Prof. Vincent Tassion
Laurin Köhler-Schindler

# Wahrscheinlichkeit & Statistik

## Serie 8

**Abgabe bis Mittwoch (27.04.2022) um 10:15 Uhr**

Diese Serie beschäftigt sich mit Anwendungen des Gesetz der grossen Zahlen und des zentralen Grenzwertsatz.

Weitere Informationen und Instruktionen zur Abgabe unter
*https://metaphor.ethz.ch/x/2022/fs/401-0614-00L/*

---

**Aufgabe 8.1 [Normalverteilung]**
Sei $n \geq 1$ und $(X_i)_{1\leq i\leq n}$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit $X_i \sim \mathcal{N}(1,4)$ für jedes $1 \leq i \leq n$. Wir definieren die folgenden Zufallsvariablen:

$$S_n = X_1 + X_2 + \cdots + X_n$$

und

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{S_n}{n}.$$

Zur Erinnerung: Wir schreiben $\Phi(a) = \mathbb{P}[Z \leq a]$ für $Z \sim \mathcal{N}(0,1)$. Numerische Werte für $\Phi$ finden sich beispielsweise unter `https://de.wikipedia.org/wiki/Standardnormalverteilungstabelle`.

(a) Bestimme die Verteilung von $S_n$ sowie $\bar{X}_n$.
*Hinweis: Nutze die Eigenschaften der Normalverteilung aus Kapitel 3.6.*

(b) Berechne die Wahrscheinlichkeit $\mathbb{P}[\mathbb{E}[X_1] - 1 \leq X_1 \leq \mathbb{E}[X_1] + 1]$.

(c) Berechne $\mathbb{P}[\mathbb{E}[S_n] - 1 \leq S_n \leq \mathbb{E}[S_n] + 1]$ für $n = 50$.

(d) Berechne $\mathbb{P}[\mathbb{E}[\bar{X}_n] - 1 \leq \bar{X}_n \leq \mathbb{E}[\bar{X}_n] + 1]$ für $n = 50$.

**Aufgabe 8.2 [Gesetz der grossen Zahlen I: Empirische Verteilungsfunktion]**
Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Verteilungsfunktion $F$. Die empirische Verteilungsfunktion $F_n : \mathbb{R} \times \Omega \to [0,1]$ ist gegeben durch

$$F_n(t,\omega) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i(\omega)\leq t}.$$

Der Wert $F_n(t,\omega)$ beschreibt also die relative Häufigkeit der $X_i(\omega)$ mit Werten $\leq t$ unter den ersten $n$. Damit ist $F_n(t) := F_n(t,\cdot)$ selbst eine Zufallsvariable für jedes $t \in \mathbb{R}$.

(a) Sei $t \in \mathbb{R}$ beliebig und $Y_i := \mathbb{1}_{X_i\leq t}$ für $i \in \mathbb{N}$. Zeige, dass $Y_1, Y_2, \ldots$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen ist. Was ist $\mathbb{E}[Y_1]$?

(b) Zeige, dass für jedes $t \in \mathbb{R}$ die empirische Verteilungsfunktion $F_n(t)$ fast sicher gegen $F(t)$ konvergiert für $n \to \infty$.

**Aufgabe 8.3 [Gesetz der grossen Zahlen II: Approximation von $\pi$ mit Münzwürfen]**
In der Vorlesung wurde gezeigt, wie man die Zahl $\pi$ mit Hilfe von $\mathcal{U}([0,1])$-verteilten Zufallsvariablen approximieren kann. Ziel dieser Aufgabe ist es, dies nun mit Hilfe einer Folge $(X_i)_{i\geq 1}$ unabhängiger, Ber(1/2)-verteilter Zufallsvariablen zu tun.

(a) Sei $n \geq 1$. Für $i, j \in \{0, 1, \ldots, 2^n - 1\}$ definieren wir Quadrate mit Seitenlänge $2^{-n}$ durch

$$S_{i,j} := [i \cdot 2^{-n}, (i+1) \cdot 2^{-n}] \times [j \cdot 2^{-n}, (j+1) \cdot 2^{-n}].$$

Zeige, dass $2n$ Ber(1/2)-verteilte Zufallsvariablen ausreichen, um zufällig eines der Quadrate $S_{i,j}$ auszuwählen (sodass jedes Quadrat mit gleicher Wahrscheinlichkeit gewählt wird).

(b) Wir wollen einen Algorithmus finden, der die (zufälligen) Bits $X_1, X_2, \ldots$ als Input nimmt, und als Output eine Ber($\pi/4$)-verteilte Zufallsvariable ausgibt. Gebe einen solchen Algorithmus an, sodass

$$\mathbb{P}[T > 2n] \leq \frac{1}{2^{n-1}},$$

wobei $T$ die Anzahl der benötigten Input Bits bezeichnet.

*Hinweis: Ein Punkt $(x, y) \in [0,1]^2$ liegt innerhalb des Einheitskreises genau dann, wenn $x^2 + y^2 \leq 1$.*

(c) Wie können wir den Algorithmus aus Teilaufgabe $(b)$ nutzen, um die Zahl $\pi$ zu approximieren?

**Aufgabe 8.4 [Zentraler Grenzwertsatz I: Zufällige Irrfahrt]**
Seien $(X_i)_{i\geq 1}$, $(Y_i)_{i\geq 1}$ und $(Z_i)_{i\geq 1}$ Folgen unabhängiger, identisch verteilter Zufallsvariablen mit

$$\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = -1] = 1/2$$

und analog $\mathbb{P}[Y_1 = 1] = \mathbb{P}[Y_1 = -1] = 1/2$ sowie $\mathbb{P}[Z_1 = 1] = \mathbb{P}[Z_1 = -1] = 1/2$. Wir definieren

$$S_n^{(x)} := \sum_{i=1}^n X_i, \quad S_n^{(y)} := \sum_{i=1}^n Y_i, \quad \text{and} \quad S_n^{(z)} := \sum_{i=1}^n Z_i.$$

Die Folge $\big((S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\big)_{n\geq 1}$ wird zufällige Irrfahrt in $\mathbb{Z}^3$ genannt. Sei $\alpha > 1/2$. Zeige, dass

$$\mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\|_2 \leq n^\alpha\right] \to 1 \text{ as } n \to \infty,$$

wobei $\|(x, y, z)\|_2 := \sqrt{x^2 + y^2 + z^2}$ die euklidische Norm ist.
*Hinweis: Betrachte zuerst die Folge $(S_n^{(x)})_{n\geq 1}$ und wende den zentralen Grenzwertsatz an.*

**Aufgabe 8.5 [Zentraler Grenzwertsatz II]**
In dieser Aufgabe berechnen wir den Grenzwert

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!}.$$

(a) Sei $X \sim$ Poisson($\lambda$) für ein $\lambda > 0$. Berechne $\mathbb{E}[X]$ und $\sigma_X^2$.

(b) Zeige mit dem zentralen Grenzwertsatz, dass der obige Grenzwert 1/2 ist.

*Hinweis: Für zwei unabhängige Zufallsvariablen $X, Y$ mit $X \sim$ Poisson($\lambda$), $\lambda > 0$, und $Y \sim$ Poisson($\mu$), $\mu > 0$, ist $X + Y \sim$ Poisson($\lambda + \mu$).*

**Lösung 8.1**

(a) Die Zufallsvariablen $(X_i)_{1\leq i\leq n}$ sind unabhängig und normalverteilt mit $m_i = \mathbb{E}[X_i] = 1$ und $\sigma_i^2 = \mathbb{E}[(X_i - m_i)^2] = 4$ für alle $1 \leq i \leq n$. Aus den Eigenschaften der Normalverteilung (siehe Kapitel 3.6) folgt, dass $S_n = X_1 + \ldots + X_n$ normalverteilt ist mit Erwartungswert $m_S = \sum_{i=1}^n m_i = n$ und Varianz $\sigma_S^2 = \sum_{i=1}^n \sigma_i^2 = 4n$.

Aus den Eigenschaften der Normalverteilung folgt ebenfalls, dass $\bar{X}_n = S_n/n$ normalverteilt ist mit Erwartungswert $m_{\bar{X}} = m_S/n = 1$ und Varianz $\sigma_{\bar{X}}^2 = (1/n)^2\sigma_S^2 = 4/n$.

Wir verwenden nun die folgende Eigenschaft der Normalverteilung: Sei $Y$ eine $\mathcal{N}(m, \sigma^2)$-verteilte Zufallsvariable. Dann ist

$$Z = \frac{Y - m}{\sqrt{\sigma^2}}$$

eine $\mathcal{N}(0, 1)$-verteilte Zufallsvariable.

(b) $X_1$ ist normalverteilt mit $\mathcal{N}(\mu_{X_1} = 1, \sigma_{X_1}^2 = 4)$. Durch Standardisieren erhalten wir $Z \sim \mathcal{N}(0, 1)$ wobei $Z = \frac{X_1-1}{2}$. Wir erhalten dann

$$\begin{aligned}
\mathbb{P}[E[X_1] - 1 \leq X_1 \leq \mathbb{E}[X_1] + 1] &= \mathbb{P}[0 \leq X_1 \leq 2] = \mathbb{P}\Big[-\frac{1}{2} \leq Z \leq \frac{1}{2}\Big] \\
&= \Phi\Big[\frac{1}{2}\Big] - \Phi\Big[-\frac{1}{2}\Big] = \Phi\Big[\frac{1}{2}\Big] - \Big[1 - \Phi\Big[\frac{1}{2}\Big]\Big] \\
&= \Phi\Big[\frac{1}{2}\Big] + \Phi\Big[\frac{1}{2}\Big] - 1 = 2 \cdot \Phi\Big[\frac{1}{2}\Big] - 1 \\
&= 2 \cdot 0.69146 - 1 = 0.38292,
\end{aligned}$$

wobei wir den numerischen Wert von $\Phi(1/2)$ aus der Tabelle (`https://de.wikipedia.org/wiki/Standardnormalverteilungstabelle`) abgelesen haben.

(c) Für $n = 50$ erhalten wir, dass $S_{50}$ den Erwartungswert $m_S = 50$ und die Varianz $\sigma_S^2 = 200$ hat. Durch Standardisieren ist $Z = \frac{S_{50}-50}{\sqrt{200}}$ eine $\mathcal{N}(0, 1)$-verteilte Zufallsvariable. Es folgt:

$$P[49 \leq S_n \leq 51] = P[-0.07 \leq Z \leq 0.07] = 2 \cdot \Phi[0.07] - 1 = 2 \cdot 0.5279 - 1 = 0.0558.$$

(d) Für $n = 50$ erhalten wir, dass $\bar{X}_{50}$ den Erwartungswert $m_{\bar{X}} = 1$ und die Varianz $\sigma_{\bar{X}}^2 = 0.08$ hat. Durch Standardisieren ist $Z = \frac{\bar{X}_{50}-1}{\sqrt{0.08}}$ eine $\mathcal{N}(0, 1)$-verteilte Zufallsvariable. Es folgt:

$$P[0 \leq \bar{X}_{50} \leq 2] = P[-3.5 \leq Z \leq 3.5] = 2 \cdot \Phi[3.5] - 1 = 2 \cdot 0.99977 - 1 = 0.99954.$$

**Lösung 8.2**

(a) Für jedes $i \in \mathbb{N}$ kann die Zufallsvariable $Y_i$ nur die Werte 1 und 0 annehmen. Insbesondere ist

$$\mathbb{P}[Y_i = 0] = \mathbb{P}[X_i > t] = \mathbb{P}[X_1 > t]$$

und

$$\mathbb{P}[Y_i = 1] = \mathbb{P}[X_i \leq t] = \mathbb{P}[X_1 \leq t],$$

da $X_1, X_2, \ldots$ identisch verteilt sind. Also ist auch die Folge $Y_1, Y_2, \ldots$ identisch verteilt und es gilt

$$\mathbb{E}[Y_1] = 0 \cdot \mathbb{P}[Y_i = 0] + 1 \cdot \mathbb{P}[Y_i = 1] = \mathbb{P}[X_1 \leq t] = F(t),$$

wobei $F$ die Verteilungsfunktion von $X_1$ bezeichnet. Die Unabhängigkeit folgt aus der Grouping Eigenschaft (Proposition 2.6) oder kann direkt mithilfe der Definition 2.5 gezeigt werden.

(b) Sei $t \in \mathbb{R}$ beliebig und seien $Y_1, Y_2, \ldots$ wie in der vorherigen Teilaufgabe definiert. Wir wissen, dass $Y_1, Y_2, \ldots$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen ist. Mit dem Gesetz der grossen Zahlen (Theorem 6.1) folgt, dass

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le t} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{n\to\infty} \mathbb{E}[Y_1] = F(t) \quad \text{fast sicher.}$$

**Lösung 8.3**

(a) Eine Zahl $x \in \{0, 1, \ldots, 2^n - 1\}$ entspricht im Binärsystem einer Folge $x_0 \cdots x_{n-1}$ der Länge $n$. Sei $(X_k)_{k=1}^{2n}$ Folge unabhängiger Ber(1/2)-verteilter Zufallsvariablen. Entsprechend sind $(X_{2k-1})_{k=1}^{n}$ und $(X_{2k})_{k=1}^{n}$ zwei unabhängige Folgen unabhängiger Ber(1/2)-verteilter Zufallsvariablen. Wir definieren

$$I = \sum_{k=1}^{n} X_{2k-1} \cdot 2^{k-1} \quad \text{und} \quad J = \sum_{k=1}^{n} X_{2k} \cdot 2^{k-1}.$$

Dies sind zwei unabhängige Zufallsvariablen mit Werten in $\{0, 1, \ldots, 2^n - 1\}$ und sie sind gleichverteilt, d.h.

$$\mathbb{P}[I = i] = \mathbb{P}[J = j] = 2^{-n} \quad \forall i, j \in \{0, 1, \ldots, 2^n - 1\}.$$

Entsprechend ist das Quadrat $S_{I,J} = S_{I,J}^{(n)}$ ebenfalls gleichverteilt.

(b) Der folgende Algorithmus gibt eine Ber($\pi/4$)-verteilte Zufallsvariable $Z$ aus:

$k = 1$
$I = X_1$
$J = X_2$
**While** $I^2 + J^2 \le 4^k \le (I+1)^2 + (J+1)^2$ **do**
  $k = k + 1$
  $I = 2I + X_{2k-1}$
  $J = 2J + X_{2k}$
**endwhile**
**If** $I^2 + J^2 > 4^k$
  $Z = 0$
**Else**
  $Z = 1$
**Return** $Z$

Idee: Wir wollen einen zufälligen Punkt in $[0, 1]^2$ auswählen und prüfen, ob dieser innerhalb des Einheitskreises liegt oder nicht. Die Auswahl eines zufälligen Punkts erfolgt iterativ wie folgt: Wir wählen zufällig ein Quadrat $S_{i,j}^{(n)}$ mit Seitenlänge $2^{-n}$ wie in Aufgabenteil (a) beschrieben. $S_{i,j}$ liegt vollständig ausserhalb des Einheitskreises, falls

$$(i \cdot 2^{-n})^2 + (j \cdot 2^{-n})^2 > 1$$

und vollständig innerhalb des Einheitskreises, falls

$$((i+1) \cdot 2^{-n})^2 + ((j+1) \cdot 2^{-n})^2 < 1.$$

In diesen Fällen wissen wir also, dass alle Punkte im Quadrat ausserhalb resp. innerhalb des Einheitskreises liegen. Andernfalls teilen wir $S_{i,j}^{(n)}$ in 4 gleich grosse Quadrate auf und wählen zufällig eines der 4 Quadrate aus. Dies entspricht der zufälligen Wahl eines Quadrats $S_{j',j'}^{(n+1)}$ und wir iterieren diesen Prozess, bis das gewählte Quadrat vollständig innerhalb oder ausserhalb des Einheitskreises liegt.

Somit gibt der Algorithmus $Z = 1$ genau dann aus, wenn der zufällige Punkt in $[0, 1]^2$ innerhalb des Einheitskreises liegt. Da die Auswahl der Quadrate (und somit des Punkts) gleichverteilt erfolgt und der Einheitskreis die Fläche $\pi$ hat, gilt $\mathbb{P}[Z = 1] = \pi/4$.

Wir stellen fest, dass $T > 2n$, falls das Quadrat $S_{I,J}^{(n)}$ weder vollständig innerhalb noch vollständig ausserhalb des Einheitskreises liegt. Man kann leicht prüfen, dass höchstens $2 \cdot 2^n$ Quadrate $S_{i,j}^{(n)}$ mit $0 \le i, j \le 2^n - 1$ den Einheitskreis schneiden. Da es insgesamt $2^n \cdot 2^n$ Quadrate gibt und die Wahl gelichverteilt erfolgt, gilt

$$\mathbb{P}[T > 2n] \le \frac{2 \cdot 2^n}{2^n \cdot 2^n} = \frac{1}{2^{n-1}}.$$

(c) Wir können den Algorithmus aus Aufgabenteil (b) mehrfach hintereinander ausführen. Wir bezeichnen den Output bei der $\ell$-ten Ausführung mit $Z_\ell$. Aufgrund des Gesetzes der grossen Zahlen gilt

$$\frac{Z_1 + \ldots + Z_\ell}{\ell} \xrightarrow{\ell\to\infty} \mathbb{E}[Z_1] = \pi/4 \quad \text{fast sicher.}$$

Somit ist $\frac{4}{\ell}(Z_1 + \ldots + Z_\ell)$ eine Approximation der Zahl $\pi$, die für $\ell \to \infty$ gegen $\pi$ konvergiert.

**Lösung 8.4**

*Schritt 1:* Für alle $\alpha > 1/2$, $\mathbb{P}\left[|S_n^{(x)}| \le n^\alpha\right] \to 1$ as $n \to \infty$.

Wir verwenden den zentralen Grenzwertsatz. Da $\mathbb{E}[X_1] = 0$ und $\sigma_{X_1}^2 = 1$, erhalten wir für beliebiges $a \in \mathbb{R}$,

$$\mathbb{P}\left[S_n^{(x)} \le a\sqrt{n}\right] = \mathbb{P}\left[\frac{S_n^{(x)}}{\sqrt{n}} \le a\right] \xrightarrow{n\to\infty} \Phi(a)$$

und somit auch

$$\mathbb{P}\left[|S_n^{(x)}| \le a\sqrt{n}\right] = \mathbb{P}\left[S_n^{(x)} \le a\sqrt{n}\right] - \mathbb{P}\left[S_n^{(x)} \le -a\sqrt{n}\right] \xrightarrow{n\to\infty} \Phi(a) - \Phi(-a) = 2\Phi(a) - 1.$$

Da $a\sqrt{n} \le n^\alpha$ für fixes $a$ und grosses $n$, folgt aus der Monotonizität, dass

$$\lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \le n^\alpha\right] \ge \lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \le a\sqrt{n}\right] = 2\Phi(a) - 1.$$

Da $\Phi(a) \to 1$ für $a \to \infty$, erhalten wir das gewünschte Resultat. Analog folgt dies auch für $S_n^{(y)}$ und $S_n^{(z)}$.

*Schritt 2:* Für alle $\alpha > 1/2$, $\mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\|_2 \le n^\alpha\right] \to 1$ as $n \to \infty$.

Wir wählen $\alpha' \in (1/2, \alpha)$ und stellen fest, dass

$$\{|S_n^{(x)}| \le n^{\alpha'}\} \cap \{|S_n^{(y)}| \le n^{\alpha'}\} \cap \{|S_n^{(z)}| \le n^{\alpha'}\} \subseteq \left\{\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\|_2 \le \sqrt{3} \cdot n^{\alpha'}\right\}.$$

Da $n^\alpha \ge \sqrt{3} \cdot n^{\alpha'}$ für grosse $n$, folgt somit

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\|_2 \le n^\alpha\right] &\ge \lim_{n\to\infty} \mathbb{P}\left[\|(S_n^{(x)}, S_n^{(y)}, S_n^{(z)})\|_2 \le \sqrt{3} \cdot n^{\alpha'}\right] \\ &\ge \lim_{n\to\infty} \mathbb{P}\left[|S_n^{(x)}| \le n^{\alpha'}, |S_n^{(y)}| \le n^{\alpha'}, |S_n^{(z)}| \le n^{\alpha'}\right] = 1,\end{aligned}$$

wobei die Gleichheit im letzten Schritt aus dem Union Bound folgt (vergleiche Aufgabe 2.1), da für $n \to \infty$

$$\mathbb{P}\left[|S_n^{(x)}| \leq n^{\alpha'}\right], \mathbb{P}\left[|S_n^{(y)}| \leq n^{\alpha'}\right], \mathbb{P}\left[|S_n^{(z)}| \leq n^{\alpha'}\right] \to 1.$$

**Lösung 8.5**

(a) Nach Definition des Erwartungswertes für diskrete Zufallsvariablen ist

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} = \lambda.$$

Analog erhält man

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.$$

Also ist

$$\sigma_X^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \lambda.$$

(b) Seien $X_1, X_2, \dots$ unabhängige und Poisson(1)-verteilte Zufallsvariablen. Aus der Teilaufgabe (a) ist bekannt, dass $\mathbb{E}[X_1] = \sigma_X^2 = 1$. Ausserdem ist $S_n = X_1 + \cdots + X_n$ wieder Poisson-verteilt mit Parameter $n$ (siehe Hinweis). Mit dem zentralen Grenzwertsatz folgt

$$\begin{aligned} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \sum_{k=0}^{n} e^{-n} \frac{n^k}{k!} &= \sum_{k=0}^{n} \mathbb{P}[S_n = k] = \mathbb{P}[S_n \leq n] \\ &= \mathbb{P}\left[\frac{S_n - n}{\sqrt{n}} \leq 0\right] \xrightarrow{n \to \infty} \Phi(0) = \frac{1}{2}. \end{aligned}$$