## Tipps zu Hausaufgabe 2:

Mit einem fairen Würfel wird genau so lange gewürfelt, bis jede der Zahlen $1, \ldots, 6$ das erste Mal mindestens zweimal vorgekommen ist. Der Wert der Zufallsvariablen $X$ sei durch die Anzahl der Würfe bestimmt.

Berechnen Sie $\mathbb{E}[X]$!

**Lösungsvorschlag**

*Vorbemerkung*: Betrachten wir eine Folge $2, 1, 5, 5, 3, 5, \ldots$ von geworfenen Augenzahlen und die Aussage $A(5)$ für jeden Wurf $w_n$, dass nämlich die Zahl 5 mindestens 2-mal vorgekommen ist. Dann trifft diese Aussage für die Würfe $w_4, w_5$ und $w_6$ zu. Das erste Mal trifft die Aussage aber für $w_4$ zu.

Wie bei der Lösung von VA 1, Blatt 4 identifizieren wir gewisse Phasen, in denen sich der Algorithmus befindet. In dem aktuellen Fall unterscheiden wir die Phasen in Abhängigkeit davon, dass eine Augenzahl bereits 0-mal bzw. 1-mal bzw. 2-mal gewürfelt wurde.
Wir protokollieren diese Phasen mit disjunkten Mengen $g_0, g_1, g_2 \subseteq [1, 6]$, wobei $g_0$ bzw. $g_1$ bzw. $g_2$ stets die Augenzahlen aus $[1, 6]$ enthalten, die 0-mal bzw. 1-mal bzw. 2-mal vorgekommen sind.
Beim Start des Algorithmus gilt natürlich $g_0 = [1, 6]$, $g_1 = \emptyset$, $g_2 = \emptyset$. Der Algorithmus endet mit $g_0 = \emptyset$, $g_1 = \emptyset$, $g_2 = [1, 6]$.
In der Phase $(g_0, g_1, g_2)$ mit $g_2 \neq [1, 6]$ würfeln wir so lange bis eine Zahl aus $g_0$ oder $g_1$ erscheint. Der Wahrscheinlichkeitsraum für diese Phase ist $W_{g_0,g_1,g_2} = \langle \Omega_{g_0,g_1,g_2}, \Pr_{g_0,g_1,g_2} \rangle$ mit

$$\begin{aligned} \Omega_{g_0,g_1,g_2} &= \{x_1 x_2 \ldots x_{n-1} z\,;\, n \in \mathbb{N}, x_i \in g_2, z \in [1, 6] \setminus g_2\}\,, \\ \Pr_{g_0,g_1,g_2}[x_1 x_2 \ldots x_{n-1} z] &= \left(\frac{|g_2|}{6}\right)^{n-1} \cdot \frac{|g_0| + |g_1|}{6} \cdot \begin{cases} \frac{|g_0|}{|g_0|+|g_1|} & : z \in g_0 \\ \frac{|g_1|}{|g_0|+|g_1|} & : z \in g_1 \end{cases} . \end{aligned}$$

Für jede Phase $(g_0, g_1, g_2)$ sei $X_{g_0,g_1,g_2}$ die Zufallsvariable, die die Länge $n$ des Ergebnisses in dieser Phase ausgibt. Der Erwartungswert von $X_{g_0,g_1,g_2}$ ist nichts anderes als die „durchschnittliche Anzahl von Schritten“, die erforderlich sind, um in den nächsten Zustand überzugehen. Es gilt

$$\mathbb{E}[X_{g_0,g_1,g_2}] = \frac{6}{|g_0| + |g_1|}\,.$$

Die Übergänge aller Phasen können nun in einem Binärbaum dargestellt werden. Die Berechnung von $\mathbb{E}[X]$ kann durch Programm erfolgen.

***Hinweis:*** *Auf den Übungsblättern in diesem Semester wird es grundsätzlich die drei Aufgabentypen Vorbereitungsaufgabe, Tutoraufgabe und Hausaufgabe geben. Die als Vorbereitung bezeichneten Aufgaben dienen der häuslichen Vorbereitung der Tutoraufgaben. Tutoraufgaben werden in den Übungsgruppen bearbeitet. Dabei wird die Lösung der Vorbereitungsaufgaben vorausgesetzt. Die Vorbereitungsaufgaben werden in der Zentralübung unterstützt.*

## Vorbereitung 1

$X$ sei Poisson-verteilt. Berechnen Sie $\mathbb{E}[(X+1)^{-1}]$.

### Lösungsvorschlag

Nach Voraussetzung gilt $X \sim \mathrm{Po}(\lambda)$ für irgendein bestimmtes $\lambda \geq 0$. Die diskrete Dichtefunktion von $X$ ist dann $f_X : \mathbb{R} \to [0,1]$ mit $f_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}$ für $i \in \mathbb{N}_0$ und $f_X(i) = 0$ sonst. Nun gilt

$$
\begin{aligned}
\mathbb{E}[(X+1)^{-1}] &= \sum_{i \in W_X} \frac{1}{i+1} f_X(i) \\
&= \sum_{i=0}^{\infty} \frac{1}{i+1} \cdot \frac{e^{-\lambda}\lambda^i}{i!} \\
&= e^{-\lambda}\lambda^{-1} \sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{(i+1)!} \\
&= e^{-\lambda}\lambda^{-1} \left( -1 + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \right) \\
&= e^{-\lambda}\lambda^{-1}(-1 + e^{\lambda}) \\
&= \frac{1 - e^{-\lambda}}{\lambda} .
\end{aligned}
$$

*Bemerkung*: Für eine Poisson-verteilte Variable $X$ können systematisch alle Erwartungswerte für beliebige Polynome in $X$ in geschlossener Form berechnen. Beispiel: $X^{\underline{n}}$. Die vorliegende Aufgabe zeigt, dass diese Methodik sogar auf die Funktion $(X+1)^{-1}$ anwendbar ist.

## Vorbereitung 2

Eine Firma stellt Kuchen mit Rosinen her. Hierfür werden $\lambda \cdot N$ Rosinen in den Teig für $N$ Kuchen gegeben, woraufhin die ganze Masse so gut durchmischt wird, dass wir annehmen können, dass jede einzelne Rosine mit derselben Wahrscheinlichkeit in einem der Kuchen landet. $N$ ist unbekannt und groß.

Wie groß muss die durchschnittliche Zahl $\lambda$ von Rosinen pro Kuchen sein, wenn höchstens durchschnittlich jeder hundertste Kuchen keine Rosinen enthalten darf?

### Lösungsvorschlag

Aus der Problemstellung entnimmt man, dass die durchschnittliche Anzahl der Rosinen pro Kuchen mit $\lambda$ gegeben ist. Über $N$ bzw. die Anzahl $\lambda \cdot N$ der Rosinen ist nur bekannt, dass diese Zahlen sehr groß sind.

Wir betrachten das Problem aus der Sicht eines einzelnen Kuchens und stellen uns vor, dass dem Kuchen eine (potentiell unendliche) Anzahl $X$ von Rosinen zugeteilt werden, wobei die einzelnen Rosinen immer mit gleicher Wahrscheinlichkeit zugeteilt werden.

Die Verteilung der Rosinen auf den Kuchen, d. h. der Zufallsvariablen $X$, nehmen wir approximativ als Poisson-Verteilung an, d. h. $f_X : \mathbb{R} \to [0,1]$ mit $f_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}$ für $i \in \mathbb{N}_0$ und $f_X(i) = 0$ sonst. Dann erhalten wir

$$\Pr[X = 0] = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}\,.$$

Wenn durchschnittlich höchstens jeder hundertste Kuchen frei von Rosinen sein darf, dann bedeutet dies, dass $\Pr[X = 0] \leq \frac{1}{100}$ erfüllt werden muss. Es folgt

$$\lambda \geq \ln 100 \approx 4,605\,.$$

## Vorbereitung 3

Zwei Arbeiter $A$ und $B$ kontrollieren unabhängig eine Tagesproduktion. $A$ und $B$ protokollieren $k_1$ bzw. $k_2$ tatsächliche Fehler. Beide Protokolle stimmen in $k_{1,2}$ Einträgen überein. Es sei $n$ die Anzahl der tatsächlich aufgetretenen Produktionsfehler. Wir nehmen an, dass die Arbeiter jeden der $n$ Fehler mit Wahrscheinlichkeit $p_1$ bzw. $p_2$ registriert haben.

Es seien $X_1$ bzw. $X_2$ die Zufallsvariablen, die die Anzahl der von Arbeiter $A$ bzw. $B$ gefundenen Fehler angeben. Wie sind die $X_i$ verteilt? Für welche Werte von $n$ gilt

$$\Pr[|\frac{1}{n}X_i - p_i| \geq 0.1] \leq 0.01\,?$$

**Lösungsvorschlag**

Es gilt $X_i \sim \mathrm{Bin}(n; p_i)$. Daraus folgt $\mathbb{E}[X_i] = n \cdot p_i$, d. h. $\mathbb{E}[\frac{1}{n}X_i] = p_i$.
Wir wenden die Chebyshev-Ungleichung wie folgt an.

$$\Pr[|\frac{1}{n}X_i - p_i| \geq 0.1] \leq \frac{\mathrm{Var}[\frac{1}{n}X_i]}{0.01}\,.$$

Weiter gilt $\mathrm{Var}[\frac{1}{n}X_i] = \frac{1}{n^2} \cdot n \cdot p_i(1-p_i)$. Wegen $p_i \in [0,1]$ gilt $p_i(1-p_i) \leq \frac{1}{4}$ und es folgt

$$\Pr[|\frac{1}{n}X_i - p_i| \geq 0.1] \leq \frac{\mathrm{Var}[\frac{1}{n}X_i]}{0.01} \leq \frac{1}{0,04n}\,.$$

Damit haben wir die hinreichende Bedingung

$$n \geq 2500\,.$$