Stochastik für die Informatik WiSe 2018/19

Klausur samt Lösungsvorschlägen

Lösungsvorschläge von Florin Boenkost und Jan Lukas Igelbrink

1. Wie beim Geburtstagsproblem werden n Objekte rein zufällig in g Boxen einsortiert.

a) Wir betrachten zuerst den Fall $g = n$, und interessieren uns für Kollisionen in Box 1.
(i) Was ist die W'keit, dass Objekt 1 und Objekt 2 in Box 1 landen?
(ii) Was ist der Erwartungswert der Anzahl der Paare von Objekten, die in Box 1 landen? (Objekt 1 und Objekt 2 bilden dabei dasselbe Paar wie Objekt 2 und Objekt 1.)
(iii) Sei X eine Poisson(1)-verteilte Zufallsvariable. Bestimmen Sie $\mathbf{E}[\frac{1}{2}X(X-1)]$.
(iv) Sehen Sie einen Zusammenhang zwischen den Fragen in (ii) und (iii)? Wenn ja, erklären Sie ihn.

b) Jetzt betrachten wir den Fall $g = n^3$ und interessieren uns für "Kollisionen irgendwo".
(i) Was ist der Erwartungswert der Anzahl Y der Paare von Objekten, die kollidieren, d.h. in ein-und derselben Box landen?
(ii) Begründen Sie die Abschätzung $\mathbf{P}(Y > 0) \le \mathbf{E}[Y]$.
(iii) Was folgt aus (i) und (ii) im Grenzwert $n \to \infty$ für die Wahrscheinlichkeit des Ereignisses, dass es zu mindestens einer Kollision kommt?

Lösungshinweise:

a)
(i) Objekt 1 und 2 landen mit gleicher Wahrscheinlichkeit in Box 1, die Ereignisse {Objekt 1 in Box 1} und {Objekt 2 in Box 1} sind unabhängig:

$$\mathbb{P}\left(\{\text{Objekt 1 und 2 in Box 1}\}\right) = \mathbb{P}\left(\{\text{Objekt 1 in Box 1}\}\right)\cdot\mathbb{P}\left(\{\text{Objekt 2 in Box 1}\}\right) = \frac{1}{n}\cdot\frac{1}{n} = \frac{1}{n^2}$$

(ii) Erwartungswert als Summe über alle Paare darstellen:

$$\mathbf{E}\left[\#\text{Paare in Box 1}\right] = \sum_{1\le i<j\le n} \mathbb{P}\left(\{\text{Objekt } i \text{ und } j \text{ in Box 1}\}\right) = \binom{n}{2}\cdot\frac{1}{n^2} = \frac{1}{2}\left(1-\frac{1}{n}\right)$$

(iii) Hilfreiche Formel für die Varianz:

$$\begin{aligned}\mathbf{E}\left[\frac{1}{2}X(X-1)\right] &= \frac{1}{2}\left(\mathbf{E}[X^2]-\mathbf{E}[X]^2+\mathbf{E}[X]^2-\mathbf{E}[X]\right)\\ &= \frac{1}{2}\left(\mathbf{Var}[X]+\mathbf{E}[X]^2-\mathbf{E}[X]\right)\\ &= \frac{1}{2}\left(1+1^2-1\right) = \frac{1}{2}\end{aligned}$$

(iv) Mit Wahrscheinlichkeit $1/n$ landet jedes Objekt in Box 1, wir betrachten n Objekte, folglich ist die Anzahl der Objekte in Box 1 für große n annähernd Poisson(1)-verteilt. $\frac{1}{2}X(X-1)$ ist die Anzahl der Paare, die man aus X Objekten bilden kann. Für Poisson(1)-verteiltes X ist $\mathbf{E}\left[\frac{1}{2}X(X-1)\right] = \frac{1}{2}$ der Grenzwert des Ergebnisses aus (ii) für $n \to \infty$.

b)

(i) Wir überlegen uns wie in a)(i)

$$\mathbb{P}\left(\{\text{Objekt } i \text{ und } j \text{ in Box } k\}\right) = \frac{1}{g^2}$$

und da für $k \neq l$ die Ereignisse $\{\text{Objekt } i \text{ und } j \text{ in Box } k\}$ und $\{\text{Objekt } i \text{ und } j \text{ in Box } l\}$ disjunkt sind, erhalten wir

$$\mathbb{P}\left(\{\text{Objekt } i \text{ und } j \text{ in einer gemeinsamen Box }\}\right) = \sum_{k=1}^{g} \mathbb{P}\left\{\text{Objekt } i \text{ und } j \text{ in Box } k\right\} = \frac{1}{g} = \frac{1}{n^3}.$$

Alternativ (und kürzer): Unabhängig davon, in welcher Box Objekt i landet, ist die W'keit, dass Objekt j in derselben Box landet, gleich $\frac{1}{g} = \frac{1}{n^3}$. Also:

$$\mathbb{P}\left(\{\text{Objekt } i \text{ und } j \text{ landen in einer gemeinsamen Box }\}\right) = \frac{1}{n^3}.$$

Damit gilt dann

$$\begin{aligned}\mathbf{E}\left[\#\text{Paare in gemeinsamer Box}\right] &= \mathbf{E}\left[\sum_{1\leq i<j\leq n} I_{\{\text{Objekt } i \text{ und } j \text{ in einer gemeinsamen Box }\}}\right] \\ &= \binom{n}{2} \cdot \frac{1}{n^3} = \frac{n-1}{2n^2}.\end{aligned}$$

(ii) Da Y Wertebereich $\mathbb{N}_0$ besitzt gilt die Markov-Ungleichung und somit

$$\mathbb{P}(Y > 0) = \mathbb{P}(Y \geq 1) \leq \frac{\mathbf{E}[Y]}{1} = \mathbf{E}[Y].$$

(iii) Wir betrachten den Grenzwert von $\mathbf{E}[Y] = \frac{n-1}{2n^2}$:

$$\lim_{n\to\infty}\left[\frac{n-1}{2n^2}\right] = \lim_{n\to\infty}\frac{1}{2n} - \lim_{n\to\infty}\frac{1}{2n^2} = 0 - 0 = 0$$

2. a) Wir betrachten den (klassischen) Anteilschätzer $H := K/n$ für die unbekannte Erfolgwahrscheinlichkeit $p \in (0,1)$. (Dabei ist n die Anzahl der Versuche (bzw. der Stichprobenumfang) und K die zufällige Anzahl der Erfolge.)
(i) Was ist die Standardabweichung von H (in Abhängigkeit von p und n?)
(ii) Begründen Sie die folgende Aussage: $\frac{1}{2\sqrt{n}}$ ist eine obere Schranke für die Standadabweichung von H.
(iii) Für welches p wird die in (ii) angegebene Schranke erreicht?
(iv) Wie groß sollte (unter Verwendung von (ii)) der Stichprobenumfang n mindestens sein, wenn das Konfidenzintervall $[H-0.02, H+0.02]$ den Parameter p mit einer Wahrscheinlichkeit $\gtrapprox 95\%$ überdecken soll?

b) Geben Sie mittels der in der Vorlesung (für passendes X) hergeleiteten Ungleichung

$$\mathbf{P}(X \geq \alpha n) \leq e^{-n\left(\alpha \ln\left(\frac{\alpha}{p}\right)+(1-\alpha)\ln\left(\frac{1-\alpha}{1-p}\right)\right)}$$

eine Abschätzung für die Wahrscheinlichkeit an, dass man bei einem 10000-fachen fairen Münzwurf mindestens 5100 mal Kopf bekommt.

Lösungshinweise:

a)

(i) Es ist $K \sim \text{Binom}(n,p)$, also gilt $\mathbf{Var}[K] = npq$ mit $q := 1-p$, womit wir

$$\mathbf{Var}\,[H] = \mathbf{Var}\,[K/n] = \frac{1}{n^2}\mathbf{Var}[K] = \frac{pq}{n}$$

und somit $\sigma_H = \sqrt{\frac{pq}{n}} = \frac{1}{\sqrt{n}} \cdot \sqrt{p(1-p)}$ erhalten (vgl. dazu Vorlesung 12a, Folie 5).

(ii) Die Idee ist für festes n den Term $\frac{1}{\sqrt{n}} \cdot \sqrt{p(1-p)}$ zu maximieren, also maximieren wir $f:(0,1) \to \mathbb{R}, f(p) = p(1-p)$. Da $f'(p) = 1-2p$ gilt und somit $f'(1/2) = 0$ sowie $f''(p) = -2$, liegt bei $p = \frac{1}{2}$ offenbar ein Maximum vor, also gilt

$$\forall p \in (0,1): \quad \sigma_h \leq \frac{1}{\sqrt{n}} \cdot \sqrt{f(1/2)} = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{1}{4}} = \frac{1}{2\sqrt{n}}.$$

(iii) In (ii) haben wir bereits $p = \frac{1}{2}$ ermittelt.

(iv) Die Forderung, dass mit W'keit 0.95 der Abstand von H und p nicht größer als 0.02 sein soll, führt auf die Bedingung

$$2\sigma_H \leq 0.02,$$

vgl. Folie 6 aus Vorlesung 12.a). Da $\sigma_H \leq \frac{1}{2\sqrt{n}}$ gilt suchen wir $n \in \mathbb{N}$ mit

$$\frac{1}{2\sqrt{n}} \leq 0.01 \Leftrightarrow n \geq \frac{1}{0.02^2} \Leftrightarrow n \geq 2.500$$

b) Gemeint ist die Chernoff-Schranke aus Vorlesung 7c. X sei also die Anzahl der Kopfwürfe in n unabhängigen Münzwürfen mit $p = \frac{1}{2}$, dann gilt $X \sim \text{Binom}(n, p)$ und für $n = 10.000$

$$\mathbb{P}\left(X \geq 5100\right) = \mathbb{P}\left(X \geq 0,51 \cdot 10.000\right)$$

und somit für $n = 10.000$, $p = \frac{1}{2}$ und $\alpha = 0,51$

$$\mathbb{P}(X \geq 5.100) \leq e^{-n\left(\alpha \ln\left(\frac{\alpha}{p}\right) + (1-\alpha)\ln\left(\frac{1-\alpha}{1-p}\right)\right)} \approx 0,1353.$$

Nebenbei bemerkt: Würde man hier mit der Normalapproximation rechnen, so erhielte man wegen $\sigma_X = \sqrt{\frac{1}{4}10000} = 50$ die Näherung $\mathbf{P}(X \geq \mu_X + 2\sigma_X) \approx 0.025$, was um einiges schärfer ist als die Chernoff-Schranke. Letztere entwickelt ihre Kraft bei "großen Abweichungen, hier ist man noch "in der Nähe des Zentrums".

3. X und Y seien reellwertige Zufallsvariable mit $\mu_X = \mu_Y = 5$, $\sigma_X = 1$, $\sigma_Y = 3$ und Regressionsgerade $\hat{Y} = -5 + 2X$.
a) *Wir drehen den Spieß um und fragen nach der besten affin linearen Vorhersage von X auf der Basis von Y:*
Finden Sie die Koeffizienten α_0 und α_1 der Regressionsgeraden $\hat{X} = \alpha_0 + \alpha_1 Y$.
b) Angenommen die bedingte Verteilung von Y gegeben $\{X = x\}$ ist von der Form $\mathrm{N}(\beta_0 + \beta_1 x, v)$ mit einer nicht von x abhängenden Varianz v.
(i) Was sind dann die Werte von β_0, β_1 und v?
(ii) Geben Sie ein (möglichst kurzes) Intervall J an, für welches gilt: $\mathbf{P}(Y \in J \mid X = 3) \approx 0.95$.

Lösungshinweise:

a) Wir lesen aus der Regressionsgerade $\hat{Y}$, welche Y auf der Basis von X prognostiziert

$$2 = \beta_1 = \frac{\mathbf{Cov}(X,Y)}{\sigma_X^2} = \mathbf{Cov}(X,Y)$$

ab und erhalten damit

$$\alpha_1 = \frac{\mathbf{Cov}(X,Y)}{\sigma_Y^2} = \frac{2}{3^2} = \frac{2}{9}, \quad \alpha_0 = \mu_x - \alpha_1 \mu_y = \frac{35}{9}.$$

b)
(i) Wir lesen die Werte für β_0, β_1 aus der Regressionsgeraden $\hat{Y}$ ab:

$$\beta_0 = -5, \quad \beta_1 = 2$$

Nach Voraussetzung hängt die bedingte Varianz von Y gegeben X nicht von X ab:

$$\mathbf{Var}[Y|X] = v.$$

Für die bedingte Erwartung von Y gegeben X gilt laut Angabe:

$$\mathbf{E}[Y|X] = -5 + 2X.$$

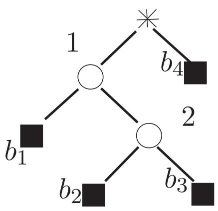
Damit ergibt sich mit dem Satz von der Zerlegung der Varianz:

$$9 = \sigma_Y^2 = \mathbf{Var}[-5 + 2X] + \mathbf{E}[v] = 4\sigma_X^2 + v = 4 + v,$$

also $v = 5$.
(ii) Mit $\mathbf{E}[Y|X = 3] = -5 + 2 \cdot 3 = 1$ liefert uns die 2σ-Regel

$$\left[1 - 2 \cdot \sqrt{5}, 1 + 2 \cdot \sqrt{5}\right] = [-3.47, 5.47].$$

4.



a) Wir betrachten die Irrfahrt "in Richtung der Blätter" auf dem skizzierten Baum: von jedem inneren Knoten κ führt der nächste Schritt zu einem zufälligen gewählten Nachfolger von κ, und für jedes Blatt b gilt $P(b, b) = 1$.
(i) Was ist die Wahrscheinlichkeit, dass die Irrfahrt bei Start in $*$ das Blatt b_1 (siehe Skizze) erreicht?
(ii) Geben Sie mindestens zwei verschiedene Gleichgewichtsverteilungen der Irrfahrt an.

b) Jetzt betrachten wir den skizzierten Baum als *ungerichteten Graphen*, und die gewöhnliche Irrfahrt auf diesem.
(i) Was ist die Wahrscheinlichkeit, dass diese Irrfahrt bei Start in $*$ von allen Blättern als erstes das Blatt b_1 erreicht?
(ii) Finden Sie die Gleichgewichtsverteilung der Irrfahrt.

Lösungshinweise:

a)

(i) Die Irrfahrt verläuft von oben nach unten, es gibt somit genau einen Pfad von $*$ nach b_1, also

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

(ii) Die Irrfahrt endet immer in einem der Blätter und verlässt dieses nicht mehr, für beliebiges $i \in \{1, 2, 3, 4\}$ erhalten wir also eine Gleichgewichtsverteilung π mit $\pi(b_i) = 1$ und $\pi(b_j) = 0$ für $j \neq i$ sowie $\pi(1) = \pi(2) = \pi(*) = 0$. Ähnlich dazu ist Aufgabe 48.ii).

b)

(i) Bezeichne $f(a)$ die Wahrscheinlichkeit, dass die Irrfahrt ausgehend von a von allen Blättern das Blatt b_1 zuerst erreicht, dann gilt folgendes Gleichungssystem

$$f(*) = \frac{1}{2} f(1), \quad f(1) = \frac{1}{3} + \frac{1}{3} f(2) + \frac{1}{3} f(*), \quad f(2) = \frac{1}{3} f(1),$$

wobei wir bereits $f(b_1) = 1$ und $f(b_2) = f(b_3) = f(b_4) = 0$ eingesetzt haben. Auflösen des Gleichungssystems ergibt

$$f(*) = \frac{3}{13}$$

.

(ii) Es handelt sich um eine gewöhnliche Irrfahrt auf einem ungerichteten Graphen, wir wissen also, dass eine reversible Gleichgewichtsverteilung π existiert, so dass $\pi(a) = \frac{\#\text{Nachbarn von } a}{2 \cdot \#\text{Kanten}}$ gilt, also

$$\pi(*) = \frac{2}{12}, \quad \pi(1) = \pi(2) = \frac{3}{12}, \quad \pi(b_1) = \pi(b_2) = \pi(b_3) = \pi(b_4) = \frac{1}{12}.$$

5. a) Wir betrachten ein zweistufiges Zufallsexperiment. In Stufe 1 wird mit W'keit 0.9 eine faire Münze gewählt, und mit W'keit 0.1 eine gezinkte, die eine Erfolgswahrscheinlichkeit 0.6 aufweist. In der zweiten Stufe wird mit der gewählten Münze ein fortgesetzer Münzwurf durchgeführt. Wie wahrscheinlich ist es, dass dann auch der 7. Versuch ein Erfolg ist, gegeben die ersten 6 waren Erfolge?

b) Es seien $U_0, U_1, \ldots$ unabhängige, auf $[0,1]$ uniform verteilte Zufallsvariable. Wie wahrscheinlich ist es, dass U_0 das größte der $U_0, U_1, \ldots, U_8$ ist, gegeben dass U_0 das größte der $U_0, U_1, \ldots, U_7$ ist? *(Argumentieren Sie mit zufälligen Permutationen.)*

c) Wir stellen die selbe Frage wie in a), wobei jetzt die in der ersten Stufe gewählte zufällige Erfolgswahrscheinlichkeit der Münze uniform auf [0,1] verteilt ist.[1]

Lösungshinweise:

a) Wir suchen eine bedingte Wahrscheinlichkeit, es liegt also nahe mit dem Satz von Bayes zu arbeiten, dieser liefert uns

$$\begin{aligned}\mathbb{P}\,(\text{7. ist ein Erfolg } \mid \text{1. bis 6. sind Erfolge }) &= \frac{\mathbb{P}\,(\text{1. bis 7. sind Erfolge })}{\mathbb{P}\,(\text{1. bis 6. sind Erfolge })}\\ &= \frac{0.9 \cdot 0.5^7 + 0.1 \cdot 0.6^7}{0.9 \cdot 0.5^6 + 0.1 \cdot 0.6^6} \approx 0,5249\end{aligned}$$

b) Wir erinnern uns an Aufgabe 44). Es gibt 7! Permutationen der Zahlen $0, \ldots, 7$, so dass die 0 ganz rechts steht. Außerdem gibt es 8! Permutationen der Zahlen $0, \ldots, 8$, so dass die 0 ganz rechts steht. Damit erhalten wir wieder mit dem Satz von Bayes

$$\begin{aligned}&\mathbb{P}\,(U_0 \text{ größte der } U_0, \ldots, U_8 \mid U_0 \text{ größte der } U_0, \ldots, U_7)\\ =\;& \frac{\mathbb{P}\,(\{U_0 \text{ größte der } U_0, \ldots, U_8\} \cap \{U_0 \text{ größte der } U_0, \ldots, U_7\})}{\mathbb{P}\,(U_0 \text{ größte der } U_0, \ldots, U_7)}\\ =\;& \frac{\mathbb{P}\,(U_0 \text{ größte der } U_0, \ldots, U_8)}{\mathbb{P}\,(U_0 \text{ größte der } U_0, \ldots, U_7)}\\ =\;& \frac{8!/9!}{7!/8!} = \frac{8}{9}.\end{aligned}$$

Alternativ (und kürzer)
U_8 fällt mit gleicher W'keit in einen von 9 möglichen Slots (links vom kleinsten, zwischen dem kleinsten und dem zweitkleinsten,$\ldots$, rechts vom größten der $U_0, \ldots, U_7$. Gegeben dass U_0 das größte der $U_0, \ldots, U_7$ ist, führen genau 8 dieser 9 Möglichkeiten auf das Ereignis $\{U_0$ ist das größte der $U_0, \ldots, U_8\}$.

c) Verstehen wir U_0 als das zufällige p, so sind die Zufallsvariablen $\mathbf{1}_{U_i \le U_0}$ Münzwürfe mit

[1] Gegenüber der Angabe in der Klausur ist diese Formulierung sprachlich leicht verbessert.

zufälliger Erfolgswahrscheinlichkeit U_0. Wir befinden uns also genau in der Situation aus b), nur dass es jetzt um den Schritt von 6 auf 7 anstelle des Schrittes von 7 auf 8 geht, und erhalten somit das Ergebnis $\frac{7}{8}$ (anstelle des Ergebnisses 8/9).

6. X und Y seien unabhängig und $\mathrm{N}(0,1)$-verteilt. Wir fassen (X,Y) auf als Koordinaten eines in der Ebene $\mathbb{R}^2$ standard-normalverteilten Punktes P.
a) Finden Sie den Radius r eines Kreises um den Ursprung so, dass P mit W'keit 10^{-6} außerhalb des Kreises liegt.
b) Es sei R der Abstand des Punktes P vom Ursprung. Was ist die Wahrscheinlichkeit des Ereignisses $\{R \geq 2\}$, gegeben das Ereignis $\{R \geq \sqrt{2}\}$?

Hinweis: In den Übungen haben wir bewiesen, dass $R^2/2 := \frac{X^2+Y^2}{2}$ *standard-exponentialverteilt ist. Das dürfen Sie hier verwenden.*

Lösungshinweise:

a) Bezeichne R den Abstand von P zu Ursprung, dann gilt $R = \sqrt{X^2+Y^2}$ und damit

$$\begin{aligned}
\mathbb{P}(R \geq r) &= \mathbb{P}(R^2 \geq r^2) \\
&= \mathbb{P}\left(\frac{X^2+Y^2}{2} \geq \frac{r^2}{2}\right) \\
&\overset{\text{Hinweis}}{=} \exp(-\frac{r^2}{2}).
\end{aligned}$$

Wir setzen $\exp(-\frac{r^2}{2}) \overset{!}{=} 10^{-6}$ und erhalten $r \approx 5{,}2565$.

b) Wir berechnen die bedingte Wahrscheinlichkeit:

$$\begin{aligned}
\mathbb{P}\left(R \geq 2 \,|\, R \geq \sqrt{2}\right) &= \mathbb{P}\left(\frac{R^2}{2} \geq 2 \,|\, \frac{R^2}{2} \geq 1\right) \\
&= \frac{\mathbb{P}\left(\frac{R^2}{2} \geq 2 \cap \frac{R^2}{2} \geq 1\right)}{\mathbb{P}\left(\frac{R^2}{2} \geq 1\right)} \\
&= \frac{\mathbb{P}\left(\frac{R^2}{2} \geq 2\right)}{\mathbb{P}\left(\frac{R^2}{2} \geq 1\right)} \\
&\overset{\text{Hinweis}}{=} \frac{e^{-2}}{e^{-1}} = \frac{1}{e} \approx 0{,}3678
\end{aligned}$$

Alternativ mithilfe Gedächtnislosigkeit der Exponentialverteilung:

$$\mathbb{P}\left(\frac{R^2}{2} \geq 2 \,|\, \frac{R^2}{2} \geq 1\right) = \mathbb{P}\left(\frac{R^2}{2} \geq 2-1\right) = \mathbb{P}\left(\frac{R^2}{2} \geq 1\right) = \frac{1}{e} \approx 0{,}3678$$