10. Übung aus Maß- und Wahrscheinlichkeitstheorie
WS 2015

1. Sind $X_1, \ldots, X_n$ unabhängig, $B_{\frac{1}{2}}$- verteilte Zufallsvariable, so nennt man eine Folge $(X_{i+1}(\omega), \ldots, X_{i+k}(\omega))$ einen *Run* der Länge k, wenn $i = 0$ oder $X_i(\omega) \neq X_{i+1}(\omega)$, wenn weiters $X_{i+1}(\omega) = \ldots = X_{i+k}(\omega)$ und wenn $i + k = n$ oder $X_{i+k}(\omega) \neq X_{i+k+1}(\omega)$. Man berechne den Erwartungswert und die Varianz von R, der Anzahl der *Runs* in $X_1, \ldots, X_n$. Außerdem versuche man zu beurteilen, ob die Folge 10101101001011000101100101011000101010011001010100 durch das Werfen einer Münze entstanden ist, oder, ob es sich eher um eine willkürlich hingeschriebene Folge handelt.

   Lösung: Mit $p := \frac{1}{2}\,, q := 1-p\,, Z_i := \begin{cases} 1\,, & X_i \neq X_{i-1} \\ 0\,, & \text{sonst} \end{cases}$ $\quad 2 \leq i \leq n$ gilt $P(Z_i = 1) = P(X_i = 1, X_{i-1} = 0 \vee X_i = 0, X_{i-1} = 1) = 2\,p\,(1-p) = \frac{1}{2}$, sowie $R = 1 + \sum\limits_{i-2}^{n} Z_i \;\Rightarrow\; \mathbb{E}R = 1 + \sum\limits_{i-2}^{n} \mathbb{E}Z_i = \frac{n+1}{2} = 25.5$ für $n = 50$. Für $|j - i| > 1$ sind $Z_i\,, Z_j$ klarerweise unabhängig, aber es gilt auch

$$P([Z_i = 1] \cap [Z_{i+1} = 1]) = P(\vec{X}_{i-1}^{i+1} = (1,0,1) \vee \vec{X}_{i-1}^{i+1} = (0,1,0))$$
$$= \frac{1}{4} = P(Z_i = 1)\,P(Z_{i+1} = 1) \quad \Rightarrow \quad Z_i\,, Z_{i+1} \text{ sind unabhängig}\,.$$

   Daher gilt $\operatorname{Var} R = \sum\limits_{i=2}^{n} \operatorname{Var} \mathbb{Z}_i = \frac{n-1}{4} = \frac{49}{4} \Rightarrow \sigma = \frac{7}{2} = 3.5$. Bei $R(\omega) = 36$ und $|R - \mathbb{E}R| = 10.5$ folgt aus der Tschebyscheff'schen Ungleichung $P(|R - \mathbb{E}R| \geq 10.5 = 3\,\sigma) \leq \frac{1}{9}$.

2. Ein Würfel wird solange geworfen bis zum 3-ten Mal eine 6 erscheint. Die Anzahl der notwendigen Würfe soll erraten werden. Wie würden Sie tippen wenn

   (a) nur ein richtiger Tipp honoriert wird?

   (b) ein Pönale im Ausmaß der absoluten Differenz zwischen Ihrem Tipp und dem tatsächlichen Ausgang zu bezahlen ist?

   (c) ein Pönale im Ausmaß des Quadrats zwischen Ausgang und Tipp zu zahlen ist?

   Lösung: Die Anzahl der notwendigen Würfe X ist ${}_{neg}B_{3,\frac{1}{6}}$ verteilt, d.h. $P(X = k) = \binom{k-1}{2} \left(\frac{5}{6}\right)^{k-3} \left(\frac{1}{6}\right)^3, \quad k \geq 3$.
   Bei (a) ist der Modus der sinnvollste Tipp

$$\frac{P(X = k)}{P(X = k+1)} = \binom{k-1}{2}\binom{k}{2}^{-1} \frac{6}{5} = \frac{6(k-2)}{5k} > 1 \Leftrightarrow k > 12.$$

Daher ist $P(X=k)$ für $k=12,13$ maximal, d.h. Modus = 12 oder13.
Bei (b) ist der Median der sinnvollste Tipp.

$$F_X(15) \approx 0.46 < frac12 \leq F_X(16) \approx 0.51 \Rightarrow med = 16.$$

Bei (c) ist der Erwartungswert $\mathbb{E}\, X = 18$ der sinnvollste Tipp.

3. Man zeige, dass für unabhängige Zufallsvariable $X_n \sim Ex_{\tau_n}$ gilt

$$\sum_{n=1}^{\infty} \frac{1}{\tau_n} = \infty \;\Rightarrow\; S := \sum_{n=1}^{\infty} X_n = \infty \quad P\text{–fs.}$$

Lösung 1: Es gilt $\mathbb{E} S_n = N_n := \sum\limits_{i=1}^{n} \frac{1}{\tau_i}$, $\operatorname{Var} S_n = \sigma_n^2 := \sum\limits_{i=1}^{n} \frac{1}{\tau_i^2}$.
Gibt es eine Teilfolge (n_k) mit $\frac{1}{\tau_{n_k}} \geq 1 \quad \forall\ k \in \mathbb{N}$, so gilt klarerweise $P\left(X_{n_k} \geq \frac{1}{\tau_{n_k}}\right) = \mathrm{e}^{-1} \quad \Rightarrow \quad \sum\limits_{k} P\left(X_{n_k} \geq \frac{1}{\tau_{n_k}}\right) = \infty$. Da die X_{n_k} unabhängig sind, folgt aus dem 2-ten Lemma von Borel-Cantelli, dass mit Wahrscheinlichkeit 1 für unendlich viele k gilt $X_{n_k} \geq \frac{1}{\tau_{n_k}} \geq 1$ Das impliziert offensichtlich $P(S=\infty) = 1$.
Gibt es hingegen ein $N : \frac{1}{\tau_i} < 1 \quad \forall\ i > N$, so kann man die ersten N Summanden vernachlässigen und o.E.d.A. annehmen $N=0$. Dann gilt $\frac{1}{\tau_i^2} \leq \frac{1}{\tau_i} \quad \forall\ i \in \mathbb{N} \quad \Rightarrow \quad \sigma_n^2 \leq N_n$. Aus der Tschebyscheff'schen Ungleichung folgt für alle $M \in \mathbb{N}$ und alle n mit $N_n \geq 2\,M$

$$P\,(S \leq M) \leq P\left(S_n \leq \frac{N_n}{2}\right) \leq P\left(|S_n - N_n| \geq \frac{N_n}{2}\right) \leq \frac{4\,\sigma_n^2}{N_n^2} \leq \frac{4}{N_n}\,.$$

Daher gilt $P(S \leq M) = 0 \quad \forall\ M \in \mathbb{N}$, und daraus folgt offensichtlich $P(S<\infty) = P\Big(\bigcup\limits_{M \in \mathbb{N}} [S \leq M]\Big) \leq \sum\limits_{M \in \mathbb{N}} P(S \leq M) = 0$.

Lösung 2: Aus $\left[S := \sum\limits_{i=1}^{\infty} X_i \leq c\right] \subseteq [S_n \leq c] = \left[\mathrm{e}^{-S_n} \geq \mathrm{e}^{-c}\right]$, der Markoff-Ungleichung und der Unabhängigkeit der X_i folgt

$$P(S \leq c) \leq \mathrm{e}^c\,\mathbb{E}\mathrm{e}^{-S_n} = \mathrm{e}^c \prod_{i=1}^{n} \mathbb{E}\mathrm{e}^{-X_i} = \mathrm{e}^c \prod_{i=1}^{n} \frac{\tau_i}{\tau_i + 1} = \mathrm{e}^c \prod_{i=1}^{n} \frac{1}{1+\tau_i^{-1}}\,.$$

Gibt es eine Teilfolge (n_k) : $\tau_{n_k}^{-1} \geq 1$, so gilt $\prod\limits_{i=1}^{n} \frac{1}{1+\tau_i^{-1}} \leq \prod\limits_{n_k \leq n} \frac{1}{2} \to 0$.
Sonst $\exists\ N : \ \forall\ i \geq N : \ 0 \leq \tau_i^{-1} < 1 \quad \Rightarrow \quad \frac{1}{1+\tau_i^{-1}} \leq 1 - \frac{1}{2\,\tau_i}$. Auch dann gilt $\prod\limits_{i=1}^{n} \frac{1}{1+\tau_i^{-1}} \leq \prod\limits_{i=N}^{n} \left(1 - \frac{1}{2\,\tau_i}\right) = \mathrm{e}^{\sum\limits_{i=N}^{n} \ln\left(1-\frac{1}{2\,\tau_i}\right)} \leq \mathrm{e}^{-\frac{1}{2} \sum\limits_{i=N}^{n} \frac{1}{\tau_i}} \to 0$.

Lösung 3: Aus $1 \geq \mathrm{e}^{-S_n} \searrow \mathrm{e}^{-S} \geq 0$ und dem Satz von Lebesgue folgt $\mathbb{E}\mathrm{e}^{-S} = \lim_n \mathbb{E}\mathrm{e}^{-S_n} = 0$ (wie in Lösung 2. gezeigt). Daher gilt $\mathrm{e}^{-S} = 0$ P–fs $\Rightarrow$ $S = \infty$ P–fs.

4. Ein Spieler soll auf die möglichen Werte $x_1, \ldots, x_m$ einer diskreten Zufallsvariablen X, die mit den Wahrscheinlichkeiten $(p_1, \ldots, p_m)$ angenommen werden, wetten. Dabei beginnt er mit einem Startkapital $S_0 > 0$, und er kann Anteile $0 \leq b_i \leq 1\,, 1 \leq i \leq m\,,\ \sum_{i=1}^{m} b_i = 1$ dieses Kapitals auf die einzelnen Werte setzen. Für den richtigen Tipp bekommt er das m-fache seines Einsatzes ausbezahlt, die Einsätze, die er auf andere Ausgänge gesetzt hat, gehen verloren. Bezeichnet X_1 den Ausgang des 1-ten Spiels, so verfügt er nach diesem Spiel über ein Kapital $S_1 = m\, b(X_1)\, S_0$, das er wieder im Verhältnis $b_1 : b_2 : \ldots : b_m$ aufteilen kann. Dieser Vorgang wird n-mal wiederholt.

   (a) Wie groß ist das Kapital S_n des Spielers nach n Runden?

   (b) Man bestimme $\lim_n \frac{1}{n} \log S_n$.

   (c) Man bestimme $\lim_n \frac{1}{n} \log S_n$, wenn der Spieler seine Anteile gemäß den Wahrscheinlichkeiten wählt, also $b_i = p_i\,,\ \forall\, 1 \leq i \leq m$.

   (d) Welche Aufteilung ist optimal? Vergleichen Sie die Punkte (b) und (c) in Hinblick auf das Ergebnis von Bsp. 6. aus Übung 7.

   (e) Wie sollte man aufteilen, um $\mathbb{E}S_1$ zu maximieren?

Lösung:

   (a) vollständige Induktion ergibt $S_n = m^n\, S_0 \prod_{j=1}^{n} b(X_j)\,.$

   (b) Aus Punkt (a) folgt $\frac{1}{n} \log S_n = \log m + \frac{\log S_0}{n} + \frac{1}{n} \sum_{j=1}^{n} \log b(X_j)\,.$

   Daher gilt $\mathbb{E}\frac{1}{n} \log S_n = \log m + \frac{\log S_0}{n} + \frac{1}{n} \sum_{j=1}^{n} \mathbb{E} \log b(X_j)$, und aus dem Gesetz der großen Zahlen folgt

$$\lim_n \frac{1}{n} \log S_n = \log m + \mathbb{E}(\log b(X_1)) = \log m + \sum_{i=1}^{m} p_i \, \log b_i\,. \quad (1)$$

   (c) Gemäß Punkt (b) gilt natürlich nun

$$\lim_n \frac{1}{n} \log S_n = \log m + \mathbb{E}(\log p(X_1)) = \log m + \sum_{i=1}^{m} p_i \, \log p_i\,. \quad (2)$$

(d) Nach Bsp. 6. aus Übung 7. ist der Limes in (2) größer als in (1). Daher ist es am sinnvollsten das Kapital proportional zu den Wahrscheinlichkeiten aufzuteilen.

(e) Ist $p_{i_0} = \max_{1\le i\le m} p_i$ und $c_{i_0} := 1, c_i := 0 \quad \forall i \ne i_0$, so gilt

$$\mathbb{E}\frac{S_1}{m} = \sum_{i=1} b_i\, p_i \le \sum_{i=1} b_i\, p_{i_0} = p_{i_0} = \sum_{i=1} c_i\, p_i.$$

Es ist also sinnvoll zu hasardieren.

5. Gilt ein Gesetz der großen Zahlen für die unabhängigen Zufallsvariablen X_n mit $X_1 \equiv 0$ und $P(X_n = n) = P(X_n = -n) = \dfrac{1}{2n\log n}$ sowie $P(X_n = 0) = 1 - \frac{1}{n\log n}$ für $n \ge 2$?

Lösung: $P(|\overline{X}_{n-1}| > \epsilon) \le \dfrac{\operatorname{Var}\left(\sum_{i=2}^{n} X_i\right)}{\epsilon^2 n^2} = \dfrac{\sum_{i=2}^{n} \frac{i}{\log i}}{n^2\,\epsilon^2} \le \dfrac{\frac{n^2}{\log n}}{n^2\,\epsilon^2} = \frac{1}{\epsilon^2 \log n} \to 0.$ Aber aus $\sum_{i=2}^{\infty} P(|X_i| \ge i) = \sum_{i=2}^{\infty} \frac{1}{i\log i} \ge \int_2^\infty \frac{dx}{x\log x} = \log\log x\Big|_2^\infty = \infty$ und der Unabhängigkeit der X_i folgt nach dem 2-ten Lemma von Borel-Cantelli $P\left(\limsup_i [|X_i| \ge i]\right) = 1$, d.h. $|X_i| \ge i$ für unendlich viele i. Aus $\lim_n \overline{X}_n = c \in \mathbb{R}$ würde jedoch im Widerspruch dazu folgen $\left|\frac{X_n}{n}\right| = \left|\frac{n-1}{n}\,\overline{X}_{n-1} - \overline{X}_n\right| \le \left|\frac{n-1}{n}\,\overline{X}_{n-1} - \overline{X}_{n-1}\right| + \left|\overline{X}_{n-1} - \overline{X}_n\right| \to 0.$

6. Man zeige, dass für eine Folge von unabhängigen Ereignissen A_n mit $P(A_n) = \frac{1}{n}$ die Summen $S_n := \sum_{i=2}^{n} \frac{\mathbb{1}_{A_i} - \frac{1}{i}}{\log i}$ P–fs konvergieren. Damit beweise man $\frac{1}{\log n} \sum_{i=1}^{n} \left(\mathbb{1}_{A_i} - \frac{1}{i}\right) \to 0$ P–fs und zeige schließlich $\frac{1}{\log n} \sum_{i=1}^{n} \mathbb{1}_{A_i} \to 1$ P–fs.

Beweis: Für $X_i := \frac{\mathbb{1}_{A_i} - \frac{1}{i}}{\log i}$, $i \ge 2$ gilt $\operatorname{Var} X_i = \frac{\frac{1}{i}\left(1-\frac{1}{i}\right)}{\log^2 i} \le \frac{1}{i\,\log^2 i}$. Wegen $\sum_{i=3}^{n} \frac{1}{i\,\log^2 i} \le \int_2^n \frac{1}{x\,\log^2 x}\,dx = -\frac{1}{\log x}\Big|_3^n \le \frac{1}{\log 3} < \infty$ gilt $\sum_i \operatorname{Var} X_i < \infty$. Daraus folgt $S_n \to S \in \mathbb{R}$ P–fs (siehe VO). Aus dem Kronecker-Lemma mit $b_i := \log i$, $a_i := \frac{\mathbb{1}_{A_i} - \frac{1}{i}}{\log i}$ folgt nun $\frac{1}{\log n} \sum_{i=1}^{n} \left(\mathbb{1}_{A_i} - \frac{1}{i}\right) \to 0$ P–fs. Wegen $\sum_{i=2}^{n} \frac{1}{i} \le \int_1^n \frac{1}{x}\,dx = \log n$ und $\sum_{i=1}^{n} \frac{1}{i} \ge \int_1^{n+1} \frac{1}{x}\,dx = \log(n+1)$ gilt auch $\lim_n \frac{1}{\log n} \sum_{i=1}^{n} \frac{1}{i} = 1$. Zusammen mit $\lim_n \frac{1}{\log n} \sum_{i=1}^{n} \left(\mathbb{1}_{A_i} - \frac{1}{i}\right) = 0$ ergibt das schließlich $\lim_n \frac{1}{\log n} \sum_{i=1}^{n} \mathbb{1}_{A_i} = 1$ P–fs.

7. Man berechne für eine Folge unabhängig, identisch verteilter Zufallsvariabler $X_n \in \mathcal{L}_2(\Omega, \mathfrak{S}, P)$ den Erwartungswert der Stichprobenvarianz $S_n^2 := \frac{1}{n-1} \sum\limits_{i=1}^{n} \left(X_i - \frac{1}{n} \sum\limits_{i=1}^{n} X_i \right)^2$ und zeige weiters, dass gilt $\lim\limits_n S_n^2 = \operatorname{Var} X_1 \quad P$–fs.

   Lösung: Sei $\eta := \mathbb{E} X_n$, $\sigma^2 := \operatorname{Var} X_n$

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \overline{X}_n^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - 2n\overline{X}_n^2 + n\overline{X}_n^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2 \right). \quad (3) \end{aligned}$$

   Aus $\mathbb{E} X_i^2 = \sigma^2 + \eta^2, \mathbb{E} \overline{X}_n^2 = \operatorname{Var} \overline{X}_n + \left(\mathbb{E} \overline{X}_n \right)^2 = \frac{\sigma^2}{n} + \eta^2$ und (3) folgt

$$\mathbb{E} S_n^2 = \frac{1}{n-1} \left(n\,\sigma^2 + n\,\eta^2 - \sigma^2 - n\,\eta^2 \right) = \sigma^2 .$$

   Aus dem GGZ folgt $\frac{1}{n-1} \sum\limits_{i=1}^{n} X_i^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum\limits_{i=1}^{n} X_i^2 \right) \to \mathbb{E} X_1^2 = \sigma^2 + \eta^2$, sowie $\overline{X}_n = \frac{1}{n} \sum\limits_{i=1}^{n} X_i \to \eta \quad P$–fs. $\Rightarrow \quad \overline{X}_n^2 \to \eta^2 \quad P$–fs. Diese Beziehungen zusammen mit (3) ergeben $S_n^2 \to \sigma^2 \quad P$-fs.