln(u) - \ln(1)}{l} = \frac{\ln(u)}{l}.$$

Damit erhalten wir die Verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & \text{falls } x < 1 \\ \frac{\ln(x)}{l} & \text{falls } 1 \le x \le e^l \\ 1 & \text{falls } x > e^l. \end{cases}$$

b) Berechnen Sie das k-te Moment von X.

Das k-te Moment der Zufallsvariablen X ist definiert als $E[X^k]$. Dieser Ausdruck berechnet sich wie folgt.

$$\mathrm{E}[X^k] = \int_{-\infty}^{\infty} x^k \cdot f_X(x)\, dx = \int_1^{e^l} x^k \cdot \frac{1}{x \cdot l}\, dx = \frac{1}{l} \cdot \int_1^{e^l} x^{k-1}\, dx = \frac{1}{l} \cdot \left[\frac{x^k}{k}\right]_1^{e^l} = \frac{e^{kl} - 1}{k \cdot l}$$

c) Alison bezeichnet eine Giraffe als *riesig*, wenn ihre Halslänge mehr als 1 m über dem Erwartungswert liegt. Zeigen Sie, dass dies mit Wahrscheinlichkeit $p := 1 - F_X(\frac{e^2+1}{2})$ (Var. 1), $p := 1 - F_X(\frac{e^3+2}{3})$ (Var. 2), $p := 1 - F_X(\frac{e^4+3}{4})$ (Var. 3), bzw. $p := 1 - F_X(\frac{e^5+4}{5})$ (Var. 4) eintritt.

Nach Definition ist eine Giraffe der Größe X riesig, falls $X > \mathrm{E}[X] + 1$. Der vorherigen Teilaufgabe können wir den Erwartungswert $\mathrm{E}[X] = \mathrm{E}[X^1] = \frac{e^l - 1}{l}$ entnehmen. Folglich gilt

$$p = \Pr[X > E[X] + 1] = 1 - F_X(E[X] + 1) = 1 - F_X\left(\frac{e^l - 1 + l}{l}\right).$$

d) Sei E das Ereignis, dass von $n \geq 1$ unabhängig ausgewählten Giraffen mindestens die Hälfte riesig ist.
Variante 1: Zeigen Sie, dass $\Pr[E] \leq 0.56$, falls $p = 0.28$.
Variante 2: Zeigen Sie, dass $\Pr[E] \leq 0.66$, falls $p = 0.33$.
Variante 3: Zeigen Sie, dass $\Pr[E] \leq 0.66$, falls $p = 0.33$.
Variante 4: Zeigen Sie, dass $\Pr[E] \leq 0.64$, falls $p = 0.32$.

0 1 2

Sei Y_i eine Indikatorvariable für das Ereignis, dass die i-te Giraffe riesig ist. Die Gesamtanzahl riesiger Giraffen $Y := \sum_{i=1}^{n} Y_i$ ist dann binomialverteilt und es gilt $E[Y] = np$.
Als nächstes wollen wir die Wahrscheinlichkeit $\Pr[Y \geq n/2]$ mittels einer geeigneten Schranke nach oben hin abschätzen. Hierzu kommt insbesondere die Markov-Ungleichung in Betracht: Y ist eine diskrete Zufallsvariable mit Wertebereich $W_Y \subseteq \mathbb{N}_0$ und es gilt $n/2 > 0$. Demnach gilt

$$\Pr[Y \geq n/2] \leq \frac{E[Y]}{n/2} = \frac{np}{n/2} = 2p\,.$$

Ergebnisse der Varianten:

- $l = 2$: $\quad p = 1 - \frac{1}{2} \cdot \ln\left(\frac{e^2 - 1 + 2}{2}\right) = 0.28, \quad 2p = 0.56$
- $l = 3$: $\quad p = 1 - \frac{1}{3} \cdot \ln\left(\frac{e^3 - 1 + 3}{3}\right) = 0.33, \quad 2p = 0.66$
- $l = 4$: $\quad p = 1 - \frac{1}{4} \cdot \ln\left(\frac{e^4 - 1 + 4}{4}\right) = 0.33, \quad 2p = 0.66$
- $l = 5$: $\quad p = 1 - \frac{1}{5} \cdot \ln\left(\frac{e^5 - 1 + 5}{5}\right) = 0.32, \quad 2p = 0.64$

Aufgabe 4 Kontinuierliche Gleichverteilung (8 Punkte)

Konditor Alfred backt köstliche Blaubeermuffins nach seinem Spezialrezept und bemerkt, dass diese unterschiedlich groß geraten sind. Die Höhe eines Muffins in cm sei dabei, unabhängig von der Höhe anderer Muffins, durch eine gleichverteilte kontinuierliche Zufallsvariable X gegeben.
Variante 1: Die minimale Höhe eines Muffins beträgt 6 cm und die maximale Höhe beträgt 7 cm.
Variante 2: Die minimale Höhe eines Muffins beträgt 5 cm und die maximale Höhe beträgt 7 cm.
Variante 3: Die minimale Höhe eines Muffins beträgt 6 cm und die maximale Höhe beträgt 9 cm.
Variante 4: Die minimale Höhe eines Muffins beträgt 5 cm und die maximale Höhe beträgt 9 cm.

Sei h die min. Höhe eines Muffins und a so gewählt, dass $a + h$ die max. Höhe eines Muffins ist. Für (b) sei sei n die Anzahl der Muffins und c die gefragte Gesamthöhe. In (c) ist Y auf $[0, h]$ gleichverteilt.

0 1

a) Berechnen Sie den Erwartungswert und die Standardabweichung von X.

Da X gleichverteilt auf $[a, a + h]$ ist, erhalten wir $\mu = a + \frac{h}{2}$ für den Erwartungswert und $\sigma = \frac{h}{\sqrt{12}}$ für die Standardabweichung von X.

0 1 2 3

b) Variante 1: Alfred hat 108 Muffins gebacken. Mit welcher Wahrscheinlichkeit übersteigt die Gesamthöhe aller Muffins einen Wert von 705 cm?
Variante 2: Alfred hat 48 Muffins gebacken. Mit welcher Wahrscheinlichkeit übersteigt die Gesamthöhe aller Muffins einen Wert von 294 cm?
Variante 3: Alfred hat 108 Muffins gebacken. Mit welcher Wahrscheinlichkeit übersteigt die Gesamthöhe aller Muffins einen Wert von 828 cm?
Variante 4: Alfred hat 300 Muffins gebacken. Mit welcher Wahrscheinlichkeit übersteigt die Gesamthöhe aller Muffins einen Wert von 2125 cm?

Berechnen Sie die gesuchte Wahrscheinlichkeit näherungsweise mit dem Zentralen Grenzwertsatz.

Sei X_i die Höhe des i-ten Muffins. Da $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen sind und dieselbe Verteilung wie X besitzen, folgt für die Gesamthöhe der n Muffins $\sum_{i=1}^n X_i$ aus dem Zentralen Grenzwertsatz, dass

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$$

näherungsweise standardnormalverteilt ist. Dabei ist $\mu = a + \frac{h}{2}$ der Erwartungswert und $\sigma = \frac{h}{\sqrt{12}}$ die Standardabweichung aus der vorherigen Teilaufgabe. Damit approximieren wir die gesuchte Wahrscheinlichkeit:

$$\Pr\left[\sum_{i=1}^n X_i > c\right] = 1 - \Pr\left[\sum_{i=1}^n X_i \leq c\right] = 1 - \Pr\left[\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq \frac{c - n\mu}{\sigma\sqrt{n}}\right]$$
$$\approx 1 - \Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right).$$

Ergebnisse der Varianten:

a	h	n	c	μ	σ	$n\mu$	$\sigma\sqrt{n}$	$\frac{c-n\mu}{\sigma\sqrt{n}}$	1-$\Phi\left(\frac{c-n\mu}{\sigma\sqrt{n}}\right)$
6	1	108	705	6,5	$\frac{1}{2\sqrt{3}}$	702	3	1	$1 - 0{,}841 = 0{,}159$
5	2	48	294	6	$\frac{1}{\sqrt{3}}$	288	4	1,5	$1 - 0{,}933 = 0{,}067$
6	3	108	828	7,5	$\frac{\sqrt{3}}{2}$	810	9	2	$1 - 0{,}977 = 0{,}023$
5	4	300	2125	7	$\frac{2}{\sqrt{3}}$	2100	20	1,25	$1 - 0{,}894 = 0{,}106$

c) Alfred ist mit der Höhe der Muffins unzufrieden und verwendet daher beim nächsten Blech Muffins eine Extraportion Backpulver. Die Höhe der Muffins erhöht sich dabei um Y cm, wobei Y eine auf dem Intervall $[0,1]$ (Var. 1), $[0,2]$ (Var. 2), $[0,3]$ (Var. 3), $[0,4]$ (Var. 4), gleichverteilte, von X unabhängige Zufallsvariable ist. Bestimmen Sie die Dichte der Zufallsvariablen Z, welche die Höhe eines neuen Muffins angibt.

Da X gleichverteilt auf $[a, a+h]$ und Y gleichverteilt auf $[0, h]$ ist, sind die Dichtefunktionen von X und Y gegeben durch

$$f_X(x) = \begin{cases} \frac{1}{h} & \text{falls } x \in [a, a+h], \\ 0 & \text{sonst} \end{cases} \quad \text{und} \quad f_Y(y) = \begin{cases} \frac{1}{h} & \text{falls } y \in [0, h], \\ 0 & \text{sonst.} \end{cases}$$

Die Höhe der neuen Muffins ist $Z = X + Y$, wobei X und Y laut Angabe unabhängig sind. Die Dichte von Z berechnet sich dann mit Hilfe der Faltungsformel:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \mathrm{d}x.$$

Für $z < a$ ist $f_Z(z) = 0$, denn für alle $x < a$ ist $f_X(x) = 0$ während für alle $x \geq a$ dann $z - x < 0$ und somit $f_Y(z-x) = 0$ ist.
Auch für $z > a + 2h$ ist $f_Z(z) = 0$, denn für alle $x > a + h$ ist $f_X(x) = 0$ während für alle $x \leq a + h$ dann $z - x > h$ und somit $f_Y(z-x) = 0$ ist.
Betrachten wir im Folgenden den Fall $z \in [a, a+2h]$. Die Bedingung $z - x \in [0, h]$ können wir umformen zu $0 \leq z - x \leq h \Leftrightarrow z \geq x \geq z - h$. Damit ergibt sich für das Integral:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \mathrm{d}x = \int_a^{a+h} \frac{1}{h} f_Y(z-x) \mathrm{d}x = \int_{\max\{a, z-h\}}^{\min\{a+h, z\}} \frac{1}{h^2} \mathrm{d}x \\ &= \frac{\min\{a+h, z\} - \max\{a, z-h\}}{h^2} \\ &= \frac{h + a + \min\{0, z-h-a\} - a - \max\{0, z-h-a\}}{h^2} \\ &= \frac{h - |z-h-a|}{h^2}. \end{aligned}$$

Ergebnisse der Varianten:

$a = 6, h = 1$:

$$f_Z(z) = \begin{cases} z - 6 & \text{falls } z \in [6,7], \\ 8 - z & \text{falls } z \in [7,8], \\ 0 & \text{sonst.} \end{cases}$$

$a = 5, h = 2$:

$$f_Z(z) = \begin{cases} \frac{z-5}{4} & \text{falls } z \in [5,7], \\ \frac{9-z}{4} & \text{falls } z \in [7,9], \\ 0 & \text{sonst.} \end{cases}$$

$a = 6, h = 3$:

$$f_Z(z) = \begin{cases} \frac{z-6}{9} & \text{falls } z \in [6,9], \\ \frac{12-z}{9} & \text{falls } z \in [9,12], \\ 0 & \text{sonst.} \end{cases}$$

$a = 5, h = 4$:

$$f_Z(z) = \begin{cases} \frac{z-5}{16} & \text{falls } z \in [5,9], \\ \frac{13-z}{16} & \text{falls } z \in [9,13], \\ 0 & \text{sonst.} \end{cases}$$

Aufgabe 5 Schätzvariablen (4 Punkte)

Als Aufgabe 5 in der Klausur wurde entweder Aufgabenteil (a) oder (b) gestellt.

0 1 2

a) Seien $X_1, \dots, X_n$ Stichprobenvariablen einer diskreten Verteilung mit unbekanntem Erwartungswert $\mu \neq 0$ und Varianz ${\sigma_X}^2 > 0$. Weiterhin seien $Y_1, \dots, Y_n$ Stichprobenvariablen einer anderen diskreten Verteilung, die den gleichen Erwartungswert μ, aber die Varianz ${\sigma_Y}^2 > 0$ hat. Alle Stichprobenvariablen seien unabhängig. Als Schätzer U für den Parameter μ betrachten wir eine gewichtete Summe der Stichprobenmittel

$$U = \alpha_X \cdot \overline{X} + \alpha_Y \cdot \overline{Y}\,.$$

Ermitteln Sie Koeffizienten α_X und α_Y, sodass U ein erwartungstreuer Schätzer ist und die Effizienz von U maximiert wird.

Zunächst untersuchen wir die Erwartungstreue von U. Es gilt

$$\mathrm{E}[U] = \mathrm{E}\left[\alpha_X \cdot \overline{X} + \alpha_Y \cdot \overline{Y}\right] = \alpha_X \cdot \mathrm{E}\left[\overline{X}\right] + \alpha_Y \cdot \mathrm{E}[\overline{Y}] = \alpha_X \cdot \mu + \alpha_Y \cdot \mu\,,$$

wobei wir ausgenutzt haben, dass die Stichprobenmittel jeweils erwartungstreue Schätzer für den Erwartungswert einer einzelnen Stichprobenvariablen sind. Somit muss $\alpha_X + \alpha_Y = 1$ gelten, damit U ein erwartungstreuer Schätzer ist.

Bei erwartungstreuen Schätzern ist das Maß für die Effizienz, die mittlere quadratische Abweichung, gerade die Varianz. Im Folgenden suchen wir also einen Koeffizienten $\alpha := \alpha_X$ (daraus folgt $\alpha_Y = 1 - \alpha$), der die Varianz von U minimiert. Für das Stichprobenmittel $\overline{X}$ gilt, unter Ausnutzung der Unabhängigkeit der X_i,

$$\mathrm{Var}[\overline{X}] = \mathrm{Var}\left[\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n} \cdot \sigma_X^2\,.$$

und analog $\mathrm{Var}[\overline{Y}] = \frac{1}{n} \cdot \sigma_Y^2$ für $\overline{Y}$. Es gilt

$$\begin{aligned}
\mathrm{Var}[U] &= \mathrm{Var}[\alpha \cdot \overline{X} + (1-\alpha) \cdot \overline{Y}] \\
&= \alpha^2 \cdot \mathrm{Var}[\overline{X}] + (1-\alpha)^2 \cdot \mathrm{Var}[\overline{Y}] && \text{Unabhängigkeit} \\
&= \alpha^2 \cdot \frac{1}{n} \cdot {\sigma_X}^2 + (1 - 2\alpha + \alpha^2) \cdot \frac{1}{n} \cdot {\sigma_Y}^2 \\
&= \frac{1}{n} \cdot \left(\alpha^2 \cdot ({\sigma_X}^2 + {\sigma_Y}^2) - 2\alpha{\sigma_Y}^2 + {\sigma_Y}^2\right)\,.
\end{aligned}$$

Fassen wir $\mathrm{Var}[U]$ als Funktion von α auf, so beträgt ihre erste Ableitung $\frac{1}{n} \cdot (2\alpha \cdot ({\sigma_X}^2 + {\sigma_Y}^2) - 2{\sigma_Y}^2)$. Da die zweite Ableitung $\frac{1}{n} \cdot 2 \cdot ({\sigma_X}^2 + {\sigma_Y}^2)$ stets positiv ist, handelt es sich bei der Nullstelle $\alpha = \frac{{\sigma_Y}^2}{{\sigma_X}^2 + {\sigma_Y}^2}$ der ersten Ableitung um eine Minimalstelle und damit um den gesuchten Faktor α_X.

b) Aiko möchte die Anzahl $m > 0$ der Fische in ihrem Aquarium schätzen. Zur Vorbereitung markiert sie einmalig eine konstante Zahl k (mit $0 < k \leq m$) von Fischen. An jedem der folgenden $n \geq 1$ Tage fängt Aiko dann solange Fische, bis sie ein markiertes Tier gefangen hat. Dabei entlässt sie jedes Tier sofort wieder in das Aquarium, es befinden sich also stets m Fische im Aquarium. Jedes Tier wird mit der gleichen Wahrscheinlichkeit gefangen. Für jeden Tag notiert sich Aiko die Anzahl der Versuche, die sie jeweils gebraucht hat.
Finden Sie einen erwartungstreuen Schätzer U für m auf Basis von Aikos Aufzeichnungen, der konsistent im quadratischen Mittel ist.

0 1 2

Sei X_i eine Zufallsvariable für die Anzahl der gefangenen Fische am i-ten Tag. Nach Angabe legt Aiko jeden gefangenen Fisch wieder in das Aquarium zurück, es handelt sich also um Ziehen mit Zurücklegen aus einer m-elementigen Grundmenge. Da jeder Fisch mit der gleichen Wahrscheinlichkeit gezogen wird, handelt es sich bei X_i um eine geometrisch verteilte Zufallsvariable mit Parameter $p := k/m$. Somit wissen wir, dass

$$\mathrm{E}[X_i] = \frac{1}{p} = \frac{m}{k} \quad \text{und} \quad \mathrm{Var}[X_i] = \frac{1-p}{p^2} \qquad \text{für } 1 \leq i \leq n \text{ gilt.}$$

Als Ausgangspunkt für einen erwartungstreuen, effizienten Schätzer betrachten wir $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Hierfür gilt

$$\mathrm{E}[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] = \frac{1}{n} \cdot n \cdot \frac{m}{k} = \frac{m}{k}.$$

Setzen wir also $U := k\overline{X}$, erhalten wir einen erwartungstreuen Schätzer für m. Schließlich untersuchen wir die Effizienz des Schätzers. Als erwartungstreuer Schätzer gilt für U

$$\mathrm{MSE}[U] = \mathrm{Var}[U] = k^2 \cdot \mathrm{Var}[\overline{X}].$$

Weiterhin sind die Zufallsvariablen X_i unabhängig, weswegen

$$\mathrm{Var}[\overline{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n} \cdot \frac{1-p}{p^2}$$

und insgesamt

$$\mathrm{MSE}[U] = k^2 \cdot \frac{1}{n} \cdot \frac{1-p}{p^2}.$$

Konsistenz im quadratischen Mittel fordert, dass $\mathrm{MSE}[U] \to 0$ für $n \to \infty$ gilt. Dies ist offensichtlich im oberen Fall gegeben, da es sich bei k^2 und $\frac{1-p}{p^2}$ um Konstanten, unabhängig von n, handelt.

Aufgabe 6 Markov-Ketten (9 Punkte)

Poltergeist Ambrosius hat im Keller des FMI-Gebäudes ein in Vergessenheit geratenes Bücherregal entdeckt. In dem Regal stehen $n \geq 1$ fortlaufend nummerierte Bücher nebeneinander aufgereiht. Jeden Morgen bevor Ambrosius schlafen geht, wählt er ein zufälliges Buch als Gutenachtgeschichte aus, welches er am Abend des Tages, wenn er aufsteht, links (Variante 1), bzw. rechts (Variante 2) von allen anderen Büchern zurück ins Regal stellt. Dabei wählt Ambrosius unabhängig früherer Gutenachtgeschichten jedes Buch mit der gleichen Wahrscheinlichkeit. Sei X_t eine Zufallsvariable, die die Reihenfolge der Bücher im Regal am Abend des t-ten Tages festhält.

0 1 2 3

a) Sei $n = 3$. Zeichnen Sie das Übergangsdiagramm der Markov-Kette $(X_t)_{t\geq 0}$, sodass sich die Kanten in Ihrer Zeichnung nicht überkreuzen. Geben Sie die Bedeutung der Zustände explizit an.

Die Zustände der Markov-Kette $(X_t)_{t\geq 0}$ entsprechen den möglichen Permutationen der Bücher in dem Regal. Für $n = 3$ erhalten wir somit $3! = 6$ Zustände, denen wir wie folgt Permutationen zuordnen:

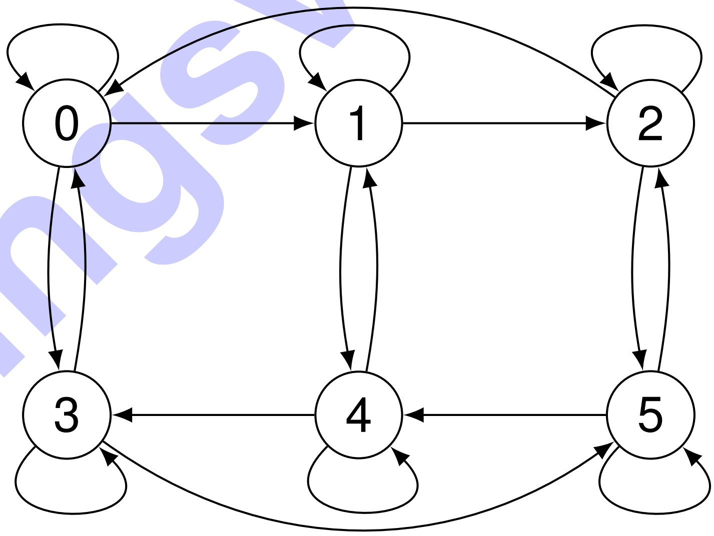
Variante: links zurückstellen

Zustand	0	1	2	3	4	5
Permutation	(123)	(312)	(231)	(213)	(132)	(321)

Variante: rechts zurückstellen

Zustand	0	1	2	3	4	5
Permutation	(123)	(231)	(312)	(132)	(213)	(321)

Aus diesen Zuständen ergibt sich das folgende Übergangsdiagramm, wobei jede Kante eine Übergangswahrscheinlichkeit von $1/3$ aufweist.



b) Zeigen Sie, dass $(X_t)_{t\geq 0}$ für jedes $n \geq 1$ ergodisch ist.

0
1
2
3

Um zu beweisen, dass $(X_t)_{t\geq 0}$ unabhängig von n stets ergodisch ist, gilt es zu zeigen, dass $(X_t)_{t\geq 0}$ aperiodisch und irreduzibel ist.
Die Tatsache, dass $(X_t)_{t\geq 0}$ aperiodisch ist, folgt unmittelbar aus der Beobachtung, dass jede Permutation $\rho : [n] \to [n]$ von Ambrosius Büchern mit positiver Wahrscheinlichkeit am nächsten Tag erhalten bleibt, nämlich genau dann, wenn Ambrosius das Buch $\rho(1)$ (bei Variante *links zurückstellen*) bzw. $\rho(n)$ (bei Variante *rechts zurückstellen*) auswählt. Folglich hat jeder Zustand des Übergangsdiagramms eine Schleife, was die Aperiodizität von $(X_t)_{t\geq 0}$ impliziert.
Für die Irreduzibilität von $(X_t)_{t\geq 0}$ müssen wir zeigen, dass ρ von jeder anderen Permutation $\rho' : [n] \to [n]$ mit positiver Wahrscheinlichkeit erreichbar ist. Unabhängig von ρ' lässt sich ρ jedoch stets durch die Büchersequenz $\rho(n), \rho(n-1), \rho(n-2), \dots, \rho(1)$ (Variante *links zurückstellen*) bzw. $\rho(1), \rho(2), \dots, \rho(n-1), \rho(n)$ (Variante *rechts* zurückstellen) erreichen. Somit ist $(X_t)_{t\geq 0}$ irreduzibel.

c) Finden Sie eine stationäre Verteilung für $(X_t)_{t\geq 0}$ in Abhängigkeit von n.

Um eine stationäre Verteilung $\pi^T \in [0,1]^{n!}$ zu finden, argumentieren wir zunächst, dass die Übergangsmatrix P der Markov-Kette doppeltstochastisch ist. Da die Übergangsmatrix einer Markov-Kette immer stochastisch ist, bleibt zu zeigen, dass alle Spaltensummen gleich 1 sind. Die i-te Spalte der Übergangsmatrix enthält die Übergangswahrscheinlichkeiten aller eingehenden Kanten des Zustands i, welche jeweils $1/n$ betragen. Die dem Zustand i zugeordnete Permutation $\rho : [n] \to [n]$ von Büchern lässt sich von n Permutationen in exakt einem Zeitschritt erreichen, denn es gibt n Positionen im Bücherregal, an denen sich das Buch $\rho(1)$ (Variante *links* zurükstellen) bzw. $\rho(n)$ (Variante *rechts zurückstellen*), welches Ambrosius zuletzt gelesen hat, am Morgen befunden haben könnte. D.h. jeder Zustand im Graph des Übergangsdiagramms besitzt genau n eingehende Kanten und somit ist jede Spaltensumme der Übergangsmatrix gleich 1.
Für die doppeltstochastische Matrix P gilt

$$(1/n!, \dots, 1/n!)P = \left(\sum_{i=1}^{n} \frac{1}{n!}\cdot\frac{1}{n}, \dots, \sum_{i=1}^{n} \frac{1}{n!}\cdot\frac{1}{n}\right) = (1/n!, \dots, 1/n!).$$

Demnach handelt es sich bei $\pi = (1/n!, \dots, 1/n!)$ um eine eine stationäre Verteilung.
(Argumentation auch ohne die letzte Rechnung direkt mit Satz 144 möglich.)

d) Angenommen die Bücher sind am Abend des t-ten Tages in aufsteigender Reihenfolge sortiert. Wie viele Tage vergehen erwartungsgemäß, bis die Bücher zum nächsten Mal in aufsteigender Reihenfolge sortiert sind?

Gesucht ist die erwartete Rückkehrzeit h_i, wobei i der Permutation $\rho = (123\dots n)$ zugeordnet ist. Da es sich bei $(X_t)_{t\geq 0}$ um eine irreduzible Markov-Kette handelt, wissen wir gemäß Satz 136, dass die stationäre Verteilung von $(X_t)_{t\geq 0}$ eindeutig ist und dem Vektor $(1/h_1, \dots, 1/h_{n!})$ entspricht. Aus dem Ergebnis der dritten Teilaufgabe folgt unmittelbar, dass $h_i = n!$. Es vergehen somit erwartungsgemäß $n!$ Tage, bis ρ wieder hergestellt ist.

Verteilungswerte der Standardnormalverteilung

Diese Tabelle der Standardnormalverteilung enthält die Werte von $\Phi(x)$ für $0 \leq x \leq 2{,}99$. Beispielsweise gilt $\Phi(1{,}55) \approx 0{,}939$. Diese Seite muss nicht abgegeben werden.

$\Phi(x)$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995
2,6	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,996	0,996	0,996
2,7	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
2,8	0,997	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998
2,9	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999	0,999