Wahrscheinlichkeit und Statistik

Musterlösung Serie 3

1. **Das Geburtstagsparadox.** Wir betrachten eine Urne mit N Kugeln aus denen wir n Kugeln mit Zurücklegen ziehen.

   a) Sei A_n das Ereignis "Alle n Kugeln haben unterschiedliche Nummern."
   Berechnen Sie $\mathbb{P}(A_n)$.

   b) Beweisen Sie die beiden Ungleichungen

   $$1 - \frac{n(n-1)}{2N} \leq \mathbb{P}(A_n) \leq \exp\left(-\frac{n(n-1)}{2N}\right).$$

   c) Berechnen Sie $n_{\min} = \inf\{n \in \mathbb{N} : \mathbb{P}(A_n) < \frac{1}{2}\}$ für $N = 365$.

   Bemerkung: Im Fall $N = 365$ entspricht die Nummer der Kugel in Ziehung i dem Geburtstag der i-ten Person in einer Gruppe von n Personen. Daher der Name *Geburtstagsparadox.*

   Lösung:

   a) Die Wahrscheinlichkeit, dass alle n Kugeln unterschiedliche Nummern haben berechnet sich als

   $$\begin{aligned}\mathbb{P}(A_n) &= \frac{|A_n|}{\Omega} \\ &= \frac{N \cdot (N-1) \cdot (N-2) \cdot \ldots \cdot (N-n+1)}{N^n} \\ &= \prod_{j=1}^{n-1} \left(1 - \frac{j}{N}\right).\end{aligned}$$

   b) • Für die untere Schranke betrachten wir $\mathbb{P}(A_n^c)$. Wir bezeichnen mit k_j die Nummer der j-ten Kugel. Dann

   $$\begin{aligned}\mathbb{P}(A_n^c) &= \mathbb{P}(\exists\, i \neq j \in \{1, \ldots, n\} : k_j = k_i) \\ &= \mathbb{P}\left(\bigcup_{j=1}^{n} \bigcup_{i=1}^{j-1} \{k_j = k_i\}\right) \\ &\leq \sum_{j=1}^{n} \sum_{i=1}^{j-1} \underbrace{\mathbb{P}(k_j = k_i)}_{=\frac{1}{N}} \\ &= \frac{1}{N} \sum_{j=1}^{n} (j-1) = \frac{n(n-1)}{2N},\end{aligned}$$

   und mit $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n)$ können wir schliessen.

   • Für die obere Schranke ist zu beachten, dass für alle $x \in \mathbb{R}$ gilt

   $$1 - x \leq \exp(-x). \tag{1}$$

Dies sieht man durch Betrachtung der Funktion $f(x) := \exp(-x) - (1-x)$. Es gilt: $f(0) = 0$ und $f'(0) = 0$, und f ist konvex, da $f''(x) = \exp(-x) > 0$ für alle $x \in \mathbb{R}$. Daher ist $f(x) \geq 0$ für alle $x \in \mathbb{R}$, was (1) liefert. Unter Verwendung des Ausdrucks für $\mathbb{P}(A_n)$ vom Aufgabenteil a) erhalten wir

$$\begin{aligned}\mathbb{P}(A_n) &= \prod_{j=1}^{n-1}\left(1 - \frac{j}{N}\right) \\ &\overset{(1)}{\leq} \prod_{j=1}^{n-1} \exp\left(-\frac{j}{N}\right) = \exp\left(-\sum_{j=1}^{n-1}\frac{j}{N}\right) = \exp\left(-\frac{n(n-1)}{2N}\right).\end{aligned}$$

c) Wir bezeichnen mit $n_{\min}$ das minimale $n \in \mathbb{N}$, sodass

$$\mathbb{P}(A_n) < \frac{1}{2}. \tag{2}$$

Aus Aufgabenteil b) wissen wir wegen $\mathbb{P}(A_n) \leq \exp\left(-\frac{n(n-1)}{2N}\right)$, dass für das kleinste $n \in \mathbb{N}$ mit

$$\exp\left(-\frac{n(n-1)}{2N}\right) < \frac{1}{2} \tag{3}$$

auch die Ungleichung (2) erfüllt ist. Aus (3) erhalten wir die Ungleichung

$$-n^2 + n - 2 \cdot 365 \cdot \ln\left(\frac{1}{2}\right) < 0.$$

Es gilt: $730 \cdot \ln\left(\frac{1}{2}\right) \approx 505.997$. Deshalb ist

$$\frac{1 + \sqrt{1 + 4 \cdot 506}}{2} = 23$$

die kleinste Zahl in $\mathbb{N}$, sodass Ungleichung (3) erfüllt ist. Nun wissen wir, dass $n_{\min} \leq 23$. Aufgabenteil b) liefert auch eine untere Schranke für $n_{\min}$: Wegen

$$1 - \frac{n(n-1)}{2N} \leq \mathbb{P}(A_n),$$

wissen wir, dass

$$1 - \frac{n(n-1)}{2 \cdot 365} < \frac{1}{2}$$

Das grösste solche n ist

$$n = \left\lfloor \frac{1 + \sqrt{1 + 4 \cdot 365}}{2} \right\rfloor = 19.$$

Deshalb muss $n_{\min} \in \{20, 21, 22, 23\}$ sein. In der Tat ist $n_{\min} = 23$, was man durch Berechnen des Wertes von $\mathbb{P}(A_n)$ für $n \in \{20, 21, 22, 23\}$ einsieht.

2. Wir untersuchen die Erfolgswahrscheinlichkeiten bei einer Aufnahmeprüfung an zwei Departementen einer Universität und betrachten folgende Ereignisse:

$$\begin{aligned}
A &= \text{“Kandidat ist männlich.”}\\
A^c &= \text{“Kandidat ist weiblich.”}\\
B &= \text{“Kandidat bewirbt sich bei Departement I.”}\\
B^c &= \text{“Kandidat bewirbt sich bei Departement II.”}\\
C &= \text{“Kandidat wird aufgenommen.”}\\
C^c &= \text{“Kandidat wird abgelehnt.”}
\end{aligned}$$

Wir nehmen an, dass die folgenden Wahrscheinlichkeiten gelten (diese Zahlen entsprechen relativen Häufigkeiten in Berkeley 1973):

$$\begin{aligned}
&\mathbb{P}[A] = 0.73,\\
&\mathbb{P}[B|A] = 0.69,\ \mathbb{P}[B|A^c] = 0.24,\\
&\mathbb{P}[C|A\cap B] = 0.62,\ \mathbb{P}[C|A^c\cap B] = 0.82,\ \mathbb{P}[C|A\cap B^c] = 0.06,\ \mathbb{P}[C|A^c\cap B^c] = 0.07.
\end{aligned}$$

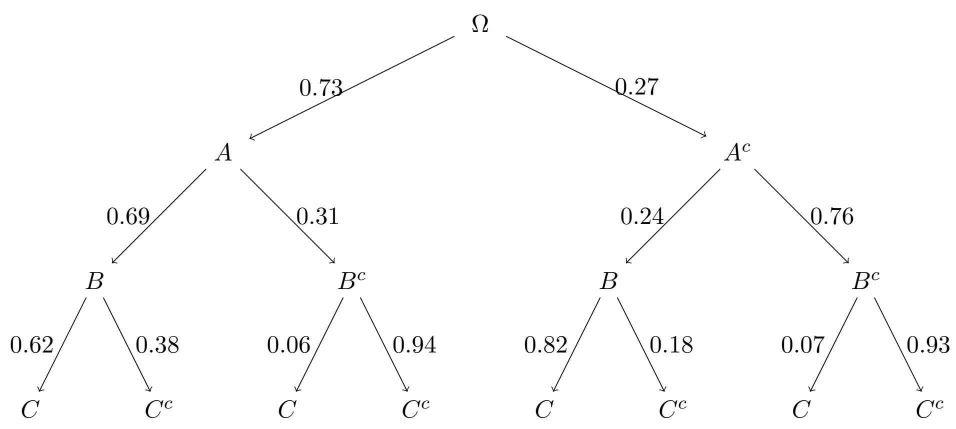
a) Zeichnen Sie den zu dieser Situation gehörigen Baum.

b) Erläutern Sie mit Ihren eigenen Worten: $\mathbb{P}[C|A\cap B] = 0.62$, $\mathbb{P}[C|A^c\cap B] = 0.82$, $\mathbb{P}[C|A\cap B^c] = 0.06$, $\mathbb{P}[C|A^c\cap B^c] = 0.07$. Finden Sie, dass aufgrund dieser Zahlen Frauen benachteiligt sind?

c) Berechnen Sie $\mathbb{P}[C|A]$ und $\mathbb{P}[C|A^c]$ und erläutern Sie mit Ihren eigenen Worten. Vergleichen Sie Ihre Resultate mit Ihrer Antwort in b).

Lösung:

a) Wir bekommen den folgenden binären Baum der Höhe 3 mit 8 Blättern.



b)
$$\mathbb{P}[C|A\cap B] = 0.62$$

ist die Wahrscheinlichkeit, aufgenommen zu werden unter der Bedingung, dass man männlich ist und sich bei Department I beworben hat.

$$\mathbb{P}[C|A^c\cap B] = 0.82$$

ist die Wahrscheinlichkeit, aufgenommen zu werden unter der Bedingung, dass man weiblich ist und sich bei Department I beworben hat.

$$\mathbb{P}[C|A \cap B^c] = 0.06$$

ist die Wahrscheinlichkeit, aufgenommen zu werden unter der Bedingung, dass man männlich ist und sich bei Department II beworben hat.

$$\mathbb{P}[C|A^c \cap B^c] = 0.07$$

ist die Wahrscheinlichkeit, aufgenommen zu werden unter der Bedingung, dass man weiblich ist und sich bei Department II beworben hat.

Frauen sind nicht benachteiligt, eher im Gegenteil, die Wahrscheinlichkeit, dass sie aufgenommen werden gegeben, dass sie sich bei Department I bewerben, ist höher als diejenige bei den Männern.

c) Wir berechnen die folgenden Wahrscheinlichkeiten

$$\begin{aligned}\mathbb{P}[C|A^c] &= \frac{\mathbb{P}[C \cap A^c]}{\mathbb{P}[A^c]} = \frac{\mathbb{P}[C \cap A^c \cap B] + \mathbb{P}[C \cap A^c \cap B^c]}{\mathbb{P}[A^c]} \\ &= \frac{0.82 \cdot 0.27 \cdot 0.24 + 0.07 \cdot 0.27 \cdot 0.76}{0.27} \sim 0.25\end{aligned}$$

und

$$\begin{aligned}\mathbb{P}[C|A] &= \frac{\mathbb{P}[C \cap A]}{\mathbb{P}[A]} = \frac{\mathbb{P}[C \cap A \cap B] + \mathbb{P}[C \cap A \cap B^c]}{\mathbb{P}[A]} \\ &= \frac{0.62 \cdot 0.69 \cdot 0.73 + 0.06 \cdot 0.31 \cdot 0.73}{0.73} \sim 0.45\end{aligned}$$

Die Wahrscheinlichkeit, aufgenommen zu werden, gegeben, dass man männlich ist, ist 0.45, die Wahrscheinlichkeit, aufgenommen zu werden, gegeben, dass man weiblich ist, ist 0.25. Hieraus könnte man voreilig schliessen, dass Frauen benachteiligt werden. Allerdings berücksichtigt man hier nicht, dass sich Frauen überwiegend bei Department II bewerben, dass eine viel grössere Ablehnungsquote als Department I aufweist.

3. **Einführung in die Bayes-Statistik.** Wir betrachten m Urnen mit folgender Zusammensetzung: Jede Urne enthält $2m$ Kugeln, und für jedes $i \in \{1, \dots, m\}$ enthält Urne i genau $2i-1$ rote, und $2m-2i+1$ weisse Kugeln. Wir wählen zufällig eine Urne und ziehen daraus n-mal mit Zurücklegen. Es bezeichne $(X_1, \dots, X_n)$ das Ergebnis der n Ziehungen, wobei

$$X_j = \begin{cases} 1 \Leftrightarrow j\text{-te Kugel ist rot} \\ 0 \Leftrightarrow j\text{-te Kugel ist weiss.} \end{cases} \tag{4}$$

Die Zusammensetzung der gewählten Urne ist nicht direkt ersichtlich, aber man kann versuchen auf Grund der Ergebnisse $(X_1, \dots, X_n)$ Rückschlüsse zu ziehen.

a) Berechnen Sie $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$, wobei $x_j \in \{0, 1\}$ für $j \in \{1, \dots, m\}$ ist. Sind $X_1, \dots, X_n$ unabhängig?

b) Berechnen Sie

$$\mathbb{P}\left(\text{Gewählte Urne enthält } 2i-1 \text{ rote Kugeln} \,|\, X_1 = x_1, \ldots, X_n = x_n\right)$$

und zeigen Sie, dass diese Wahrscheinlichkeit nur von $k := \sum_{j=1}^n x_j$ (d.h. von der Gesamtzahl der roten Kugeln in der Stichprobe) abhängt. Geben Sie die Zahlenwerte an für $n = 3,\ m = 3$ in den Fällen $k \in \{0, 1, 2, 3\}$

	$i = 1$	$i = 2$	$i = 3$
$k = 0$			
$k = 1$			
$k = 2$			
$k = 3$			

und vergleichen Sie diese mit Ihrer Vermutung vor dem Rechnen.

Lösung:

a) Sei $k := \sum_{j=1}^n x_j$ die Gesamtzahl der roten Kugeln in der gegebenen Stichprobe $X_1 = x_1, \ldots, X_n = x_n$, wobei n den Umfang der Stichprobe bezeichnet. Dann ist

$$\begin{aligned}
&\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \\
&= \sum_{i=1}^m \mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n \,|\, \text{Urne } i \text{ gewählt}\right) \mathbb{P}\left(\text{Urne } i \text{ gewählt}\right) \\
&= \sum_{i=1}^m \left(\frac{2i-1}{2m}\right)^k \left(\frac{2(m-i)+1}{2m}\right)^{n-k} \frac{1}{m},
\end{aligned}$$

wobei $\left(\frac{2i-1}{2m}\right)^k \left(\frac{2(m-i)+1}{2m}\right)^{n-k}$ die Wahrscheinlichkeit beschreibt, k rote und $n-k$ weisse Kugeln in der vorgegebenen Reihenfolge $X_1 = x_1, \ldots, X_n = x_n$ zu ziehen, wenn wir Urne i gewält haben, und $\frac{1}{m}$ ist die Wahrscheinlichkeit, Urne i auszuwählen.
Um Unabhängigkeit verwerfen zu können, genügt es, ein Gegenbeispiel zu finden. Wir betrachten die Ereignisse $\{X_1 = 1\}$ und $\{X_2 = 1\}$. Dann gilt es, dass

$$\begin{aligned}
\mathbb{P}\left(\{X_1 = 1\} \cap \{X_2 = 1\}\right) &= \mathbb{P}\left(X_1 = 1, X_2 = 1\right) \\
&= \sum_{i=1}^m \left(\frac{2i-1}{2m}\right)^2 \frac{1}{m} \\
&\neq \left(\sum_{i=1}^m \frac{2i-1}{2m} \frac{1}{m}\right)^2 = \mathbb{P}\left(\{X_1 = 1\}\right) \mathbb{P}\left(\{X_2 = 1\}\right),
\end{aligned}$$

deshalb sind $X_1, \ldots, X_n$ nicht unabhängig.

b) Um die gesuchte bedingte Wahrscheinlichkeit zu berechnen, kann man folgende Überlegungen anstellen

$$\begin{aligned}
&\mathbb{P}\left(\text{Gewählte Urne enthält } 2i-1 \text{ rote Kugeln} \,|\, X_1 = x_1, \ldots, X_n = x_n\right) \\
&= \frac{\mathbb{P}\left(\{\text{Urne } i \text{ gewählt}\} \cap \{X_1 = x_1, \ldots, X_n = x_n\}\right)}{\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n\right)} \\
&= \frac{\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n \,|\, \text{Urne } i \text{ gewählt}\right)}{\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n\right)} \mathbb{P}\left(\text{Urne } i \text{ gewählt}\right) \\
&= \frac{\left(\frac{2i-1}{2m}\right)^k \left(\frac{2(m-i)+1}{2m}\right)^{n-k}}{\sum_{j=1}^m \left(\frac{2j-1}{2m}\right)^k \left(\frac{2(m-j)+1}{2m}\right)^{n-k} \frac{1}{m}} \frac{1}{m}
\end{aligned}$$

Deshalb berechnen sich die Zahlenwerte an für $n = 3,\ m = 3$ in den Fällen $k \in \{0, 1, 2, 3\}$ als

	$i = 1$	$i = 2$	$i = 3$
$k = 0$	0.817	0.176	0.007
$k = 1$	0.439	0.474	0.088
$k = 2$	0.088	0.474	0.439
$k = 3$	0.007	0.176	0.817

.

4. **Das Ziegen Problem.** Sie befinden sich in einer Spielshow in der es ein Auto zu gewinnen gibt. Dem Spielablauf liegen folgende Regelnmzugrunde:

   1. Ein Auto und zwei Ziegen sind auf drei Tore verteilt, alle Tore sind verschlossen, sodass Auto und Ziegen nicht sichtbar sind.
   2. Sie wählen ein Tor aus, welches aber vorerst verschlossen bleibt.
   3. Falls Sie das Tor mit dem Auto gewählt haben, dann öffnet der Moderator eines der beiden anderen Tore, hinter dem sich immer eine Ziege befindet. Falls Sie ein Tor mit einer Ziege gewählt haben, dann öffnet der Moderator dasjenige der beiden anderen Tore, hinter dem die zweite Ziege steht.
   4. Der Moderator bietet Ihnen an, die erste Entscheidung zu überdenken und das andere ungeöffnete Tor zu wählen.
   5. Das letztlich gewählte Tor wird geöffnet und Sie gewinnen das Auto, falls es sich hinter diesem Tor befindet.

   Um im vorletzten Schritt die richtige Entscheidung zu treffen überlegen wir uns folgendes: Wir bezeichnen das von Ihnen gewählte Tor mit 1 und betrachten die Ereignisse

$$
\begin{aligned}
B_i &= \text{“Auto ist hinter Tor i.”} \qquad && (i = 1, 2, 3),\\
A_j &= \text{“Moderator öffnet Tor j.”} \qquad && (j = 2, 3).
\end{aligned}
$$

   a) Welche Wahl der Wahrscheinlichkeiten $\mathbb{P}(B_i)$ für $i \in \{1, 2, 3\}$ sowie $\mathbb{P}(A_j|B_i)$ für $i \in \{1, 2, 3\}$ und $j \in \{1, 2\}$ ist angebracht?

   b) Berechnen Sie mit der Bayes-Formel $\mathbb{P}(B_1|A_j)$ für $j \in \{2, 3\}$. Hat das Öffnen des anderen Tores mit einer Ziege dahinter die Wahrscheinlichkeit verändert, dass sich das Auto hinter Tor 1 befindet?

   c) Drücken Sie das Ereignis “Auto gewonnen” mit Hilfe von A_i und B_j aus in den beiden Fällen, wo Sie in Schritt 4 immer, bzw. nie wechseln.

   d) Würden Sie im vorletzten Schritt das Tor wechseln?

Lösung:

a) $\mathbb{P}(B_1) = \mathbb{P}(B_2) = \mathbb{P}(B_3) = \frac{1}{3}$.
$\mathbb{P}(A_2|B_1) = \mathbb{P}(A_3|B_1) = \frac{1}{2}$.
$\mathbb{P}(A_2|B_2) = 0 \quad \mathbb{P}(A_3|B_2) = 1$.
$\mathbb{P}(A_2|B_3) = 1 \quad \mathbb{P}(A_3|B_3) = 0$.

b)

$$\begin{aligned}\mathbb{P}(B_1|A_2) &= \frac{\mathbb{P}(A_2|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A_2|B_1)\mathbb{P}(B_1)+\mathbb{P}(A_2|B_2)\mathbb{P}(B_2)+\mathbb{P}(A_2|B_3)\mathbb{P}(B_3)}\\ &= \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}\frac{1}{3}+0\frac{1}{3}+1\frac{1}{3}} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.\end{aligned}$$

$$\begin{aligned}\mathbb{P}(B_1|A_3) &= \frac{\mathbb{P}(A_3|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A_3|B_1)\mathbb{P}(B_1)+\mathbb{P}(A_3|B_2)\mathbb{P}(B_2)+\mathbb{P}(A_3|B_3)\mathbb{P}(B_3)}\\ &= \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}\frac{1}{3}+1\frac{1}{3}+0\frac{1}{3}} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.\end{aligned}$$

D.h. die Wahrscheinlichkeit dass sich das Auto tatsächlich hinter Tor 1 befindet, liegt also auch nach dem Öffnen eines der anderen beiden Tore, nach wie vor bei $\frac{1}{3}$, unabhängig davon welches Tor geöffnet wurde.

c) Falls man das Tor nie wechselt, gewinnt man das Auto genau dann wenn die erste Wahl auf das Auto fällt. Das Ereignis "Auto gewonnen" tritt also im Fall der Strategie "nie wechseln" genau dann ein, wenn das Ereignis B_1 eintritt.
Andererseits, falls man das Tor immer wechselt, gewinnt man das Auto genau dann, wenn die erste Wahl auf eine Ziege fällt (und der Moderator die andere Ziege zeigt), d.h. das Ereignis "Auto gewonnen" tritt im Fall der Strategie "immer wechseln" genau dann ein, wenn das Ereignis $(B_2 \cap A_3) \cup (B_3 \cap A_2)$ eintritt.

d) Da $\mathbb{P}(B_1) = \frac{1}{3}$ und $\mathbb{P}((B_2 \cap A_3) \cup (B_3 \cap A_2)) = \frac{2}{3}$, ist es ratsam, das Tor zu wechseln.