Musterlösung zu Übungsblatt 4

erstellt von Jonas Strauch

13.S *Erwartungswert des Stichprobenmittels beim Ziehen ohne Zurücklegen.*
Für diese Aufgabe können wir das Argument der Austauschbarkeit, das wir in Aufgabe 12 auf dem letzten Blatt entdeckt hatten verwenden und folgern so, dass $h(T_1)$, $h(T_2)$ und $h(T_3)$ identisch verteilt sind.
Ein alternatives Argument wäre es, die 3 gezogenen Bäume nachträglich "durchzumischen" und statt T_1, T_2 und T_3 nun $T_{\Pi(1)}$, $T_{\Pi(2)}$ und $T_{\Pi(3)}$ zu betrachten, mit einer rein zufälligen Permutation Π von $1, 2, 3$. Das arithmetische Mittel der drei Höhen wird sich hierbei nicht ändern, die Höhen $h(T_{\Pi(1)})$, $h(T_{\Pi(2)})$ und $h(T_{\Pi(3)})$ sind aufgrund der rein zufälligen Permutation aber identisch verteilt.
Dank dieser Überlegung und der Linearität des Erwartungswertes können wir berechnen:

$$\begin{aligned}\mathbf{E}\left[\frac{1}{3}(h(T_1)+h(T_2)+h(T_3))\right] &= \frac{1}{3}\mathbf{E}[h(T_1)+h(T_2)+h(T_3)] = \frac{1}{3}\mathbf{E}[3\cdot h(T_1)] = \mathbf{E}[h(T_1)] \\ &= 0,4\cdot 1 + 0,3\cdot 2 + 0,3\cdot 3 = 1,9.\end{aligned}$$

14. *Wie wahrscheinlich ist eine so große Abweichung vom Erwartungswert?*
Wir ziehen ohne Zurücklegen, haben es also mit einer hypergeometrischen Verteilung zu tun.
R ist hypergeometrisch zu den Parametern $n = 15$ (wie oft wird gezogen?), $r + b = 50$ (wie viele Kugeln sind insgesamt in der Urne?) und $r = 30$ (wie viele „günstige" Kugeln?) verteilt. Aus Vorlesung 3a4 wissen wir, dass für den Erwartungswert von R gilt:

$$\mathbf{E}[R] = n\frac{r}{r+b} = 15\cdot\frac{30}{50} = 9.$$

a) R wird mindestens den Wert 0 und höchstens den Wert 15 annehmen (wir ziehen genau 15 Kugeln), außerdem ist R ganzzahlig.
Je nachdem ob a größer oder kleiner als $\mathbf{E}[R]$ ist können wir die Bedingung, mindestens eine Differenz von 6 zum Erwartungswert zu haben zu $a \geq 15$ bzw $a \leq 3$ umstellen.
Genau die Zahlen 0, 1, 2, 3 und 15 erfüllen also die Bedingung. Wir fassen sie als Ergebnis in der Menge $A := \{0, 1, 2, 3, 15\}$ zusammen.
b)(i) Wir wissen, welche Werte R annehmen darf damit die Bedingung $|R - \mathbf{E}[R]| \geq 6$ erfüllt ist, nun können wir die einzelnen Wahrscheinlichkeiten aufaddieren und erhalten:

$$\begin{aligned}\mathbf{P}(|R-\mathbf{E}[R]|\geq 6) &= \sum_{a\in A}\mathbf{P}(R=a) \\ &= \sum_{a\in A}\frac{\binom{r}{a}\binom{b}{n-a}}{\binom{r+b}{n}} \\ &= \sum_{a\in A}\frac{\binom{30}{a}\binom{20}{15-a}}{\binom{50}{15}} \\ &= \frac{\binom{30}{0}\binom{20}{15}}{\binom{50}{15}} + \frac{\binom{30}{1}\binom{20}{14}}{\binom{50}{15}} + \frac{\binom{30}{2}\binom{20}{13}}{\binom{50}{15}} + \frac{\binom{30}{3}\binom{20}{12}}{\binom{50}{15}} + \frac{\binom{30}{15}\binom{20}{0}}{\binom{50}{15}} \\ &\approx 0,00031\end{aligned}$$

(ii) Mithilfe von `dhyper()` können wir das Verteilungsgewicht für ein a unter einer hypergeometrischen Verteilung berechnen, hierfür müssen a und die definierenden Parameter der Verteilung an die Funktion übergeben werden, d.h.:

$$\texttt{dhyper(a, r, b, n)} = \mathbf{P}(R = a)$$

Die Parameter der hypergeometrischen Verteilung sind hier wie in den Folien benannt.
Nun können wir wie beim Rechnen per Hand die einzelnen Wahrscheinlichkeiten aufaddieren, es bietet sich zB eine for-Schleife an, oder wie folgt mit Hilfe von Vektoren:

```
sum(dhyper(c(0,1,2,3,15), 30, 20, 15))
0.000311643
```

15. *Die Wahrscheinlichkeit für Fixpunktfreiheit einer rein zufälligen Abbildung.*
Wir berechnen hier die Wahrscheinlichkeit, dass es in einer rein zufälligen Abbildung $\{1 \dots n\} \rightarrow \{1 \dots n\}$ mindestens einen Fixpunkt gibt. Das selbe haben wir schon in der Vorlesung gemacht, dort aber für eine rein zufällige Permutation.
Aufgrund der reinen Zufälligkeit der Permutation gilt $\mathbf{P}(E_i) = \mathbf{P}(E_1) = \frac{1}{n} \quad \forall i \in \{1 \dots n\}$
a)(i) Mit der de-morganschen Regel gilt:

$$\begin{aligned}
\mathbf{P}(E_1 \cup \dots \cup E_n) &= \mathbf{P}((E_1^c \cap \dots \cap E_n^c)^c) \\
&= 1 - \mathbf{P}(E_1^c \cap \dots \cap E_n^c) \\
&= 1 - \prod_{i=1}^{n} \mathbf{P}(E_i^c) && \left(\begin{array}{c}\text{Hier nutzen wir die Unabhängigkeit der } E_i \\ \text{mehr dazu im weiteren Verlauf der Vorlesung}\end{array}\right) \\
&= 1 - \mathbf{P}(E_1^c)^n \\
&= 1 - \left(1 - \frac{1}{n}\right)^n
\end{aligned}$$

(ii)Nun verwenden wir die Einschluss-Ausschluss-Regel. Wie in (i) nutzen wir aus, dass $\mathbf{P}(E_i)$ für alle E_i gleich ist und die E_i unabhängig (*).

$$\begin{aligned}
\mathbf{P}(E_1 \cup \dots \cup E_n) &\overset{\text{E-A-Regel}}{=} \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} \left((-1)^{|I|+1} \cdot \mathbb{P}\left(\bigcap_{i \in I} E_i \right) \right) \\
&\overset{(*)}{=} \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} \left((-1)^{|I|+1} \cdot \left(\mathbb{P}\left(E_1\right)\right)^{|I|} \right) \\
&\overset{\mathbb{P}(E_1)=\frac{1}{n}}{=} \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} \left((-1)^{|I|+1} \cdot \left(\frac{1}{n}\right)^{|I|} \right) \\
&\overset{\text{Umsummieren}}{=} \sum_{i=1}^{n} \left(\sum_{I \subseteq \{1,\dots,n\} : |I| = i} \left((-1)^{i+1} \cdot \left(\frac{1}{n}\right)^{i} \right) \right) \\
&\overset{\text{Def. von } \binom{n}{i}}{=} \sum_{i=1}^{n} \left(\binom{n}{i} \left((-1)^{i+1} \cdot \left(\frac{1}{n}\right)^{i} \right) \right) \\
&= \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} \left(\frac{1}{n}\right)^{i}
\end{aligned}$$

b) Wir rechnen mit dem Ergebnis aus (ii) weiter und vereinfachen, bis wir das Ergebnis aus (i) erhalten. So zeigen wir die geforderte Gleichheit, die Rechnung „in diese Richtung" ist vielleicht einsichtiger.
Wir lassen die Summe bei $i = 0$, der 0-te Summand ist -1, um das Ergebnis nicht zu verändern addieren wir 1 direkt wieder.

$$\begin{aligned}
\sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}\left(\frac{1}{n}\right)^{i} &= 1+\sum_{i=0}^{n}(-1)^{i-1}\binom{n}{i}\left(\frac{1}{n}\right)^{i} \\
&= 1-\sum_{i=0}^{n}(-1)^{i}\binom{n}{i}\left(\frac{1}{n}\right)^{i} \\
&= 1-\sum_{i=0}^{n}\binom{n}{i}\left(-\frac{1}{n}\right)^{i} \\
&= 1-\sum_{i=0}^{n}\binom{n}{i}1^{n-i}\left(-\frac{1}{n}\right)^{i}
\end{aligned}$$

Diese Summe können wir nun dank folgender Idee zu $(1-\frac{1}{n})^n$ umformen (um so unser Ergebnis aus a) (i) zu erhalten): Mit dem Binomialkoeffizienten zählen wir jeweils, wie viele Möglichkeiten es gibt, aus den Faktoren von $\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{1}{n}\right)$ je i-mal $-\frac{1}{n}$ und $n{-}i$-mal 1 auszuwählen. (Mittels dieser Beobachtung wurde in der Vorlesung 2a2 der binomischen Lehrsatz bewiesen.)
c) Wir berechnen

$$\lim_{n\to\infty} 1-\left(1-\frac{1}{n}\right)^{n} = 1-\frac{1}{e}$$

Interessanterweise ist das der gleiche Grenzwert wie bei der Wahrscheinlichkeit, dass eine rein zufällige Permutation mindestens einen Fixpunkt hat.

16.S *Poisson-und Exponentialapproximation.*
Sei n beliebig aber fest.
Um die Anzahl der Treffer und die Trefferzeit zu zählen definieren wir uns für jedes geworfene X_i eine Indikatorvariable I_i, die anzeigt, ob die Diagonale getroffen wurde. (Formal ist I_i Indikator zum Ereignis $\{X_i \in D\}$.)
Wir wählen ein rein zufälliges Pixel, die Wahrscheinlichkeit die Diagonale zu treffen lässt sich also mit $\frac{\#\text{günstige}}{\#\text{mögliche}}$ berechnen. Das Quadrat hat n^2 Pixel, n liegen auf der Diagonalen.
Somit ergibt sich:

$$\mathbf{P}(I_i = 1) = \frac{n}{n^2} = \frac{1}{n}.$$

a) Bezeichnen wir mit Z_m die Anzahl der „Diagonalentreffer" bei m Würfen. Pixel können mehrfach getroffen werden. Es handelt sich also um ein Ziehen mit Zurücklegen und Z_m ist binomialverteilt zu den Parametern m (wie oft wird gezogen/gewählt?) und $p = \frac{1}{n}$ (Erfolgswahrscheinlichkeit pro Zug). Nochmal formal:
$Z_m = \sum_{i=1}^{m} I_i$, Z_m ist also die Summe unabhängiger, identisch Bernoulli-verteilter Zufallsvariablen und somit binomialverteilt.

(i) Wir berechnen $\mathbf{P}(Z_m = 0)$ näherungsweise mithilfe der Poissonapproximation, diese lautet (bei großem m und kleinem p mit $\lambda := mp$, dem Erwartungswert der Binomialverteilung)

$$\mathbf{P}(Z_m = k) \approx e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2\ldots$$

Für $m := 2n$ haben wir $\lambda = 2n \cdot \frac{1}{n} = 2$ und damit ist:

$$\mathbf{P}(Z_{2n} = 0) \approx e^{-\lambda}\frac{\lambda^0}{0!} = \frac{1}{e^2} \approx 0,135.$$

(ii) Hier machen wir uns die Gegenwahrscheinlichkeit zu Nutze

$$\mathbf{P}(Z_{3n} \geq 1) = 1 - \mathbf{P}(Z_{3n} < 1) = 1 - \mathbf{P}(Z_{3n} = 0)$$

Für die Poissonapproximation benötigen wir wieder den Erwartungswert, $\lambda = \mathbf{E}[Z_{3n}] = 3$

$$\mathbf{P}(Z_{3n} \geq 1) = 1 - \mathbf{P}(Z_{3n} = 0) \approx 1 - e^{-3}\frac{3^0}{0!} = 1 - \frac{1}{e^3} \approx 0,950.$$

b) Sei T der Zeitpunkt, zu dem das erste Mal ein Punkt auf die Diagonale fällt, also
$T = \min\{i \in \mathbb{N} : X_i \in D\}$.
T ist geometrisch zum Parameter $p = \frac{1}{1000}$ verteilt, dabei ist p die Erfolgswahrscheinlichkeit pro Zug.
Wir ziehen 1000 mal pro Sekunde, wenn der erste Treffer also zwischen zweiter und dritter Sekunde stattfinden soll muss dieser zwischen den Würfen 2000 und 3000 passiert sein. Wir berechnen also

$$\mathbf{P}(2000 \leq T \leq 3000) = \mathbf{P}(2000 \leq T) - \mathbf{P}(T > 3000) = \mathbf{P}(T > 1999) - \mathbf{P}(T > 3000).$$

Wir haben die Schreibweise $\mathbf{P}(T > t)$ gewählt um nun die Exponentialapproximation anwenden zu können. Diese lautet (bei großem $\mathbf{E}[T]$):

$$\mathbf{P}\left(\frac{T}{\mathbf{E}[T]} > t\right) \approx e^{-t}.$$

Wir berechnen also $\mathbf{E}[T] = \frac{1}{p} = \frac{1}{\frac{1}{1000}} = 1000$ und damit dann

$$\begin{aligned}\mathbf{P}(2000 \leq T \leq 3000) &= \mathbf{P}(T > 1999) - \mathbf{P}(T > 3000)\\ &= \mathbf{P}\left(\frac{T}{1000} > 1,999\right) - \mathbf{P}\left(\frac{T}{1000} > 3\right)\\ &\approx e^{-1,999} - e^{-3} \approx 0,0857.\end{aligned}$$