Technische Universität Berlin

Skript zur Vorlesung, Übung und Tutorium

Wahrscheinlichkeitstheorie I

gelesen von Prof. Dr. Wilhelm Stannat im Sommersemester 2019

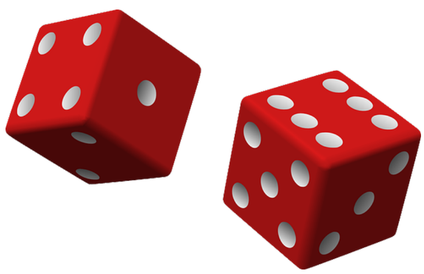


Abb. 1: Würfel. [Quelle:Wiki]

Inhaltsverzeichnis

Mitschrift von Viktor Glombik.

Zuletzt geändert am 9. Oktober 2020.

1 Mathematische Modellierung von Zufallsexperimenten

1.1 Axiomatischer Aufbau der Wahrscheinlichkeitstheorie

11.04.2019

Unter einem Zufallsexperiment versteht man einen zeitlich und örtlich festgelegten Vorgang mit unbestimmtem Ausgang. Beispiele hierfür sind

Zufallsexperiment

- Werfen eines Würfels oder einer Münze,
- zufälliges Ziehen von Kugeln aus einer Urne,
- Kartenspiele,
- Wahlergebnis der nächsten Europawahl,
- Temperatur auf dem Alexanderplatz am 11. April 2019, 12:00,
- Lebensdauern.

DEFINITION 1.1.1 (ERGEBNIS, ERGEBNISRAUM)

Die Gesamtheit aller möglichen Ausgänge eines Zufallsexperiments heißt Ergebnisraum. Seine Elemente heißen Ergebnisse und stellen einen möglichen Ausgang des Zufallsexperiments dar.

Ergebnisraum

Beispiel 1.1.2 (diskrete und kontinuierliche Ergebnisräume)
Beim k-maligen Würfeln ist der Ergebnisraum $\Omega := \{1, \dots, 6\}^k$, es gilt $|\Omega| = 6^k$. Bei der Temperatur an einem bestimmten Ort zu einer festgelegten Zeit bietet sich $\Omega := [10, 30]$ an. ◊

DEFINITION 1.1.3 ((ELEMENTAR)EREIGNIS)

Teilmengen $A \subset \Omega$ heißen Ereignisse. Die Gesamtheit der Ereignisse ist somit $\mathcal{P}(A)$. Wir nennen Ω das sichere und $\varnothing$ das unmögliche Ereignis. Die Menge $\{\omega\}$ nennt man Elementarereignis.

Ereignis

Elementarereignis

Bermerkung 1.1.4 Elementarereignisse sind keine Ergebnisse!

Beispiel 1.1.5 (Ereignisse) Die Menge $\{1, 3, 5\}$ bezeichnet das Ereignis „ungerade Augenzahl", die Menge $[40, \infty)$ eine hohe Temperatur. ◊

Seien $(A_k)_{k=1}^n \subset \mathcal{P}(\Omega)$ Ereignisse.

Sprache der Ereignisse	Mengen-schreibweise
$\bigcup_{k=1}^n A_k$	mind. ein A_k tritt ein
$\bigcap_{k=1}^n A_k$	alle A_k treten ein
$A^\complement$	A tritt nicht ein
$A \subset B$	A impliziert B

Abb. 2: Mengenoperation auf Ereignissen

Der Satz A.1.1 zeigt, dass es im Allgemeinen unmöglich ist, jedem Ereignis $A \in \mathcal{P}(\Omega)$ in konsistenter Weise eine Wahrscheinlichkeit zuzuordnen. Deshalb schränkt man sich auf kleinere Mengensysteme $\mathcal{A} \subset \mathcal{P}(\Omega)$ ein, welche unter den in Abb. 2 genannten Mengenoperationen abgeschlossen sind, die σ-Algebren.

σ-Algebren

Wahrscheinlichkeiten

Wir wollen für jedes messbare Ereignis $A \in \mathcal{A}$ eine Wahrscheinlichkeit $\mathbb{P}(A) \in [0,1]$ festlegen, welche ein Maß dafür sein soll, dass A eintritt; tritt A niemals (sicher) ein, so setzt man $\mathbb{P}(A) = 0$ $(= 1)$. Insbesondere gilt $\mathbb{P}(\varnothing) = 0$ und $\mathbb{P}(\Omega) = 1$.

Zusätzlich soll für disjunkte Ereignisse A und B

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \qquad \text{(Additivität)}$$

gelten. Daraus folgt unmittelbar für eine Familie $(A_k)_{k=1}^n$ paarweise disjunkter Ereignisse

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathbb{P}(A_k). \qquad \text{(endliche Additivität)}$$

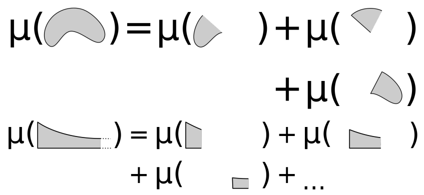


Abb. 3: endliche und abzählbare Additivität [Quelle: Wiki]

Gilt für jede abzählbare Familie paarweise disjunkter Ereignisse $(A_k)_{k \in \mathbb{N}}$

$$\mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right) = \sum_{k \in \mathbb{N}} \mathbb{P}(A_k), \qquad \text{(abzählbare / } \sigma\text{-Additivität)}$$

so heißt $\mathbb{P}$ Wahrscheinlichkeitsmaß:

Definition 1.1.6 (Kolmogorovsche Axiome, 1933)
Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine Abbildung $\mathbb{P} : \mathcal{A} \to [0,1]$ heißt Wahrscheinlichkeitsmaß auf $\mathcal{A}$, falls sie normiert und σ-additiv ist.

Messraum
Wahrscheinlichkeitsmaß

Bemerkung Ein Wahrscheinlichkeitsmaß kann als Verallgemeinerung der relativen Häufigkeit gesehen werden. [Mehr dazu im anderen Skript, todo]

Definition 1.1.7 (Wahrscheinlichkeitsraum)
Sei Ω eine nichtleere Menge, $\mathcal{A}$ eine σ-Algebra auf Ω und $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$. Dann heißt das Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ Wahrscheinlichkeitsraum.

Wahrscheinlichkeitsraum

Sei im Folgenden $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Definition 1.1.8 (Stetigkeit von Massen)
Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(A_k)_{k \in \mathbb{N}} \subset \mathcal{A}$ und $A \in \mathcal{A}$.

Gilt $A_1 \subset A_2 \subset \ldots$ und $A = \bigcup_{n\in\mathbb{N}} A_n$, so folgt $\mathbb{P}(A_n) \xrightarrow{n\to\infty} \mathbb{P}(A)$ (σ-stetig von unten).

Gilt $A_1 \supset A_2 \supset \ldots$, $\mathbb{P}(A_1) < \infty$ und $A = \bigcap_{k\in\mathbb{N}} A_k$, so folgt $\mathbb{P}(A_n) \xrightarrow{n\to\infty} \mathbb{P}(A)$ (σ-stetig von oben).

Lemma 1.1.9 (Eigenschaften von $\mathbb{P}$)

Jedes Wahrscheinlichkeitsmaß ist endlich additiv, monoton, subadditiv und stetig von oben / unten. Es gilt $\mathbb{P}(\varnothing) = 0$ und $\mathbb{P}(A^\complement) = 1 - \mathbb{P}(A)$.

Beispiel 1.1.10 (Erste Wahrscheinlichkeiten: Würfeln)

Beim einmaligen (zweimaligen) fairen Würfeln ist jede der sechs (36) möglichen Augenzahlen gleich wahrscheinlich.

Man setzt daher $\mathbb{P}(\{\omega\}) := \frac{1}{6}$ $(:= \frac{1}{36})$ für $\omega \in \Omega := \{1, \ldots, 6\}^{(2)}$. Die Wahrscheinlichkeit, eine ungerade Zahl zu Würfeln ist $3 \cdot \frac{1}{6} = \frac{1}{2}$ und, dass die Augensumme größer als zehn ist, $3 \cdot \frac{1}{36} = \frac{1}{12}$. $\Diamond$

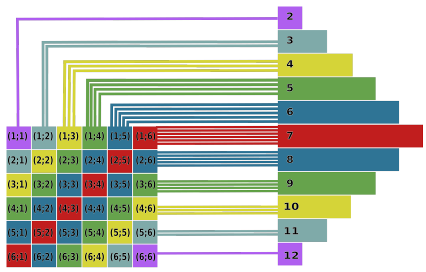


Abb. 4: Wahrscheinlichkeitsverteilung beim zweimaligen fairen Würfeln.

Satz 1.1.1: Einschluss-Ausschluss-Prinzip

Für $(A_k)_{k=1}^n \subset \mathcal{A}$ gilt die Siebformel von Sylvester:

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{\substack{I\subset\{1,\ldots,n\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I} A_k\right)$$
$$= \sum_{j=1}^n (-1)^{j+1} \sum_{1\leqslant i_1\leqslant\ldots\leqslant i_j\leqslant n} \mathbb{P}\left(\bigcap_{k=1}^j A_{i_k}\right)$$

Beweis. Wir beweisen mittels vollständiger Induktion über $n \in \mathbb{N}_{>1}$.

<u>Induktionsanfang: $n = 2$.</u> Da $A_1 \cup A_2 = A_1 \amalg (A_2 \backslash A_1)$ und $A_2 = (A_2 \backslash A_1) \amalg (A_2 \cap A_1)$ disjunkte Zerlegungen sind, gilt aufgrund der Additivität

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2 \backslash A_1) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2).$$

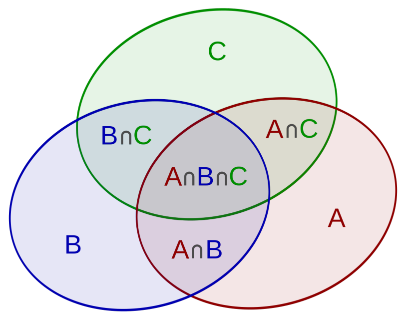


Abb. 5: Visualisierung der Formel von Sylvester für $n = 3$. [Quelle: Wiki]

<u>Induktionsschritt: $n \to n+1$.</u> Es gilt

$$\mathbb{P}\left(\bigcup_{k=1}^{n+1} A_k\right) = \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{k=1}^n A_k\right) \cap A_{n+1}\right)$$
$$= \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{k=1}^n (A_k \cap A_{n+1})\right).$$

Nach Induktionsannahme gilt

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{\substack{I\subset\{1,\ldots,n\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I} A_k\right)$$

und

$$\mathbb{P}\left(\bigcup_{k=1}^{n}(A_k \cap A_{n+1})\right) = \sum_{\substack{I\subset\{1,\dots,n\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I}(A_k \cap A_{n+1})\right)$$

$$= \sum_{\substack{I\subset\{1,\dots,n\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I\cup\{n+1\}} A_k\right).$$

Einsetzen ergibt

$$\mathbb{P}\left(\bigcup_{k=1}^{n+1} A_k\right) = \sum_{\substack{I\subset\{1,\dots,n\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I} A_k\right) + \mathbb{P}(A_{n+1})$$

$$- \sum_{\substack{I\subset\{1,\dots,n\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I\cup\{n+1\}} A_k\right)$$

$$=??$$

$$= \sum_{\substack{I\subset\{1,\dots,n+1\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I} A_k\right). \qquad \square$$

Bemerkung (V) Der Satz gilt sogar für endliche Inhalte $\mathbb{P}$ auf Ringen.

Beispiel 1.1.11 (Anwendung: Fixpunktfreie Permutation)

Es treffen sich $n \in \mathbb{N}$ Studenten und legen ihre Jacken auf einen Haufen. Vor dem Heimweg nimmt jeder zufällig eine Jacke. Wie groß ist die Wahrscheinlichkeit, mindestens einer der Studenten seine eigene Jacke bekommt?

Sei $\mathcal{N} := \{1,\dots n\}$. Wir wählen den Wahrscheinlichkeitsraum

$$\Omega := \{(\varphi(k))_{k\in\mathcal{N}} : \varphi : \mathcal{N} \to \mathcal{N} \text{ bijektiv}\},\ \mathcal{A} := \mathcal{P}(\Omega)$$

und die Gleichverteilung $\mathbb{P} : \mathcal{P}(A) \to [0,1],\ A \mapsto \frac{|A|}{|\Omega|}$. Dann gilt $|\Omega| = n!$.

Definiere nun für $i \in \mathcal{N}$ die Ereignisse

$$A_i := \{\texttt{Student } i \texttt{ geht mit seiner Jacke nach Hause}\} = \{\varphi(i) = i\}$$

Nach dem obigen Satz gilt (vgl. Abb. 2) ist die Wahrscheinlichkeit, dass mindestens einer der Studenten seine eigene Jacke bekommt,

$$\mathbb{P}\left(\bigcup_{i\in\mathcal{N}} A_i\right) = \sum_{\substack{I\subset\{1,\dots,n\}\\ I\neq\varnothing}} (-1)^{|I|-1}\,\mathbb{P}\left(\bigcap_{k\in I} A_k\right)$$

$$\overset{(\star)}{=} \sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\,\mathbb{P}\left(\bigcap_{k=1}^{j} A_k\right)$$

$$= \sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\prod_{k=1}^{j}\frac{1}{(n-k+1)}$$

$$= \sum_{j=1}^{n}\frac{(-1)^{j-1}}{j!} \xrightarrow{n\to\infty} 1-\frac{1}{e} \approx 63.21\%,$$

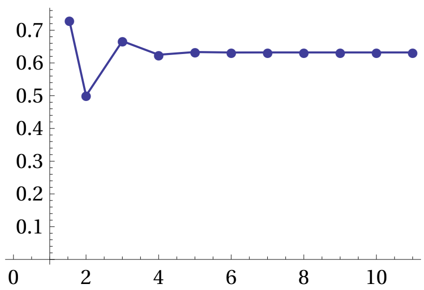


Abb. 6: Die linken Wahrscheinlichkeiten für $n \leqslant 12$. [Quelle: WolframAlpha]

wobei wir in ($\star$) benutzen, dass die Wahrscheinlichkeiten aller Schnittmengen mit derselben Anzahl an Teilmengen gleich sind. $\Diamond$

1.2 Diskrete Wahrscheinlichkeitsräume

Ist Ω höchstens abzählbar also diskret, so ist jede Teilmenge diskret

$$A := \{\omega_1, \dots\} = \bigcup_{n\in\mathbb{N}} \{\omega_n\} \subset \Omega$$

abzählbar und somit eine abzählbare Vereinigung von Elementarereignissen. Deswegen ist für diskrete Wahrscheinlichkeitsräume deren Potenzmenge die natürliche σ-Algebra.

Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf Ω kann man

$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}} \mathbb{P}(\{\omega_n\}) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}).$$

als absolut konvergente Reihe darstellen.

DEFINITION 1.2.1 (ZÄHLDICHTE, ENGL.: PMF)

Sei Ω diskret. Eine Funktion $p : \Omega \to [0,1]$ mit der Eigenschaft engl.: *probability mass function*

$$\sum_{\omega\in\Omega} p(\omega) = 1. \qquad (1)$$

heißt Zähldichte bzw. Wahrscheinlichkeitsfunktion. Zähldichte

Durch

$$p : \Omega \to [0,1],\ \omega \mapsto \mathbb{P}(\{\omega\})$$

erhalten wir aus $\mathbb{P}$ eine Zähldichte auf Ω. Eine einfache Methode zur Konstruktion von Wahrscheinlichkeitsmaßen auf diskreten Wahrscheinlichkeitsräumen mittels Funktionen mit Eigenschaft (1) liefert der

SATZ 1.2.1: 1-1-BEZIEHUNG ZWISCHEN ZÄHLDICHTEN UND WAHRSCHEINLICHKEITSMASSEN

Seien Ω diskret und p eine Zähldichte. Dann definiert

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0,1],\ A \mapsto \sum_{\omega\in A} p(\omega)$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$ und heißt Verteilung auf Ω.

Beweis. Die Nichtnegativität und Normiertheit sind klar, also bleibt nur die σ-Additivität nachzuweisen. Seien

$$\left(A_n := (\omega_{nk})_{k=1}^{|A_n|+1}\right)_{n\in\mathbb{N}} \subset \mathcal{P}(\Omega)$$

paarweise disjunkt. Dann ist

$$\{\omega_{nk} : n \in \mathbb{N},\ k \in \{1, \dots |A_n| + 1\}\}$$

eine Aufzählung aller Elemente von $A := \bigcup_{n\in\mathbb{N}} A_n$, wobei jedes Element aus A genau einmal aufgezählt wird, da die Ereignisse A_n paarweise disjunkt sind. Dann folgt

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega) = \sum_{n\in\mathbb{N}} \sum_{k=1}^{|A_n|+1} p(\omega_{nk}) = \sum_{n\in\mathbb{N}} \sum_{\omega\in A_n} p(\omega) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n). \quad \square$$

Zur vollständigen Beschreibung des Wahrscheinlichkeitsraums genügt also *im diskreten Fall* die Angabe von (Ω, p).

Beispiel 1.2.2 (n-facher Münzwurf) Für $\mathcal{N} := \{1, \dots, n\}$ wählen wir

$$\Omega := \{(i_k)_{k\in\mathcal{N}} : i_j \in \{0,1\}, j \in \mathcal{N}\} = \{0,1\}^n,$$

wobei wir 0 als Kopf und 1 als Zahl interpretieren. Da alle Ergebnisse gleich wahrscheinlich sind, setzen wir

$$p(\omega) = \frac{1}{|\Omega|} = 2^{-n}.$$

Für $k \in \mathcal{N}$ ist die Wahrscheinlichkeit im k-ten Wurf Kopf zu werfen durch

$$\frac{|\{(i_1, \dots, i_{k-1}, 0, i_{k+1}, \dots, i_n) : i_j \in \{0,1\}, j \in \mathcal{N}\}|}{2^n} = \frac{2^{n-1}}{2^n} = \frac{1}{2}$$

gegeben. $\Diamond$

Das obige Beispiel ist Spezialfall eines

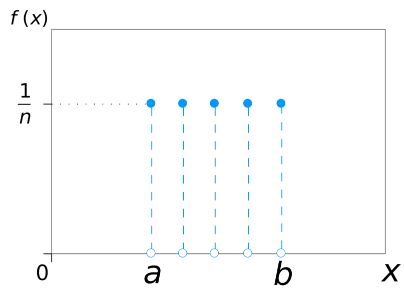


Abb. 7: Zähldichte einer diskreten Gleichverteilung. [Quelle: Wiki]

Definition 1.2.3 (Laplace-Wahrscheinlichkeitsraum)
Sei Ω endlich. Dann heißt das durch $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ definierte Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$ Gleichverteilung auf Ω.

Gleichverteilung

(Man sagt, $\mathbb{P}$ werde durch die Zähldichte $p : \Omega \to [0,1],\ \omega \mapsto \frac{1}{|\Omega|}$ induziert.)

Die Berechnung von $\mathbb{P}(A)$ führt auf Abzählprobleme, deren wichtigste Vertreter wir anhand einfacher Urnenmodelle illustrieren:

Abzählproblem

Beispiel 1.2.4 (Abzählprobleme mit Urnen)
Eine Urne enthält mit den Zahlen $1, \dots, n$ beschriftete Kugeln, von denen wir $k \leqslant n$ ziehen. Sei $\mathcal{N} := \{1, \dots, n\}$.

1. **in Reihenfolge mit Zurücklegen**. Wir setzen

$$\Omega := \{(x_1, \dots, x_k) \subset \mathcal{N}^k\}.$$

Dann gilt $|\Omega| = n^k$.

2. **in Reihenfolge ohne Zurücklegen.** Wir setzen

$$\Omega := \{(x_1, \dots, x_k) \subset \mathcal{N}^k, x_i \neq x_j \text{ für } i \neq j\}.$$

Dann gilt

$$|\Omega| = n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

Für $k = n$ erhält man den Spezialfalls $|\Omega| = n!$, welche also die Anzahl aller Permutationen von $\mathcal{N}$ angibt.

Permutationen

3 **ohne Reihenfolge mit Zurücklegen.** Wir setzen

$$\Omega := \{\{x_1, \ldots, x_k\} \subset \mathcal{N}^k, x_i \neq x_j \text{ für } i \neq j\}.$$

Im Unterschied zum Ziehen unter Beachtung der Reihenfolge werden alle $k!$ Stück k-Tupel, welche zu der selben Menge an Kugeln führen, zu einem Elementarereignis zusammengefasst. Also gilt

$$|\Omega| = \frac{1}{k!} \frac{n!}{(n-k)!} =: \binom{n}{k},$$

was genau der Anzahl aller k-elementigen Teilmengen einer n-elementigen Grundmenge entspricht.

Eine alternative Darstellung von Ω erhält man, da es unter allen k-Tupeln, welche zur selben Menge führen, genau eines gibt, in welchem die Elemente ihrer Größe nach geordnet sind:

$$\Omega = \{(x_1, \ldots, x_k) \subset \mathcal{N}^k, x_i < x_{i+1} \ \forall i \in \{1, \ldots, k-1\}\}.$$

4 **ohne Reihenfolge ohne Zurücklegen.**

Analog zum vorherigen Fall ordnen wir wieder die Nummern der gezogenen Kugeln der Größe nach an:

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(k)} \tag{2}$$

wobei wegen des Zurücklegens Kugeln mehrfach gezogen werden können. Durch den (bijektiven) Übergang von $x_{(i)}$ zu $x_{(i)} + i - 1$ erhält man aus (2) eine streng monoton aufsteigende Folge

$$x_{(1)} \leqslant x_{(2)} + 1 \leqslant x_{(3)} + 2 \leqslant \ldots \leqslant x_k + k - 1.$$

Wir erhalten also

$$\Omega := \{(x_1, \ldots, x_k) :\subset \{1, \ldots, n+k-1\}^k, x_i < x_{i+1} \ \forall i \in \{1, \ldots, k-1\}\}$$

und somit nach 3 $|\Omega| = \binom{n+k-1}{k}$.

[König-Skript] Alternativ kann man die Formel auch so herleiten: *Wie viele Möglichkeiten gibt es, k ununterscheidbare Murmeln auf n Zellen zu verteilen?* Seien die k Murmeln in einer Reihe gelegt. Sie in n Zellen einzuteilen ist äquivalent dazu, $n-1$ Trennwände zwischen die k Murmeln zu setzen. Dadurch erhalten wir eine Reihe von $n+k-1$ Objekten, und nach 3 gibt es $\binom{n+k-1}{n}$ Möglichkeiten, diese anzuordnen. ◊

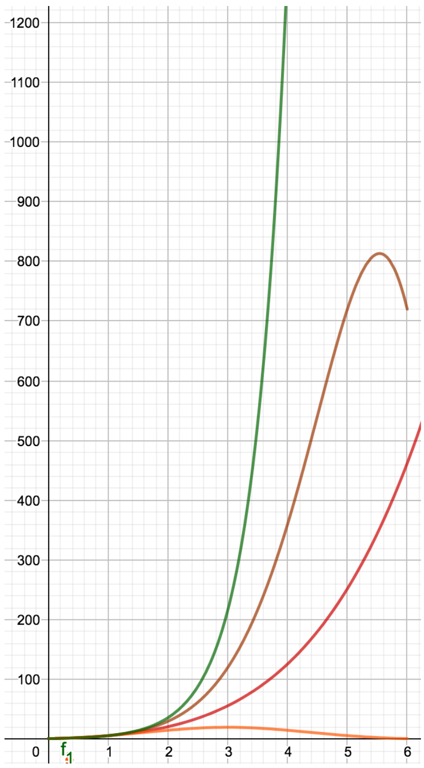


Abb. 8: Graphen der verschiedenen Urnenmodelle (stetig fortgesetzt).

Eine Anwendung ist der

Satz 1.2.2: Binomischer Lehrsatz

Für Elemente x, y eines kommutativen unitären Rings gilt

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}, \qquad n \in \mathbb{N}.$$

Beweis. Schreibe $(x+y)^n = \prod_{k=1}^n (x_k + y_k)$ mit $x_i = x$ und $y_i = y$. Beim Ausmultiplizieren tritt der Term $x^k y^{n-k}$ immer dann auf wenn in k Klammern der Faktor x_i und in $n-k$ Klammern der Faktor y_i gewählt wird, also nach 3 in $\binom{n}{k}$ Fällen. □

Korollar 1.2.5

Es gilt $\sum_{k=0}^n \binom{n}{k} = 2^n$, $\sum_{k=0}^n (-1)^k \binom{n}{k} = 0$ *und* $\sum_{k=0}^n k\binom{n}{k} = n2^{n-1}$.

Beispiel 1.2.6 (Anwendung Abzählprobleme: Paar Schuhe)

In einem Karton befinden sich n Paar Schuhe. Man nimmt zufällig $r \leqslant n$ Schuhe heraus. Wie groß ist die Wahrscheinlichkeit $\mathbb{P}_n^{(r)}(p)$, dass darunter genau $p \in \{1, \dots, \lfloor \frac{r}{2} \rfloor\}$ Paare sind?

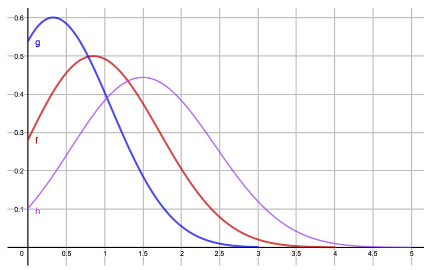


Abb. 9: Die Funktionen $\mathbb{P}_{15}^{(r)}(p)$ für $r \in \{8, 6, 10\} \cong \{f, g, h\}$.

Wir betrachten ein Abzählproblem ohne Reihenfolge und ohne Zurücklegen über dem Ereignisraum $\Omega := \{x_1, y_1, \dots, x_n, y_n\}$, in welchem sich n Paar Schuhe (x_i, y_i) befinden. Somit ist $|\Omega| = \binom{2n}{r}$. Wir betrachten die Möglichkeiten genau p aus n Paare zu ziehen, also $\binom{n}{p}$. Somit blieben uns dann $n-p$ Paare, die es möglich wären zu ziehen. Da wir jedoch kein weiteres ziehen wollen, betrachten wir dann $\binom{n-p}{r-2p}$ Möglichkeiten kein weiteres zu ziehen. Zuallerletzt muss noch beachtet werden, dass wir für jedes Paar, die Möglichkeit linker Schuh oder rechter Schuh, als 2^{r-2p} Kombinationsmöglichkeiten haben. Somit gilt

$$\mathbb{P}_n^{(r)}(p) = \frac{\binom{n}{p}\binom{n-p}{r-2p}2^{r-2p}}{\binom{2n}{r}}.$$ ◊

Beispiel 1.2.7 (Geburtstags-Paradox)

Man fragt zufällig $n \in \mathbb{N}_{\geqslant 2}$ Studenten nach ihrem Geburtstag. Wie groß ist die Wahrscheinlichkeit $\wp_n$, dass mindestens zwei der angesprochenen Stunden am gleichen Tag Geburtstag haben?

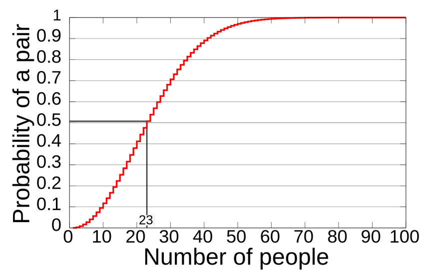


Abb. 10: [Quelle: Wiki]

Wir benutzen das Gegenereignis $\hat{\wp}$: jeder Student hat an einem anderen Tag Geburtstag: Für den ersten Studenten ist die Wahrscheinlichkeit, mit keinem der vorherigen Studenten einen Geburtstag zu teilen, genau eins. Für den zweiten Studenten ist die Wahrscheinlichkeit an dem Geburtstag des vorherigen Studenten Geburtstag zu haben, $\frac{364}{365}$. Somit erhalten wir $\hat{\wp} = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \dots \cdot \frac{364-n}{365}$, also $\wp_n = 1 - \prod_{k=0}^{n-1}\left(1 - \frac{k}{365}\right)$. Somit gilt schon $\wp_{23} > 0.5$. ◊

Beispiel 1.2.8 (Multiple-Choice Test)

Ein Student legt einen multiple-choice Test mit 20 Fragen ab und wählt bei jeder Frage zufällig eine der $k \in \mathbb{N}$ vorgeschlagenen Antworten, von denen nur eine richtig ist. Für jede richtige Antwort erhält der Student einen Punkt. Wie groß ist die Wahrscheinlichkeit, dass der Student am Ende genau $\ell \in \mathbb{N}$ Punkte hat?

Ist die Antwort falsch (Wahrscheinlichkeit $\frac{k-1}{k}$) so erhält der Student Null Punkte, mit einer Wahrscheinlichkeit von $\frac{1}{k}$ die richtige Antwort und somit einen Punkt. Wir wählen also den diskreten Wahrscheinlichkeitsraum mit $\Omega := \{0,1\}^{20}$ und $\mathcal{F} := \mathcal{P}(\Omega)$. Somit gilt für ein $\omega \in \Omega$

$$\mathbb{P}(\{\omega\}) := p(\omega) = \left(\frac{1}{k}\right)^{\sum_{j=1}^{20}\omega_j}\left(\frac{k-1}{k}\right)^{20-\sum_{j=1}^{20}\omega_j} = \frac{(k-1)^{20-\sum_{j=1}^{20}\omega_j}}{k^{20}}.$$

Für $S_\ell := \{\omega \in \Omega : \sum_{j=1}^{20}\omega_j = \ell\}$ gilt somit

$$\mathbb{P}(\S_\ell) = \sum_{\omega \in S_\ell} p(\omega) = |S_\ell|\frac{(k-1)^{20-\ell}}{k^{20}} = \binom{20}{\ell}\frac{(k-1)^{20-\ell}}{k^{20}}. \qquad \Diamond$$

Beispiel 1.2.9 (Multinominalkoeffizient)

Sei $n \in \mathbb{N}$ und A eine n-elementige Menge. Sei weiter $n_1, \ldots, n_k \in N$ mit $\sum_{\ell=1}^{k} n_\ell = n$, für $k \in \mathbb{N}_{\leqslant n}$. Wie viele verschiedene Möglichkeiten gibt es, die Menge A in k disjunkte Teilmengen A_j der Größe $n_1, \ldots, n_k$ zu unterteilen, wobei wir davon ausgehen, dass $n_i \neq n_j$ für $i \neq j$ gilt?

Wir setzen $\mathcal{N}_1$ als die Anzahl der Möglichkeiten, aus A n_1 Kugeln auszuwählen. Dann gilt $\mathcal{N}_1 = \binom{n}{n_1}$. Analog setzen wir $\mathcal{N}_2$ als die Anzahl der Möglichkeiten, aus $A \backslash A_1$ n_2 Kugeln auszuwählen. Dann gilt $\mathcal{N}_2 = \binom{n-n_1}{n_2}$.

Also ist Wahrscheinlichkeit gegeben durch

$$\begin{aligned}\prod_{j=1}^{k}\binom{n-\sum_{k=1}^{j-1} n_k}{n_j} &= \binom{n}{n_1}\cdot\binom{n-n_1}{n_2}\cdot\binom{n-(n_1+n_2)}{n_3}\cdot\ldots\cdot\binom{n_k+n_{k-1}}{n_{k-1}}\cdot 1\\ &= \frac{n!}{(n_1!)\cancel{(n-n_1)!}}\cdot\frac{\cancel{(n-n_1)!}}{(n_2!)(n-(n_1+n_2)!}\cdot\ldots\cdot\frac{\cancel{(n_k+n_{k-1})!}}{(n_{k-1}!)(n_k!)} = \frac{n!}{\prod_{j=1}^{k}(n_j!)} \quad \Diamond\end{aligned}$$

Beispiel 1.2.10 (Vier Asse)

Wir betrachten ein Standard-52- Kartenspiel. Jede Position einer Karte im Deck ist gleich wahrscheinlich. Wie groß ist die Wahrscheinlichkeit, dass genau vier Asse hintereinander zu finden sind?

Wir haben

$$\Omega := \{(\omega_1, \ldots, \omega_{52}) : \omega_i \in \{1, \ldots, 52\}, \omega_i \neq \omega_j\ \forall i \neq j\}, \quad \mathcal{F} := \mathcal{P}(\Omega).$$

Und wählen als Wahrscheinlichkeitsmaß $\mathbb{P}$ die Gleichverteilung. Ohne Beschränkung der Allgemeinheit seien die vier Asse $\{1,2,3,4\}$.

Wir setzen

$$A := \{\omega \in \Omega : \exists k \leqslant 49 : \{k, k+1, k+2, k+3\} = \{1,2,3,4\}\}. \qquad \Diamond$$

Da es 49 Möglichkeiten für die Position der vier Asse gibt, die auf 4! Weisen permutiert werden können und für die verbleibenden Karten 48! Permutation möglich sind, gilt $|A| = 49 \cdot 4! \cdot 48! = 49! \cdot 4!$. Oder auch (vgl. 1.2.9) $49 \cdot \frac{1}{\frac{52!}{48! \cdot 4!}}$.

1.3 Stetige Wahrscheinlichkeitsräume

Für viele Zufallsexperimente kann der Ergebnisraum nicht diskret gewählt werden, bei anderen ergibt er sich natürlicherweise bei unendlichen Wiederholungen von diskreten Zufallsexperimenten.

Beispiel 1.3.1 (∞-facher Münzwurf)
Der überabzählbare Ergebnisraum ist $\Omega := \{0,1\}^{\mathbb{N}}$. Wir können die ersten n Münzwürfe $\{\omega_1, \dots, \omega_n\}$ als Teilmenge von Ω auffassen, indem wir die zugehörige Zylindermenge

$$\{\omega_1, \dots, \omega_n\} \times \{0,1\} \times \{0,1\} \times \dots$$

betrachten. ◊

Gibt es eine Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathcal{P}(\Omega)$, sodass für jede solche Zylindermenge

$$\mathbb{P}(\{\omega_1, \dots, \omega_n\} \times \{0,1\} \times \{0,1\}, \dots) = 2^{-n} \tag{3}$$

gilt, also dass die Wahrscheinlichkeit eines Ereignisses, welche nur von den ersten n Würfen abhängt, gerade der Wahrscheinlichkeit dieses Ereignisses bezüglich des in Beispiel 1.2.2 betrachteten Wahrscheinlichkeitsmaßes für n faire Münzwürfe entspricht? Nach Satz A.1.1 ist die Antwort Nein.

Der Ausweg besteht in der Verkleinerung der Ereignissysteme auf eine σ-Algebra, welche strikt kleiner als die Potenzmenge ist, aber immer noch alle von endlich vielen Münzwürfen erzeugten Zylindermengen enthält. Es existiert ein Wahrscheinlichkeitsmaß mit der Eigenschaft (3) auf diesem Mengensystem.

DEFINITION 1.3.2 (BOREL-σ-ALGEBRA)
Sei

$$\mathcal{C} := \left\{ (a,b] := \prod_{k=1}^{n} (a_k, b_k] : a = (a_k)_{k=1}^{n}, b = (b_k)_{k=1}^{n} \right\},$$

wobei $-\infty \leqslant a_i \leqslant b_i < \infty$ für alle $i \in \{1, \dots, n\}$ gelten soll. Dann ist $\sigma(\mathcal{C}) =: \mathcal{B}(\mathbb{R}^d)$ die Borel-σ-Algebra.

Borel-σ-Algebra

Bermerkung 1.3.3 Es gilt $\mathcal{B}(\mathbb{R}^d) \subsetneq \mathcal{P}(\mathbb{R}^d)$. $\mathcal{B}(\mathbb{R}^d)$ enthält u.a. alle offenen, abgeschlossenen und kompakten Teilmengen des $\mathbb{R}^d$.

An die Stelle der Zähldichte tritt im stetigen Fall eine

Definition 1.3.4 (Dichte (engl. PDF))

Ein integrierbare Funktion $f : \mathbb{R}^d \to [0, \infty)$ heißt Dichte, wenn $\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = 1$ gilt.

Dichte

engl. probability density function

Satz 1.3.1: Kontinuierliches Analogon zu 1.2.1

Sei f eine Dichte auf $\mathbb{R}^d$. Dann definiert

$$\mathbb{P} : \mathcal{B}(\mathbb{R}^d) \to \mathbb{R}, \; A \mapsto \int_A f(x)\,\mathrm{d}x$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}^d)$.

Definition 1.3.5 (Gleichverteilung im $\mathbb{R}^d$)

Sei $\Omega \in \mathcal{B}(\mathbb{R}^d)$ endlich. Das zu der Dichte $\frac{\mathbb{1}_\Omega}{|\Omega|}$ gehörende Wahrscheinlichkeitsmaß auf Ω heißt Gleichverteilung auf Ω.

Gleichverteilung

Es gilt $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \mathcal{B}(\mathbb{R})$. Insbesondere ist die Dichte der Gleichverteilung auf $[a, b]$ durch $\frac{\mathbb{1}_{[a,b]}}{\prod_{k=1}^d (b_i - a_i)}$ gegeben.

Gleichverteilung

Beispiel 1.3.6 (Betrandsches Paradox)

TODO. nicht so wichtig, vgl. zweite Übung. ◊

2 Bedingte Wahrscheinlichkeit und Unabhängigkeit

2.1 Bedingte Wahrscheinlichkeit

Sind über den Ausgang eines Zufallsexperiments bereits Informationen verfügbar, ändern sich die Wahrscheinlichkeiten.

Sei im Folgenden $(\Omega, \mathcal{A}, \mathbb{P})$ stets ein Wahrscheinlichkeitsraum.

Beispiel 2.1.1 (Bedingte Wahrscheinlichkeit beim Würfeln)
Nach Beispiel 1.1.10 ist die Wahrscheinlichkeit mit zwei fairen Würfeln die Augensumme größer als 10 zu würfeln, $\frac{1}{12}$.

Wie ändert sich die Wahrscheinlichkeiten, wenn zuerst eine sechs gewürfelt wird? Für den zweiten Wurf gibt es sechs gleichwahrscheinliche Möglichkeiten, von denen zwei das Gewünschte erreichen. Somit ist die Wahrscheinlichkeit unter dieser Annahme gegeben durch

$$\begin{aligned}\mathbb{P}(\text{Augensumme} > 10 \mid 1.\text{ Wurf } 6) &= \frac{\mathbb{P}(\text{Augensumme} > 10,\ 1.\text{ Wurf } 6)}{\mathbb{P}(1.\text{ Wurf } 6)} \\ &= \frac{\frac{2}{36}}{\frac{6}{36}} = \frac{1}{3}.\end{aligned} \qquad \lozenge$$

Die obige Notation erklären wir mit der folgenden

Definition 2.1.2 (Bedingte Wahrscheinlichkeit)
Seien und $A, B \in \mathcal{A}$ Ereignisse. Dann heißt

$$\mathbb{P}(A \mid B) := \begin{cases} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, & \text{für } \mathbb{P}(B) > 0, \\ 0, & \text{für } \mathbb{P}(B) = 0. \end{cases}$$

die bedingte Wahrscheinlichkeit von A gegeben B.

bedingte Wahrscheinlichkeit

Korollar 2.1.3 (Eigenschaften bedingter Wahrscheinlichkeiten)
Seien $A, B \in \mathcal{A}$. Dann gilt

- $B \subset A \implies \mathbb{P}(A \mid B) = 1$ *für* $\mathbb{P}(B) > 0$,
- $\mathbb{P}(A \mid B) = 0$ *für* $A \cap B = \varnothing$.
- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ *für alle* $B \in \mathcal{A}$ *und* $\mathbb{P}(A) = 1$.

Satz 2.1.1: Bedingtes Wahrscheinlichkeitsmass

Seien $B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ und $\widetilde{\mathbb{P}} : \mathcal{A} \to [0,1],\ A \mapsto \mathbb{P}(A \mid B)$. Dann ist $(\Omega, \mathcal{A}, \widetilde{\mathbb{P}})$ ein Wahrscheinlichkeitsraum.

Beweis. Die Nichtnegativität und Wohldefiniert folgt direkt aus den entsprechenden Eigenschaften von $\mathbb{P}$. Aus der Existenz eines $A \in \mathcal{A}$ mit $\widetilde{P}(\mathcal{A}) > 1$ folgt $\mathbb{P}(A \cap B) > \mathbb{P}(B)$, was im Widerspruch zu der Monotonie von $\mathbb{P}$ steht.

Sei nun $(A_k)_{k\in\mathbb{N}} \subset \mathcal{A}$ eine paarweise disjunkte Familie. Dann ist auch $(A_k \cap B)_{k\in\mathbb{N}} \subset \mathcal{A}$ eine paarweise disjunkte Familie und aufgrund der σ-Additivität von $\mathbb{P}$ $(\star)$ gilt für $A \in \mathcal{A}$

$$\widetilde{\mathbb{P}}(A) = \frac{\mathbb{P}\left(\bigcup_{k\in\mathbb{N}} A_k \cap B\right)}{\mathbb{P}(B)} \overset{(\star)}{=} \frac{\sum_{k\in\mathbb{N}} \mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \sum_{k\in\mathbb{N}} \widetilde{\mathbb{P}}(A_k). \qquad \square$$

Beispiel 2.1.4 (Bedingte Laplace Wahrscheinlichkeit)
Seien Ω endlich, $\mathbb{P}$ die Gleichverteilung auf Ω und $B \neq \varnothing$. Dann ist die bedingte Laplace-Wahrscheinlichkeit die Gleichverteilung auf B:

$$\mathbb{P}(A \mid B) = \frac{\frac{|A\cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{|A \cap B|}{|B|}. \qquad \Diamond$$

Beispiel 2.1.5 (Verkehrsunfälle)
Seien die Ereignisse $U :=$ „Unfall“, $M :=$ „Versicherter männlich“ und $W :=$ „Versicherte weiblich“ definiert und die Wahrscheinlichkeiten $\mathbb{P}(U \mid W) = 0.002$, $\mathbb{P}(U \mid M) = 0.005$ sowie $\mathbb{P}(W) = 0.4 = 1 - \mathbb{P}(M)$ gegeben. Dann gilt

$$\begin{aligned}\mathbb{P}(U) &= \mathbb{P}(U, W) + \mathbb{P}(U, M) \overset{2.1.2}{=} \mathbb{P}(U \mid W)\,\mathbb{P}(W) + \mathbb{P}(U \mid M)\,\mathbb{P}(M) \\ &= 0.002 \cdot 0.4 + 0.005 \cdot 0.6 = 0.0038. \qquad \Diamond\end{aligned}$$

Die Berechnung dieser „totalen“ Wahrscheinlichkeit für einen Arbeitsunfall ist ein Spezialfall des folgenden

Satz 2.1.2: von der totalen Wahrscheinlichkeit

Seien $A \in \mathcal{A}$ und $(B_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ eine paarweise disjunkte Familie mit $A \subset \bigcup_{n\in\mathbb{N}} B_n$. Dann gilt

$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}} \mathbb{P}(A \mid B_n) \cdot \mathbb{P}(B_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A \cap B_n).$$

Beweis. Für $\mathbb{P}(B_n) > 0$ folgt $\mathbb{P}(A \mid B_n)\,\mathbb{P}(B_n) = \mathbb{P}(A \cap B_n)$ $(\star)$ aus der Definition. Für $\mathbb{P}(B_n) = 0$ folgt, da $\mathbb{P}(A \cap B_n) \leqslant \mathbb{P}(B_n) = 0$ mit der Monotonie auch $(\star)$. Somit gilt

$$\sum_{n\in\mathbb{N}} \mathbb{P}(A \mid B_n)\,\mathbb{P}(B_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A \cap B_n) \overset{1.1.9}{=} \mathbb{P}\left(A \cap \bigcup_{n\in\mathbb{N}} B_n\right) = \mathbb{P}(A). \square$$

Beispiel 2.1.6 (Aus $\mathbb{P}(U)$ die W-keit $\mathbb{P}(M \mid U)$ berechnen)
Es gilt

$$\mathbb{P}(M \mid U) = \frac{\mathbb{P}(M,U)}{\mathbb{P}(U)} = \frac{\mathbb{P}(U \mid M)\,\mathbb{P}(M)}{\mathbb{P}(U)} = \frac{0.005 \cdot 0.6}{0.0038} = 0.789. \quad \Diamond$$

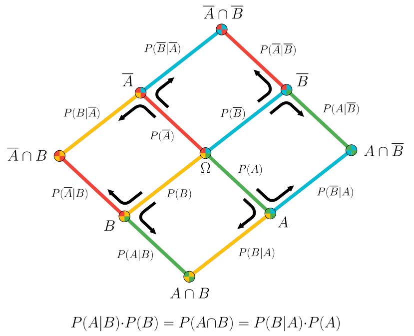


Abb. 11: Darstellung von des Satzes von Bayes für $n = 1$ mit Entscheidungsbäumen. [Quelle: Wiki]

Der folgende Satz verallgemeinert die Rechnung aus dem letzten Beispiel.

Satz 2.1.3: Bayes (1763)

Unter den Voraussetzungen von Satz 2.1.2 und $P(A) > 0$ gilt

$$\mathbb{P}(B_n \mid A) = \frac{\mathbb{P}(A \mid B_n) \cdot \mathbb{P}(B_n)}{\sum_{k\in\mathbb{N}} \mathbb{P}(A \mid B_k) P(B_k)} \left(= \frac{\mathbb{P}(A \cap B_n)}{\sum_{k\in\mathbb{N}} \mathbb{P}(A \cap B_k)} \right).$$

Beweis. Aus Satz 2.1.2 folgt $P(A) = \sum_{k\in\mathbb{N}} P(A \mid B_k) \cdot P(B_k)$ und somit

$$\begin{aligned} \mathbb{P}(B_n \mid A) &= \frac{\mathbb{P}(B_n \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B_n)}{\mathbb{P}(B_n)} \cdot \frac{\mathbb{P}(B_n)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A \mid B_n) \cdot \mathbb{P}(B_n)}{\sum_{k\in\mathbb{N}} \mathbb{P}(A \mid B_k)\,\mathbb{P}(B_k)}. \end{aligned} \qquad \square$$

Beispiel 2.1.7 (Anwendung: Das Ziegenproblem)
In einer Gameshow gibt es drei Türen, hinter zwei stehen Ziegen, hinter der anderen ein Sportwagen. Wir dürfen eine Tür wählen und behalten, was dahinter ist. Wir wählen eine Tür. Anstatt unsere Tür zu öffnen, öffnet der Moderator eine andere und es kommt eine Ziege zum Vorschein. Uns wird angeboten, die Tür zu wechseln, sollten wir das tun?

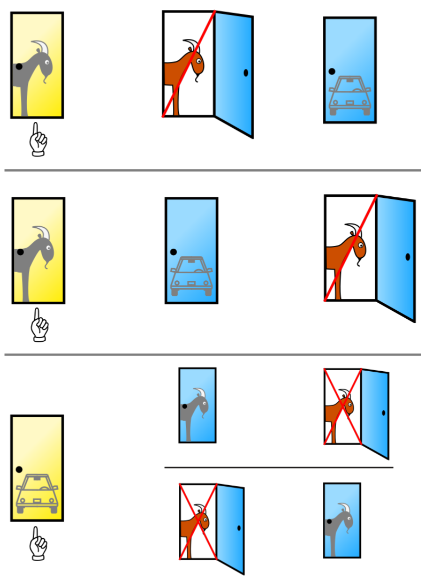


Abb. 12: [Quelle: Wiki]

Zusätzliche Annahmen: Die Position des Sportwagens ist zufällig gemäß der Gleichverteilung auf die drei Türen verteilt. Der Moderator weiß, wo der Sportwagen ist. Wählt der Kandidat eine Tür mit dem Sportwagen, so wählt er gleichverteilt eine andere Tür.

Sei $A_i :=$ `"Der Sportwagen ist hinter Tür` i`"` für $i \in \{1,2,3\}$. Dann gilt $\bigsqcup_{k=1}^{3} A_i =: \Omega$. Ohne Einschränkung wählt der Kandidat Tür eins und der Moderator Tür zwei. Sei $B :=$ "Moderator öffnet Tür mit Ziege"$\subset \Omega$.

Da der Moderator gleichverteilt wählt, wenn wir den Sportwagen gewählt haben, gilt $\mathbb{P}(B \mid A_1) = \frac{1}{2}$. Der Moderator wähle die Tür mit der anderen Ziege, wenn wir eine gewählt haben, es gilt $\mathbb{P}(B \mid A_2) = 0$ und $\mathbb{P}(B \mid A_3) = 1$.

Also ist die Gewinnchance ohne Wechsel $P(A_1) = \frac{1}{3}$ und mit

$$\mathbb{P}(A_3 \mid B) = \frac{\mathbb{P}(B \mid A_3)\,\mathbb{P}(A_3)}{\sum_{k=1}^{3} \mathbb{P}(B \mid A_k)\,\mathbb{P}(A_k)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{2}{3}$$

mit der Formel von Bayes. Es ist also sinnvoll, zu wechseln. ◊

Weitere Anwendung: Ruinprobleme. **TODO**

Beispiel 2.1.8 (Falsch-Positive)

Angenommen, 5‰ der Bevölkerung haben eine Krankheit K. Ein Test zeigt bei 99% der Erkrankten eine positive Reaktion (P): $\mathbb{P}(P \mid K) = 0.99$, allerdings zeigt er bei 2 % der Gesunden eine positive Reaktion: $\mathbb{P}(P \mid K^c) = 0.02$.

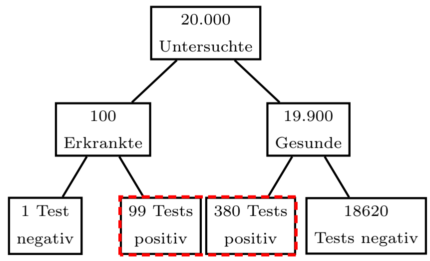


Abb. 13: Von 479 positiv getesteten sind nur 99 (knapp jeder fünfte) wirklich erkrankt. [Quelle: selber]

Die Formel von Bayes liefert mit $B_1 := K$ und $B_2 := K^\complement$

$$\begin{aligned}\mathbb{P}(K \mid P) &= \frac{\mathbb{P}(P \mid K)\,\mathbb{P}(K)}{\mathbb{P}(P \mid K) \cdot \mathbb{P}(K) + \mathbb{P}(P \mid K^\complement)\,\mathbb{P}(K^\complement)} \\ &= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.02 \cdot 0.995} \approx 0.2\end{aligned}$$

Also ist in nur zwei von zehn positiven Tests die getestet Person wirklich erkrankt. ◊

2.2 Unabhängigkeit

DEFINITION 2.2.1 (STOCHASTISCHE UNABHÄNGIGKEIT)

Eine Familie von Ereignissen $(A_i)_{i\in I}$ heißt stochastisch unabhängig wenn für alle endlichen nichtleeren Teilmengen $J \subset I$ gilt

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j). \tag{4}$$

stochastisch unabhängig

Gegenbeispiel 2.2.2 (Vollständig $\iff$ paarweise unabhängig)

Seien $A \in \mathcal{A}$ mit $\mathbb{P}(A) \in (0,1)$ und $n \in \mathbb{N}_{>2}$. Definiere $A_1 = \ldots = A_{n-1} :=$ A und $A_n = \varnothing$. Dann gilt $\mathbb{P}\left(\bigcap_{k\in\mathbb{N}} A_k\right) = \mathbb{P}(A)^{n-1} \cdot \mathbb{P}(A_n) = 0$ aber

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A) \neq \mathbb{P}(A)^2 = \mathbb{P}(A_1)\,\mathbb{P}(A_2).$$

Definiere $A :=$ „1. Wurf Zahl“, $B :=$ „2. Wurf Zahl“ und $C :=$ „1. und 2. Wurf gleich“. Es gilt $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ und $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$ aber $\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)$.◊

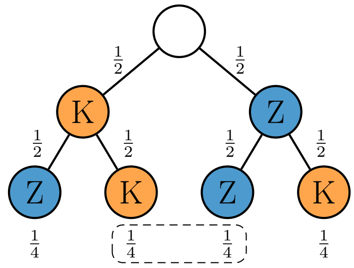


Abb. 14: Wahrscheinlichkeiten beim zweimaligen fairen Münzwurf. [Quelle: selbst]

SATZ 2.2.1: UNABHÄNGIGKEIT ENDLICHER FAMILIEN

Die Familie $(A_k)_{k=1}^n$ ist genau dann unabhängig, wenn für jede Wahl von $B_k \in \{A_k, A_k^\complement\}$, $k \in \{1,\ldots,n\}$

$$\mathbb{P}\left(\bigcap_{k=1}^n B_k\right) = \prod_{k=1}^n \mathbb{P}(B_k)$$

gilt.

Beweis. „ $\Longrightarrow$ “: Wir führen eine vollständige Induktion über $m := |\{i \in \{1,\ldots,n\} : B_i = A_i\}|$.

Induktionsanfang: $m = 0$. Für $m = 0$ gilt $B_i = A_i$ für alle $i \in \{1,\ldots,n\}$.

Induktionsannahme: Ist $m := |\{i \in \{1,\ldots,n\} : B_i = A_i\}|$, so sind die Ereignisse $(B_k)_{k=1}^n$ unabhängig.

Induktionsschritt: $n \to n+1$. Durch Umnummerierung können wir $B_1 = A_1^\complement$ annehmen. Nach Annahme sind die Ereignisse $A_1, B_2, \ldots, B_n$ unabhängig, also gilt für eine nichtleere Teilmenge $J \subset \{2,\ldots,n\}$

$$\mathbb{P}\left(\bigcap_{j\in J} B_j\right) = \prod_{j\in J} \mathbb{P}(B_j) \quad \text{und} \quad \mathbb{P}\left(A_1 \cap \bigcap_{j\in J} B_j\right) = \mathbb{P}(A_1) \prod_{j\in J} \mathbb{P}(B_j).$$

Mit einer Verallgemeinerung von (5) folgt durch Umstellen

$$\begin{aligned}\mathbb{P}\left(\bigcap_{j\in J\cup\{1\}} B_j\right) &= \mathbb{P}\left(\bigcap_{j\in J} B_j\right) - \mathbb{P}\left(A_1 \cap \bigcap_{j\in J} B_j\right)\\ &= \prod_{j\in J}\mathbb{P}(B_j) - \mathbb{P}(A_1)\prod_{j\in J}\mathbb{P}(B_j)\\ &= (1-\mathbb{P}(A_1))\prod_{j\in J}\mathbb{P}(B_j) = \mathbb{P}(B_1)\prod_{j\in J}\mathbb{P}(B_j).\end{aligned}$$

„ $\Longleftarrow$ “: Nach Annahme gilt

$$\mathbb{P}\left(\bigcap_{k=1}^n B_k\right) = \prod_{k=1}^n \mathbb{P}(B_k) \quad \text{und} \quad \mathbb{P}\left(\bigcap_{k=2}^n B_k \cap B_1^{\complement}\right) = \prod_{k=2}^n \mathbb{P}(B_k)\cdot B_1^{\complement}$$

Addition ergibt

$$\prod_{k=1}^n \mathbb{P}(B_k) + \prod_{k=2}^n \mathbb{P}(B_k)\cdot B_1^{\complement} = \prod_{k=2}^n \mathbb{P}(B_k)\left[\mathbb{P}(B_1)+\mathbb{P}(B_1^{\complement})\right] = \prod_{k=2}^n \mathbb{P}(B_k).$$

Aufgrund von

$$\bigcap_{k=2}^n B_k = \bigcap_{k=2}^n B_k \cap \left(B_1 \cup B_1^{\complement}\right) = \left(\bigcap_{k=1}^n B_k\right) \cup \left(\bigcap_{k=2}^n B_k \cap B_1^{\complement}\right), \qquad (5)$$

wobei die letzte Vereinigung aufgrund von $B_1 \cap B_1^{\complement}$ disjunkt ist, folgt

$$\mathbb{P}\left(\bigcap_{k=2}^n B_j\right) = \prod_{k=2}^n \mathbb{P}(B_k).$$

Somit erhält man die Produktformel für Durchschnitte von $n-1$ Mengen und iterativ für $(k)_{k=1}^{n-2}$ Mengen. □

Beispiel 2.2.3 („Kopf“ im k-ten von n Würfen)

Betrachte beim n-maligen Wurf einer fairen Münze die Familie der Ereignisse $(A_k :=$ „k-ter Wurf Kopf“$)_{k=1}^n$. Sie ist unabhängig, denn für jede Wahl $B_k \in \{A_k, A_k^{\complement}\}$

$$\mathbb{P}\left(\bigcap_{k=1}^n B_k\right) = 2^{-n} = \prod_{k=1}^n \mathbb{P}(B_k). \qquad \Diamond$$

Der n-malige Münzwurf ist ein Spezialfall für die unabhängige Hintereinanderausführung von Teilexperimenten, welche im nächsten Kapitel betrachtet werden.

2.3 Produktwahrscheinlichkeitsräume

Definition 2.3.1 (Produktraum und -maß)

Seien $((\Omega_k, \mathbb{P}_k))_{k=1}^n$ diskrete Wahrscheinlichkeitsräume,

$$\Omega := \prod_{k=1}^n \Omega_k \text{ und } \mathbb{P} : \mathcal{P}(\Omega) \to [0,1],\ A \mapsto \sum_{\substack{\omega := \{\omega_1, \ldots, \omega_n\} \\ \omega \in A}} \prod_{k=1}^n \mathbb{P}_k(\{\omega_k\})$$

Dann heißt $(\Omega, \mathbb{P})$ Produktraum und $\mathbb{P} := \bigotimes_{k=1}^n \mathbb{P}_k$ Produktmaß. Produktmaß

Korollar 2.3.2 (Zylindermengen)

Für Zylindermengen $A := \prod_{k=1}^n A_k \in \mathcal{P}(\Omega)$ mit $A_k \in \Omega_k$ gilt

$$\begin{aligned}\mathbb{P}(A) &= \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega_1 \in A_1} \cdots \sum_{\omega_n \in A_n} \prod_{k=1}^n \mathbb{P}_k(\{\omega_k\}) \\ &= \sum_{\omega_1 \in A_1} \cdots \sum_{\omega_{n-1} \in A_{n-1}} \prod_{k=1}^{n-1} \mathbb{P}_k(\{\omega_k\}) \mathbb{P}_n(A_n) = \ldots = \prod_{k=1}^n \mathbb{P}_k(A_k).\end{aligned}$$

Beispiel 2.3.3 (n-maliger Münzwurf als Produktraum)

Der n-malige Münzwurf $(\Omega, \mathbb{P})$ ist der Produktraum der Wahrscheinlichkeitsräume $\Omega_i := \{0,1\}$ mit $\mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = \frac{1}{2}$ für $i \in \{1, \ldots, n\}$. ◊ Produktraum

Eine Verallgemeinerung ist das

Definition 2.3.4 (Bernoulli-Experiment)

Seien $p \in [0,1]$, $n \in \mathbb{N}$ und $\Omega := \{0,1\}$. Ein durch $\Omega := \prod_{i=1}^n \Omega_i$ und das Produkt der Wahrscheinlichkeitsmaße $\mathbb{P}_i(\{1\}) = p = 1 - \mathbb{P}_i(\{0\})$ für $i \in \{1, \ldots, n\}$ über Ω beschriebene Zufallsexperiment heißt Bernoulli-Experiment der Länge n mit Erfolgswahrscheinlichkeit p. Bernoulli-Experiment

Lemma 2.3.5 ($\prod_{k=1}^n$ Laplace = Laplace)

Endliche Produkte Laplace-Wahrscheinlichkeitsräume sind Laplace-Wahrscheinlichkeitsräume.

Beweis. Seien $(\Omega_k)_{k=1}^n$ eine Familie nichtleere endlicher Mengen und $\mathbb{P}_k$ die Gleichverteilung auf Ω_k für $k \in \{1, \ldots, n\}$. Dann gilt für das Produktmaß

$$\mathbb{P}(\{\omega_1, \ldots, \omega_n\}) = \prod_{k=1}^n \mathbb{P}_k(\{\omega_k\}) = \prod_{k=1}^n \frac{1}{|\Omega_k|} = \frac{1}{|\prod_{k=1}^n \Omega_k|}.$$

Also ist das Produktmaß $\mathbb{P} = \bigotimes_{k=1}^n \mathbb{P}_k$ die Gleichverteilung auf $\prod_{k=1}^n \Omega_k$. □

Der folgende Satz formalisiert die Intuition, dass der Produktraum die unabhängige Hintereinanderausführung von Zufallsexperimenten entspricht.

SATZ 2.3.1

Seien $(A_i \in \mathcal{P}(\Omega_i))_{i=1}^n$ Ereignisse. Die Ereignisse im Produktraum, welche nur von Ereignis A_i im i-ten Zufallsexperiment abhängen

$$A_i^{(i)} := \{\omega = (\omega_1, \dots, \omega_n) \in \Omega : \omega_i \in A_i\} = \prod_{k=1}^{i-1} \Omega_k \times A_i \times \prod_{k=i+1}^{n} \Omega_k$$

sind unabhängig.

Beweis. Aufgrund von $(A_i^{(i)})^{\complement} = \prod_{k=1}^{i-1} \Omega_k \times A_i^{\complement} \times \prod_{k=i+1}^{n} \Omega_k$ genügt es für die Anwendung von Satz 2.2.1

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k^{(k)}\right) = \prod_{k=1}^{n} A_k^{(k)}$$

für beliebige $A_k \in \mathcal{P}(\Omega_k)$ zu zeigen. Aus Korollar 2.3.2 ($\star$) folgt

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{k=1}^{n} A_k^{(k)}\right) &\overset{\text{Def.}}{=} \mathbb{P}\left(\bigcap_{k=1}^{n} \prod_{j=1}^{i-1} \Omega_j \times A_k \times \prod_{j=i+1}^{n} \Omega_j\right) = \mathbb{P}\left(\prod_{k=1}^{n} A_k\right) \\
&\overset{(\star)}{=} \prod_{k=1}^{n} \mathbb{P}_k(A_k) = \prod_{k=1}^{n} \mathbb{P}\left(\prod_{j=1}^{i-1} \Omega_j \times A_k \times \prod_{j=i+1}^{n} \Omega_j\right) \\
&= \prod_{k=1}^{n} \mathbb{P}\left(A_k^{(k)}\right).
\end{aligned}$$

□

Bemerkung 2.3.6 (kontinuierliche Produkträume)

Analog kann man auf Familien von beliebigen Wahrscheinlichkeitsräumen $(\Omega_k, \mathcal{A}_k, \mathbb{P}_k)_{k=1}^n$ das Produkt $(\Omega, \mathcal{A}, \mathbb{P}) := (\prod_{k=1}^n \Omega_k, \bigotimes_{k=1}^n \mathcal{A}_i, \bigotimes_{k=1}^n \mathbb{P}_k)$ definieren. Sei $A := \prod_{k=1}^n A_k$ mit $A_k \in \mathcal{A}_k$ ein Zylindermenge. Dann ist $\mathcal{A}$ die kleinste σ-Algebra, welche von allen Zylindermengen erzeugt wird und $\mathbb{P}$ das eindeutig bestimmte Wahrscheinlichkeitsmaß ist, für das $\mathbb{P}(\prod_{k=1}^n A_k) = \prod_{k=1}^n \mathbb{P}(A_k)$ gilt.

Lemma 2.3.7 (Produktmaße und -dichten)

Sind $(f_i : \mathbb{R} \to [0, \infty))_{i=1}^n$ *Dichten auf* $\mathbb{R}$ *und* $\mathbb{P}_i(A) := \int_A f_i(x)\,\mathrm{d}x$ *für* $A \in \mathcal{B}(\mathbb{R})$ *die zugehörigen Wahrscheinlichkeitsmaße auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, *so ist*

$$f : \mathbb{R}^n \to \mathbb{R},\ (x_1, \dots, x_n) \mapsto \prod_{k=1}^{n} f_k(x_k)$$

eine Dichte auf $\mathbb{R}^n$ *und das zugehörige Wahrscheinlichkeitsmaß*

$$\mathbb{P} : \mathcal{B}(\mathbb{R}^n) \to [0,1],\ A \mapsto \int_A f(x)\,\mathrm{d}x$$

genau das Produktmaß von $(P_i)_{i=1}^n$.

Beweis. Für Zylindermengen $A := \prod_{k=1}^n A_k$ gilt

$$\mathbb{P}(A) = \int\limits_{\prod_{k=1}^n A_k} f(x)\,\mathrm{d}x = \int\limits_{A_1} f_1(x_1)\,\mathrm{d}x_1 \cdots \int\limits_{A_n} f_n(x_n)\,\mathrm{d}x_n = \prod_{k=1}^n \mathbb{P}_k(A_k)$$

und das beendet den Beweis?? **WARUM GENÜGT ES, FÜR DEN ZYLINDER-FALL ZU ZEIGEN??** □

Bemerkung 2.3.8 (diskreter Fall)

Aus dem Lemma folgt analog für den diskreten Fall, dass für endliche BOREL-Mengen $(\Omega_i)_{i=1}^n \subset \mathbb{R}$ und die Gleichverteilung $\mathbb{P}_i(A_i) = \frac{|A_i|}{|\Omega_i|}$ auf Ω_i das zugehörige Produktmaß $\mathbb{P} = \bigotimes_{k=1}^n \mathbb{P}_k$ die Gleichverteilung auf $\prod_{k=1}^n \Omega_i$ ist.

3 Zufallsvariablen und Verteilungen

3.1 Grundlegende Konzepte

Anstatt einzelne Ergebnisse $\omega \in \Omega$ ist man häufig nur am Wert $X(\omega)$ einer Messgröße X interessiert, z.B. Temperatur oder Aktienkurs. Messgröße

Seien im Folgenden $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(E, \mathcal{E})$ ein Messraum.

DEFINITION 3.1.1 (ZUFALLSVARIABLE(VEKTOR), REALISIERUNG)

- Eine Abbildung $X : \Omega \to E$ heißt Zufallsvariable, wenn Zufallsvariable

$$\{X \in B\} := X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \tag{6}$$

  für alle $B \in \mathcal{E}$ gilt.

- Der Wert $X(\omega)$ heißt Realisierung von X zum Ergebnis ω. Realisierung
- Ist $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so heißt X reellwertig und für $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ d-dimensionaler Zufallsvektor. Zufallsvektor

Bermerkung 3.1.2 (Notation) Es hat sich durchgesetzt, $\mathbb{P}(X \in B)$ anstatt $\mathbb{P}(\{X \in B\})$, $\mathbb{P}(X \in A, Y \in B)$ anstatt $\mathbb{P}(\{X \in A\} \cap \{Y \in B\})$ usw. zu schreiben.

Lemma 3.1.3 (Zufallsvariablen)

- *Ist Ω diskret, so ist jede Abbildung $X : \Omega : E$ eine Zufallsvariable.*
- *Für jedes $A \in \mathcal{A}$ definiert die Indikatorfunktion $\mathbb{1}_{\mathcal{A}}$ eine Zufallsvariable.* *Indikatorfunktion*
- *Sind $X_1, \dots, X_n$ reellwertige Zufallsvariablen und ist $f : \mathbb{R}^n \to \mathbb{R}$ BOREL-messbar, so ist auch $f(X_1, \dots, X_n)$ wieder eine Zufallsvariable. Insbesondere sind Summen, Produkte etc. reellwertiger Zufallsvariablen reellwertige Zufallsvariablen.*

DEFINITION 3.1.4 (VERTEILUNG)

Sei $X : \Omega \to E$ eine Zufallsvariable. Dann nennt man

$$\mu := \mathbb{P} \circ X^{-1} = \mathbb{P}_X : \mathcal{E} \to \mathbb{R}, \ B \mapsto \mathbb{P}(X \in B)$$

Verteilung von X und schreibt $X \sim \mu$. Verteilung

Satz 3.1.1: $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ Wahrscheinlichkeitsraum

Die Verteilung (einer Zufallsvariable) ist ein Wahrscheinlichkeitsmaß.

Beweis. Da X ein Zufallsvariable ist, gilt $\{X \in B\} \in \mathcal{A}$ für alle $B \in \mathcal{E}$, somit ist $\mathbb{P}_X$ wohldefiniert. Weil $\mathbb{P}$ nichtnegativ und normiert ist, ist es auch $\mathbb{P}_X$. Es bleibt nur noch die σ-Additivität zu zeigen. Sei $(B_n)_{n\in\mathbb{N}} \subset \mathcal{E}$ eine paarweise disjunkte Folge, dann ist die Urbildfolge $(\{X \in B_n\})_{n\in\mathbb{N}} \subset \mathcal{A}$ auch paarweise disjunkt. Aus der σ-Additivität von $\mathbb{P}$ (‡) und da Urbilder auch unter Vereinigungen invariant ($\star$) sind, gilt

$$\mathbb{P}_X\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \mathbb{P}\left(X \in \bigcup_{n\in\mathbb{N}} B_n\right) \overset{(\star)}{=} \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} \{X \in B_n\}\right)$$
$$\overset{(\ddagger)}{=} \sum_{n\in\mathbb{N}} \mathbb{P}(X \in B_n) = \sum_{n\in\mathbb{N}} \mathbb{P}_X(B_n). \qquad \square$$

Diskrete Zufallsvariablen und Verteilungen

Die Verteilung einer diskreten Zufallsvariablen X wird vollständig durch die zugehörige Zähldichte

$$p_X : E \to [0,1],\ b \mapsto \mathbb{P}(X = b)$$

beschrieben. Dies gilt auch, wenn nur das Bild $X(\Omega)$ diskret ist.

Beispiel 3.1.5 (Zähldichte beim zweimaligen Würfeln)
Beim zweimaligen fairen Würfeln sei $X : \Omega \to \{2, \dots, 12\}$ die Augensumme. Dann gilt

$$p_X : \{2, \dots, 12\} \to [0,1],\ b \mapsto \mathbb{P}(\{(k,\ell) \in \Omega : k + \ell = b\}). \qquad \diamond$$

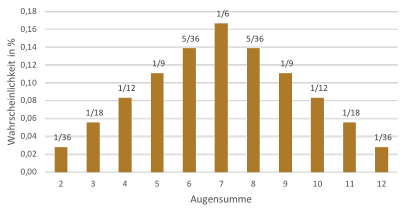


Abb. 15: Ein Stabdiagram der Wahrscheinlichkeiten.

Reellwertige Zufallsvariablen und Verteilungen

Definition 3.1.6 (Verteilungsfunktion (CDF))
Sei X eine reellwertige Zufallsvariable. Dann heißt

$$F_X : \mathbb{R} \to [0,1],\ x \mapsto \mathbb{P}(X \leqslant x) = \mathbb{P}_X((-\infty, x])$$

Verteilungsfunktion von X.

engl.: cummulative distribution function

Verteilungsfunktion

Definition 3.1.7 (entartete Zufallsvariable)

Eine Zufallsvariable, welche nur einen Wert $c \in \mathbb{R}$ annimmt, heißt deterministisch oder entartet, da ihr Wert vom Ausgang des Zufallsexperiments unabhängig ist. Es gilt $F_X(x) = \mathbb{1}_{x \geqslant c}(x)$.

deterministisch

Lemma 3.1.8 (Eigenschaften der diskreten Verteilungsfunktion)

Sei X eine diskrete Zufallsvariable und F_X ihre Verteilungsfunktion. Dann gilt

1. *F_X ist eine stückweise konstante monoton wachsende Treppenfunktion mit Werte zwischen Null und Eins.*
2. *Ihre Sprungstellen sind genau diejenigen Werte x, sodass $\mathbb{P}(X = x) > 0$ gilt. Die Höhe der Sprungstellen ist $\mathbb{P}(X = x)$.*

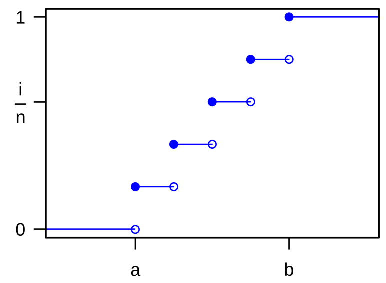


Abb. 16: Verteilungsfunktion der diskreten Gleichverteilung. [Quelle: Wiki]

Beweis. Das folgt aus der Darstellung

$$F_X(x) = \sum_{y \leqslant x} \mathbb{P}_X(\{y\}) = \sum_{y \leqslant x} \mathbb{P}(X = y). \qquad \square$$

Beispiel 3.1.9 (Verteilungsfunktion der Gleichverteilung)

Ist X gleichverteilt auf $[a, b]$, schreibt man $X \sim \mathcal{U}([a, b])$ und es gilt für $x \in \mathbb{R}$ nach Beispiel 1.3.5

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \frac{|(-\infty, x] \cap [a, b]|}{|[a, b]|} = \frac{x - a}{b - a} \cdot \mathbb{1}_{[a,b)} + \mathbb{1}_{x \geqslant b}. \quad \Diamond$$

Satz 3.1.2: Eigenschaften der kontinuierlichen Verteilungsfunktion

Sei X eine reellwertige Zufallsvariable. Dann gilt

1. F_X ist monoton wachsend.
2. F_X ist rechtsstetig.
3. $F \in [0, 1]$, $F(-\infty+) = 0$, $F(\infty-) = 1$.

Bemerkung 3.1.10 (Verteilungsfunktion $\iff$ W-Maß)

Alternativ kann man eine Verteilungsfunktion auch genau über die oben genannten Eigenschaften definieren und dann zeigen, dass zu jeder Verteilungsfunktion F genau ein Wahrscheinlichkeitsmaß $\mathbb{P}(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ existiert, sodass $\mathbb{P}((-\infty, x]) = F(x)$ gilt.

Ebenfalls kann man zu jeder Verteilungsfunktion F einen Wahrscheinlichkeitsraum und eine reellwertige Zufallsvariable konstruieren, deren Verteilungsfunktion mit F übereinstimmt.

Daraus folgt

Lemma 3.1.11

Eine Verteilung ist durch die Angabe ihrer Verteilungsfunktion eindeutig bestimmt: Sind X und Y zwei reellwertige Zufallsvariablen mit $F_X = F_Y$, so folgt $P_X = P_Y$.

also insbesondere: gilt $F_\mu = F_\nu$ für zwei Wahrscheinlichkeitsmaße μ und ν auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ so folgt $\mu \equiv \nu$.

Lemma 3.1.12

Eine Verteilungsfunktion F_X von X ist genau dann stetig in x wenn $\mathbb{P}(X = x) = 0$ gilt.

Beweis. Leicht, vgl. HA V Aufg 4 $\Longleftarrow$. □

Ist $f : \mathbb{R} \to [0, \infty)$ eine Dichte, so ist die Funktion

$$F : \mathbb{R} \to [0, 1], \; x \mapsto \int_{-\infty}^{x} f(y) \, \mathrm{d}y \tag{7}$$

eine Verteilungsfunktion. Das motiviert die

Definition 3.1.13 (absolutstetige Verteilungsfkt.)

Jede Verteilungsfunktion F, für die eine Dichte existiert, sodass (7) gilt, heißt absolutstetige Verteilungsfunktion.

absolutstetige Verteilungsfunktion

Ist X eine Zufallsvariable mit absolutstetiger Verteilungsfunktion F_X und zugehöriger Dichte f_X, so folgt aus (7) für alle $B \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(X \in B) = \int_B f_X(y) \, \mathrm{d}y. \tag{8}$$

3.2 Diskrete Verteilungen

Bernoulli-Verteilung

Die Verteilung der Indikatorfunktion $\mathbb{1}_A$ des Ereignisse $A \in \mathcal{A}$ nimmt zwei Werte an. Wir interpretieren das Ereignis $\{\mathbb{1}_A = 1\} = A$ als Erfolg und bezeichnen $p := \mathbb{P}(X = 1) = \mathbb{P}(A)$ als Erfolgswahrscheinlichkeit. Für die Wahrscheinlichkeit des Misserfolgs gilt $\mathbb{P}(X = 0) \overset{1.1.9}{=} 1 - p$.

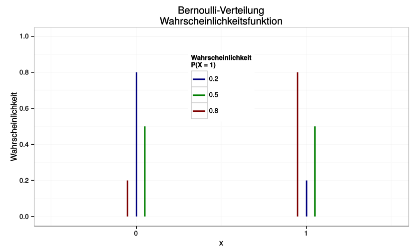


Abb. 17: unnamed figure

Definition 3.2.1 (Bernoulli-Verteilung/Experiment)
Sei $p \in [0,1]$. Das durch die Zähldichte

$$\{0,1\} \to [0,1],\ 1 \mapsto p,\ 0 \mapsto 1-p$$

definierte Wahrscheinlichkeitsmaß auf $\{0,1\}$ heißt Bernoulli-Verteilung zu p und wird mit $\mathrm{Ber}(p)$ bezeichnet. Zufallsexperimente mit nur zwei Ausgängen nennt man Bernoulli-Experimente.

Bernoulli-Verteilung zu p
Bernoulli-Experimente

Beispiel 3.2.2 (Bernoulli-Experimente)
Beim Werfen einer Münze, dem Geschlecht eines Neugeborenen oder dem Ziehen einer Kugel aus einer Urne mit zwei verschiedenen Arten von Kugeln handelt es sich um Bernoulli-Experimente. ◊

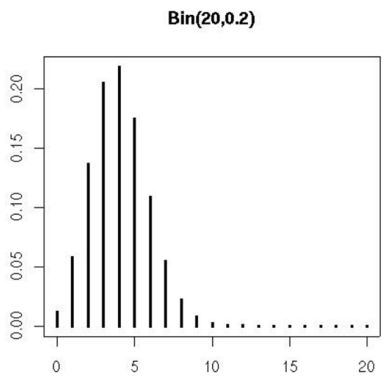


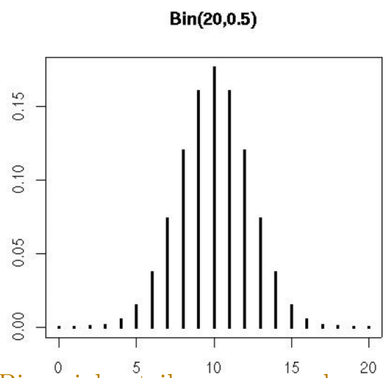
Abb. 18: unnamed figure

Binomialverteilung

Seien $(X_k \sim \mathrm{Ber}(p))_{k=1}^n$ unabhängige Zufallsvariablen und $p \in [0,1]$. Die Zufallsvariable $S_n := \sum_{k=1}^n S_k \in \{0,\dots,n\}$ zählt die Gesamtanzahl der Erfolge. Dann gilt für $k \in \mathbb{N}$

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

da $\binom{n}{k}$ die Anzahl der n-Tupel mit k Einsen und $n-k$ Nullen ist.



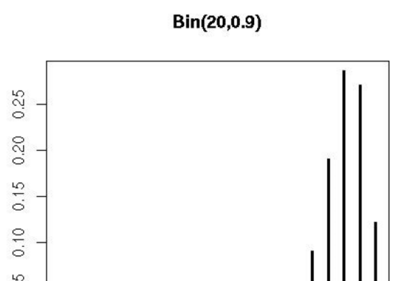
Definition 3.2.3 (Binomialverteilung)
Sei $n \in \mathbb{N}$ und $p \in [0,1]$. Das durch die Zähldichte

$$b(\cdot, n, p) : \{0,\dots,n\} \to [0,1],\ k \mapsto \binom{n}{k} p^k (1-p)^{n-k}$$

definierte Wahrscheinlichkeitsmaß auf $\{0,\dots,n\}$ heißt Binomialverteilung zu n und p und wird mit $\mathrm{Bin}(n,p)$ bezeichnet.

Binomialverteilung zu n und p

Abb. 19: unnamed figure



Geometrische Verteilung

Seien $(X_k \sim \mathrm{Ber}(p))_{k\in\mathbb{N}}$ unabhängige Zufallsvariablen mit $p > 0$. Die Wartezeit auf den ersten Erfolg definieren wir als $T := \min\limits_{X_k=1} k$. Wir erhalten für $k \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}(T = k) &= \mathbb{P}((X_n = 0)_{n=0}^{k-1}, X_k = 1)\\ &= \prod_{n=0}^{k-1} \mathbb{P}(X_n = 0) \cdot \mathbb{P}(X_k = 1) = (1-p)^{k-1}p.\end{aligned}$$

DEFINITION 3.2.4 (GEOMETRISCHE VERTEILUNG)

Sei $p \in (0, 1]$. Das durch die Zähldichte

$$\mathbb{N} \to [0, 1],\ k \mapsto (1-p)^{k-1}p$$

definierte Wahrscheinlichkeitsmaß auf $\mathbb{N}$ heißt geometrische Verteilung zu p und wird mit $\mathrm{Geo}(p)$ bezeichnet.

geometrische Verteilung

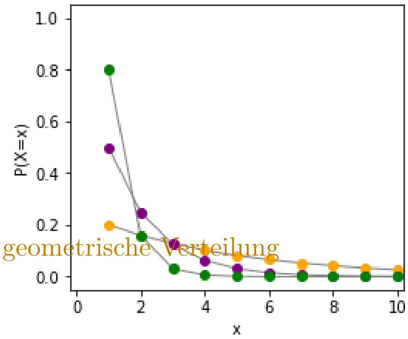


Abb. 21: Die geometrische Verteilung für $p \in \{0.2, 0.5, 0.8\}$.

Ihre wohl wichtigste Eigenschaft ist die

Satz 3.2.1: Gedächtnislosigkeit

① Sei T geometrisch verteilt. Dann gilt für alle $k \geqslant 2$, $n \in \mathbb{N}$

$$\mathbb{P}(T \geqslant n + k - 1 \mid T \geqslant n) = \mathbb{P}(T \geqslant k) \tag{9}$$

② Ist $T : \Omega \to \mathbb{N}$ eine Zufallsvariable mit Eigenschaft (9), so ist sie geometrisch verteilt.

Beweis. ① Sei T geometrisch verteilt. Dann gilt für $m \in \mathbb{N}$

$$\begin{aligned}\mathbb{P}(T \geqslant m) = \sum_{k=m}^{\infty} \mathbb{P}(T = k) &= \sum_{k=m}^{\infty} (1-p)^{k-1}p\\ &= (1-p)^{m-1} \cdot p \cdot \sum_{k=0}^{\infty} (1-p)^k\\ &= (1-p)^{m-1} \cdot p \cdot \frac{1}{1-(1-p)} = (1-p)^{m-1}\end{aligned}$$

und somit

$$\begin{aligned}\mathbb{P}(T \geqslant k - 1 + n \mid T \geqslant n) &= \frac{\mathbb{P}(T \geqslant k - 1 + n)}{\mathbb{P}(T \geqslant n)} = \frac{(1-p)^{k-1+n-1}}{(1-p)^{n-1}}\\ &= (1-p)^{k-1} = \mathbb{P}(T \geqslant k).\end{aligned}$$

② Da $\mathbb{P}(T \geqslant k) = \mathbb{P}(T > k - 1)$ gilt, ist (9) äquivalent zu

$$\mathbb{P}(T \geqslant k + n \mid T > n) = \mathbb{P}(T > k). \tag{10}$$

Setzt man $n = 2$ in (10) ein, so erhält man

$$\mathbb{P}(T \geqslant k+2 \mid T > 2) = \mathbb{P}(T > k) \; \forall k \in \mathbb{N}\,.$$

Nach Definition der bedingten Wahrscheinlichkeit ($\star$) gilt nun

$$\begin{aligned}
\mathbb{P}(T > k+1) = \mathbb{P}(T \geqslant k+2) &\overset{(\star)}{=} \mathbb{P}(T \geqslant k+2 \mid T > 2)\,\underbrace{\mathbb{P}(T > 2)}_{>0} \\
&\overset{(10)}{=} \mathbb{P}(T > k)\,\mathbb{P}(T > 2) = \mathbb{P}(T \geqslant k+1)\,\mathbb{P}(T > 2) \\
&\overset{(\star)}{=} \mathbb{P}(T \geqslant k+1 \mid T > 2)\,\mathbb{P}(T > 2)^2 \\
&\overset{(10)}{=} \mathbb{P}(T \geqslant k-1)\,\mathbb{P}(T > 2)^2.
\end{aligned}$$

Iteration ergibt $\mathbb{P}(T > k+1) = \mathbb{P}(T > 2)^{k+1}$ ($\star$) für alle $k \in \mathbb{N}_0$. Mit $q := \mathbb{P}(T > 2)$ folgt

$$\mathbb{P}(T = k) = \mathbb{P}(T > k-1) - \mathbb{P}(T > k) = q^{k-1} - q^k = q^{k-1}(1-q).$$

Nun bleibt noch $q < 1$ zu zeigen. Wir nehmen an, dass $q = 1$ gilt. Durch Umstellen folgt mit ($\star$) $\mathbb{P}(T > k+1) = 1$ für alle $k \in \mathbb{N}_0$ und somit $\mathbb{P}(T \leqslant k+1) = 0$ bzw. $\mathbb{P}(T \leqslant n) = 0$ für alle $n \in \mathbb{N}$. Mit der Stetigkeit von Maßen folgt

$$\mathbb{P}(T \in \mathbb{N}) = \lim_{n\to\infty} \mathbb{P}(T \leqslant n) = 0,$$

was ein Widerspruch zu $\mathbb{P}(T \in \mathbb{N}) = 1$ darstellt. $\square$

Beispiel 3.2.5

Seien $W := (\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X, Y : W \to (\mathbb{N}, \mathcal{P}(\mathbb{N})$ zwei unabhängige Zufallsvariablen mit $X, Y \sim \mathrm{Geo}(p)$ für $p \in (0,1]$. Dann gilt $\min(X,Y) \sim \mathrm{Geo}(1-q^2)$ für $q := p - 1$. (vgl. A.2.5). Bestimmen Sie die Verteilung von $Z := \min(X,Y)$. $\lozenge$

Negative Binomialverteilung

Als Verallgemeinerung der geometrischen Verteilung können wir für jedes $n \geqslant 1$ die Wartezeit auf den n-ten Erfolg

$$T_n := \min_{\sum_{j=1}^k X_j = n} k \in \mathbb{N}_{\geqslant n}$$

betrachten. Für die Verteilung ergibt sich analog zu den obigen Überlegungen für $k \geqslant n$

$$\mathbb{P}(T_n = k) = \binom{k-1}{n-1}(1-p)^{k-n}p^n$$

Definition 3.2.6 (Negative Binomialverteilung)

Sei $p \in (0,1]$ und $n \in \mathbb{N}$. Das durch die Zähldichte

$$\mathbb{N}_{\geqslant n} \to [0,1],\ k \mapsto \binom{k-1}{n-1}(1-p)^{k-n}p^n$$

definierte Wahrscheinlichkeitsmaß auf $\mathbb{N}_{\geqslant n}$ heißt negative Binomialverteilung zu den Parametern n und p.

negative Binomialverteilung

Poisson-Verteilung

Definition 3.2.7 (Poisson-Verteilung)

Sei $\lambda > 0$. Das durch die Zähldichte

$$\pi_\lambda : \mathbb{N}_0 \to [0,1],\ k \mapsto e^{-\lambda}\frac{\lambda^k}{k!}$$

definierte Wahrscheinlichkeitsmaß auf $\mathbb{N}_0$ heißt Poisson-Verteilung zu λ und wird mit Poiss(λ) bezeichnet.

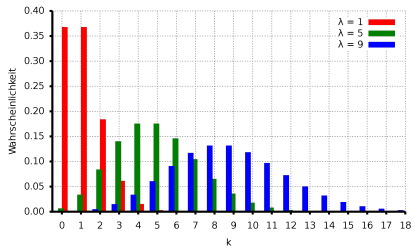
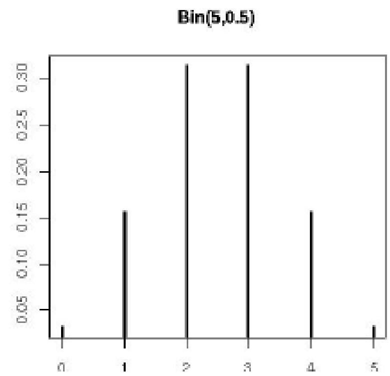


Abb. 22: von Wiki

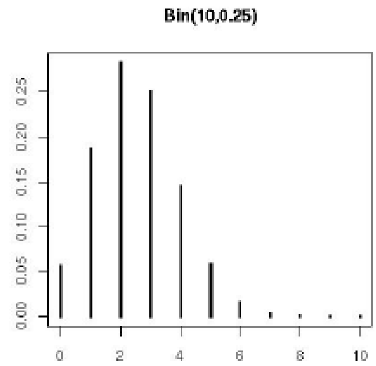
Der nächste Satz zeigt, dass die Poisson-Verteilung sich als Näherung der Binomialverteilung für große n und kleiner p (genauer: kleine np^2) eignet. Eine näherungsweise Berechnung von Wahrscheinlichkeiten gewisser Ereignisse mit Hilfe einer Poisson-Verteilung ist immer dann gerechtfertigt, wenn es sich um seltene Ereignisse handelt.



Satz 3.2.2: Poissonscher Grenzwertsatz

Sei $(p_n)_{n\in\mathbb{N}} \subset [0,1]$ eine Folge von Erfolgsparametern mit $np_n \xrightarrow{n\to\infty} \lambda > 0$. Dann gilt für alle $k \in \mathbb{N}_0$

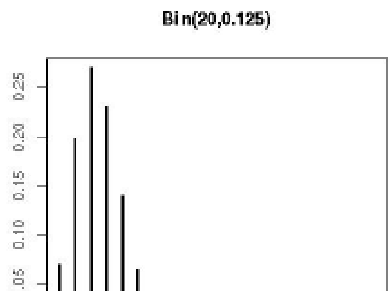
$$b(k;n,p_n) \xrightarrow{n\to\infty} \pi_\lambda.$$



Beweis. Es gilt

$$\begin{aligned} b(k;n,p_n) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!}\cdot \underbrace{\frac{n}{n}}_{=1} \cdot \underbrace{\frac{n-1}{n}}_{\to 1} \cdot \ldots \cdot \underbrace{\frac{n-k+1}{n}}_{\to 1} \underbrace{(np_n)^k}_{\to \lambda^k} \underbrace{\left(1-\frac{np_n}{n}\right)^{n-k}}_{\sim\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}} \\ &\xrightarrow{n\to\infty} \frac{\lambda^k}{k!}e^{-\lambda} = \pi_\lambda(k). \end{aligned}$$

□



Hypergeometrische Verteilung

Sei ein Grundmenge Ω mit $|\Omega| = N$, von denen K Elemente die Eigenschaft E besitzen. Es werde n-mal ohne Zurücklegen gezogen. Sei X die Zufallsvariable, welche die Anzahl der gezogenen Elemente mit Eigenschaft E angibt.

Beispiel 3.2.8 (Hochrechnungen)
Wir wollen die Anzahl der Fische in einem See, N, schätzen. Wir markieren zunächst K Fische rot. Danach ziehe man $n \leqslant N$ Fische. Dann ist X die Anzahl der markierten Fische aus dieser Stichprobe und $\tilde{N} := \frac{nK}{X}$ ist eine natürliche Schätzung für N, da $\frac{X}{n} \sim \frac{K}{N}$ gilt.

<u>Fall 1: $\frac{n}{N}$ ist klein.</u> Dann gibt es keinen großen Unterschied zwischen Ziehen mit und ohne Zurücklegen. Deswegen können wir die Verteilung von X durch $\operatorname{Bin}\left(n, \frac{K}{N}\right)$ approximieren, also $\mathbb{P}(X = k) \approx b\left(k; n, \frac{K}{N}\right)$. Dies sieht man so: Für $N, K_N \to \infty$ mit $p_N := \frac{K_N}{N} \xrightarrow{N\to\infty} p$ gilt mit $K = K_N$

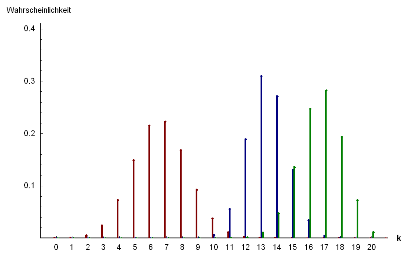


Abb. 24: Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung für $k = 20$; $N = 20, n = 30$ (blau), $N = 50, n = 60$ (grün) und $N = 20, n = 60$ (rot)

$$\begin{aligned}
\mathbb{P}(X = k) &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \binom{n}{k} \frac{K!}{(K-k)!} \frac{(N-K)!}{((N-K)-(n-k))!} \frac{(N-n)!}{N!} \\
&= \binom{n}{k} \prod_{\ell=N-K-n-k+1}^{K} \frac{\ell}{N} \cdot \prod_{\ell=N-n+1}^{N} \frac{N}{\ell} \xrightarrow{N\to\infty} \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

<u>Fall 2: $\frac{n}{N}$ ist groß.</u> Dann muss man die Verteilung von X exakt berechnen: für $k \in \{0, \dots, n\}$ gilt die obige Formel. $\Diamond$

Definition 3.2.9 (Hypergeometrische Verteilung)
Seien $K, n \leqslant N$. Das durch die Zähldichte

$$\operatorname{Hyp}(\cdot; n, N, k) : \{0, \dots, n\} \to [0,1], \quad k \mapsto \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

definierte Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ heißt Hypergeometrische Verteilung und wird mit $\operatorname{Hyp}(n, N, k)$ bezeichnet.

Hypergeometrische Verteilung

Es ergibt sich die folgende Übersicht diskreter Verteilungen:

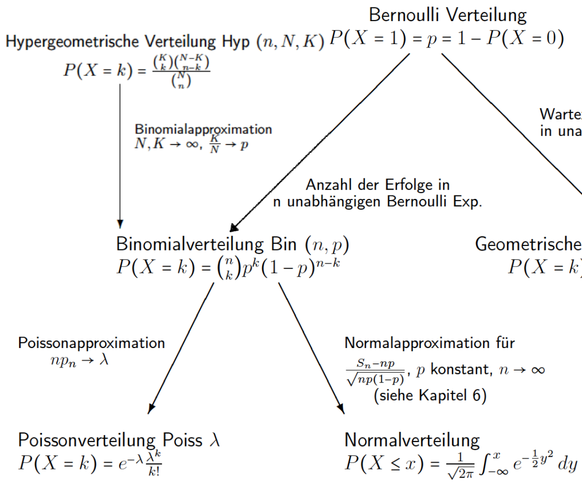


Abb. 25: Übersicht über die diskreten Verteilungsfunktionen und ihre Approximationsmöglichkeiten.

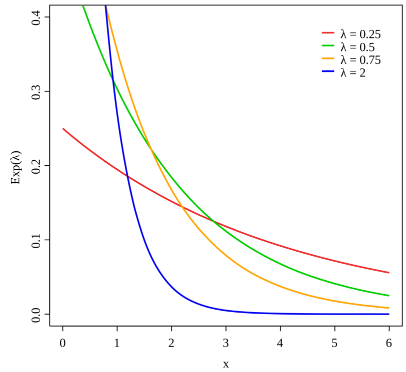
3.3 Stetige Verteilung

Definition 3.3.1 (Gleichverteilung)

Für $a < b$ heißt eine Zufallsvariable mit Dichte $f := \frac{\mathbb{1}_{[a,b]}}{b-a}$ und Verteilungsfunktion ist $\frac{x-a}{b-a} \cdot \mathbb{1}_{[a,b)} + \mathbb{1}_{x \geqslant b}$ gleichverteilt auf $[a, b]$.

Definition 3.3.2 (Exponentialverteilung)

Für $\lambda > 0$ heißt die zu der Dichte $f_\lambda(x) := \lambda e^{-\lambda x} \mathbb{1}_{x \geqslant 0}$ zugehörende Verteilung, $F(x) := 1 - e^{-\lambda x} \mathbb{1}_{x \geqslant 0}$, Exponentialverteilung zum Parameter λ und wird $\operatorname{Exp}(\lambda)$ bezeichnet.



Exponentialverteilung

Abb. 26: Die Dichte der Exponentialverteilung [Quelle: Wiki]

Die Exponentialverteilung ist das stetige Analogon der geometrischen Verteilung, dementsprechend verwendet man die Exponentialverteilung zur Modellierung stetig verteilter Wartezeiten.

Beispiel 3.3.3 (Vielfache / Summen der Exponentialverteilung)
Sei $X \sim \exp(\vartheta)$ für $\vartheta > 0$. Für $a > 0$ gilt $aX \sim \exp\left(\frac{\vartheta}{a}\right)$ (vgl. A.2.2). $\diamondsuit$

Satz 3.3.1: Exponentialverteilung gedächtnislos

1. Sei T exponentialverteilt. Dann gilt für alle $s > 0$ und $t > 0$
$$\mathbb{P}(T > s + t \mid T > t) = \mathbb{P}(T > s) \tag{11}$$
2. Ist T eine nichtnegative Zufallsvariable mit den Eigenschaften (11) und $\mathbb{P}(T > 0) > 0$, so ist T exponentialverteilt.

Beweis. 1. Sei T exponentialverteilt mit Parameter $\lambda > 0$ und F die zugehörige Verteilungsfunktion. Dann gilt $\mathbb{P}(T > t) \overset{1.1.9}{=} 1 - F(t) = e^{-\lambda t}$ für $t > 0$ und somit
$$\mathbb{P}(T > s{+}t \mid T > t) = \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(T > s).$$

2. Sei T eine nichtnegative Zufallsvariable mit der Eigenschaft (11), F ihr Verteilungsfunktion und $\overline{F} := 1 - F$. Dann folgt für $t, s > 0$
$$\begin{aligned}\overline{F}(t + s) &= \mathbb{P}(T > s + t) = \mathbb{P}(T > s + t \mid T > t)\, \mathbb{P}(T > t) \\ &\overset{(11)}{=} \mathbb{P}(T > s)\, \mathbb{P}(T > t) = \overline{F}(t)\overline{F}(s).\end{aligned} \tag{12}$$

Aus Analysis I folgt, dass jede strikt positive rechtsstetige Funktion, welche der Funktionalgleichung (12) genügt, von der Form $\overline{F}(t) = e^{-\lambda t}$ für ein $\lambda \in \mathbb{R}$ ist. ($\star$) Da $\overline{F}(t) \xrightarrow{t \to \infty} 0$ gilt, folgt $\lambda > 0$. Somit ist T exponentialverteilt.

($\star$): Induktiv folgt aus (12) $\overline{F}(1/n) = \overline{F}(1)^n$, also $F(1) = \overline{F}(1/n)^{1/n}$. Da $\overline{F}(1/n) > 0$ und $\overline{F}(1) \in (0,1]$, existiert ein $\lambda \in \mathbb{R}$ mit $\overline{F}(1) = e^{-\lambda}$. Ähnluch folgt $\overline{F}(1) = \overline{F}(p/q)^{p/q}$ für alle $p, q \in \mathbb{Z}$ und somit $\overline{F}(t) = e^{-\lambda t}$ für $t \in \mathbb{Q}$. Aufgrund der Rechtsstetigkeit folgt $\overline{F}(t) = e^{-\lambda t}$ für $t \in \mathbb{R}$. □

Analog zum diskreten Fall können wir auch die Familie der Exponentialverteilungen in eine Familie allgemeinerer Wartezeitenverteilungen einbetten:

Definition 3.3.4 (Γ-Verteilung)
Die zu der Dichte $f(x) := \frac{\lambda^\vartheta}{\Gamma(\vartheta)} x^{\vartheta-1} e^{-\lambda x} \mathbb{1}_{x>0}(x)$ für $\lambda, \vartheta > 0$ gehörende Verteilung heißt Γ-Verteilung $\Gamma_{\vartheta,\lambda}$.

Γ-Verteilung

Hierbei ist $\Gamma(\vartheta) := \int_0^\infty x^{\vartheta-1} e^{-x}\, dx$ für $\vartheta > 0$ die Γ-Funktion. Insbesondere gilt $\mathrm{Exp}(\lambda) = \Gamma_{1,\lambda}$.

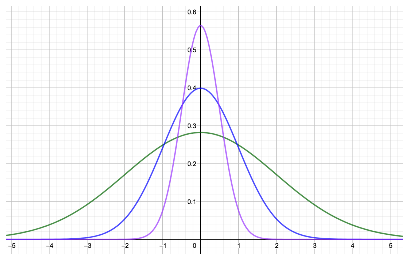


Abb. 27: Die bei Null zentrierte ($m = 0$) Normalverteilung für $\sigma \in \{0.5, 1, 2\}$.

Die Summe n unabhängig $\mathrm{Exp}(\lambda)$-verteilter Wartezeiten ist $\Gamma_{n,\lambda}$-verteilt (siehe Beispiel 3.30) **TODO**.

Definition 3.3.5 (Normalverteilung)
Die für $m \in \mathbb{R}$ und $\sigma > 0$ zu der $f_{m,\sigma^2}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right)$ gehörenden Verteilung heißt Normalverteilung $\mathcal{N}(m, \sigma^2)$ mit Mittel m und Varianz σ^2.

Normalverteilung

Varianz

Bermerkung 3.3.6 Für $m = 0$ und σ^2 spricht man von der Standardnormalverteilung. Es gilt $X \sim \mathcal{N}(m, \sigma^2) \implies \frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$.

Beispiel 3.3.7 ($(aX + b) \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$)
Seien $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma > 0$, $a \in \mathbb{R}\backslash\{0\}$ und $b \in \mathbb{R}$. Dann ist $(aX + b) \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$ (vgl. A.2.3). ◊

Beispiel 3.3.8 (Summe von Normalverteilungen)
Seien $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ unabhängige Zufallsvariablen mit $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ mit $m_i \in \mathbb{R}$ und $\sigma_i > 0$ für $i \in \{1,2\}$. Dann gilt $Z := X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. (vgl. A.2.4) ◊

3.4 Gemeinsame Verteilung und Unabhängigkeit von Zufallsvariablen.

Sei im Folgenden immer $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Definition 3.4.1 (Unabhängige Zufallsvariablen)

Die $((E_k, \mathcal{E}_k))_{k=1}^n$-wertige Zufallsvariablen $(X_k)_{k=1}^n$ heißen (stochastisch) unabhängig, falls für alle Teilmengen $B_i \in \mathcal{E}_i$ die Ereignisse $(\{X_k \in B_k\})_{k=1}^n$ (stochastisch) unabhängig sind.

(stochastisch) unabhängig

Korollar 3.4.2 (Produktformel)

Die obigen Zufallsvariablen sind (vgl. 2.2.1) genau dann unabhängig, wenn $\mathbb{P}((X_k \in B_k)_{k=1}^n) = \prod_{k=1}^n \mathbb{P}(X_k \in B_k)$ *für alle* $(B_k \in \mathcal{E}_k)_{k=1}^n$ *gilt.*

Lemma 3.4.3 (Unabhängigkeit transformationsinvariant)

Sind $(f_k : (E_k, \mathcal{E}_k) \to (D_k, \mathcal{D}_k))_{k=1}^n$ *messbare Abbildungen, so sind die transformierten Zufallsvariablen* $(f_k(X_k))_{k=1}^n$ *unabhängig.*

Beweis. Aufgrund der Messbarkeit der Transformationen gilt $f^{-1}(B_i) \in \mathcal{E}_i$ für alle $B_i \in \mathcal{D}_i$ für alle $i \in \{1, \dots, n\}$. Aufgrund der Unabhängigkeit von $(X_k)_{k=1}^n$ sind ist auch $(\{X_k \in f_k^{-1}(B_k)\})_{k=1}^n$ stochastisch unabhängig. Aus $\{X_k \in f_k^{-1}(B_k)\} = \{f_k(X_k) \in B_k\}$ folgt die Unabhängigkeit. □

Wir wollen die Unabhängigkeit von Zufallsvariablen mittels ihrer Verteilungen charakterisieren.

Definition 3.4.4 (gemeinsame Verteilung)

Seien $(X_k)_{k=1}^n$ $\left((E_k, \mathcal{E}_k)\right)_{k=1}^n$-wertige Zufallsvariablen und $E := \prod_{k=1}^n$, $\mathcal{E} := \bigotimes_{k=1}^n \mathcal{E}_k$ und

$$\mathbb{X} := (X_k)_{k=1}^n : \Omega \to E,\ \omega \mapsto \left((X_k(\omega))_{k=1}^n\right).$$

Dann heißt die Verteilung

$$\mathbb{P}_{\mathbb{X}} : \mathbb{E} \to E,\ B \mapsto \mathbb{P}(\mathbb{X} \in B) = \mathbb{P}((X_1, \dots, X_n) \in B)$$

gemeinsame Verteilung von $(X_k)_{k=1}^n$.

gemeinsame Verteilung

Bermerkung 3.4.5 Für Zylindermengen $B := \prod_{k=1}^n B_k \in \mathcal{E}$ gilt insbesondere $\mathbb{P}_{\mathbb{X}}(B) = \mathbb{P}\left((X_k \in B_k)_{k=1}^n\right)$.

Gemeinsame Verteilung und Unabhängigkeit diskreter Zufallsvariablen

Seien $(X_k)_{k=1}^n$ E_i-wertige Zufallsvariablen, wobei die E_i höchstens abzählbar sind. Dann ist auch $\mathbb{X} := (X_k)_{k=1}^n$ eine $E := \prod_{k=1}^n E_k$-wertige Zufallsvariable, wobei E höchstens abzählbar ist. Daher ist die gemeinsame Verteilung der $(X_k)_{k=1}^n$ eindeutig durch ihre Zähldichte

$$p_{\mathbb{X}} : E \to [0,1],\ \left((e_k)_{k=1}^n\right) = \mathbb{P}\left((X_k = e_k)_{k=1}^n\right)$$

bestimmt.

Satz 3.4.1: Unabhängig mit gemeins. Verteilung

Unter den obigen Bedingungen sind die folgenden Aussage äquivalent.

1. $(X_k)_{k=1}^n$ sind unabhängig.
2. Es gilt $\mathbb{P}\left((X_k = e_k)_{k=1}^n\right) = \prod_{k=1}^n \mathbb{P}(X_k = e_k)$ für alle $(e_k)_{k=1}^n \in E$.
3. Es gilt $p_{\mathbb{X}}\left((e_k)_{k=1}^n\right) = \prod_{k=1}^n p_{X_k}(e_k)$ für alle $(e_k)_{k=1}^n \in E$.

Beweis. Wir zeigen nur 2 $\implies$ 1. Seien $(B_k \subset E_k)_{k=1}^n$, $B := \prod_{k=1}^n B_k$ und $e := \left((e_k)_{k=1}^n\right)$. Dann gilt

$$\begin{aligned}\mathbb{P}\left((X_k \in B_k)_{k=1}^n\right) &= \sum_{e\in B} \mathbb{P}\left((X_k = e_k)_{k=1}^n\right) = \sum_{e\in B}\prod_{k=1}^n \mathbb{P}(X_k = e_k)\\ &= \prod_{k=1}^n \sum_{e_k\in B_k} \mathbb{P}(X_k = e_k) = \prod_{k=1}^n \mathbb{P}(X_k \in B_k).\end{aligned}$$ □

Beispiel 3.4.6 (Bernoulli-Experimente)

Seien $(A_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Ereignisse mit $\mathbb{P}(A_n) = p$ und $X_n := \mathbb{1}_{A_n}$ der Ausgang des n-ten Versuchs. Dann ist $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger $\{0,1\}$-wertiger Zufallsvariablen, denn für alle $(i_k)_{k=1}^n \in \{0,1\}$ gilt

$$\mathbb{P}\left((X_k = i_k)_{k=1}^n\right) = \mathbb{P}\left(\bigcap_{k=1}^n A_k^{(i_k)}\right) = \prod_{k=1}^n \mathbb{P}(A_k^{(i_k)}) = \prod_{k=1}^n \mathbb{P}(X_k = i_k),$$

wobei $A_j^{(i_j)} = A_j\, \mathbb{1}_{i_j=0} + A_j^{\complement} i_j = 0$ ist. Wir haben benutzt, dass nach Satz 2.2.1 mit $(A_k)_{k=1}^n$ auch die Ereignisse $(A_k^{(i_k)})_{k=1}^n$ unabhängig sind.

Für die Zähldichte der gemeinsamen Verteilung der $\mathbb{X} := (X_k)_{k=1}^n$ gilt

$$p_{\mathbb{X}}\left((i_k)_{k=1}^n\right) = p^{\sum_{j=1}^n i_j}(1-p)^{\sum_{j=1}^n 1-i_j}$$

Sie stimmt also mit dem Produkt der Wahrscheinlichkeitsmaße $(\mathbb{P}_j(1) = p = 1 - \mathbb{P}_j(\{0\}))_{j=1}^n$ (vgl. 2.3.4) überein. ◊

3.5 Erwartungswert und Varianz reellwertiger Zufallsvariablen

Erwartungswert und Varianz sind die beiden wichtigsten Kennzahlen einer Zufallsvariablen.

Erwartungswert diskreter Zufallsvariablen

Definition 3.5.1 (diskreter Erwartungswert)

Sei X eine diskrete Zufallsvariable mit Werten $x_n \in \mathbb{R}$ für $n \in \mathbb{N}$.

- Ist $x_n \geqslant 0$ für alle $n \in \mathbb{N}$, so heißt
$$\mathbb{E}[X] := \sum_{n \in \mathbb{N}_0} x_n \, \mathbb{P}(X = x_n) \in [0, \infty]$$
der Erwartungswert von X.
- Der Erwartungswert von X existiert, wenn $\mathbb{E}[|X|] < \infty$ gilt.

Erwartungswert

Bermerkung 3.5.2 Der Erwartungswert ist das mit den Wahrscheinlichkeiten gewichtete Mittel der Funktionswerte und lässt sich daher als Schwerpunkt von X interpretieren. Der Erwartungswert hängt nur von der Verteilung und nicht der Zufallsvariablen ab.

Beispiel 3.5.3 (Erste Erwartungswerte)

- Für den fairen Münzwurf gilt $\mathbb{E}[X] = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = \frac{1}{2}$.
- Beim Würfeln gilt $\mathbb{E}[X] = \sum_{k=1}^{6} \frac{k}{6} = \frac{7}{2}$.
- Für $A \in \mathcal{A}$ gilt $\mathbb{E}[\mathbb{1}_A] = 0 \cdot \mathbb{P}(A^{\complement}) + 1 \cdot \mathbb{P}(A) = \mathbb{P}(A)$. $\Diamond$

Lemma 3.5.4 (Rechenregeln für Erwartungswert)

Seien X und Y beliebige diskrete Zufallsvariablen, deren Erwartungswert existieren. Dann gilt

$$\mathbb{E}[aX + Y] = a\,\mathbb{E}[X] + \mathbb{E}[Y] \ \forall a, b \in \mathbb{R} \qquad \text{(Linearität)}$$
$$X \geqslant 0 \implies \mathbb{E}[X] \geqslant 0 \qquad \text{(Nichtnegativität)}$$
$$|\,\mathbb{E}[X]| \leqslant \mathbb{E}[|X|] \qquad (\triangle \neq)$$

Ferner gilt für eine beliebige (messbar???) Funktion $h : \mathbb{R} \to \mathbb{R}$ der Transformationssatz:

Transformationssatz

$$\mathbb{E}(h(X)) = \sum_{n \in \mathbb{N}_0} h(x_n) \, \mathbb{P}(X = x_n)$$

Aus der Linearität und Nichtnegativität folgt die Monotonie des Erwartungswerts.

Satz 3.5.1: Erwartungswert und Unabhängigkeit

Seien $(X_k)_{k=1}^n$ unabhängige diskrete Zufallsvariablen, deren Erwartungswerte existieren. Dann existiert der Erwartungswert von $X := \prod_{k=1}^n X_k$ und es gilt $\mathbb{E}[X] = \prod_{k=1}^n \mathbb{E}[X_k]$.

Beweis. Für $k \in \{1, \ldots, n\}$ seien x_ℓ^k die Werte Zufallsvariablen X_k und $\ell \in \mathbb{N}$. Dann gilt

$$\begin{aligned}
\mathbb{E}[|X|] &= \sum_{\ell_1,\ldots,\ell_n} \left|\prod_{k=1}^n x_{\ell_k}^k\right| \mathbb{P}((X_k = x_{\ell_k}^k)_{k=1}^n) \\
&\overset{(\star)}{=} \sum_{\ell_1,\ldots,\ell_n} \left|\prod_{k=1}^n x_{\ell_k}^k\right| \prod_{k=1}^n \mathbb{P}(X_k = x_{\ell_k}^k) = \sum_{\ell_1,\ldots,\ell_n} \prod_{k=1}^n |x_{\ell_k}^k| \mathbb{P}(X_k = x_{\ell_k}^k) \\
&= \prod_{k=1}^n \sum_{\ell_1,\ldots,\ell_n} |x_{\ell_k}^k| \mathbb{P}(X_k = x_{\ell_k}^k) = \prod_{k=1}^n \mathbb{E}[|X_k|] < \infty,
\end{aligned}$$

wobei wir in $(\star)$ die Unabhängigkeit nutzen. Also existiert $\mathbb{E}[X]$ und Wiederholung derselben Rechnung ohne den Betrag beweist die Behauptung. □

Beispiel 3.5.5

- Sind $(X_k)_{k=1}^n$ unabhängig Bernoulli-verteilt mit Erfolgswahrscheinlichkeit p, so gilt $\mathbb{E}[X_k] = p$ für alle $k \in \{1, \ldots, n\}$. Insbesondere gilt $\mathbb{E}[S_n] = \sum_{k=1}^n \mathbb{E}[X_k] = np$ für $S_n := \sum_{k=1}^n X_k \sim \mathrm{Bin}(n, p)$.

  Aus dem Transformationssatz folgt für $\alpha \in \mathbb{R}$ und alle $k \in \{1, \ldots, n\}$

  $$\mathbb{E}[\exp(\alpha X_k)] = e^0\, \mathbb{P}(X_k = 0) + e^\alpha\, \mathbb{P}(X_k = 1) = (1-p) + pe^\alpha$$

  und somit

  $$\begin{aligned}
  \mathbb{E}[\exp(\alpha S_n)] = \mathbb{E}\left[\prod_{k=1}^n \exp(\alpha X_k)\right] &= \prod_{k=1}^n \mathbb{E}[\exp(\alpha X_k)] \\
  &= (1 - p + pe^\alpha)^n.
  \end{aligned}$$

- Ähnlich zu oben folgt $\mathbb{E}[X] = \lambda$ und $\mathbb{E}[\exp(\alpha X)] = \exp(-\lambda(1-e^\alpha))$ für $X \sim \mathrm{Poiss}(\lambda)$. ◊

Erwartungswert reellwertiger Zufallsvariablen

Definition 3.5.6 (allgemeiner Erwartungswert)

Sei X eine Zufallsvariable. Dann nennen wir

$$\mathbb{E}[X] := \int_\Omega X \, \mathrm{d}\mathbb{P} \quad \text{falls} \quad \int_\Omega |X| \, \mathrm{d}\mathbb{P} < \infty$$

Erwartungswert von X.

Bermerkung 3.5.7 Das Lemma 3.5.4 (vgl. Transformationssatz und der Satz 3.5.1) übertragen sich auf den reellwertigen Fall.

Die Konstruktion des Bildmaßes $\mathrm{d}\mathbb{P}$ wird im Anhang (vgl. A.1.4) ausgeführt.

Beispiel 3.5.8

- Sei $X \sim \mathcal{U}([a,b])$. Dann existiert der Erwartungswert und ist gleich $\int_a^b \frac{x \, \mathrm{d}x}{b-a} = \frac{a+b}{2}$, also der Mittelpunkt des Intervalls.
- Ist $X \sim \operatorname{Exp}(\lambda)$, so existiert der Erwartungswert und ist gleich λ^{-1}.
- Ist $X \sim \mathcal{N}(0,1)$ so existiert der Erwartungswert und ist gleich Null. Aus Bemerkung 3.3.6 folgt $\mathbb{E}[X] = m$ für $X \sim \mathcal{N}(m, \sigma^2)$. ♢

Varianz reellwertiger Zufallsvariablen

Seien im folgenden alle Zufallsvariablen stets reellwertig.

Definition 3.5.9 (Momente)

Seien X eine Zufallsvariable und $p \in \mathbb{N}$. Dann heißt $\mathbb{E}[X^p]$ das p-te Moment von X.

p-te Moment

Satz 3.5.2: Cauchy-Bunjakowski-Schwarz

Seien X und Y quadratintergriebare Zufallsvariablen. Dann gilt $\mathbb{E}[|XY|] < \infty$ und

$$|\mathbb{E}[XY]| \leqslant \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}.$$

Beweis. Aus der binomischen Formel folgt $|XY| \leqslant \frac{X^2}{2} + \frac{Y^2}{2}$ und somit $\mathbb{E}[|XY|] < \infty$. Seien o.B.d.A ($E[X^2] > 0$ oder) $E[Y^2] > 0$ und definiere $\alpha := \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$. Dann gilt

$$0 \leqslant \mathbb{E}[(X - \alpha Y)^2] = \mathbb{E}[X^2] - 2\alpha\,\mathbb{E}[XY] + \alpha^2\,\mathbb{E}[Y^2] = \mathbb{E}[X^2] - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]},$$

woraus durch Ziehen der Wurzel die Behauptung folgt. □

DEFINITION 3.5.10 (VARIANZ)

Sei X eine Zufallsvariable mit $|\mathbb{E}[X]| < \infty$. Dann heißt die mittlere quadratische Abweichung von X um ihren Erwartungswert

$$\operatorname{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] \in [0, \infty]$$

Varianz von X. Varianz

Bemerkung 3.5.11

Aus Satz 3.5.2 folgt insbesondere $\operatorname{Var}[X] < \infty \iff \mathbb{E}[X^2] < \infty$.

DEFINITION 3.5.12 (STANDARDABWEICHUNG)

Die Zahl $S[X] := \sqrt{\operatorname{Var}[X]}$ heißt Standardabweichung von X. Standardabweichung

DEFINITION 3.5.13 (KOVARIANZ, UNKORRELIERTHEIT)

Seien X und Y quadratintegrierbare Zufallsvariablen. Dann heißt

$$\operatorname{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

die Kovarianz von X und Y. Gilt $\operatorname{Cov}(X, Y) = 0$ so heißen X und Y unkorreliert. Kovarianz

Satz 3.5.3: Rechenregeln für Varianzen

Seien X, Y und $(X_k)_{k=1}^n$ quadratintegrierbare Zufallsvariablen.

- Es gilt

$$\begin{aligned}
&\operatorname{Var}[aX + b] = a^2 \operatorname{Var}[X] \quad \text{und} && \text{(lokale Invarianz)}\\
&\operatorname{Cov}[aX + b, cY + d] = ac \operatorname{Cov}(X, Y)\ \forall a, b, c, d \in \mathbb{R} &&\\
&\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad \text{und} && \text{(Verschiebungssatz)}\\
&\operatorname{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] &&
\end{aligned}$$

- Die Kovarianz ist eine positiv semidefinite symmetrische Bilinearform.
- Identität von Bienaymé (B). Sind $(X_k)_{k=1}^n$ paarweise unkorreliert, so folgt $\operatorname{Var}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \operatorname{Var}(X_k)$.
- Sind X und Y unabhängig, so sind sie unkorreliert.

Beweis. Leichtes Rechnen. □

Gegenbeispiel 3.5.14 (Unkorreliert $\implies$ unabhängig)

Seien $A_1, A_2 \subset A_3$ disjunkt mit $0 < \mathbb{P}(A_1) = \mathbb{P}(A_2) < \mathbb{P}(A_3) < 1$. Dann sind die Zufallsvariablen $X := \mathbb{1}_{A_1} - \mathbb{1}_{A_2}$ und $Y := \mathbb{1}_{A_3}$ unkorreliert, da

$$XY = \mathbb{1}_{A_1 \cap A_3} - \mathbb{1}_{A_2 \cap A_3} = \mathbb{1}_{A_1} - \mathbb{1}_{A_2} = X$$

sowie $\mathbb{E}[X] = \mathbb{P}(A_1) - \mathbb{P}(A_2) = 0$ und somit

$$\operatorname{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \mathbb{E}[X] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0$$

gilt, aber nicht unabhängig, da

$$\begin{aligned}\mathbb{P}(X = 1, Y = 1) &= \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1)\\ &\neq \mathbb{P}(A_1)\,\mathbb{P}(A_3) = \mathbb{P}(X = 1)\,\mathbb{P}(Y = 1)\end{aligned}$$

gilt. ◊

Beispiel 3.5.15 (Varianzen der wichtigen Verteilungen)

- Für $(X_k \sim \operatorname{Ber}(p))_{k=1}^n$ gilt $\operatorname{Var}[S_n] = \sum_{k=1}^n \operatorname{Var}[X_1] = np(1-p)$ für $S_n := \sum_{k=1}^n X_k \sim \operatorname{Bin}(n, p)$.
- Für $X \sim \operatorname{Poiss}(\lambda)$ gilt $\mathbb{E}[X^2] = \lambda^2 + \lambda$ und somit $\operatorname{Var}[X] = \lambda$.
- Sei $X \sim \operatorname{Exp}(\lambda)$. Induktiv erhält man mit partieller Integration $\mathbb{E}[X^k] = \frac{k!}{\lambda^k}$ und somit $\operatorname{Var}[X] = \lambda^{-2}$.
- Sei $X \sim \mathcal{N}(0, 1)$. Dann gilt $\mathbb{E}[X^2] = 1$ und somit $\operatorname{Var}[X] = 1$. Sei nun $X \sim \mathcal{N}(m, \sigma^2)$. Nach Bemerkung 3.3.6 folgt $\operatorname{Var}[X] = \sigma^2$. ◊

4 Erzeugenden & charakteristische Funktionen

4.1 Erzeugenden Funktionen

Definition 4.1.1 (Erzeugenden Funktion, engl. PGF)

probability generating function

Seien X eine diskrete Zufallsvariable mit Werten in $\mathbb{N}_0$ und $p_n := \mathbb{P}(X = n)$ für $n \in \mathbb{N}$ die Zähldichte der zugehörigen Verteilung.

- Dann heißt

$$\mathcal{G}_X : [0,1] \to [0,1], \ \ s \mapsto \mathbb{E}[s^X] \overset{3.5.4}{=} \sum_{n=0}^{\infty} p_n s^n.$$

Allgemeiner ist der Definitionsbereich von $\mathcal{G}_X$ der Konvergenzradius der Potenzreihe.

erzeugende Funktion von X.

erzeugende Funktion

- Ist $\mathbb{P} = (p_n)_{n\in\mathbb{N}_0}$ eine beliebige Verteilung auf $\mathbb{N}_0$, so heißt $\mathcal{G}(s) := \sum_{n=0}^{\infty} p_n s^n$ erzeugenden Funktion von $\mathbb{P}$.

Bemerkung Die Potenzreihe ist wohldefiniert, da

$$\sum_{n\in\mathbb{N}_0} |p_n s^n| \leqslant \sum_{n\in\mathbb{N}_0} p_n |s|^n \leqslant \sum_{n\in\mathbb{N}_0} p_n = 1$$

gilt.

Lemma 4.1.2

Die erzeugenden Funktion ist für $|s| < 1$ gliedweise differenzierbar mit

$$\frac{\mathcal{G}^{(n)}(0)}{n!} = p_n.$$

Korollar 4.1.3 ($\mathbb{P} \iff \mathcal{G}$)

Jede Verteilung $\mathbb{P} = (p_n)_{n\in\mathbb{N}_0}$ ist eindeutig durch ihre erzeugenden Funktion bestimmt.

Beispiel 4.1.4 (erzeugenden Funktionen wichtiger Verteilungen)

- Für $\mathrm{Bin}(n,p)$ gilt $\mathcal{G}(s) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k \overset{1.2.2}{=} (sp+1-p)^n$.
- Sind $X \sim \mathrm{Bin}(n,p)$ und $Y \sim \mathrm{Bin}(m,p)$ unabhängige Zufallsvariablen für $n,m \in \mathbb{N}$ und $p \in [0,1]$. So gilt $\mathcal{G}_{X+Y}(t) = (tp+1-p)^{n+m}$, also $X+Y \sim \mathrm{Bin}(n+m,p)$ (s. A.2.6).
- Für $\mathrm{Geo}(p)$ gilt $\mathcal{G}(s) = \frac{ps}{1-(1-p)s}$ mit der geometrischen Reihe.

- Seien $X, Y \sim \mathrm{Geo}(p)$ unabhängige Zufallsvariablen mit $p \in (0,1)$. Dann gilt $\mathcal{G}_{X+Y} = \sum_{k=0}^{\infty} \binom{n}{k} t^k p^k (1-p)^{n-k}$ und somit $\mathbb{P}(X+Y = n) = p^2(n-1)(1-p)^{n-2}$ für alle $n \in \mathbb{N}_{>1}$.
- Für $\mathrm{Poiss}(\lambda)$ gilt $\mathcal{G}(s) = \sum_{k \in \mathbb{N}_0} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{\lambda(s-1)}$. $\Diamond$

Bemerkung (Notation) Wir schreiben $F(a\pm) := \lim_{x \to a^\pm} F(x)$.

Lemma 4.1.5 (Grenzwertsatz von Abel)
Sei $A := \sum_{n \in \mathbb{N}_0} a_n$ eine konvergente Reihe reeller Zahlen. Dann konvergiert die Potenzreihe $f(x) := \sum_{n \in \mathbb{N}_0} a_n x^n$ auf $[0,1]$ und die Funktion $f(x)$ ist stetig auf $[0,1]$ mit $f(1) = A$.

Satz 4.1.1: Momentenberechnung

(X diskret?)

1. Es gilt $\mathbb{E}[X] = \mathcal{G}'_X(1-)$.
2. Ist $\mathcal{G}'_X(1-) < \infty$, so gilt

$$\mathrm{Var}[X] = \mathcal{G}''_X(1-) + \mathcal{G}'_X(1-)\big(1 - \mathcal{G}'_X(1-)\big).$$

Beweis. 1. Für $|s| < 1$ ist $\mathcal{G}'_X(s) = \sum_{n \in \mathbb{N}} n p_n s^{n-1}$. Für $\mathbb{E}[X] = \infty$ ist die Identität klar, sonst folgt sie aus dem Satz von Abel.

2. Für $|s| < 1$ gilt

$$\mathcal{G}''_X(s) = \sum_{n=2}^{\infty} n(n-1) p_n s^{n-2} = \underbrace{\sum_{n \in \mathbb{N}} n^2 p_n s^{n-2}}_{=:\mathcal{S}_1(s)} - \underbrace{\sum_{n \in \mathbb{N}} n p_n s^{n-2}}_{=:\mathcal{S}_2(s)}.$$

Ist $\mathcal{G}'_X(1-) < \infty$ so folgt

$$\lim_{s \nearrow 1} \mathcal{S}_2(s) = \lim_{s \nearrow 1} \frac{\mathcal{G}'_X(s)}{s} = \mathcal{G}'_X(1-) = \mathbb{E}[X]$$

und analog zu 1

$$\lim_{s \nearrow 1} \mathcal{S}_1(s) = \sum_{n \in \mathbb{N}} n^2 p_n = \mathbb{E}[X^2]$$

und somit

$$\begin{aligned} \mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \lim_{s \nearrow 1} \mathcal{G}''_X(s) + \frac{\mathcal{G}'_X(s)}{s} - \mathcal{G}'_X(s)^2 \\ &= \mathcal{G}''_X(1-) + \mathcal{G}'_X(1-)\big(1 - \mathcal{G}'_X(1-)\big). \end{aligned}$$ $\square$

Beispiel 4.1.6 (Momentenberechnung)

- Aus dem letzten Beispiel wissen wir, dass für $X \sim \mathrm{Bin}(n,p)$ die erzeugende Funktion $\mathcal{G}_X$ ein Polynom und somit überall differenzierbar ist. Aus dem Satz 4.1.1 folgt $\mathbb{E}[X] = np$ und $\mathrm{Var}[X] = np(1-p)$.

- Für $X \sim \text{Poiss}(\lambda)$ eine global konvergente Potenzreihe und somit differenzierbar. Somit folgt $\mathbb{E}[X] = \lambda = \lambda^2 + \lambda(1-\lambda) = \text{Var}[X]$. ◊

SATZ 4.1.2: TODO

Seien $(X_k)_{k=1}^n$ unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen und $S_n := \sum_{k=1}^n X_k$. Dann gilt für $|s| < 1$

$$\mathcal{G}_{S_n}(s) = \prod_{k=1}^n \mathcal{G}_{X_k}(s)$$

Beweis. Da $(X_k)_{k=1}^n$ unabhängig sind, ist es nach ... auch $\left(s^{X_k}\right)_{k=1}^n$ für beliebige $s \in \mathbb{R}$. Nach Satz 3.33 (!!!) gilt

$$\mathcal{G}_{S_n}(s) = \mathbb{E}\left[s^{\sum_{k=1}^n X_k}\right] = \mathbb{E}\left[\prod_{k=1}^n s^{X_k}\right] = \prod_{k=1}^n \mathbb{E}[s^{X_k}] = \prod_{k=1}^n \mathcal{G}_{X_k}(s). \qquad \square$$

Beispiel 4.1.7 (Anwendung auf BERNOULLI-Experimente)
Seien $(Y_k \sim \text{Ber}(p))_{k=1}^n$ unabhängig. Dann gilt $X := \sum_{k=1}^n Y_k \sim B(n,p)$. Mit dem obigen Satz können wir unser Ergebnis aus Beispiel 4.1.4 verifizieren: $\mathcal{G}_X(s) = \prod_{k=1}^n \mathcal{G}_{Y_k}(s) = (sp + 1 - p)^n$. ◊

SATZ 4.1.3: KONVERGENZSATZ FÜR CH. FUNKTIONEN

Seien $\left(P_n = \left(p_k^{(n)}\right)_{k\in\mathbb{N}_0}\right)_{n\in\mathbb{N}_0}$ eine Folge von Verteilungen auf $\mathbb{N}_0$ mit erzeugenden Funktionen $(\mathcal{G}_n)_{n\in\mathbb{N}_0}$. Dann sind die beiden folgenden Aussage äquivalent.

- Der Grenzwert $p_k := \lim\limits_{n\to\infty} p_k^{(n)}$ existiert für alle $k \in \mathbb{N}_0$.
- Der Grenzwert $\mathcal{G}(s) := \lim\limits_{n\to\infty} \mathcal{G}_n(s)$ existiert für alle $|s| < 1$.

In diesem Fall gilt $p_k \geqslant 0$ für $k \geqslant 0$ mit $\sum_{k\in\mathbb{N}_0} p_k \leqslant 1$ und $\mathcal{G}(s) = \sum_{k\in\mathbb{N}_0} p_k s^k$.

Beweis. TODO. □

Das folgende Beispiel zeigt, dass im Allgemeinen $\mathfrak{P} := \sum_{k\in\mathbb{N}_0} p_k < 1$ gilt, dass also die Grenzverteilung ein defektes Wahrscheinlichkeitsmaß auf $\mathbb{N}_0$ ist: $p_k^{(n)} = \mathbb{1}_{k=n}$. Dann gilt $p_k^{(n)} \xrightarrow{n\to\infty} 0 =: p_k$ für $k \in \mathbb{N}_0$ also sogar $\mathfrak{P} = 0$. Im Falle von $\mathfrak{P} = 1$ gilt in der Situation des obigen Satzes $\mathcal{G}(s) = \lim_{n\to\infty} \mathcal{G}_n(s)$ für $|s| \leqslant 1$.

Beispiel 4.1.8 (Anwendung: POISSONscher Grenzwertsatz)
Sei $(p_n)_{n\in\mathbb{N}}$ eine Folge von Erfolgsparametern mit $np_n \xrightarrow{n\to\infty} \lambda > 0$ und $(\mathcal{G}_n(s) := (sp_n + 1 - p_n)^n)_{n\in\mathbb{N}_0}$ die erzeugenden Funktion der Binomial-

verteilung $\mathrm{Bin}(n,p)$. Dann folgt mit $\lambda_n := np_n$

$$\mathcal{G}_n(s) = (sp_n + 1 - p_n)^n = \left(1 - (1-s)\frac{\lambda_n}{n}\right)^n \xrightarrow{n\to\infty} e^{(s-1)\lambda},$$

da $\left(1-(1-s)\frac{\lambda}{n}\right)^n \xrightarrow{n\to\infty} e^{(s-1)\lambda}$ für alle $\lambda \in \mathbb{R}$ und

$$\lim_{n\to\infty}\left|\left(1-(1-s)\frac{\lambda_n}{n}\right)^n - \left(1-(1-s)\frac{\lambda}{n}\right)^n\right| \leqslant \lim_{n\to\infty}|1-s||\lambda_n - \lambda| = 0$$

gilt. Die letzte Ungleichung sieht man so ein: Seien $a := 1-(1-s)\frac{\lambda_n}{n}$ und $b := 1-(1-s)\frac{\lambda}{n}$. Aufgrund von $s \in [0,1]$ gilt $a, b \in [0,1]$ für große $n \in \mathbb{N}$. Nun gilt

$$\begin{aligned}
|a^n - b^n| &= \left|\sum_{k=1}^{n} a^k b^{n-k} - \sum_{k=0}^{n-1} a^k b^{n-k}\right| = \left|\sum_{k=0}^{n-1} a^{k+1} b^{n-1-k} - \sum_{k=0}^{n-1} a^k b^{n-k}\right| \\
&= \left|\sum_{k=0}^{n-1} \underbrace{a^k b^{n-1-k}}_{<1}\right| |a-b| \leqslant n|a-b| = \cancel{n}|1-s|\left|\frac{\lambda_n - \lambda}{\cancel{n}}\right|.
\end{aligned}$$

Warum ist das eine Anwendung des Satzes? ◊

4.2 Charakteristische Funktionen

Vorbetrachtung

Die Funktion $\mathcal{G}(s) = \sum_{n\in\mathbb{N}_0} p_n s^n$ ist analytisch in $|s| < 1$ und kann stetig auf dem Rand fortgesetzt werden, (vgl. Satz von Abel). Nach komplexer Analysis ist $\mathcal{G}$ durch seine Randwerte $\chi_X(t) := \mathcal{G}(e^{it})$ für $t \in \mathbb{R}$ eindeutig bestimmt.

Definition 4.2.1 (Charakteristische Funktion)

Sei X eine reellwertige Zufallsvariable. Dann heißt die Funktion

$$\chi_X : \mathbb{R} \to \mathbb{C},\ t \mapsto \mathbb{E}[\exp(itX)]$$

charakteristische Funktion von X.

charakteristische Funktion

Korollar 4.2.2 (Eigenschaften der charakteristischen Funktion)

Sei X eine reellwertige Zufallsvariable und χ ihre charakteristische Funktion. Dann gilt

- $\chi(0) = 1$ *und* $|\chi(t)| \overset{L\ 3.5.4}{\leqslant} \mathbb{E}[|\exp(itX)|] = 1$.
- *Ist X diskret (stetig) verteilt mit den Werten x_n (Dichte f), so gilt*

$$\chi_X(t) = \sum_n \mathbb{P}(X = x_n)e^{itx_n} \left(= \int_{\mathbb{R}} e^{itx} f(x)\,\mathrm{d}x\right).$$

den Integralausdruck bezeichnet man als Fourier-Transformierte von f.

- *Die charakteristische Funktion bestimmt die Verteilung einer Zufallsvariablen eindeutig, es gilt $\chi_X = \chi_Y \iff F_X = F_Y$.*
- *Besitzt X endliches für $n \in \mathbb{N}_0$ n-tes Moment, so folgt $\mathbb{E}[X^n] = i^{-n}\chi_X^{(n)}(0)$ (vgl. Satz 4.1.1).*
- *Seien $(X_k)_{k=1}^n$ unabhängige Zufallsvariablen. Dann gilt für $t \in \mathbb{R}$ (vgl. Satz 4.1.2)*

$$\chi_{(\sum_{k=0}^n X_k)}(t) = \prod_{k=1}^n \chi_{X_k}(t).$$

- *Seien $X, (X_k)_{k\in\mathbb{N}}$ Zufallsvariablen. Dann sind die beiden folgenden Aussagen äquivalent: (1) $F_{X_n}(x) \xrightarrow{n\to\infty} F_X(x)$ für alle Stetigkeitspunkte x von F und (2) $\chi_{X_n}(t) \xrightarrow{n\to\infty} \chi_X(t)$ für alle $t \in \mathbb{R}$ (vgl. Satz 4.1.3).*

Beispiel 4.2.3 (Charakteristische Funktion wichtiger Verteilungen)

- Ist $X \sim \mathrm{Ber}(p)$, so gilt

$$\chi_X(t) = \mathbb{E}[e^{itX}] = e^{it\cdot 0}(1-p) + e^{it}p = (1-p) + e^{it}p =: B.$$

- Ist $X \sim \mathrm{Bin}(n,p)$, so gilt $\chi_X(t) = \sum_{k=0}^n e^{itk}\binom{n}{k}p^k(1-p)^{n-k} = B^n$.
- Ist $X \sim \mathcal{N}(0,1)$, so gilt $\chi_X = e^{-\frac{t^2}{2}}$. Also folgt für $Y \sim \mathcal{N}(m,\sigma)^2$ $\chi_Y = \exp\left(itm - \frac{(t\sigma)^2}{2}\right)$. (s. A.2.7) $\diamondsuit$

Lemma 4.2.4 (Linearkombinationen von Normalverteilungen)

Seien $(X_k \sim \mathcal{N}(m_k, \sigma_k^2)_{k=1}^n$ *unabhängig und* $(\alpha_k)_{k=1}^n \subset \mathbb{R}$. *Dann gilt* $Z := \sum_{k=1}^n \alpha_k X_k \sim \mathcal{N}(m, \sigma^2)$ *mit* $m := \langle \alpha, m \rangle := \sum_{k=1}^n \alpha_k m_k$ *und* $\sigma^2 := \sum_{k=1}^n \alpha_k^2 \sigma_k^2$.

Beweis. Es gilt

$$\begin{aligned}
\chi_Z(t) &= \mathbb{E}\left[it \sum_{k=1}^n \alpha_k X_k\right] \overset{\text{L}}{\underset{(\circledast)}{=}} \prod_{k=1}^n \mathbb{E}[\exp(it\alpha_k X_k)] \\
&= \prod_{k=1}^n \exp\left(it\alpha_k m_k - \sigma_k^2 \frac{(t\alpha_k)^2}{2}\right) \\
&= \exp\left(\sum_{k=1}^n \alpha_k m_k - \left(\sum_{k=1}^n \alpha_k^2\sigma_k^2\right)\frac{t^2}{2}\right) = \exp\left(itm - \frac{(t\sigma)^2}{2}\right),
\end{aligned}$$

wobei wir in $(\circledast)$ Lemma 3.4.3 benutzt haben. Aufgrund der Eindeutigkeit der charakteristischen Funktion folgt die Behauptung. $\square$

5 Gesetze der großen Zahlen

5.1 Markovsche Ungleichung

Satz 5.1.1: Markovsche Ungleichung

Seien X eine reellwertige Zufallsvariable und $\Phi : \mathbb{R} \to [0, \infty)$ monoton wachsend. Dann gilt

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}[\Phi(X)]}{\Phi(a)}$$

für alle $a \in \mathbb{R}$ mit $\Phi(a) > 0$.

Beweis. Sei $a \in \mathbb{R}$. Dann Φ monoton wachsend ist, folgt $\Phi(X(\omega)) < \Phi(a)$ aus $X(\Omega) \geqslant a$ für alle $\omega \in \Omega$. Das impliziert (Notation aus (6))

$$\{X \geqslant a\} \subset \{\Phi(X) \geqslant \Phi(a)\} \implies \mathbb{P}(X \geqslant a) \leqslant \mathbb{P}(\Phi(X) \geqslant \Phi(a)). \tag{13}$$

Nun folgt

$$\begin{aligned}\Phi(a)\,\mathbb{P}(X \geqslant a) &\overset{(13)}{\leqslant} \Phi(a)\,\mathbb{P}\left(\Phi(X) \geqslant \Phi(a)\right) \underset{\text{(L)}}{\overset{3.5.3}{=}} \mathbb{E}[(\Phi(a)\,\mathbb{1}_{\{\Phi(X)\geqslant\Phi(a)}] \\ &\overset{\text{(M)}}{\leqslant} \mathbb{E}[(\Phi(X)\,\mathbb{1}_{\{\Phi(X)\geqslant\Phi(a)}] \overset{\text{(M)}}{\leqslant} \mathbb{E}[\Phi(X)]\end{aligned}$$

Da $\Phi(a) > 0$ ist, folgt die Behauptung durch Umstellen. □

Korollar 5.1.1 (Chebyshevsche Ungleichung)

Für $p \in [1, \infty)$ betrachte die monoton wachsende Funktion $\Phi(t) := t^p\,\mathbb{1}_{t>0}(t)$. Also folgt für alle $a > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant a) \leqslant \frac{\mathbb{E}[|X - \mathbb{E}[X]|^p]}{a^p}$$

Der Spezialfall $p = 2$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant a) \leqslant \frac{\mathrm{Var}[X]}{a^2}$$

heißt Chebyshevsche Ungleichung.

Chebyshev*sche Ungleichung*

Korollar 5.1.2 (Generische Chernoff-Schranke)

Für $\Phi(t) := e^{\lambda t}$ mit $\lambda > 0$ folgt für alle $a \in \mathbb{R}$

$$\mathbb{P}\left(X - \mathbb{E}[X] \geqslant a\right) \leqslant e^{-\lambda a}\,\mathbb{E}[\exp(\lambda(X - \mathbb{E}[X])].$$

5.2 Das schwache Gesetz der großen Zahlen

Definition 5.2.1 (stochastische Konvergenz)

Seien $X, (X_k)_{k\in\mathbb{N}}$ Zufallsvariablen. Die Folge $(X_k)_{k\in\mathbb{N}}$ konvergiert in Wahrscheinlichkeit gegen X gegen X, wenn für alle $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n\to\infty} 0$$

gilt und schreiben $X_n \xrightarrow{\mathbb{P}} X$.

Satz 5.2.1: Schwaches Gesetz der grossen Zahlen

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unkorrellierter identische verteilter Zufallsvariablen mit endlicher Varianz. Dann gilt

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1].$$

Beweis. Sei $m := \mathbb{E}[X_1]$ und $\varepsilon > 0$. Dann gilt $\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] = m$ und

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - m\right| > \varepsilon\right) &\overset{5.1.1}{\leqslant} \frac{1}{\varepsilon^2}\operatorname{Var}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] \\ &\overset{(B)}{=} \frac{1}{(\varepsilon n)^2}\sum_{k=1}^{n}\operatorname{Var}[X_k] \\ &\overset{\text{i.v.}}{=} \frac{1}{\varepsilon^2\cdot n}\operatorname{Var}[X_1] \xrightarrow{n\to\infty} 0.\end{aligned}$$

□

Beispiel 5.2.2 (empirisches → theoretisches Mittel)

Seien $(X_n \sim \operatorname{Ber}(p))_{n\in\mathbb{N}}$ unabhängig. Dann gilt $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\mathbb{P}} p$. ◊

Satz 5.2.2: Schwaches GGZ (Verallgemeinerung)

Sei $(X_k)_{k=1}^{n}$ eine Familie paarweise unkorrelierter Zufallsvariablen mit endlicher Varianz. Gilt

$$\frac{1}{n^2}\sum_{k=1}^{n}\operatorname{Var}[X_k] \xrightarrow{n\to\infty} 0,$$

so folgt

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k]) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Beweis. Für $\varepsilon > 0$ gilt

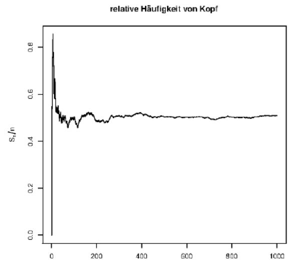
$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}[X_k]\right| > \varepsilon\right) &\leqslant \frac{1}{\varepsilon^2}\operatorname{Var}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] \\ &= \frac{1}{(\varepsilon n)^2}\sum_{k=1}^{n}\operatorname{Var}[X_k] \xrightarrow{n\to\infty} 0.\end{aligned}$$

□

5.3 Das starke Gesetz der großen Zahlen

Wir betrachten eine Folge fairer Münzwürfe $(X_k)_{k\in\mathbb{N}}$ (u.i.v) mit $\mathbb{P}(X_k = 0) = \mathbb{P}(X_k = 1) = 0.5$. Nach dem schwachen Gesetz der großen Zahlen gilt für $S_n := \sum_{k=1}^n X_k$

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1] = 0.5.$$



Betrachten wir nun einen Folge von Realisierungen $(X_k(\omega))_{k\in\mathbb{N}}$ der Münzwürfe erhalten wir das nebenstehende Verhalten für die Folge der relativen Häufigkeiten. Für ein „typisches" ω sollte als $\frac{S_n}{n}(\omega) \to 0.5$ gelten.

Daher benötigen wir den stärkeren Konvergenzbegriff der

Definition 5.3.1 (fast sichere Konvergenz)

Seien X und $(X_k)_{k\in\mathbb{N}}$ Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$. Die Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert $\mathbb{P}$-fast sicher gegen X falls

$$X_n \xrightarrow{\text{f.s.}} X :\iff X = \lim_{n\to\infty} X_n :\iff \mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Satz 5.3.1: f.s.e und stochastische Konvergenz

Fast sichere Konvergenz impliziert stochastische Konvergenz.

Beweis. Seien $\varepsilon > 0$ und

$$B_N := \left\{|X_n(\omega) - X(\omega)| < \varepsilon \ \forall n \geqslant N\right\}.$$

Die B_N sind aufsteigend und es gilt

$$B := \bigcup_{N\in\mathbb{N}} B_N \supset \{X_n(\omega) \xrightarrow{n\to\infty} X(\omega)\} =: A$$

Nach Annahme folgt $\mathbb{P}(A) = 1$ und somit $\mathbb{P}(B) = 1$. Aufgrund der Stetigkeit von $\mathbb{P}$ von unten folgt

$$\mathbb{P}(B_N) \xrightarrow{N\to\infty} 1$$

und somit

$$\lim_{N\to\infty} \mathbb{P}(\{|X_N(\omega) - X(\omega)| \geqslant \varepsilon\}) \leqslant \lim_{N\to\infty} \mathbb{P}(B_N^{\complement}) = 0. \qquad \square$$

Gegenbeispiel 5.3.2 (stochastische $\implies$ f.s. Konvergenz)

Wir betrachten den Wahrscheinlichkeitsraum mit $\Omega := [0,1]$, $\mathcal{A} := \mathcal{B}([0,1])$ und der Gleichverteilung $\mathbb{P}$. Seien

$$c_n := \sum_{k=1}^n \frac{1}{k} \quad \text{und} \quad A_n := \{\omega : [c_{n-1} \bmod 1, c_n \bmod 1]\}.$$

Für $\varepsilon \in (0,1)$ folgt

$$\mathbb{P}(\mathbb{1}_{A_n} \geqslant \varepsilon) \stackrel{?}{=} \mathbb{P}(A_n) = \frac{1}{n} \xrightarrow{n\to\infty} 0.$$

Somit gilt $\mathbb{1}_{A_n} \xrightarrow[n\to\infty]{\text{f.s}} 0$. Andererseits gilt für alle $\omega \in (0,1)$ **Warum??**

$$\limsup_{n\to\infty} \mathbb{1}_{A_n}(\omega) = 1 \neq 0 = \liminf_{n\to\infty} \mathbb{1}_{A_n}(\omega). \qquad \Diamond$$

Definition 5.3.3 (lim sup und lim inf von Mengen)

Sei $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ eine Folge von Ereignissen. Wir definieren

$$\limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{k=n}^{\infty} A_k = \{\omega \in A_k \text{ für } \infty \text{ viele } k\}$$

$$\liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{k=n}^{\infty} A_k = \{\omega \in A_k \text{ für alle bis auf endlich viele } k\}$$

Satz 5.3.2: Borel-Cantelli

Für $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ gilt

1. aus $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$ folgt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.
2. Sind die Ereignisse $(A_n)_{n\in\mathbb{N}}$ unabhängig, so folgt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$ aus $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$.

Beweis. 1 Aus $\bigcup_{k\geqslant n} A_k \searrow \bigcap_{n\in\mathbb{N}} \bigcup_{k\geqslant n} A_k$ für $n \to \infty$ folgt aus der σ-Stetigkeit von oben von $\mathbb{P}$

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geqslant n} A_k\right) \leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0.$$

2 Es genügt zu zeigen, dass für alle $n \in \mathbb{N}$

$$\mathbb{P}\left(\bigcup_{k\geqslant n} A_k\right) = 1 \quad \text{bzw.} \quad \mathbb{P}\left(\bigcap_{k\geqslant n} A_k^{\complement}\right) = 0$$

gilt. Für alle $\alpha \in \mathbb{R}$ gilt (Taylor!) $1 - \alpha \leqslant e^{-\alpha}$ und somit

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{k\geqslant n} A_k^{\complement}\right) &= \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{m} A_k^{\complement}\right) = \lim_{m\to\infty} \prod_{k=n}^{m} \mathbb{P}(A_k^{\complement}) \\
&= \lim_{m\to\infty} \prod_{k=n}^{m} 1 - \mathbb{P}(A_k) \leqslant \lim_{m\to\infty} \prod_{k=n}^{m} e^{-\mathbb{P}(A_k)} \\
&= \lim_{m\to\infty} \exp\left(\sum_{k=n}^{m} \mathbb{P}(A_k)\right) = 0,
\end{aligned}$$

weil die Exponentialfunktion stetig ist. $\square$

Korollar 5.3.4 (Bank-Klausur 2014 / 02)

Seien $(X_n \sim \mathrm{Ber}(p_n))_{n\in\mathbb{N}}$ *unabhängig mit* $p_n \in [0,1]$ *für alle* $n \in \mathbb{N}$. *Dann gilt* $\sum\limits_{n\in\mathbb{N}} p_n < \infty$ *genau dann wenn* $X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}f.s.} 0$ *gilt.*

Beweis. " $\Longrightarrow$ ": Definiere $A_n := \{X_n = 1\}$. Dann gilt $\mathbb{P}(A_n) = p_n$. Aus dem Lemma von Borel-Cantelli folgt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$ und somit

$$1 = \mathbb{P}((\limsup_{n\to\infty} A_n)^{\complement}) = \mathbb{P}(\liminf_{n\to\infty} A_n^{\complement}) = \mathbb{P}(X_n = 0 \text{ für schließlich alle } n).$$

" $\Longleftarrow$ ": Angenommen $\sum_{k\in\mathbb{N}} p_n = \infty$. Aus dem Lemma von Borel-Cantelli folgt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$, was ein Widerspruch zur Voraussetzung ist. □

Lemma 5.3.5 (schnelle stochastische Konvergenz)

Gilt für alle $\varepsilon > 0$ *schnelle stochastische Konvergenz gegen 0*

$$\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| \geqslant \varepsilon) < \infty$$

so folgt fast sichere Konvergenz gegen 0

$$\mathbb{P}\left(\{X_n(\omega) \to 0\}\right) = 1.$$

Beweis. Fehlt. □

Satz 5.3.3: Starkes GGZ (Kolmogorov, 1930)

Sei $(X_k)_{k\in\mathbb{N}}$ u.i.v. mit endlichem Erwartungswert. Dann gilt

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\text{f.s}} \mathbb{E}[X_1]$$

Unter dem obigen Voraussetzungen gilt insbesondere das schwache GGZ.

6 Der Zentrale Grenzwertsatz

bla

7 Multivariate Verteilungen

DEFINITION 7.0.1 (MULTIVARIATE KENNGRÖSSEN)

Sei $\mathbb{X}$ ein n-dim. Zufallsvektor. Dann ist $\mathbb{E}[X] := ((\mathbb{E}(X_k))_{k=1}^n)^T$ der multivariate Erwartungswert. Die Varianz verallgemeinert die Kovarianzmatrix

Erwartungswert

Kovarianzmatrix

$$\operatorname{Cov}(\mathbb{X},\mathbb{X}) := (\operatorname{Cov}(X_i,X_j))_{1\leqslant i,j\leqslant n} = \mathbb{E}[(X-\mathbb{E}[X])\cdot(\mathbb{X}-\mathbb{E}[X])^T]$$

und

$$\chi_{\mathbb{X}} : \mathbb{R}^n \to \mathbb{R},\ \lambda \mapsto \mathbb{E}[\exp(i\langle\,\lambda\,\mathbb{X}\,\rangle)]$$

die charakteristische Funktion.

charakteristische Funktion

Lemma 7.0.2 (Eigenschaften von Cov$(\mathbb{X},\mathbb{X})$)

Die Kovarianzmatrix ist positiv semidefinit, symmetrisch und genau dann eine Diagonalmatrix wenn die Komponenten $(X_k)_{k=1}^n$ paarweise unkorrelliert sind.

semidefinit, symmetrisch

Beweis. Für beliebige $(\lambda_k)_{k=1}^n$ gilt

$$\begin{aligned}\langle\,\lambda\operatorname{Cov}(\mathbb{X},\mathbb{X})\lambda\,\rangle &= \sum_{i,j=1}^n \lambda_i \operatorname{Cov}(X_i,X_j)\lambda_j\\ &\overset{\text{Def.}}{=} \sum_{i,j=1}^n \lambda_i\,\mathbb{E}[(X_i-\mathbb{E}[X_i])(X_j-\mathbb{E}[X_j])]\lambda_j\\ &\overset{\text{(L)}}{=} \mathbb{E}\left[\left(\sum_{k=1}^n \lambda_k\big(X_j-\mathbb{E}[X_k]\big)\right)^2\right] \geqslant 0.\end{aligned}$$ □

Bermerkung 7.0.3 Die (gemeinsame) Verteilung ist eindeutig durch ihre charakteristische Funktion bestimmt.

7.1 Die Multivariate Normalverteilung

Definition 7.1.1 (n-dim. Standardnormalverteilung)
Ein Zufallsvektor $\mathbb{X} = ((X_k)_{k=1}^n)^T$ heißt standardnormalverteilt, wenn $X_k \sim \mathcal{N}(0,1)$ gilt und diese unabhängig sind.

Nach Satz ... (3.26) besitzt $\mathbb{X}$ die (gemeinsame) Dichte

$$f_{\mathbb{X}} : \mathbb{R}^d \to \mathbb{R},\ (x_1, \dots, x_n) \mapsto \prod_{k=1}^n \frac{\exp\left(-\frac{x_k^2}{2}\right)}{\sqrt{2\pi}} = \frac{\exp\left(-\frac{|x|^2}{2}\right)}{(2\pi)^{\frac{n}{4}}}.$$

Korollar 7.1.2 (Eigenschaften der n-dim. SNV)
Ist $\mathbb{X}$ standardnormalverteilt, so gilt $\operatorname{Cov}(\mathbb{X}, \mathbb{X}) = \mathbb{E}_n$ *und*

$$\chi_{\mathbb{X}} : \mathbb{R} \to \mathbb{R},\ \lambda \mapsto \mathbb{E}[\exp(i\langle \lambda, \mathbb{X} \rangle)] = \prod_{k=1}^n e^{-\frac{\lambda_k^2}{2}} = e^{-\frac{|\lambda|^2}{2}}.$$

Beweis. Für $i \neq j \in \{0, \dots, n\}$ gilt $\mathbb{E}[X_i] = \mathbb{E}[X_j] = 0$ und somit

$$\operatorname{Cov}[X_i, X_j] \overset{\text{Def.}}{=} \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \overset{\text{unabh.}}{=} \mathbb{E}[X_i]\,\mathbb{E}[X_j] = 0.$$

Für $i = j$ gilt $\operatorname{Cov}[X_i, X_j] = \operatorname{Var}[X_i] = 1^2 = 1$. □

Definition 7.1.3 (n-dim. Normalverteilung)
Ein n-dim. Zufallsvektor $\mathbb{X}$ heißt normalverteilt, wenn

$$\mathbb{X} = A\,\mathbb{Y} + m \qquad \text{mit } A \in \mathbb{R}^{n \times k}, m \in \mathbb{R}^n, \mathbb{Y} \in \mathbb{R}^k, \mathbb{Y} \sim \mathcal{N}(0,1)$$

gilt.

Korollar 7.1.4 (Eigenschaften der n-dim. NV)
Es gilt $\mathbb{E}[\mathbb{X}] = m$, $\operatorname{Cov}[\mathbb{X}, \mathbb{X}] = AA^T$ *und* $\chi_{\mathbb{X}}(\lambda) = \exp\left(i\langle \lambda, m \rangle - \frac{\langle \lambda, \Sigma^2 \lambda \rangle}{2}\right)$.

Beweis. Es gilt

$$\mathbb{E}[\mathbb{X}] = A \underbrace{\mathbb{E}[Y]}_{=0} + m = m$$

und

$$\begin{aligned}\operatorname{Cov}[\mathbb{X}, \mathbb{X}] &= \mathbb{E}[(\mathbb{X} - \mathbb{E}[\mathbb{X}])((\mathbb{X} - \mathbb{E}[\mathbb{X}])^T] = \mathbb{E}[(A\,\mathbb{Y})(A\,\mathbb{Y})^T] \\ &= A\,\mathbb{E}[\mathbb{Y}\,\mathbb{Y}^T]A^T = AA^T =: \Sigma^2.\end{aligned}$$

Ferner gilt für $\lambda \in \mathbb{R}$

$$\begin{aligned}\chi_{\mathbb{X}}(\lambda) &= \mathbb{E}[\exp(i\langle \lambda, A\,\mathbb{Y} + m \rangle)] = \mathbb{E}[\exp(i\langle \lambda, m \rangle + i\langle A^T\lambda, \mathbb{Y} \rangle)] \\ &= \exp\left(i\langle \lambda, m \rangle - \frac{|A^T\lambda|^2}{2}\right) = \exp\left(i\langle \lambda, m \rangle - \frac{\langle \lambda, \Sigma^2\lambda \rangle^2}{2}\right).\end{aligned}$$ □

Definition 7.1.5 (multivariate Verteilung der NV)
Sei $\mathbb{X}$ normalverteilt. Die Verteilung von $\mathbb{X}$ heißt mehrdimensionale Normalverteilung mit Mittelwert(vektor) m und Kovarianzmatrix Σ^2. Man schreibt $\mathbb{X} \sim \mathcal{N}(m, \Sigma^2)$.

Korollar 7.1.6 (Eigenschaften der n-dim. Verteilung der NV)
Affin lineare Transformationen nver Zufallsvektoren sind nv.

Beweis. Seien $\mathbb{X} = A\,\mathbb{Y} + m$ wie oben, $B \in \mathbb{R}^{\ell \times n}$, $d \in \mathbb{R}^\ell$ und $\mathbb{W} := B\,\mathbb{X} + d$. Dann gilt $\mathbb{E}[\mathbb{W}] = B\,\mathbb{E}[\mathbb{X}] + d = Bm + d$ und $\operatorname{Cov}(\mathbb{W}, \mathbb{W}) = B \operatorname{Cov}(\mathbb{X}, \mathbb{X}) B^T = B\Sigma^2 B^T$. □

Satz 7.1.1: Dichte der n-dim. NV

Ist $\mathbb{X}$ ein k-dimensionaler Zufallsvektor mit charakteristische Funktion $\chi_{\mathbb{X}}(\lambda) = \exp\left(i\langle \lambda, m\rangle - \frac{\langle \lambda, \Sigma^2 \lambda\rangle}{2}\right)$, wobei $m \in \mathbb{R}^n$ und $\Sigma^2 \in \mathbb{R}^{n \times n}$ symmetrisch positiv semidefinit ist, dann besitzt die Normalverteilung $\mathcal{N}(m, \Sigma^2)$ die n-dimensionale Dichte

$$f_{m,\Sigma^2} : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \frac{\exp\left(-\frac{1}{2}\langle x - m, (\Sigma^2)^{-1}(x-m)\rangle\right)}{(2\pi)^{\frac{n}{2}}}$$

Beweis. Symmetrische Matrizen sind orthonormal diagonalisierbar; es existiert eine orthonormale Matrix $U \in \mathbb{R}^{n \times n}$ und Zahlen $(\sigma_k^2 \geqslant)_{k=1}^n$ (nichtnegative Eigenwerte von Σ^2), sodass

$$U^T \Sigma^2 U = \operatorname{diag}((\sigma_k^2)_{k=1}^n)),$$

gilt. Für den Zufallsvektor $\mathbb{Y} := U^T(\mathbb{X} - m)$ folgt dann

$$\begin{aligned}
\chi_{\mathbb{Y}}(\lambda) &= \mathbb{E}[\exp(\langle U\lambda, \mathbb{X} - m\rangle] \\
&= \exp\left(-i\langle U\lambda, m\rangle + i\langle U\lambda, m\rangle - \frac{1}{2}\langle U\lambda, \Sigma^2 U\lambda\rangle\right) \\
&= \exp\left(-\frac{1}{2}\langle U\lambda, \Sigma^2 U\lambda\rangle\right) = \exp\left(-\frac{1}{2}\sum_{k=1}^n \lambda_k^2 \sigma_k^2\right),
\end{aligned}$$

d.h. die Komponenten $(Y_k)_{k=1}^n$ sind unabhängig $\mathcal{N}(0, \sigma_k^2)$-verteilt. Gilt $\sigma_k^2 > 0$ für alle $k \in \{1, \ldots, n\}$, so ist Σ^2 positiv definit und $\mathbb{N}(m, \Sigma^2)$ besitzt eine n-dimensionale Dichte f_{m,Σ^2}: Für $B \in ?$ gilt

$$\begin{aligned}
\mathbb{P}(\mathbb{X} \in B) &= \mathbb{P}(U\,\mathbb{Y} + m \in B) = \mathbb{P}(\mathbb{Y} \in U^T(B - m)) \\
&\overset{ref3.26}{=} \int_{U^T(B-m)} \prod_{k=1}^n \frac{\exp\left(-\frac{y_k^2}{2\sigma_k^2}\right)}{\sqrt{2\pi\sigma_k^2}} \, \mathrm{d}y_n \ldots \mathrm{d}y_1 \\
&= \int_B \frac{\exp\left(-\frac{1}{2}\sum_{k=1}^n \frac{(U^T(x-m)_k)^2}{\sigma_k^2}\right)}{(2\pi)^n \frac{n}{2}\sqrt{\det(\Sigma^2)}} \, \mathrm{d}x_n \ldots \mathrm{d}x_1,
\end{aligned}$$

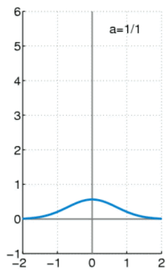


Abb. 28: Für $\sigma_k^2 = 0$ ist $\chi(t) = e^{-\sigma_k^2 \frac{t^2}{2}} = 1$ die charakteristische Funktion des Dirac-Maßes δ_0, die man als Grenzfall der Normalverteilung $\mathbb{N}(0, \sigma_k^2)$ für $\sigma_k^2 \searrow 0$ auffassen kann.

wobei wir für die letzte Gleichheit $\prod_{k=1}^{n} \sigma_k^2 = \det(\Sigma^2)$ und

$$\int f(y)\,\mathrm{d}y_n \ldots \mathrm{d}y_1 = \int f(U^T(x-m))\,\mathrm{d}x_n \ldots \mathrm{d}x_1$$

verwenden. Man sagt, dass $x = Uy + m$ mit Umkehrabbildung $y = U^T(x-m)$ maßerhaltene Transformationen sind. Da

$$U \operatorname{diag}\left(\left(\sigma_k^{-2}\right)_{k=1}^{n}\right) U^T = \Sigma^{-2}$$

folgt

$$\mathbb{P}(\mathbb{X} \in B) = \frac{\exp\left(-\frac{1}{2}\left\langle x-m, (\Sigma^2)^{-1}(x-m)\right\rangle\right)}{(2\pi)^{\frac{n}{2}}}. \qquad \square$$

8 Markov-Ketten

14.06.19

Beispiel 8.0.1 (Irrfahrt auf $\mathbb{Z}$)

Seien $(Y_k \sim \mathrm{Ber}(p))_{k=1}^n$ unabhängig und $p \in (0,1)$. Dann sind auch $(X_k := 2Y_k - 1)_{k=1}^n$ unabhängig mit

$$\mathbb{P}(X_k = 1) = p \qquad \text{und} \qquad \mathbb{P}(X_k = -1) = 1 - p =: q$$

Definiere $S_n := \sum_{k=1}^n X_n$ und $S_0 := 0$.

Irrfahrt

Wir können S_n als eine Irrfahrt, d.h. zufällige Bewegung eines Teilchens auf $\mathbb{Z}$, auffassen. Gegeben die Position S_n zur Zeit n ergibt sich der nächste Schritt zur Position S_{n+1} aus dem Inkrement X_{n+1}. Die Position S_{n+1} hängt nur von S_n und X_{n+1} ab und nicht von $(S_k)_{k=1}^n$, das Teilchen ist gedächtnislos (vgl. (14) und (15)). ◊

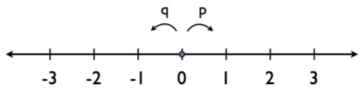


Abb. 29: TODO

DEFINITION 8.0.2 (STOCHASTISCHE MATRIX)

stochastische Matrix

Sei I eine nichtleere höchstens abzählbare Menge. Eine Matrix $P := (p_{ij})_{i,j \in I}$ heißt stochastische Matrix oder Übergangsmatrix

$$p_{ij} \in [0,1] \ \forall i,j \in I \qquad \text{und} \qquad \sum_{j \in I} p_{i,j} = 1 \ \forall i \in I$$

Die Komponenten $p_{i,j}$ heißen Übergangswahrscheinlichkeiten.

DEFINITION 8.0.3 (MARKOV-KETTE (A. A. MARKOV, 1906))

Seien I wie oben und $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Markov-Kette

- Eine Folge $(X_k)_{k \in \mathbb{N}_0}$ I-wertiger Zufallsvariablen heißt (zeitlich homogene) Markov-Kette mit Übergangsmatrix P, wenn für alle $n \in \mathbb{N}_0$ und $(i_k)_{k=0}^{n+1} \subset I$ mit $\mathbb{P}((X_k = i_k)_{k=0}^n) > 0$ gilt

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid (X_k = i_k)_{k=0}^n) = p_{i_n, i_{n+1}}. \tag{14}$$

- I heißt Zustandsraum der Markov-Kette.
- Die Startverteilung ν einer Markov-Kette ist das durch $(\nu_i := \mathbb{P}(X_0 = i))_{i \in I}$ auf I definierte Wahrscheinlichkeitsmaß (also die Verteilung $\mathbb{P}_{X_0}$ von X_0 unter $\mathbb{P}$).
- Ist ν auf einen Punkt konzentriert (d.h. $\nu_i = 1$ für ein $i \in I$), so schreibt man P_i statt P und sagt, dass die Markov-Kette in i startet.

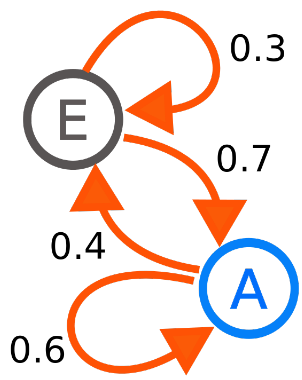


Abb. 30: unnamed figure [Quelle: Wiki]

Satz 8.0.1: Markov-Eigenschaft

Für eine Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$ mit Startverteilung ν gilt

1. Für alle $n \in \mathbb{N}_0$ und alle $(i_k)_{k=0}^n$ gilt

$$\mathbb{P}((X_k = i_k)_{k=0}^n) = \nu_0 \prod_{k=0}^{n-1} p_{i_k,i_{k+1}}.$$

2. Seien $n, m \in \mathbb{N}_0$, $i_n \in I$ und $A \subset I^n, B \subset I^m$. Ist

$$\mathbb{P}((X_k)_{k=0}^{n-1} \in A, X_n = i_n) > 0,$$

so gilt die Markov-Eigenschaft

Markov-Eigenschaft

$$\begin{aligned}\mathbb{P}((X_k)_{k=n+1}^{n+m} \in B \mid (X_k)_{k=0}^{n-1} \in A, X_n = i_n)\\ = \mathbb{P}((X_k)_{k=n+1}^{n+m} \in B \mid X_n = i_n), \qquad (15)\end{aligned}$$

die Zukunft hängt nicht von Vergangenheit, sondern nur von Gegenwart ab.

Beweis. 1 Mit Induktion über $n \in \mathbb{N}_0$. Der Induktionsanfang folgt aus der Definition der Startverteilung.

Induktionsschritt Es gilt

$$\begin{aligned}\mathbb{P}((X_k = i_k)_{k=0}^{n+1}) &= \mathbb{P}(X_{n+1} = i_{n+1} \mid (X_k = i_k)_{k=0}^n)\, \mathbb{P}((X_k = i_k)_{k=0}^n))\\ &\overset{(14)}{=} p_{i_n,i_{n+1}}\, \mathbb{P}((X_k = i_k)_{k=0}^n) \overset{\text{IS}}{=} \nu_{i_0} \prod_{k=0}^{n} p_{i_k,i_{k+1}}\end{aligned}$$

2 Seien $\tilde{A} := \{(i_k)_{k=0}^{m-1} \in A\}$ und $\tilde{B} := \{(i_k)_{k=n+1}^{n+m} \in B\}$. Dann gilt

$$\begin{aligned}&\mathbb{P}(((X_k)_{k=0}^{n-1}, X_n = i_n, (X_k)_{k=n+1}^{n+m} \in B)\\ &= \sum_{\tilde{A}}\sum_{\tilde{B}} \mathbb{P}((X_k)_{k=0}^{n+m}) \overset{1}{=} \sum_{\tilde{A}}\sum_{\tilde{B}} \nu_{i_0} \prod_{k=0}^{n+m-1} p_{i_k,i_{k+1}}\\ &= \left(\sum_{\tilde{A}} \nu_{i_0} \prod_{k=0}^{n-1} p_{i_k,i_{k+1}}\right)\left(\sum_{\tilde{B}} \prod_{k=n}^{n+m-1} p_{i_k,i_{k+1}}\right)\\ &\overset{1}{=} \mathbb{P}((X_k)_{k=0}^{n-1} \in A, X_n = i_n)\left(\sum_{\tilde{B}} \prod_{k=n}^{n+m-1} p_{i_k,i_{k+1}}\right)\end{aligned}$$

Folglich gilt

$$\mathbb{P}((X_k)_{k=n+1}^{n+m} \in B \mid (X_k)_{k=0}^{n-1} \in A, X_n = i_n) = \sum_{\tilde{B}} \prod_{k=n}^{n+m-1} p_{i_k,i_{k+1}}$$

unabhängig von A. Insbesondere gilt

$$\begin{aligned}
&\mathbb{P}((X_k)_{k=n+1}^{n+m} \in B \mid (X_k)_{k=0}^{n-1} \in A, X_n = i_n) \\
&= \mathbb{P}((X_k)_{k=n+1}^{n+m} \in B \mid \underbrace{(X_k)_{k=0}^{n-1} \in I^n}_{\text{immer wahr}}, X_n = i_n) \\
&= \mathbb{P}((X_k)_{k=n+1}^{n+m} \in B \mid X_n = i_n)
\end{aligned}$$

□

20.06.19

DEFINITION 8.0.4 (ZEITLICH INHOMOGENE MARKOV-KETTE) Als Verallgemeinerung kann man die Übergangsmatrix von der Zeit abhängen lassen und erhält eine Familie $P^{(n)} = (p_{i,j}^{(n)})_{i,j}$ für $n \in \mathbb{N}_0$ von Übergangsmatrizen. Nach Satz 8.0.1 gilt

$$\mathbb{P}((X_k = i_k)_{k=0}^{n}) = \nu_{i_0} \prod_{k=0}^{n-1} p_{i_k,i_{k+1}}^{(k)} \tag{16}$$

Eine Folge $(X_k)_{k\in\mathbb{N}}$ I-wertiger Zufallsvariablen, welche (16) erfüllt, erfüllt die MARKOV-Eigenschaft (15) und heißt zeitliche inhomogene MARKOV-Kette.

zeitliche inhomogene MARKOV-Kette

SATZ 8.0.2: EXISTENZ EINER MARKOV-KETTE

Sei P eine stochastische Matrix, ν ein Wahrscheinlichkeitsmaß auf I und $N \in \mathbb{N}$. Dann existiert ein diskreter Wahrscheinlichkeitsraum $(\Omega, \mathbb{P}_\nu)$ und Abbildungen $(X_i : \Omega \to I)_{0\leqslant i\leqslant N}$, sodass $(X_k)_{k=0}^{N}$ eine homogene MARKOV-Kette mit Startverteilung ν und Übergangsmatrix P ist.

Beweis. TODO □

Lemma 8.0.5 (U.i.d Zufallsvariablen sind MARKOV-Kette) *Seien $X := (X_k)_{k=0}^{N}$ u.i.d. I-wertige Zufallsvariablen. Dann ist X eine zeitlich homogene MARKOV-Kette mit Übergangsmatrix $p_{i,j} = p_j := \mathbb{P}(X_0 = j)$ für $j \in I$.*

Beweis. Für $(i_k)_{k=0}^{n+1}$ mit $\mathbb{P}((X_k = i_k)_{k=0}^{n}) > 0$ gilt

$$\begin{aligned}
\mathbb{P}(X_{n+1} = i_{n+1} \mid (X_k = i_k)_{k=0}^{n}) &\overset{2.1.2}{=} \frac{\mathbb{P}((X_k)_{k=0}^{n+1})}{\mathbb{P}((X_k)_{k=0}^{n})} = \frac{\prod_{k=0}^{n+1} \mathbb{P}(X_k = i_k)}{\prod_{k=0}^{n} \mathbb{P}(X_k = i_k)} \\
&= \mathbb{P}(X_{n+1} = i_{n+1}).
\end{aligned}$$

□

Beispiel 8.0.6 (Irrfahrt auf $\mathbb{Z}^d$) Wir können die Situation aus 8.0.1 verallgemeinern, indem wir eine beliebige ganze Zahl als Startpunkt zulassen. Die Übergangsmatrix ist durch $p_{i,j} := p \cdot \mathbb{1}_{j=i+1} + q \cdot \mathbb{1}_{j=i-1}$ gegeben.

Eine weitere Möglichkeit ist, die Irrfahrt auf $d > 1$ Dimensionen zu verallgemeinern. Die Übergangsmatrix ist durch $\frac{1}{2d}\mathbb{1}_{\sum_{k=1}^{d}|i_k - j_k|=1}$ gegeben. $\Diamond$

Irrfahrt

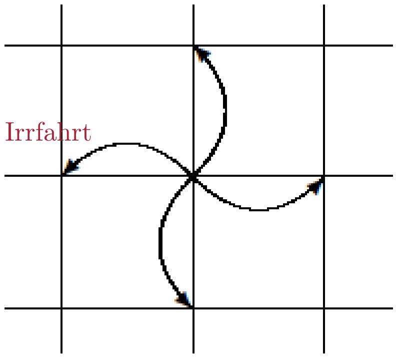
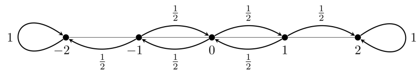


Abb. 31: Irrfahrt auf $\mathbb{Z}^2$ 4[Quelle: Stannat Skript]

Beispiel 8.0.7 (Markov-Ketten mit verschiedenen Rändern)
Sei $\{0, 1, \dots, b\}$ der Zustandsraum für $b \in \mathbb{N}_{>0}$.
Eine Markov-Kette mit Übergangsmatrix wie in Beispiel 8.0.6 für $i \in \{0, \dots, b-1\}$, $p_{0,0} = p_{b,b} = 1$ und $p_{0,j} = p_{b,j} = 0$ für $j \in \{0, \dots, b-1\}$ besitzt einen absorbierenden Rand.

absorbierenden Rand



Eine Markov-Kette mit Übergangsmatrix wie in Beispiel 8.0.6 für $i \in \{0, \dots, b-1\}$ und $p_{0,1} = p_{b,b-1} = 1$ besitzt einen reflektierenden Rand. $\Diamond$

reflektierenden Rand

Abb. 32: Eine Markov-Kette mit absorbierendem Rand. [Quelle: Wiki]

Beispiel 8.0.8 (Ehrenfest-Modell)
Seien n Kugeln auf zwei Schachteln verteilt, davon k in der linken und $n-k$ in der rechten. In jedem Schritt wird zufällig einer der Kugeln gezogen und in die jeweils andere Schachtel gelegt. Der Zustandsraum (für die Anzahl der Kugeln in der linken Schachtel) ist $\{0, \dots, n\}$ und die Übergangsmatrix durch

$$p_{i,j} = \begin{cases} \frac{i}{n}, & \text{wenn } i = j+1, i \in \{1, \dots n\}, \\ \frac{n-i}{i}, & \text{wenn } i = j-1, i \in \{0, \dots, n-1\}, \\ 0, & \text{sonst.} \end{cases} \qquad \Diamond$$

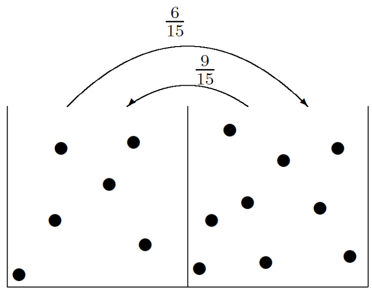


Abb. 33: Ehrenfest Modell [Quelle: Stannat Skript]

8.1 Irreduzibilität

Das folgende Lemma zeigt, dass die Einträge $p_{i,j}^{(n)}$ der Matrix P^n die n-Schritt Übergangswahrscheinlichkeiten der Markov-Kette sind.

Lemma 8.1.1 (n-Schritt Übergangswahrscheinlichkeiten)

Für $m, n \in \mathbb{N}_0$ und $i, j \in I$ mit $P(X_m = i) > 0$ gilt

$$\mathbb{P}(X_{m+n} = j \mid X_m = i) = p_{i,j}^{(n)}.$$

Beweis. Mittels vollständiger Induktion über $n \in \mathbb{N}$.

Induktionsschritt. Sei $\mathcal{K} := \{k \in I : \mathbb{P}(X_{m+1} = k, X_m = i) > 0\}$.

$$\begin{aligned}
\mathbb{P}(X_{n+m}) &\overset{2.1.2}{=} \frac{\mathbb{P}(X_{m+n+1} = j, X_m = i)}{\mathbb{P}(X_m = i)} \\
&\overset{?}{=} \sum_{\mathcal{K}} \frac{\mathbb{P}(X_{m+n+1} = j, X_{m+1} = k, X_m = i)}{\mathbb{P}(X_{m+1} = k, X_m = i)} \cdot \frac{\mathbb{P}(X_{m+1} = k, X_m = i)}{\mathbb{P}(X_m = i)} \\
&\overset{2.1.2}{=} \sum_{\mathcal{K}} \mathbb{P}(X_{m+n+1} = j \mid X_{m+1} = k, X_m = i) \cdot \mathbb{P}(X_{m+1} = k \mid X_m = i) \\
&\overset{(15)}{=} \sum_{\mathcal{K}} \mathbb{P}(X_{m+n+1} = j \mid X_{m+1} = k) p_{i,k} = \sum_{k \in I} p_{k,j}^{(n)} p_{i,k} = p_{i,j}^{(n+1)},
\end{aligned}$$

wobei wir in der vorletzten Gleichung verwendet haben, dass

$$\mathbb{P}(X_{m+1} = k, X_m = i) = \underbrace{\mathbb{P}(X_{m+1} = k \mid X_m = i)}_{=p_{i,k}} \underbrace{\mathbb{P}(X_m = i)}_{>0} > 0$$

genau dann gilt, wenn $p_{i,k} > 0$ ist. □

Beispiel 8.1.2 (Chapman-Kolmogorov Gleichung)

Seien $m, n \in \mathbb{N}_0$ und $i, j \in I$. Aus der Assoziativität des Matrizenprodukts folgt die Chapman-Kolmogorov Gleichung

$$p_{i,j}^{(n+m)} = \sum_{k \in I} p_{i,k}^{(m)} p_{k,j}^{(n)}$$

und daraus

$$p_{i,j}^{(n+m)} \geqslant p_{i,k}^{(m)} p_{k,j}^{(n)} \tag{17}$$

für alle $k \in I$. ◊

Seien im Folgenden stets P eine Übergangsmatrix auf I und $i, j \in I$.

Definition 8.1.3 (Erreichbarkeit ($i \rightsquigarrow j$))

Existiert ein $n \in \mathbb{N}_0$ mit $p_{i,j}^{(n)} > 0$, so heißt j von i aus erreichbar.

von i aus erreichbar

Korollar 8.1.4 ($\rightsquigarrow$ ist eine Quasiordnung)

Die Relation $\rightsquigarrow$ ist reflexiv, transitiv aber nicht symmetrisch.

Beweis. Reflexivität: Da $P^0 = E_I$ ist, gilt $p_{ii}^{(0)} = 1$ für alle $i \in I$.

Transitivität: Gilt $i \rightsquigarrow j$ und $j \rightsquigarrow k$ so folgt nach Definition die Existenz von $m, n \in \mathbb{N}$ mit $p_{i,j}^{(m)}, p_{j,k}^{(n)} > 0$. Aus (17) folgt $p_{i,k}^{(m+n)} > p_{i,j}^{(m)} p_{j,k}^{(n)} > 0$.

Symmetrie: Betrachte die eindimensionale Irrfahrt mit absorbierendem Rand. Die Randpunkte sind von allen inneren Punkten aus erreichbar aber nicht umgekehrt. □

DEFINITION 8.1.5 (ÄQUIVALENTE ZUSTÄNDE)

Die Zustände i, j sind äquivalent ($i \leftrightsquigarrow j$), wenn $i \rightsquigarrow j \rightsquigarrow i$ gilt. Wir schreiben $A \rightsquigarrow B$ wenn Zustände $i \in A$ und $j \in B$ mit $i \rightsquigarrow j$ existieren.

äquivalent

Korollar 8.1.6 ($\leftrightsquigarrow$ ist eine Äquivalenzrelation)

$\leftrightsquigarrow$ ist eine Äquivalenzrelation auf I. Die Notation $A \rightsquigarrow B$ für Äquivalenzklassen A, B ist unabhängig von den Repräsentanten $i \rightsquigarrow j$.

Beweis. Die erste Aussage folgt direkt aus Korollar 8.1.4. **TODO:** zweite Aussage. □

DEFINITION 8.1.7 (IRREDUZIBILITÄT)

P (und jede zugehörige MARKOV-Kette) heißt irreduzibel, wenn je zwei Elemente aus I äquivalent sind.

irreduzibel

Beispiel 8.1.8 (eindimensionale Irrfahrt ist irreduzibel)

Die eindimensionale Irrfahrt ist irreduzibel, denn für $i < j$ gilt

$$p_{i,j}^{(j-i)} = \mathbb{P}(S_{j-i} = j \mid S_0 = i) = p^{j-i} > 0$$

und analog für $j < i$

$$p_{i,j}^{(i-j)} = \mathbb{P}(S_{i-j} = j \mid S_0 = i) = q^{i-j} > 0.$$

Aus der Irreduzibilität folgt direkt, dass die symmetrische einfache Irrfahrt keine absorbierenden Zustände besitzt. ◊

DEFINITION 8.1.9 (ABGESCHLOSSENHEIT)

Eine nichtleere Menge $J \subset I$ heißt abgeschlossen, falls keine Zustände $j \in J$ und $i \in I \backslash J$ existieren, sodass $j \rightsquigarrow i$ gilt.

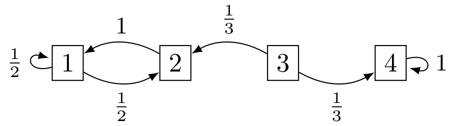
abgeschlossen

Beispiel 8.1.10 (Äquivalenzklassen der $\mathbb{Z}$-Irrfahrt mit abs. Rand)

Eine eindimensionale Irrfahrt mit absorbierendem Rand besitzt drei Äquivalenzklassen: $\{0\}$, $\{b\}$ (beide abg.) und $\{1, \dots, b-1\}$ (nicht abg.).◊

Beispiel 8.1.11 (Äquivalenzklassen am Übergangsgraph)

Anhand dem jeder Markov-Kette mit endlichem Zustandsraum zuordenbaren Übergangsgraphen sind Äquivalenzklassen oft leichter zu erkennen: Sei $I := \{1, 2, 3, 4\}$ und



$$P := \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Die drei Äquivalenzklassen sind $\{1, 2\}$, $\{3\}$ und $\{4\}$, da 3 nicht von 1 aus erreichbar ist und 3 nicht von 4. Der Zustand 4 ist absorbierend. $\diamondsuit$

8.2 Stationäre Verteilungen

Definition 8.2.1 (stationäre Markov-Kette)

Eine Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$ heißt stationär, falls für beliebige $m \in \mathbb{N}$ und $r \in \mathbb{N}_0$ die Zufallsvektoren $(X_k)_{k=0}^r$ und $(X_k)_{k=m}^{m+r}$ dieselbe Verteilung besitzen.

stationär

Korollar 8.2.2 (Rückrichtung von Lemma 8.0.5 (?))

Für eine stationäre Markov-*Kette sind insbesondere die einzelnen Zufallsvariablen identisch verteilt.*

Lemma 8.2.3 (Charakterisierung stationären Markov-Ketten)

Eine Markov-*Kette $(X_n)_{n\in\mathbb{N}_0}$ mit Anfangsverteilung ν und Übergangsmatrix P über I ist genau dann stationär, wenn*

$$\sum_{j\in I} \nu_j p_{j,i} = \nu_i$$

für alle $i \in I$ gilt.

Beweis. „$\implies$“: Sei $(X_n)_{n\in\mathbb{N}_0}$ stationär. Dann besitzen insbesondere X_0 und X_1 die selben Verteilung ν. Nach Satz 8.0.1 folgt

$$\begin{aligned}\nu_i &\overset{\text{Def.}}{=} \mathbb{P}(X_0 = i) \overset{\text{(i.v.)}}{=} \mathbb{P}(X_1 = i) \overset{(2.1.2)}{=} \sum_{j\in I} \mathbb{P}(X_1 = i, X_0 = j) \\ &\overset{(2.1.2)}{=} \sum_{j\in I} \mathbb{P}(X_1 = i \mid X_0 = j)\,\mathbb{P}(X_0 = j) \overset{\text{Def.}}{=} \sum_{j\in I} p_{j,i}\nu_j.\end{aligned}$$

„$\impliedby$“: Sei $\sum_{j\in I} \nu_j p_{j,i} = \nu_i$ für alle $i \in I$. Nach Satz 8.0.1 gilt

$$\mathbb{P}((X_k = i_k)_{k=0}^r) = \nu_{i_0} \prod_{k=0}^{r-1} p_{i_k,i_{k+1}} \tag{18}$$

sowie

$$\begin{aligned}\mathbb{P}((X_{m+k} = i_k)_{k=0}^r) &= \sum_{(j_k)_{k=0}^{m-1}\in I} \mathbb{P}((X_k = j_k)_{k=0}^{m-1}, (X_k = i_k)_{k=0}^r) \\ &= \sum_{(j_k)_{k=0}^{m-1}\in I} \nu_{j_0} \prod_{k=0}^{m-1} p_{j_k,j_{k+1}} \cdot i_0 \cdot \prod_{k=0}^{r-1} p_{i_k,i_{k+1}} \\ &\prod_{k=0}^{r-1} p_{i_k,i_{k+1}} \cdot \left(\sum_{(j_k)_{k=0}^{m-1}\in I} \nu_{j_0} \prod_{k=0}^{m-2} p_{j_k,j_{k+1}} \cdot p_{j_{m-1},i_0} \right).\end{aligned} \tag{19}$$

Nach Annahme gilt nun

$$\sum_{(j_k)_{k=0}^{m-1}\in I} \nu_{j_0} \prod_{k=0}^{m-2} p_{j_k,j_{k+1}} \cdot p_{j_{m-1},i_0} = \sum_{(j_k)_{k=m-1}^{1}\in I} \underbrace{\left(\sum_{j_0\in I} \nu_{j_0} p_{j_0,j_1}\right)}_{=\nu_{j_1}} \cdot \prod_{k=1}^{m-2} p_{j_k,j_{k+1}} \cdot p_{j_{m-1},i_0}$$
$$= \sum_{(j_k)_{k=m-1}^{1}\in I} \nu_{j_1} \cdot \prod_{k=1}^{m-2} p_{j_k,j_{k+1}} \cdot p_{j_{m-1},i_0}$$
$$= \ldots = \nu_{i_0}.$$

Einsetzen in (19) ergibt

$$\mathbb{P}((X_{k+m} = i_k)_{k=0}^{r}) = \nu_{i_0} \prod_{k=0}^{r-1} p_{i_k,i_{k+1}} \overset{(18)}{=} \mathbb{P}((X_k = i_k)_{k=0}^{r}) \qquad \square$$

Definition 8.2.4 (stationäre Verteilung)

Eine Wahrscheinlichkeitsverteilung ν auf einem Zustandsraum I heißt stationär (invariant) bezüglich der Übergangsmatrix P, wenn

$$\sum_{j\in I} \nu_j p_{j,i} = \nu_i \;\forall i \in I. \tag{20}$$

stationär

Korollar 8.2.5 (Charakterisierung st. Verteilungen)

Die stationäre Verteilungen sind genau die nichtnegativen Eigenvektoren von P^T zum Eigenwert 1, deren Komponentensumme 1 beträgt.

Beispiel 8.2.6 (stationären Verteilungen der Irrfahrt)

Betrachte die Irrfahrt für $p \in (0,1)$ und $q := 1 - p$. auf

- $I := \{0, \ldots, b\}$ mit absorbierendem Rand. Beim Lösen des Gleichungssystems

$$\begin{cases} \nu_0 + q\nu_1 = \nu_0, & p\nu_{b-3} + q\nu_{b-1} = \nu_{b-2}, \\ q\nu_2 = \nu_1, & p\nu_{b-2} = \nu_{b-1}, \\ p\nu_1 + q\nu_3 = \nu_2, & p\nu_{b-1} + \nu_b + \nu_b \\ \vdots & \end{cases}$$

  erhält man sukzessive $\nu_1 = \ldots = \nu_{n-1} = 0$. Man erhält die einparametrige Schar stationärer Verteilung

$$(\nu_\gamma := (\gamma, 0, \ldots, 0, 1-\gamma))_{\gamma\in[0,1]}$$

- $I := \mathbb{N}_0$ mit Reflexion in 0. Wir erhalten das Gleichungssystem

$$\begin{cases} q\nu_1 = \nu_0, \\ \nu_0 + q\nu_2 = \nu_1, \\ p\nu_{k-1} + q\mu_{k+1} = \nu_k. \end{cases} \implies \begin{cases} \nu_1 - \nu_0 = \frac{p}{q}\nu_0, \\ \nu_2 - \nu_1 = \frac{p}{q}(\nu_1 - \nu_0) - \nu_0, \\ \nu_{k+1} - \nu_k = \frac{p}{q}(\nu_k - \nu_{k-1}), \quad k > 1 \end{cases}$$

Mit $\gamma := \nu$ und $j := \frac{p}{1}$ folgt

$$\begin{cases} \nu_1 - \nu_0 = j\gamma, \\ \nu_2 - \nu_1 = j^2\gamma - \gamma, \\ \nu_{k+1} - \nu_k = j^{k+1}\gamma - j^{k-1}\gamma, \quad k > 1. \end{cases}$$

Für $k \geqslant 1$ impliziert das $\nu_k = j^k\gamma + j^{k-1}\gamma$. Also ist

- $\gamma > 0$, da sonst $\nu \equiv 0$ kein Wahrscheinlichkeitsmaß wäre,
- $p < q$, da sonst $\nu_k \geqslant \gamma$ für alle $k \in \mathbb{N}$ also $\sum_{k\in\mathbb{N}} \nu_k = \infty \neq 1$ gelten würde, eine Widerspruch dazu, dass ν ein Wahrscheinlichkeitsmaß ist.

Unter diesen Bedingung gilt mit Indexverschiebung

$$1 = \sum k = 0^{\infty}\nu_k = 2\gamma \sum_{k=0}^{\infty} j^k = \frac{2\gamma}{1-j} \implies \gamma = \frac{1-j}{2}.$$

Damit ist die stationäre Verteilung durch

$$\nu_k = \frac{1}{2}\begin{cases} 1-j, & k = 0, \\ j^{k-1} - j^{k+1}, & k \geqslant 1 \end{cases}$$

gegeben.

Im Allgemeinen existiert somit nicht immer eine stationäre Verteilung und auch wenn sie existiert, muss sie nicht eindeutig sein. $\Diamond$

SATZ 8.2.1: ENDLICHE MK BESITZT ST. VERTEILUNG

Sei I endlich. Dann besitzt jede MARKOV-Kette auch I mindestens eine stationäre Verteilung.

Beweis. Seien o.B.d.A. $I := \{1, \dots, N\}$ und P die Übergangsmatrix einer MARKOV-Ketteauf I. Die Menge der Wahrscheinlichkeitsmaße auf I können wir mit dem $N-1$-dimensionalen Simplex

$$\Delta_N := \left\{ (\nu_k)_{k=1}^N \in \mathbb{R}^N : \nu_j \geqslant 0, \ \sum_{k=1}^N \nu_k = 1 \right\}$$

identifizieren. Sei nun $\nu^{(0)}$ eine beliebige Startverteilung und $\nu^{(n)} := \nu^{(0)}P^n$ für $n \geqslant 1$.

Dann ist auch das CESARO-Mittel $\pi^{(n)} := \frac{1}{n}\sum_{k=0}^{n-1} \nu^{(k)}$ eine Wahrscheinlichkeitsmaß auf I, da Linearkombinationen von Maßes mit nichtnegativen Koeffizienten wieder Maße sind.

Nach dem Satz von BOLZANO-WEIERSTRASS besitzt die Folge $(\pi^{(n)})_{n\in\mathbb{N}}$ eine konvergente Teilfolge $(\pi^{(n_k)})_{k\in\mathbb{N}}$, welche gegen π konvergiert. Aufgrund der Abgeschlossenheit von Δ_N ist $\pi \in \Delta_N$ ein Wahrscheinlichkeitsmaß auf I. Aus

$$\pi^{(n_k)}P = \frac{1}{n_k}\sum_{k=0}^{n_k-1}\nu^{(0)}P^{k+1} = \frac{1}{n_k}\sum_{k=1}^{n_k}\nu^{(0)}P^{k} = \pi^{(n_k)} + \frac{\nu^{n_k} - \nu^{(0)}}{n_k} \quad (21)$$

folgt

$$\pi P \overset{\text{Def.}}{=} \lim_{k\to\infty}\lim_{k\to\infty}\pi^{(n_k)}P \overset{(21)}{=} \lim_{k\to\infty}\lim_{k\to\infty}\pi^{(n_k)} + \frac{\nu^{n_k} - \nu^{(0)}}{n_k} = \pi$$

und somit die Invarianz von π. $\square$

8.3 Der Konvergenzsatz für Markov-Ketten auf endlichen Zustandsräumen

Satz 8.3.1: Konvergenzsatz für endlichen Markov-Ketten

Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette auf einem endlichem Zustandsraum I mit Übergangsmatrix P. Existiert ein $r \in \mathbb{N}$, sodass die Matrix P^r mindestens eine Spalte mit strikt positiven Einträgen besitzt, gilt

1. Für beliebige $i, j \in I$ existieren die von i unabhängigen Grenzwerte
$$\lim_{n\to\infty} p_{i,j}^{(n)} = \nu_j.$$
2. $\nu = (\nu_i)_{i\in I}$ ist die einzige stationäre Verteilung von $(X_n)_{n\in\mathbb{N}_0}$.

Beweis. Sehr lang. □

Wir untersuchen nun, was die Voraussetzung an die Positivität einer Spalte von P^r konkret bedeutet.

Definition 8.3.1 ((Un)wesentliche Zustände)

Ein Zustand i heißt wesentlich, falls aus $i \rightsquigarrow j$ stets $j \rightsquigarrow i$ folgt. wesentlich

Lemma 8.3.2 (Äquivalente Zustände auch (un)wesentlich)

Aus der (Un)wesentlichkeit von $i \in I$ folgt die (Un)wesentlichkeit aller äquivalenten Zustände.

Beweis. Sei i wesentlich und j zu i äquivalent. Angenommen, es gilt $j \rightsquigarrow k$. Aus $i \rightsquigarrow j$ und $j \rightsquigarrow k$ folgt $i \rightsquigarrow k$ und damit $k \rightsquigarrow i$, weil i wesentlich ist. Aus $j \rightsquigarrow i$ folgt deshalb $k \rightsquigarrow j$. □

Die Einschränkung der Übergangsmatrix P auf eine wesentliche Klasse $W \subset I$, P^W, ist wieder eine Übergangsmatrix, d.h. die Zeilensummen sind 1, denn für alle $i \in W$ und $j \notin W$ gilt $p_{i,j} = 0$. Insbesondere ist die Einschränkung irreduzibel.

Lemma 8.3.2 liefert eine Zerlegung des Zustandsraums I in wesentliche (Äquivalenz)Klassen $(K_i)_{i=1}^{\omega}$ und unwesentliche Klassen $(K_i)_{i=\omega+1}^{\omega+n}$. Die Übergangsmatrix kann nach entsprechender Nummerierung der Zustände auf folgenden Gestalt transformiert werden:

$$P = \left[\begin{array}{ccc|ccc} P_{11} & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & P_{\omega,\omega} & & & \\ \hline & & & P_{\omega+1,\omega+1} & & * \\ & * & & & \ddots & \\ & & & * & & P_{\omega+n,\omega+n} \end{array}\right]$$

Dabei sind $P_{\ell,\ell}$ die zu den Klassen K_ℓ gehörenden Teilmatrizen. Insbesondere sind die Teilmatrizen $(P_{\ell,\ell})_{\ell=1}^{\omega}$ zu den wesentlichen Klassen stochastisch und somit sind alle weiteren Elemente in der oberen Hälfte der Matrix 0. Die Teilmatrizen $(P_{\ell,\ell})_{\ell=\omega+1}^{\omega+n}$ zu den unwesentlichen Klassen sind nicht stochastisch und daher können die mit $*$ markierten Einträge von Null verschieden sein.

Beispiel 8.3.3 (Motivation: Periodizität vs. Konvergenzsatz)

Für die Beurteilung des Langzeitverhaltens der Markov-Kette zu erfassen genügt es nicht, irreduzible Ketten einer wesentlichen Klasse zu betrachten: Seien $I := \{1,2\}$ und $P := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, welche irreduzibel ist. Es gilt $P^{2k+1} = P$ und $P^{2k} = E_2$ für $k \in \mathbb{N}_0$.

Es sind also weder die Voraussetzungen noch die Aussagen des Konvergenzsatzes 8.3.1 erfüllt. $\Diamond$

Definition 8.3.4 (Periode von Markov-Ketten)

Sei $i \in I$. Die Zahl $d_i := \mathrm{ggT}\{n \in \mathbb{N} : p_{i,i}^{(n)} > 0\}$ heißt Periode von i. Der Zustand heißt aperiodisch, wenn $d_i = 1$ gilt.

Haben alle Zustände der Markov-Kette Periode eins, so heißt diese aperiodisch. Haben alle Zustände dieselbe Periode $d > 1$, so heißt die Markov-Kette periodisch mit Periode d.

Lemma 8.3.5 (Äquivalente Zustände teilen Periode)

Seien i und j äquivalente Zustände. Dann folgt $d_i = d_j$.

Beweis. Nach Voraussetzung existieren $k, \ell \in \mathbb{N}$ mit $p_{i,j}^{(k)}, p_{i,j}^{(\ell)} > 0$. Sei $\Lambda_i := \{m \in \mathbb{N} : p_{i,i}^{(m)} > 0\}$.

Dann folgt für $n \in \Lambda_j$

$$p_{i,i}^{(n+d_i(k+\ell))} \overset{(17)}{\geqslant} p_{i,j}^{(k)} p_{j,j}^{(n)} p_{j,i}^{(\ell)} \cdot \prod_{c=1}^{d_i-1} p_{i,j}^{(k)} p_{i,j}^{(\ell)} > 0.$$

Also gilt $n + d_i(k+\ell) \in \Lambda_i$ und somit $d_i \mid n$. Da dies für alle $n \in \Lambda_j$ gilt, gilt $d_i \mid d_j$ und somit $d_i \leqslant d_j$. Analog zeigt man $d_j \leqslant d_i$. $\square$

Lemma 8.3.6

Gilt $d = \mathrm{ggT}\{(n_k)_{k\in\mathbb{N}}\}$*, so existieren* $K, L \in \mathbb{N}$ *sodass für jedes* $\ell \geqslant L$ *Zahlen* $c_k \in \mathbb{N}$ *und Indizes existieren, sodass* $\ell d = \sum_{k=1}^{K} c_k n_k$ *gilt.*

Beweis. Nicht relevant. □

Satz 8.3.2

Sei $(X_n)_{n\in\mathbb{N}_0}$ eine irreduzible aperiodische Markov-Kette mit endlichem Zustandsraum I. Dann existiert ein $r \in \mathbb{N}$, sodass alle Elemente der Matrix P^r strikt positiv sind.

Bemerkung Insbesondere sind die Voraussetzungen des Konvergenzsatzes erfüllt und die Markov-Kette konvergent.

A Appendix

A.1 Maßtheoretische Grundlagen

σ-Algebren, Inhalte und (Prä)maße

vgl. erste Übung

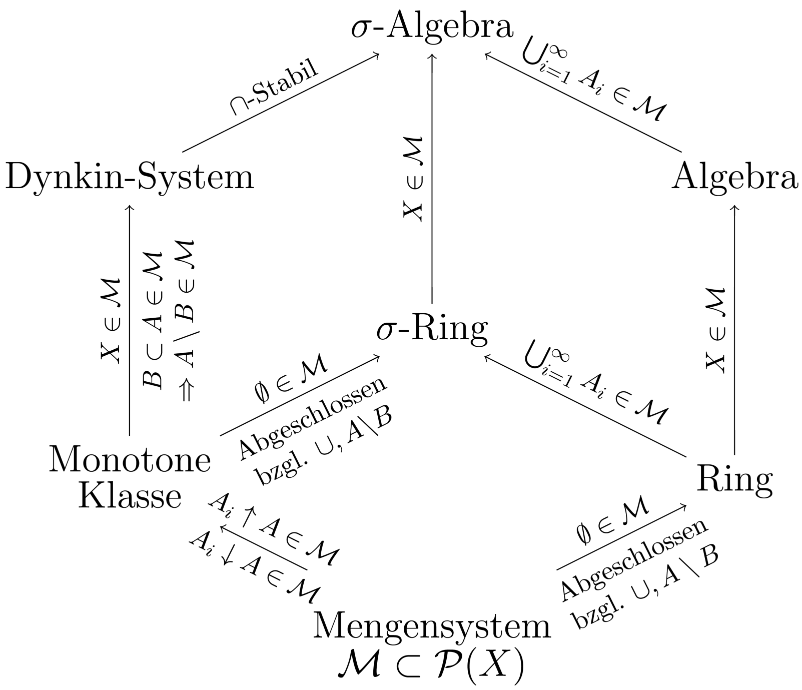


Abb. 34: Quelle: arbeitsbuch achim klenke glaub ich

Konstruktion von Maßen und Eindeutigkeitssatz

vgl. zweite Übung

Satz von Vitali

Satz A.1.1: Vitali, 1905

Hier Satz 1.15

Beweis. TODO □

Konstruktion des Integrals

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ eine Wahrscheinlichkeitsraum und $X : \Omega \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Zufallsvariable. Gilt $\mathbb{E}(|X|) < \infty$, man sagt $\mathbb{E}(X)$ „existiert", was zu äquivalent zu $X \in L^1(\Omega, \mathcal{P})$ ist, setzt man

$$\mathbb{E}[X] := \int_\Omega X(\omega) d\,\mathbb{P}(\omega)$$

Konstruktion des Bildmaßes d $\mathbb{P}$.

1. Sei $A \in \mathcal{F}$ und $X := \mathbb{1}_A$. Dann definieren wir
$$\mathbb{E}[\mathbb{1}_A] := \int_A \mathrm{d}\,\mathbb{P} = \mathbb{P}(A).$$
2. Sei $X := \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$ eine einfache Funktion, wobei $\alpha_k \geqslant 0$ und $A_k \in \mathcal{F}$ für alle $k \in \{1, \ldots, n\}$ paarweise disjunkt seien sollen und $\Omega = \bigsqcup_{k=1}^n A_k$ gilt. Dann gilt (oder Definition??)
$$\mathbb{E}[X] = \sum_{k=1}^n \alpha_k \mathbb{P}(A_k).$$
3. Sei X eine nichtnegative Zufallsvariable. Dann existiert eine Folge einfacher Zufallsvariablen $(x_n)_{n\in\mathbb{N}}$ so dass $x_n(\omega) \overset{n\to\infty}{\nearrow} X(\omega)$ für alle $\omega \in \Omega$ gilt. So eine Folge ist z.B. durch $x_n(\omega) := k2^{-n}\,\mathbb{1}_{[k2^{-n},(k+1)2^{-n}]}$ gegeben. Dann definieren wir
$$\mathbb{E}(X) := \lim_{n\to\infty} \mathbb{E}[x_n] = \sup_{n\in\mathbb{N}} \mathbb{E}[x_n].$$
4. Zerlegung in Positiv- und Negativteil.

Transformationssatz für Maße Seien $(\Omega, \mathcal{F}, \mathcal{P})$ ein Wahrscheinlichkeitsraum $X : \Omega \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Zufallsvariable und $h : \mathbb{R} \to \mathbb{R}$ messbar. Dann gilt

$$\int_{X^{-1}(A)} h(X)\,\mathrm{d}\,\mathbb{P} = \int_A h(z)\,\mathrm{d}\,\mathbb{P}_X(z) = \int_\mathbb{R} h(z) f_X(z)\,\mathrm{d}(\lambda(z)),$$

wobei die letzte Gleichheit nur gilt, wenn X eine Dichte f_X besitzt.

Diskrete Zufallsvariablen sind nur ein Spezialfall: Ist X diskret mit Werten in I und eine Zähldichte $p_k = \mathbb{P}(X = k)$ für $k \in I$ gegeben, so ist

$$\mathbb{E}[X] = \sum_{k\in I} k \cdot \mathbb{P}(X = k) \quad \text{und} \quad \mathrm{Var}[X] = \sum_{k\in I} (k - \mathbb{E}[X])^2\, \mathbb{P}(X = k)$$

Beweis des Transformationssatzes

Große Übung, 29.05.19

SATZ A.1.2: TRANSFORMATIONSSATZ, VERSION I

Sei $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine messbare Funktion.

(a) Dann gilt

$$\mathbb{E}[h(X)] \overset{\text{Def.}}{=} \int_\Omega h(X)\,\mathrm{d}\mathbb{P} = \int_\mathbb{R} h(x)\,\mathbb{P}_X(\mathrm{d}x),$$

wobei $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$ das Bildmaß von X auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist. Bildmaß

(b) Hat X (bzw. $\mathbb{P}_X$) eine Dichte f_X, so gilt

$$\int_\mathbb{R} h(x)\,\mathbb{P}_X(\mathrm{d}x) = \int_\mathbb{R} f_X(x) h(x)\,\mathrm{d}x.$$

Beispiel A.1.1 (k-tes Moment der Exponentialverteilung)
Sei $X \sim \mathrm{Exp}(\lambda)$ für $\lambda > 0$ ein exponentialverteilte Zufallsvariable. Dann gilt für alle $k \in \mathbb{N}$

$$\mathbb{E}[X^k] = \int_\Omega X^k\,\mathrm{d}\mathbb{P} \overset{\text{(a)}}{=} \int_\mathbb{R} x^k\,\mathbb{P}_X(\mathrm{d}x) \overset{\text{(b)}}{=} \int_{\mathbb{R}^+} x^k \cdot \lambda \cdot e^{-\lambda}\,\mathrm{d}x = k! \cdot \lambda^{-k},$$

wobei man das Integral über k-fache partielle Integration bzw. über vollständige Induktion ausrechnet. ◊

Bermerkung A.1.2 Der Erwartungswert $\mathbb{E}[X]$ hängt also nur von der Verteilung, nicht von der Abbildung X selbst ab!

Beweis. Wir führen eine maßtheoretische Induktion über h:

1 Sei $h := \mathbb{1}_A$ für $A \in \mathcal{B}(\mathbb{R})$.

(a) Dann gilt $h(X) = \mathbb{1}_A(X) = \mathbb{1}_{X \in A}$ (‡) und somit

$$\begin{aligned}\mathbb{E}[h(X)] &\overset{\text{Def.}}{=} \int_\Omega h(X)\,\mathrm{d}\mathbb{P} \overset{(\dagger)}{=} \int_\Omega \mathbb{1}_{X\in A}\,\mathrm{d}\mathbb{P} = \mathbb{P}(X \in A) = \mathbb{P}_X(A) \\ &= \int_\mathbb{R} \mathbb{1}_A(x)\,\mathbb{P}(\mathrm{d}x) = \int_\mathbb{R} h(x)\,\mathbb{P}(\mathrm{d}x).\end{aligned}$$

(b) Es gilt

$$\mathbb{E}[h(X)] \overset{\text{(a)}}{=} \mathbb{P}(X \in A) = \int_A f_X(x)\,\mathrm{d}x = \int_\mathbb{R} h(x) f_X(x)\,\mathrm{d}x.$$

2 Sei $n \in \mathbb{N}$, $\alpha_k \geqslant 0$ und $A_k \in \mathcal{B}(\mathbb{R})$ paarweise disjunkt für alle $k \in \{1, \ldots, n\}$ und $h := \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$. Dann ist h und somit auch $h \circ X$ messbar.

(a) Dann gilt

$$\begin{aligned}\mathbb{E}[h(X)] &= \int_\Omega \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}(X)\,\mathrm{d}\mathbb{P} \overset{\text{L}}{=} \sum_{k=1}^n \alpha_k \int_\Omega \mathbb{1}_{A_k}(X)\,\mathrm{d}\mathbb{P} \\ &\underset{\text{L}}{\overset{\text{(a)}}{=}} \int_\Omega \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}(x)\,\mathbb{P}_X(\mathrm{d}x) = \int_\mathbb{R} h(x)\,\mathbb{P}_X(\mathrm{d}x)\end{aligned}$$

(b) und

$$\begin{aligned}\mathbb{E}[h(X)] &\overset{(a)}{=} \int_{\mathbb{R}} h(x)\mathcal{P}_X(\mathrm{d}x) \overset{\mathrm{L}}{=} \sum_{k=1}^{n} \alpha_k \int_{\Omega} \mathbb{1}_{A_k}(X)\,\mathbb{P}_X(\mathrm{d}x)\\ &= \sum_{k=1}^{n} \alpha_k \int_{\mathbb{R}} \mathbb{1}_{A_k}(x) f_X(x)\,\mathrm{d}x \overset{\mathrm{L}}{=} \int_{\mathbb{R}} h(x) f_X(x)\,\mathrm{d}x.\end{aligned}$$

3 Sie $h : \mathbb{R} \to \mathbb{R}$ nichtnegativ und messbar. Dann existiert $(h_n : \mathbb{R} \to \mathbb{R})_{n\in\mathbb{N}}$ eine punktweise monoton wachsenden Folge einfacher Funktionen mit $h_n(x) \nearrow h(x)$ für alle $x \in \mathbb{R}$. Dann ist $(h_n(X) : \Omega \to \mathbb{R})_{n\in\mathbb{N}}$ eine monoton wachsende Folge nichtnegativer Elementarfunktionen und es gilt $h_n(X(\omega) \nearrow h(X(\omega))$ für alle $\omega \in \Omega$.

(a) Dann gilt

$$\begin{aligned}\int_{\Omega} h(X)\,\mathrm{d}\mathbb{P} &\overset{\text{Def.}}{=} \lim_{n\to\infty} \int_{\Omega} h_n(X)\,\mathrm{d}\mathbb{P} \overset{2}{=} \lim_{n\to\infty} \int_{\mathbb{R}} h_n(x)\,\mathbb{P}_X(\mathrm{d}x)\\ &\overset{\text{Def.}}{=} \int_{\mathbb{R}} h(x)\,\mathbb{P}_X(\mathrm{d}x)\end{aligned}$$

(b) und

$$\begin{aligned}\mathbb{E}[h(X)] &\overset{(a)}{=} \int_{\mathbb{R}} h(x)\,\mathbb{P}_X(\mathrm{d}x) = \overset{\text{Def.}}{=} \lim_{n\to\infty} \int_{\mathbb{R}} h_n(x)\,\mathbb{P}_X(\mathrm{d}x)\\ &\overset{2}{=} \lim_{n\to\infty} \int_{\mathbb{R}} h_n(x) f_X(x)\,\mathrm{d}x \overset{\text{Def.}}{=} \int_{\mathbb{R}} h(x) f_X(x)\,\mathrm{d}x.\end{aligned}$$

4 Sei $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine messbare Funktion. Falls definiert, gilt $h = h^+ - h^-$.

(a) Dann gilt

$$\begin{aligned}\mathbb{E}[h(X)] &\overset{\mathrm{L}}{=} \mathbb{E}[h^+(X)] - \mathbb{E}[h^-(X)]\\ &\underset{3}{\overset{\text{Def.}}{=}} \int_{\mathbb{R}} h^+(x)\,\mathbb{P}_X(\mathrm{d}x) - \int_{\mathbb{R}} h^-(x)\,\mathbb{P}_X(\mathrm{d}x)\\ &\overset{\mathrm{L}}{=} \int_{\mathbb{R}} h(x)\,\mathbb{P}_X(\mathrm{d}x).\end{aligned}$$

(b) und

$$\begin{aligned}\mathbb{E}[h(X)] &\overset{(a)}{=} \int_{\mathbb{R}} h^+(x)\,\mathbb{P}_X(\mathrm{d}x) - \int_{\mathbb{R}} h^-(x)\,\mathbb{P}_X(\mathrm{d}x)\\ &\overset{3}{=} \int_{\mathbb{R}} h^+(x) f_X(x)\,\mathrm{d}x - \int_{\mathbb{R}} h^-(x) f_X(x)\,\mathrm{d}x\\ &\overset{\mathrm{L}}{=} \int_{\mathbb{R}} h(x) f_X(x)\,\mathrm{d}x. \qquad \square\end{aligned}$$

Das geht natürlich allgemeiner:

Satz A.1.3: Transformationssatz, Version 2

Seien := $(F, \mathcal{F}, \mu)$ und $\tilde{G} := (G, \mathcal{G}, \nu)$ Maßräume, $f : \tilde{F} \to \tilde{G}$ und $g : \tilde{G} \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar und $g \in L^1(\tilde{G})$. Dann gilt

$$\int_G g \, \mathrm{d}\nu = \int_F g \circ f \, \mathrm{d}\mu,$$

falls $\nu = \mu \circ f$, d.h. $\nu(A) = \mu(f \in A)$[a] für alle $A \in \mathcal{G}$.

[a] vgl. absolutstetiges Maß und Satz von Radon-Nykodym.

Bei uns war $f := X$, $\tilde{F} := (\Omega, \mathcal{A}, \mathbb{P})$, $g := h$ und $\tilde{G} := (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$. Besitzt ν eine Dichte, d.h. eine messbare Funktion $f_\nu : G \to [0, \infty)$ sodass

$$\nu(A) = \int_G \mathbb{1}_A(x) f_\nu(x) \, \mathrm{d}x$$

für alle $A \in G$ gilt, so folgt

$$\int_G g \, \mathrm{d}\nu = \int_G g f_\nu \, \mathrm{d}\lambda(x), \text{d.h.} \int_G g(x)\nu(\mathrm{d}x) = \int_G g(x) f_\nu(x) \, \mathrm{d}x.$$

Bermerkung A.1.3 Seien $G, G' \subset \mathbb{R}$ offene Teilmengen, $f : G \to G'$ ein $\mathcal{C}^1$-Diffeomorphismus und $\lambda_G := \lambda|_G$. Dann gilt

$\mathcal{C}^1$-Diffeomorphismus

$$\lambda_G(\{x \in \mathbb{R} : f^{-1} \in A\}) = \int_A |\det(Df(x))| \lambda_G(\mathrm{d}x). \tag{22}$$

Sei nun $g : \mathbb{R} \to \mathbb{R}$ messbar. Dann gilt

$$\begin{aligned}
\int_{\mathbb{R}} g(x)\lambda_G(\mathrm{d}x) &= \int_R g(f \circ f^{-1}(x))\lambda_{G'}(\mathrm{d}x) = \int_R (g \circ f)(f^{-1}(x))\lambda_{G'}(\mathrm{d}x) \\
&\overset{\text{A.1.2}}{=} \int_R (g \circ f)(x)(\lambda_{G'} \circ (f^{-1})^{-1})(\mathrm{d}x) \\
&\overset{(22)}{=} \int_R (g \circ f)(x) |\det(Df(x))| \lambda_G(\mathrm{d}x).
\end{aligned}$$

A.2 Rechnungen

Beispiel A.2.1 (Aus der Übersicht 35)

1 Sei $X \sim \mathrm{Ber}(p)$. Dann gilt mit dem Transformationssatz (T)

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \overset{(T)}{=} \left(p \cdot 1^2 + (1-p) \cdot 0^2\right) - p^2 = p(1-p).$$

Ferner gilt für das k-te Moment

$$\mathbb{E}[X^k] \overset{(T)}{=} p \cdot 1^k + (1-p) \cdot 0^k = p.$$

Ferner gilt

$$\mathcal{G}_X(s) = \sum_{n \in \mathbb{N}_0} \mathbb{P}(X = n) s^n = \mathbb{P}(X=0) \cdot 1 + \mathbb{P}(X=1) s^1 = 1 - p + ps.$$

2 Sei $X \sim \mathrm{Geo}(p)$ für $p \in (0,1]$. Dann gilt mit der Ableitung der geometrischen Reihe

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} k \, \mathbb{P}(X = k) = p \sum_{k \in \mathbb{N}} k(1-p)^{k-1} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}.$$

Nun gilt $\sum_{k \in \mathbb{N}} kx^k = \frac{x}{(1-x)^2}$ und somit $\sum_{k \in \mathbb{N}} k^2 x^{k-1} = \frac{x+1}{(1-x)^3}$ und somit

$$\begin{aligned}
\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &\overset{(T)}{=} p \sum_{k \in \mathbb{N}} k^2 (1-p)^{k-1} - \frac{1}{p^2} \\
&= \frac{p(2-p)}{(1-(1-p))^3} - \frac{1}{p^2} = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}.
\end{aligned}$$

Ferner gilt

$$\begin{aligned}
\chi_X(t) &= \sum_{n \in \mathbb{N}} \mathbb{P}(X = n) e^{itn} = p \cdot \sum_{n \in \mathbb{N}} (1-p)^{n-1} e^{itn} \\
&= \frac{p}{1-p} \sum_{n \in \mathbb{N}} ((1-p)e^{it})^n = \frac{p}{\cancel{1-p}} \frac{\cancel{(1-p)}e^{it}}{1-(1-p)e^{it}} \\
&= \frac{p \cdot e^{it}}{1-(1-p)e^{it}}
\end{aligned}$$

3 Sei $X \sim \mathrm{N\,Bin}$. Es gilt

$$\begin{aligned}
\binom{-n}{k} &= \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!} \\
&= (-1)^k \frac{n(n+1)\cdots(n+k-1)}{k!} \\
&= (-1)^k \frac{(n+k-1)!}{k!(n-1)!} = (-1)^k \binom{n+k-1}{k}
\end{aligned}$$

und somit mit $q := 1-p$

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=n}^{\infty} k\binom{k-1}{n-1} q^{k-n}p^n \\
&= p^n \sum_{k\in\mathbb{N}_0} (k+n)\binom{k+n-1}{n-1} q^k \\
&= p^n \sum_{k\in\mathbb{N}_0} (k+n)\binom{k+n-1}{k} q^k \\
&= p^n \sum_{k\in\mathbb{N}_0} (k+n)\frac{(k+n-1)!}{k!(n-1)!} q^k \\
&= n\cdot p^n \sum_{k\in\mathbb{N}_0} \frac{(k+n)!}{k!n!} q^k = n\cdot p^n \sum_{k\in\mathbb{N}_0} \binom{k+n}{k} q^k \\
&= n\cdot p^n \sum_{k\in\mathbb{N}_0} \binom{-(n+1)}{k} (p-1)^k = \frac{n\cdot p^n}{(1+p-1)^{n+1}} = \frac{n}{p}.
\end{aligned}$$

Alternativ kann man $X = \sum_{k=1}^{n} G_k$ schreiben, wobei $(G_k \sim \mathrm{Geo}(p))_{k=1}^{n}$ u.i.d verteilt sind. Dann gilt $\mathbb{E}[X] \overset{(L)}{=} \sum_{k=1}^{n} \mathbb{E}[G_k] = \frac{n}{p}$.

Todo varianz: here

4 Sei $N \sim \mathrm{Poiss}(\lambda)$. Dann gilt

$$\begin{aligned}
\chi_N(t) &= \sum_{k=0}^{\infty} e^{itk}\mathcal{P}(N=k) = \sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda\cdot e^{it})^k}{k!} \\
&= e^{-\lambda}\exp\left(\lambda\cdot e^{it}\right) = \exp(\lambda(e^{it}-1)).
\end{aligned}$$

5 Sei $X \sim \mathrm{Hyp}(K,n,N)$. Es gilt

$$k\binom{n}{k} = k\frac{n!}{k!(n-k)!} = \frac{n(n-1)!}{(k-1)!(n-1-(k-1))!} = n\binom{n-1}{k-1}. \tag{23}$$

Daraus folgt

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \sum_{k=0}^{n} \frac{K\binom{K-1}{k-1}\binom{(N-1)-(K-1)}{(n-1)-(k-1)}}{\frac{N}{n}\binom{N-1}{n-1}} \\
&= \frac{nK}{N}\sum_{k=0}^{n} \underbrace{\frac{\binom{K-1}{k-1}\binom{(N-1)-(K-1)}{(n-1)-(k-1)}}{\binom{N-1}{n-1}}}_{\substack{\mathbb{P}(\hat{X}=k),\\ \hat{X}\sim\mathrm{Hyp}(K-1,n-1,N-1)}} \overset{(1)}{=} \frac{nK}{N}
\end{aligned}$$

Eine Modifikation von (23) ergibt

$$k(k-1)\binom{n}{k} = n(k-1)\binom{n-1}{k-1} = n(n-1)\binom{n-2}{k-2}.$$

und analog zu oben findet man

$$\mathbb{E}[X(X-1)] = \frac{n(n-1)K(K-1)}{N(N-1)}$$

und induktiv

$$\mathbb{E}[X(X-1)\dots(X-k+1)] = \prod_{j=0}^{k-1}\frac{(n-j)(K-j)}{(N-j)}$$

Somit gilt

$$\begin{aligned}\operatorname{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)+X] - \mathbb{E}[X]^2 \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}[X](1-\mathbb{E}[X]) \\ &= \frac{K(K-1)n(n-1)}{N(N-1)} + \frac{Kn}{N}\left(1-\frac{Kn}{N}\right) \\ &= \frac{Kn(N-K)(N-n)}{N^2(N-1)}.\end{aligned}$$

6 Sei $X \sim \mathcal{U}([a,b])$. Dann gilt

$$\mathbb{E}[X^k] \overset{(T)}{=} \frac{1}{b-a}\int_a^b x^k \,\mathrm{d}x = \frac{b^{k+1}-a^{k+1}}{(k+1)(b-a)} \overset{4.1.8}{=} \frac{1}{k+1}\sum_{j=0}^{k} a^j b^{k-j}.$$

Ferner gilt

$$\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3}(a^2+ab+b^2) - \frac{a^2+2ab+b^2}{2} = \frac{(b-a)^2}{12}.$$

Hausaufgabe 8.2(b): Nach der EULERsche Formel $e^{ix} = \cos(x) + i\sin(x)$ $(\star)$ gilt

$$\begin{aligned}\chi_X(t) &= \int_{\mathbb{R}} \exp(itx) f(x)\,\mathrm{d}x = \int_{\mathbb{R}} \exp(itx)\,\mathbb{1}_{(a,b)}(x)\,\mathrm{d}x \\ &= \int_a^b \exp(itx)\,\mathrm{d}x \overset{(\star)}{=} \int_a^b \cos(tx)\,\mathrm{d}x + i\cdot\int_a^b \sin(tx)\,\mathrm{d}x \\ &= \frac{\sin(tb)-\sin(ta)}{t} + \frac{i(-\cos(tb)+\cos(ta))}{t} \\ &= \frac{-i\,(\cos(tb)+i\sin(tb))}{t} + \frac{i\,(\cos(ta)+i\sin(ta))}{t} \\ &\overset{(\star)}{=} \frac{i(e^{ita}-e^{-itb})}{t}.\end{aligned}$$

Im vor-vorletzten Term kann man sehen, dass das Ergebnis genau dann reellwertig ist, wenn $-\cos(tb)+\cos(ta) = 0$ für alle $t \in \mathbb{R}$ gilt, also wenn $a = -b$ ist, da $a \neq b$ ist. Also ist $\chi_X(t)$ für alle Paare $(a,-a) \in \mathbb{R}^2$ reellwertig. ♢

Beispiel A.2.2 (aus Beispiel 3.3.3)

Sei $W := (\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : W \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Zufallsvariable. Zeigen Sie mit Hilfe der Dichte von X:

1 Sei $X \sim \exp(\vartheta)$ für $\vartheta > 0$ und sei $a > 0$, dann ist $aX \sim \exp\left(\frac{\vartheta}{a}\right)$.

Beweis. Für $t \geqslant 0$ gilt

$$\mathbb{P}(aX \leqslant t) = \mathbb{P}\left(X \leqslant \frac{t}{a}\right) = 1 - e^{-\vartheta\frac{t}{a}} = 1 - e^{-\frac{\vartheta}{a}t}. \qquad \square$$

2 Sei $\lambda > 0$ und $X \sim \exp(\lambda)$. Bestimmen Sie die Verteilung von $Y := \sum_{k=1}^{\infty} k\,\mathbb{1}_{[k-1,k)}(X)$.

Lösung: Weil $\operatorname{Im}(Y) = \mathbb{N}$ ist, ist Y diskret, es gilt $Y : W \to (\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Es gilt

$$\begin{aligned}\mathbb{P}(Y = n) = \mathbb{P}(x \in [n-1, n)) &= \int_{n-1}^{n} \lambda e^{-\lambda t} dt \\ &= e^{-\lambda(n-1)} - e^{-\lambda n} = (e^{\lambda} - 1)e^{-n\lambda}.\end{aligned} \qquad \Diamond$$

Beispiel A.2.3 (aus Beispiel 3.3.7) Für o.B.d.A $a > 0$ und der Substitution $\varphi(t) := \frac{t-b}{a}$ gilt

$$\begin{aligned}F_{aX+b}(t) &= \mathbb{P}(aX + b \leqslant t) = \mathbb{P}\left(x \leqslant \frac{t-b}{a}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\frac{t-b}{a}} \exp\left(\frac{-(s-\mu)^2}{2\sigma^2}\right) ds \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\varphi(t)} \exp\left(\frac{-(\varphi(s)-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{a} ds \\ &= \frac{1}{\sqrt{2\pi\sigma^2 a^2}} \int_{-\infty}^{t} \exp\left(-\frac{\left(\frac{s-n}{a} - \mu\right)^2}{2\sigma^2}\right) ds \\ &= \frac{1}{\sqrt{2\pi\sigma^2 a^2}} \int_{-\infty}^{t} \exp\left(-\frac{(s - (b + a\mu))^2}{2s^2\sigma^2}\right) ds.\end{aligned} \qquad \Diamond$$

Beispiel A.2.4 (aus Beispiel 3.3.8)

Seien $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ unabhängige Zufallsvariablen mit $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ mit $m_i \in \mathbb{R}$ und $\sigma_i > 0$ für $i \in \{1, 2\}$. Zeigen Sie, dass $Z := X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Beweis. Seien $\tilde{X}_i := X_i - m_i \sim \mathcal{N}(0, \sigma_i^2)$ für $i \in \{1, 2\}$ und $\tilde{Z} := \tilde{X}_1 + \tilde{X}_2$.

Dann gilt

$$f_{\tilde{Z}}(z) = \int_{\mathbb{R}} f_{\tilde{X}_1}(x) \cdot f_{\tilde{X}_2}(z-x) dx = \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{x^2}{2\sigma^2} + \frac{(z-x)^2}{\sigma_2^2}\right)\right).$$

Nun gilt

$$\begin{aligned}\frac{x^2}{2\sigma^2} + \frac{(z-x)^2}{\sigma_2^2} &= \frac{x^2(\sigma_1^2 + \sigma_2^2)}{\sigma_1^2\sigma_2^2} - \frac{2xz}{\sigma_2^2} + \frac{z^2}{\sigma_2^2} \\ &= \frac{x^2(\sigma_1^2 + \sigma_2^2)}{\sigma_1^2\sigma_2^2} - \frac{2xz}{\sigma_2^2} + \frac{z^2\sigma_1^2}{\sigma_2^2(\sigma_1^2 + \sigma_2^2)} - \frac{z^2\sigma_1^2}{\sigma_2^2(\sigma_1^2 + \sigma_2^2)} + \frac{z^2}{\sigma_2^2} \\ &= \Bigg(\underbrace{\frac{x\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2} - \frac{z\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}}_{=:u=:\varphi(x)}\Bigg)^2 + \frac{z^2}{\sigma_1^2 + \sigma_2^2}.\end{aligned}$$

Dann gilt

$$\begin{aligned} f_{\tilde{Z}}(z) &= \frac{1}{2\pi\cancel{\sigma_1\sigma_2}} \exp\left(-\frac{1}{2}\frac{z^2}{\sigma_1^2+\sigma_2^2}\right) \frac{\cancel{\sigma_1\sigma_2}}{\sqrt{\sigma_1^2+\sigma_2^2}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}}\,du \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}\frac{z^2}{\sigma_1^2+\sigma_2^2}\right) \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\sqrt{2\pi} \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{z^2}{\sigma_1^2+\sigma_2^2}\right) \sim \mathcal{N}(0,\sigma_1^2+\sigma_2^2) \end{aligned}$$

und daher $Z = \tilde{Z} + m_1 + m_2 \sim \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$. □

Beispiel A.2.5 (aus Beispiel 3.2.5)

Sei $W := (\Omega, \mathcal{A}, \mathbb{P}))$ ein Wahrscheinlichkeitsraum und $X, Y : W \to (\mathbb{N}, \mathcal{P}(\mathbb{N})$ zwei unabhängige Zufallsvariablen mit $X \sim \text{Geo}(p)$ und $Y \sim \text{Geo}(y)$ für $p \in (0,1]$. Bestimmen Sie die Verteilung von $Z := \min(X,Y)$.

Lösung: Es gilt

$$\mathbb{P}(Z \geqslant n) = \mathbb{P}(X \geqslant n)\,\mathbb{P}(Y \geqslant n),$$

weil X und Y unabhängig voneinander sind. Ferner gilt

$$\begin{aligned} \mathbb{P}(X \geqslant n) &= \sum_{k=n}^{\infty} (1-p)^{k-1} \cdot p = p(1-p)^{n-1} \sum_{k=0}^{\infty} (1-p)^k \\ &= p(1-p)^{n-1} \frac{1}{1-(1-p)} = (1-p)^{n-1} \end{aligned}$$

Also gilt

$$\begin{aligned} \mathbb{P}(Z = n) = \mathbb{P}(Z \geqslant n) - \mathbb{P}(Z \geqslant n+1) &= ((1-p)^{n-1})^2 - ((1-p)^n)^2 \\ &= \left((1-p)^2)\right)^{n-1} (1-(1-p)^2) \end{aligned}$$

und somit $Z \sim \text{Geo}(1-q^2)$ für $q := 1-p$. ◊

Beispiel A.2.6 (Aus Beispiel 4.1.4)

Seien $X \sim \text{Bin}(n,p)$ und $Y \sim \text{Bin}(m,p)$ unabhängige Zufallsvariablen für $n, m \in \mathbb{N}$ und $p \in [0,1]$. Berechnen Sie die erzeugende Funktion von $X+Y$.

Lösung: Mit dem binomischen Lehrsatz und $\binom{n}{k} = 0$ für $k > n$ gilt

$$\mathcal{G}_X(t) = \sum_{k=0}^{\infty} \binom{n}{k} t^k p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (tp)^k (1-p)^{n-k} = (tp+1-p)^n$$

Da X und Y unabhängig sind, folgt

$$\mathcal{G}_{X+Y}(t) = \mathcal{G}_X(t)\,\mathcal{G}_Y(t) = (tp+1-p)^n (tp+1-p)^m = (tp+1-p)^{n+m}.$$

Also gilt $X+Y \sim \text{Bin}(n+m,p)$.

Seien $X, Y \sim \operatorname{Geo}(p)$ zwei unabhängige Zufallsvariablen mit $p \in (0,1)$. Berechnen Sie die erzeugende Funktion der Zufallsvariablen $S := X + Y$ und $\mathbb{P}(S = n)$ für alle $n \in \mathbb{N}$.

Lösung: Aus der Vorlesung ist $\mathcal{G}_X(t) = \mathcal{G}_Y(t) = \frac{pt}{1-(1-p)t}$ bekannt. Weil X und Y unabhängig sind, gilt

$$\begin{aligned}\mathcal{G}_{X+Y}(t) &= \frac{p^2t^2}{(1-(1-p)t)^2} \overset{(\star)}{=} \sum_{k=0}^{\infty} p^2t^2(k+1)((1-p)t)^k \\ &= \sum_{k=2}^{\infty} p^2(k-1)(1-p)^{k-2}t^k = \sum_{k=0}^{\infty} p^2(k-1)(1-p)^{k-2}t^k,\end{aligned}$$

da $\mathbb{P}(X+Y=0) = \mathbb{P}(X+Y=1) = 0$ ist, gilt

$$\mathbb{P}(X+Y=n) = p^2(n-1)(1-p)^{n-2}$$

für $n \geqslant 2$. Die Gleichheit $(\star)$ folgt aus Differenzieren der geometrischen Reihe. $\Diamond$

Beispiel A.2.7 (Ch. Fkt. der Normalverteilung (vgl. 4.2.3)) Betrachte zunächst $X \sim \mathcal{N}(0,1)$, da dann $Y := \sigma X + m \sim \mathcal{N}(m, \sigma^2)$ und somit

$$\chi_Y(t) = \mathbb{E}[e^{itY}] = \mathbb{E}[e^{itm}e^{i\sigma X}] = e^{itm}\,\mathbb{E}[e^{i\sigma X}] = e^{itm}\chi_X(\sigma t)$$

gilt. Nach $e^{ix} = \cos(x) + i\sin(x)$ gilt

$$\begin{aligned}\mathbb{E}[e^{itX}] &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{itz}e^{-\frac{z^2}{2}}\,\mathrm{d}z \\ &= \frac{1}{\sqrt{2\pi}}\left(\int_{\mathbb{R}}\cos(tz)e^{-\frac{z^2}{2}}\,\mathrm{d}z + \underbrace{i\int_{\mathbb{R}}\sin(tz)e^{-\frac{z^2}{2}}\,\mathrm{d}z}_{=0 \text{ weil } \sin \text{ ungerade}}\right).\end{aligned}$$

Aus dem Differentiationslemma folgt mit partieller Integration

$$\begin{aligned}\chi_X'(t) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\frac{\mathrm{d}}{\mathrm{d}t}\cos(tz)e^{-\frac{z^2}{2}}\,\mathrm{d}z = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\sin(tz)(-z)e^{-\frac{z^2}{2}}\,\mathrm{d}z \\ &= \frac{1}{\sqrt{2\pi}}\left(\underbrace{\left[\sin(tz)e^{-\frac{z^2}{2}}\right]_{z=-\infty}^{\infty}}_{=0} - \int_{\mathbb{R}} t\cos(tz)e^{-\frac{z^2}{2}}\,\mathrm{d}z\right) = -t\chi_X(t).\end{aligned}$$

Also löst $\chi_X(t)$ die Differentialgleichung $\chi_X(t)' = -t\chi_X(t)$ zum Anfangswert $\chi_X(0) = 1$. Daraus folgt $\chi_X(t) = e^{-\frac{t^2}{2}}$. Also ist bis auf eine Faktor die Dichte ihre eigene charakteristische Funktion und somit ein Fixpunkt der Fourier-Transformation .

TODO: das geht auch einfacher mit quadratischer Ergänzung und der Translationsinvarianz des Integrals $\Diamond$

A.3 Übersicht über die Verteilungen

r	Def.-menge	(Zähl)Dichte	Verteilungsfkt.	$\mathbb{E}$	$\mathbb{E}[\cdot^k]$	Var	ch. Funktion	erz. Fkt.
	$\{0,1\}$	$1-p\cdot\mathbb{1}_0$	$\mathbb{1}_{[0,\infty)} - p\,\mathbb{1}_{[0,1)}$	p	p	pq	$q+pe^{it}$	$q+pt$
, 1]	$\{0,\dots,n\}$	$\binom{n}{k}p^k q^{n-k}$		np		npq	$(q+pe^{it})^n$	$(sp+q)^n$
	$\mathbb{N}$	$q^k p$	$1-q^k$	$\frac{1}{p}$		$\frac{q}{p^2}$	$\frac{p\cdot e^{it}}{1-qe^{it}}$	$\frac{ps}{1-q^s}$
1]	$\mathbb{N}_{\geqslant n}$	$\binom{k-1}{n-1}q^{k-n}p^n$		$\frac{n}{p}$		$\frac{nq}{p^2}$		
	$\mathbb{N}_0$	$e^{-\lambda}\cdot\frac{\lambda^k}{k!}$		λ		λ	$\exp(\lambda(e^{it}-1))$	$e^{\lambda(s-1)}$
N	$\{0,\dots,n\}$	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$		$\frac{nK}{N}$		$\frac{Kn(N-K)(N-n)}{N^2(N-1)}$		
b	$\mathbb{R}$	$\frac{\mathbb{1}_{[a,b]}}{b-a}$	$\frac{x-a}{b-a}\,\mathbb{1}_{[a,b)} + \mathbb{1}_{[b,\infty)}$	$\frac{a+b}{2}$	$\frac{\sum_{j=0}^k a^j b^{k-j}}{k+1}$	$\frac{(b-a)^2}{12}$	$\frac{i(e^{ita}-e^{-itb})}{t}$	
	$\mathbb{R}$	$\lambda e^{-\lambda x}\,\mathbb{1}_{[0,\infty)}(x)$	$1-e^{-\lambda x}\,\mathbb{1}_{[0,\infty)}$	λ^{-1}	$\frac{k!}{\lambda^k}$	λ^{-2}		
	$\mathbb{R}$	$\frac{\lambda^\vartheta x^{\vartheta-1}e^{-\lambda x}}{\Gamma(\vartheta)}\,\mathbb{1}_{[0,\infty)}(x)$		$\frac{\vartheta}{\lambda}$				
0	$\mathbb{R}$	$\frac{\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$		m		σ^2	$\exp\left(itm-\frac{(t\sigma)^2}{2}\right)$	

enngrößen wichtiger Verteilungen. Hierbei ist $q := 1-p$.

A.4 Nützliche Identitäten

Kombinatorik / Binomialkoeffizient

$k\binom{n}{k} = n\binom{n-1}{k-1}$ und $\binom{n}{k} = \binom{n}{n-k}$

Geometrische Reihe und ihre Verwandte

Sei $|x| < 1$. Dann gilt $\sum_{k\in\mathbb{N}_0} x^k = \frac{1}{1-x}$, $\sum_{k\in\mathbb{N}} x^k = \frac{x}{1-x}$, $\sum_{n\in\mathbb{N}_0} kx^k = \frac{x}{(1-x)^2}$, $\sum_{n\in\mathbb{N}_0} kx^{k-1} = \frac{1}{(1-x)^2}$, $\sum_{n\in\mathbb{N}_0} k^2x^{k-1} = \frac{x+1}{(1-x)^3}$.

A.5 Klausuraufgaben

Deuschel 2018 Probe Sei X eine $\mathbb{R}^d$-wertige Zufallsvariable mit Verteilungsfunktion $\mathbb{P}_X$ und charakteristischer Funktion φ_X. Zeigen Sie die Äquivalenz der folgenden Aussagen

1. $\mathbb{P}_X = \mathbb{P}_{-X}$.
2. $\varphi_X = \varphi_{-X}$.
3. $\varphi_X(t) = \mathbb{E}[\cos(\langle t, X\rangle)]$ für $t \in \mathbb{R}^d$.
4. φ_X ist reell.

Beweis. (c) $\implies$ (d). Nach Definition gilt

$$\begin{aligned}\varphi_X(t) &\overset{\text{Def.}}{=} \mathbb{E}[\exp(i\langle t, X\rangle)]\\ &\overset{\text{L}}{=} \mathbb{E}[\cos(\langle t, X\rangle)] + i\,\mathbb{E}[\sin(\langle t, X\rangle)] \overset{!}{=} \mathbb{E}[\cos(\langle t, X\rangle)].\end{aligned}$$

Somit gilt $i\,\mathbb{E}[\sin(\langle t, X\rangle)] = 0$.

(d) $\implies$ (b). Für alle $a, b \in \mathbb{R}$ gilt

$$\varphi_{aX+b}(t) = e^{itb}\varphi_X(at) \qquad \text{und} \qquad \varphi_X(-t) = \overline{\varphi_X(t)}.$$

Damit folgt

$$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$$

(b) $\implies$ (c): Nach Definition muss

$$\mathbb{E}[\cos(\langle t, X\rangle)] + i\,\mathbb{E}[\sin(\langle t, X\rangle)] = \mathbb{E}[\cos(\langle t, -X\rangle)] + i\,\mathbb{E}[\sin(\langle t, -X\rangle)]$$

gelten. Nach den Eigenschaften des Standardskalarproduktes ist das äquivalent zu

$$\mathbb{E}[\cos(\langle t, X\rangle)] + i\,\mathbb{E}[\sin(\langle t, X\rangle)] = \mathbb{E}[\cos(\langle t, X\rangle)] - i\,\mathbb{E}[\sin(\langle t, X\rangle)]$$

Das impliziert

$$i\,\mathbb{E}[\sin(\langle t, X\rangle)] = -i\,\mathbb{E}[\sin(\langle t, X\rangle)] \implies i\,\mathbb{E}[\sin(\langle t, X\rangle)] = 0$$

und somit folgt die Behauptung. □

Deuschel 2018 Probe Seien X und Y i.i.d verteilte reellwertige Zufallsvariablen und $\ln(X) \sim \text{Exp}(1)$. Bestimmen Sie die Verteilungsfunktion von $Z := XY$.

Lösung: **Here's the solution using convolution:** Let $\overline{X} := \ln(X)$ and $\overline{Y}$ analogously. We can now calculate the PDF of $\ln(XY) =$

$\overline{X}+\overline{Y}$ using convolution: For $x \geqslant 0$ we have

$$f_{\ln(XY)}(x) = f_{\overline{X}} * f_{\overline{Y}} = \int_{-\infty}^{\infty} f_{\overline{X}}(y) f_{\overline{Y}}(x-y)\, \mathrm{d}y = \int_0^x e^{-y} e^{y-x}\, \mathrm{d}y = \int_0^x e^{-x}\, \mathrm{d}y = xe^x$$

and $f_{\ln(XY)}(x) = 0$ otherwise. Therefore, for $x \geqslant 0$ we have

$$P(\ln(XY) \leqslant x) = \int_{-\infty}^{x} f_{\ln(XY)}(y)\, \mathrm{d}y = \int_0^x ye^{-y}\, \mathrm{d}y = 1 - e^{-x}(x+1).$$

For $z \geqslant 1$ this implies

$$P(XY \leqslant z) = 1 - \frac{1+\ln(z)}{z}.$$

and $P(XY \leqslant z) = 0$ for $z \leqslant 1$.

Zweite Lösung. Since $\ln(X) \sim$Exp(1), for all $k \geqslant 0$ we have

$$\mathbb{P}(X \leqslant e^k) = \mathbb{P}(\ln(X) \leqslant k) = 1 - e^{-k}$$

By substitution we obtain

$$\mathbb{P}(X \leqslant a) = 1 - \frac{1}{a}.$$

for all $a \geqslant 1$. We can now obtain the density function by calculating the derivative: $f_X(x) = \frac{1}{x^2}$.

Now, for $z \geqslant 1$ we have

$$\begin{aligned}
\mathbb{P}(XY \leqslant z) &= \int_1^{\infty} 1_{(-\infty,z]}(x,y)\, \mathrm{d}\mathbb{P}(x,y) = \int_1^{\infty} \int_{\mathbb{R}+} 1_{(1,\frac{z}{y})}(x)\, \mathrm{d}\mathbb{P}_X\, \mathrm{d}\mathbb{P}_Y \\
&= \int_1^{z} \int_{\mathbb{R}+} 1_{(1,\frac{z}{y})}(x) \frac{1}{x^2} \frac{1}{y^2}\, \mathrm{d}x\, \mathrm{d}y \\
&= \int_1^{z} \frac{1}{y^2} \left[\int_1^{\frac{z}{y}} \frac{1}{x^2}\, \mathrm{d}x \right] \mathrm{d}y \\
&= \int_1^{z} \frac{1}{y^2} \left(1 - \frac{y}{z}\right) \mathrm{d}y = 1 - \frac{1+\ln(z)}{z}.
\end{aligned}$$

Dritte Lösung mit regulärer bedingter Verteilung Für $x > 0$ gilt

$$\begin{aligned}
\mathcal{P}(\ln(XY) \leqslant x) &= \mathcal{P}(\ln(X) + \ln(Y) \leqslant x) = \int_0^{\infty} \mathcal{P}(\ln(X) + \ln(Y) \leqslant x \mid Y = y) f_Y(y)\, \mathrm{d}y \\
&= \int_0^{\infty} \mathcal{P}(\ln(X) \leqslant x - y) f_Y(y)\, \mathrm{d}y = \int_0^{x} (1 - e^{y-x}) e^{-y}\, \mathrm{d}y \\
&= \int_0^{x} e^{-y} - e^{-x}\, \mathrm{d}y = 1 - e^{-x}(1+x).
\end{aligned}$$

Daraus folgt für $z > 1$

$$\mathcal{P}(Z \leqslant z) = \mathcal{P}(\ln(Z) \leqslant \ln(z)) = 1 - e^{-\ln(z)}(1 + \ln(z)) = 1 - \frac{\ln(z)+1}{z}.$$

Für $z \leqslant 1$ gilt $\mathcal{P}(Z \leqslant z) = 0$.

1 **Ideen:** Die ZVen Y_k sind Bernoulli-verteilt zum Parameter $p_k :=$ $\mathbb{P}(X_k \leqslant t)$ und somit unabhängig und identisch verteilt. (Eigentlich nicht identischem weil p_k sich ändert, oder??) Somit gilt $\mathbb{E}[Y_1] = p$

und $\mathrm{Var}[Y_1] = p(1-p)$. Die die Verteilungsfunktion einer $\mathrm{Ber}(p)$-verteilten Zufallsvariable durch $F(t) = \mathbb{1}_{[0,\infty)}(t) - p\,\mathbb{1}_{[0,1)}(t)$ gegeben ist, impliziert $F(t) \in \{0,1\}$, dass $p \in \{0,1\}$ gilt und somit $X \in \{0,1\}$.

Ferner gilt $S_n \sim \mathrm{Bin}(n,p)$. Nach Poissonschem Approximationssatz konvergieren eine Folge binomial-Zähldichten punktweise gegen die Zähldichte der Exponentialverteilung zum Parameter $\lambda := \lim_{n\to\infty} np_n$.

Scheutzow, 07/17 Entscheiden Sie, welche der folgenden Aussagen immer wahr sind.

1 Aus $\mathbb{E}[X^6] < \infty$ folgt auch $\mathbb{E}[X^4] < \infty$.

Stimmt. Ansatz 1: Da jeder Wahrscheinlichkeitsraum Ω endlich ist, gilt $L^q(\Omega) \subset L^p(\Omega)$ für alle $1 \leqslant p \leqslant q \leqslant \infty$.
Ansatz 2: Es gilt $X^4 \leqslant X^6 + 1$.

2 Sind Z_1 und Z_2 exponentialverteilte Zufallsvariablen, so ist $Z := \min_{j=1}^{2} Z_j$ auch exponentialverteilt.

Stimmt. Seien $Z_1, Z_2 \sim \mathrm{Exp}(\lambda)$ für ein $\lambda > 0$. Dann gilt für $x \geqslant 0$

$$\begin{aligned}\mathbb{P}(Z \leqslant x) &= \sum_{k=1}^{2} \binom{2}{k} (1 - e^{\lambda x})^k (e^{-\lambda x})^{2-k} \\ &= \cancel{2e^{-\lambda x}} - 2e^{-2\lambda x} + 1 \cancel{-2e^{-\lambda x}} + e^{-2\lambda x} = 1 - e^{-2\lambda x}\end{aligned}$$

und somit $Z \sim \mathrm{Exp}(2\lambda)$.

Was macht man für $(Z_k \sim \mathrm{Exp}(\lambda_k))_{k\in\mathbb{N}}$**???**

3 Ist $X + Y$ eine int'bare Zufallsvariable, so ist $\mathbb{E}[|X+Y|] < \infty$.

Stimmt, da Z eine integrierbare Zufallsvariable ist, wenn ihr Erwartungswert existiert, also folgt die Aussage aus der Definition 3.5.1.

4 Es ist nicht möglich, dass $X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-f.s}} 0$ und $\mathbb{E}[X_n] \xrightarrow{n\to\infty} \infty$.

todo.

5 Sind X und Y unkorrelliert, so sind sie unabhängig.

Falsch, siehe Gegenbeispiel 3.5.14.

Scheutzow, 07/17 Sei $f(x) := c \cdot \ln\left(1 - \frac{x}{2}\right)$ für $c \in \mathbb{R}$.

1 Für welche $c \in \mathbb{R}$ ist f eine erzeugende Funktion?

Lösung. Assuming that X is a non-negative integer random variable, we can use that

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}.$$

These probability mass values must sum to one, which imposes a constraint on the PGF, which we can use to find the constant c. Now, substituting your specified PGF you get:

$$G_X^{(k)}(z) = c \cdot \left(\frac{d}{dz}\right)^k \ln(1 - z/2) = \begin{cases} c \cdot \ln(1 - z/2), & \text{for } k = 0, \\ -c \cdot (k-1)! \cdot (2-z)^{-k}, & \text{for } k > 0, \end{cases}$$

which gives the mass function:

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!} = \begin{cases} 0, & \text{for } k = 0, \\ -(c/k) \cdot 2^{-k}, & \text{for } k > 0. \end{cases}$$

The constraint equation therefore reduces to:

$$1 = \sum_{k=0}^{\infty} \mathbb{P}(X = k) = -c \cdot \sum_{k=1}^{\infty} \frac{1}{k \cdot 2^k} = -c \cdot \ln(2).$$

From this constraint we have $c = -\ln(2)$ so your PGF is:

$$G_X(z) = -\frac{\ln(1 - \frac{z}{2})}{\ln(2)} = 1 - \frac{\ln(2-z)}{\ln(2)},$$

and the corresponding probability mass function is:

$$p_X(k) \equiv \mathbb{P}(X = k) = \frac{1}{\ln(2)} \cdot \frac{1}{k \cdot 2^k} \qquad \text{for all } k \in \mathbb{N}.$$

2 Sei X_c eine Zufallsvariable mit der von f erzeugten Verteilung. Bestimmen Sie $\mathbb{E}[X_c]$ und $\mathbb{P}(X_c = 1)$.

Lösung. TODO

3 Geben Sie die Verteilungsfunktion von X_c an.

Lösung. TODO

Scheutzow, 07/17 Seien $X, Y \sim \operatorname{Exp}(\lambda)$ u.i.v. mit $\lambda > 0$.

1 Bestimmen Sie die Dichte von $Z := X + Y$.

Lösung. Seien f_X und f_Y die Dichten von X bzw. Y. Dann gilt für $x \in \mathbb{R}$

$$\begin{aligned}
f_Z(x) = (f_X * f_Y)(x) &= \int_{\mathbb{R}} f_X(x-y) f_Y(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}} \lambda e^{-\lambda(x-y)}\,\mathbb{1}_{[0,\infty)}(x-y)\cdot \lambda e^{-\lambda y}\,\mathbb{1}_{[0,\infty)}(y)\,\mathrm{d}y \\
&= \lambda^2 \int_0^x e^{-\lambda(x-y)}\cdot e^{-\lambda y}\,\mathrm{d}y \cdot \mathbb{1}_{[0,\infty}(x) = \lambda^2 \int_0^x e^{-\lambda x}\,\mathrm{d}y \cdot \mathbb{1}_{[0,\infty}(x) \\
&= \underline{\underline{\lambda^2 x \cdot e^{-\lambda x}\cdot \mathbb{1}_{[0,\infty}(x).}}
\end{aligned}$$

2 Bestimmen Sie die Verteilung von (X, Z).

3 Berechnen Sie $\mathbb{E}[X \mid Z]$.

Scheutzow 07/17 Seien X, Y unabhängig und identisch stetig gleichverteilt auf $[0, a]$ für $a > 0$. Bestimmen Sie die Dichte, Verteilungsfunktion, den Erwartungswert und die Varianz von $Z := X + Y$.

Lösung.

1 Die Dichte bestimmen wir wie oben mit Faltung: Für $x \in \mathbb{R}$ gilt

$$\begin{aligned}
\int_{\mathbb{R}} f_X(x-y) f_Y(y)\,\mathrm{d}y &= \frac{1}{(a-0)^2}\int_{\mathbb{R}} \mathbb{1}_{[0,a]}(x-y)\,\mathbb{1}_{[0,a]}(y)\,\mathrm{d}y \\
&= \frac{1}{a^2}\int_0^a 1\,\mathrm{d}y \cdot \mathbb{1}_{[0,2a]}(x) = \frac{\mathbb{1}_{[0,2a]}(x)}{a}
\end{aligned}$$

2 **TODO**

2016 Erstklausur Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B, C \in \mathcal{A}$. Zeigen Sie, dass gilt

1 Sind $A \cap B$ und C unabhängig und $\mathbb{P}(B) > 0$, so folgt

$$\mathbb{P}(A \cap C \mid B) = \mathbb{P}(A \mid B)\,\mathbb{P}(C) \tag{24}$$

Beweis. Nach Definition gilt

$$\begin{aligned}
\mathbb{P}(A \cap C \mid B) &\overset{\text{Def.}}{=} \frac{\mathbb{P}(A \cap C \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}((A \cap B) \cap C)}{\mathbb{P}(B)} \\
&\overset{\text{unabh.}}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\cdot \mathbb{P}(C) \overset{\text{Def.}}{=} \mathbb{P}(A \mid B)\,\mathbb{P}(C). \qquad \square
\end{aligned}$$

2 Die Gleichung (24) gilt auch, falls statt Unabhängigkeit $\mathbb{P}(C) = 1$ gilt.

Beweis. Wir müssen den Schritt, den wir oben mit der Unabhängigkeit begründet haben, nur anderweitig begründen.

Dafür zeigen wir das für $\mathbb{P}(C) \in (0, 1)$ und $\mathbb{P}(D) = 1$ gilt

$$\mathbb{P}(C \cap D) = \mathbb{P}(C)\,\mathbb{P}(D).$$

Dies folgt direkt aus $\mathbb{P}(D \mid C) = \mathbb{P}(D) = 1$. $\square$

③ Es gelte $\mathbb{P}(A \mid B) = \mathbb{P}(A \mid B^{\complement})$ und $\mathbb{P}(B) \in (0,1)$. Dann sind A und B unabhängig.

Beweis. Aus dem Satz über die totale Wahrscheinlichkeit folgt

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^{\complement}) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) + \mathbb{P}(A \mid B^{\complement})\,\mathbb{P}(B^{\complement}) \\ &= \mathbb{P}(A \mid B)\big(\mathbb{P}(B) + \mathbb{P}(B^{\complement})\big) = \mathbb{P}(A \mid B).\end{aligned}$$

und somit die Behauptung. □

2016 Erstklausur Seien $\left(X_n \sim \operatorname{Exp}\left(\frac{1}{n}\right)\right)_{n\in\mathbb{N}}$ Zufallsvariablen.

① Berechnen Sie die Verteilungsfunktion F_n der $Y_n := \frac{X_n}{2n+1}$.

Lösung: Für $x \in \mathbb{R}$ gilt

$$\begin{aligned}\mathbb{P}(Y_n \leqslant x) &= \mathbb{P}(X_n \leqslant x(2n+1)) = 1 - e^{-\frac{2n+1}{n}x}\,\mathbb{1}_{[0,\infty)}(x(2n+1)) \\ &= 1 - e^{-\frac{2n+1}{n}x}\,\mathbb{1}_{[0,\infty)}(x)\end{aligned}$$

② Bestimmen Sie für alle $x \in \mathbb{R}$ den punktweise Grenzwert $F(x) := \lim_{n\to\infty} F_n$.

Lösung. Aufgrund der Stetigkeit der Exponentialfunktion gilt für $x \geqslant 0$

$$F(x) = \lim_{n\to\infty} F_n(x) = 1 - \exp\left(\lim_{n\to\infty} -\frac{2n+1}{n}x\right) = 1 - e^{-2x}$$

und $F(x) \equiv 0$ auf $(-\infty, 0)$.

③ die Funktion F ist die Verteilungsfunktion einer Zufallsvariable Y. Um welche Verteilung handelt es sich und was können sie über die Konvergenz der Folge $(Y_n)_{n\in\mathbb{N}}$ aussagen?

Lösung. Die Funktion F ist die Verteilungsfunktion einer Exp(2)-verteilten Zufallsvariable.

TODO:zweiter Teil.

2016 Erstklausur Peter will nach der Vorlesung mit dem Bus nach Hause fahren. Dazu kann er den Bus A oder den Bus B nehmen. Die Wartezeiten auf die Busse können durch unabhängig auf dem Intervall $[0, 10]$ gleichverteilte Zufallsvariablen X und Y beschrieben werden.

① Bestimmen Sie die Verteilungsfunktionen von $Z_1 := \max(X, Y)$ und $Z_2 := \min(X, Y)$.

Lösung. Für gilt $x \in [0, 10]$ gilt mit dem binomischen Lehrsatz

$$\mathbb{P}(Z_1 \leqslant x) = 1 - (1 - \mathbb{P}(X \leqslant x))^2 = 1 - \left(1 - \frac{x}{10}\right)^2 = \frac{x}{5} - \frac{x^2}{100}.$$

und $\mathbb{P}(Z_1 \leqslant x) = 0$ für $x < 0$ sowie $\mathbb{P}(Z_1 \leqslant x) = 1$ für $x > 10$.

Ferner gilt

$$\mathbb{P}(Z_2 \leqslant x) = \mathbb{P}(X \leqslant x)^2 = \frac{x^2}{100}$$

und $\mathbb{P}(Z_2 \leqslant x) = 0$ für $x < 0$ sowie $\mathbb{P}(Z_2 \leqslant x) = 1$ für $x > 10$.

2 Wie groß ist die Wahrscheinlichkeit, dass Peter länger als zwei Minuten auf einen Bus warten muss?

Lösung. Aus den obigen Rechnungen ergibt sich sofort

$$\mathbb{P}(Z_2 > 2) = 1 - \mathbb{P}(Z_2 \leqslant 2) = 1 - \frac{4}{100} = \frac{24}{25}.$$

3 Peter kommt zur Haltestelle und wird dort für fünf Minuten abgelenkt. Wie hoch ist die Wahrscheinlichkeit, dass er beide Busse verpasst hat?

Lösung. Aus 1 ergibt sich

$$\mathbb{P}(Z_1 \leqslant 5) = \frac{5}{5} - \frac{25}{100} = \frac{3}{4}.$$

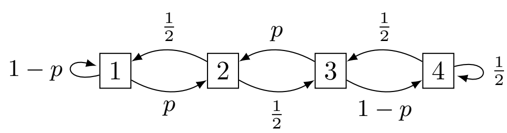
4 Zeigen Sie, dass die Dichte von (X, Y) durch

$$f_{Z_2}(z) = \frac{1}{5}\left(1 - \frac{z}{10}\right)\mathbb{1}_{[0,10]}(z)$$

gegeben ist.

5 Geben Sie die gemeinsame Dichte und die gemeinsame Verteilung des Zufallsvektors (X, Y) an.

Erstklausur 2016 Betrachte den folgenden Übergangsgraphen einer Markov-Kette.



1 Geben Sie die Übergangsmatrix P an.

Lösung. Es gilt

$$P := \begin{pmatrix} 1-p & p & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & p & 0 & 1-p \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

2 Bestimmen Sie alle Äquivalenzklassen mit ihrer wechselseitigen Erreichbarkeit in Abhängigkeit von p.

Lösung. Fall 1: $p = 0$. Die Klassen sind $\{1\}, \{2\}, \{3\}, \{4\}$ und nur $\{2\}$ ist abgeschlossen.
Fall 2: $p \in (0,1)$. Die einzige Klasse ist $\{1,2,3,4\}$ und sie ist abgeschlossen.
Fall 1: $p = 1$. Die Klassen sind $\{1,2,3\}$ und $\{4\}$ und nur letztere ist abgeschlossen.

3 Sei nun $p = 0$. Welche Zustände sind (un)wesentlich?

Lösung. $\{1\}$, $\{3\}$ und $\{4\}$ sind wesentlich, $\{2\}$ ist unwesentlich.

4 Bestimmen Sie alle stationären Verteilungen der Markov-Kette.

Lösung. Da die stationären Verteilungen genau die Eigenvektoren von P^T zum Eigenwert 1 und Komponentensumme 1 sind, lösen wir

$$\begin{pmatrix} 1-p & \frac{1}{2} & 0 & 0 \\ p & 0 & p & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1-p & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \gamma \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \gamma \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

Wir erhalten

$$\begin{aligned} \mathrm{I}: \ (1-p)\gamma + \frac{x_2}{2} = \gamma \implies x_2 = 2p\gamma, \\ \mathrm{IV}: \ (1-p)x_3 + \frac{x_4}{2} = x_4 \implies x_4 = 2(1-p)x_3 \end{aligned}$$

und somit aus III

$$x_2 + x_4 = 2x_3 \implies p\gamma + (1-p)x_3 = x_3 \implies x_3 = \gamma$$

Da die Komponentensumme 1 ergeben muss, erhalten wir

$$1 \overset{!}{=} \gamma + x_2 + x_3 + x_4 = 2\gamma + 2p\gamma + 2(1-p)\gamma = 4\gamma$$

also $\gamma = \frac{1}{4}$ und somit ist die einzige stationäre Verteilung durch

$$\left(\frac{1}{4}, \frac{p}{2}, \frac{1}{4}, \frac{1-p}{2}\right)$$

gegeben.

Bank, 2014/2 Kreuzen Sie alle wahren Aussagen an.

(1) Seien X und Y zwei unabhängige integrierbare Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Dichten f_X und f_Y bezüglich des Lebesgue-Maßes. Dann gilt

(1) X^2 und $\sqrt{|Y|}$ sind unabhängig.

Stimmt. Dies folgt daraus, dass $x \mapsto x^2$ und $x \mapsto \sqrt{|x|}$ messbare Abbildungen sind.

(2) $\mathbb{E}[X \cdot \mathbb{1}_{\{Y \leqslant 0\}}] = \mathbb{E}[X] - \mathbb{E}[X]\,\mathbb{P}(Y > 0)$.

Stimmt. Es gilt $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ und somit

$$\begin{aligned}\mathbb{E}[X] - \mathbb{E}[X]\,\mathbb{P}(Y > 0) &= \mathbb{E}[X]\big(1 - \mathbb{P}(Y > 0)\big) = \mathbb{E}[X]\,\mathbb{P}(Y \leqslant 0)\\ &= \mathbb{E}[X]\,\mathbb{E}[\mathbb{1}_{\{Y \leqslant 0\}}] = \mathbb{E}[X \cdot \mathbb{1}_{\{Y \leqslant 0\}}],\end{aligned}$$

wobei die letzte Gleichung aus der Unabhängigkeit von X und Y folgt (messbare Funktion für Y ist $\mathbb{1}_{(\infty,0]}$).

(3) $\mathbb{P}(X < Y) = \int_{\mathbb{R}} \mathbb{P}(X \leqslant y) f_Y(y)\,\mathrm{d}y$.

Stimmt. Ansatz 1: Sei $f_{X,Y}$ die gemeinsame Dichte von X und Y. Dann gilt mit dem Satz von Fubini-Tonelli

$$\begin{aligned}\mathbb{P}(X < Y) &= \int_{\mathbb{R}^2} f_{X,Y}(x,y)\,\mathbb{1}_{\{x<y\}}\,\mathrm{d}x, y = \int_{\mathbb{R}} \int_{-\infty}^{y} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \int_{\mathbb{R}} \int_{-\infty}^{y} f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}} \mathbb{P}(X \leqslant y) f_Y(y)\,\mathrm{d}y,\end{aligned}$$

wobei der vorletzte Schritt aus der Unabhängigkeit der Zufallsvariablen folgt.

Ansatz 2:

$$\mathbb{P}(X < Y) = \mathbb{P}(X - Y < 0) = \mathbb{P}(X - Y \leqslant 0) - \underbrace{\mathbb{P}(X = Y)}_{=0 \text{ (unabh.)}}.$$

So folgt

$$\begin{aligned}\mathbb{P}(X - Y \leqslant 0) &= \int_{-\infty}^{0} f_{X+(-Y)}(x)\,\mathrm{d}x = \int_{-\infty}^{0} (f_X * f_{-Y})(x)\,\mathrm{d}x\\ &= \int_{-\infty}^{0} \int_{\mathbb{R}} f_X(x-\tau) f_{-Y}(\tau)\,\mathrm{d}\tau\,\mathrm{d}x\\ &= \int_{\mathbb{R}} \int_{-\infty}^{0} f_X(x-\tau)\mathrm{d}x\; f_{-Y}(\tau)\,\mathrm{d}\tau\\ &= \int_{\mathbb{R}} \int_{-\infty}^{-\tau} f_X(x)\,\mathrm{d}x\; f_{-Y}(\tau)\,\mathrm{d}\tau\\ &= \int_{\mathbb{R}} \mathbb{P}(X \leqslant -\tau) f_{-Y}(\tau)\,\mathrm{d}\tau\\ &= \int_{\mathbb{R}} \mathbb{P}(X \leqslant -\tau) f_Y(-\tau)\,\mathrm{d}\tau\\ &= \int_{\mathbb{R}} \mathbb{P}(X \leqslant \tau) f_Y(\tau)\,\mathrm{d}\tau = \int_{\mathbb{R}} \mathbb{P}(X < \tau) f_Y(\tau)\,\mathrm{d}\tau\end{aligned}$$

(2) Für reellwertige Zufallsvariablen $X, X_1, X_2, \ldots$ auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ gilt

1 Aus $X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-f.s.}} X$ folgt $X_n \xrightarrow[n\to\infty]{L_1(\mathbb{P})} X$.

Falsch. Aus punktweise Konvergenz folgt nicht Norm-Konvergenz.

Betrachte hierzu Hütchenfunktionen mit höher werdenden Spitzen und schrumpfender Basis, deren Flächeninhalt konstant sind, aber punktweise gegen die Nullfunktion konvergieren.

2 stochastische Konvergenz $X_n \to X$ impliziert die Existenz einer Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $n_k < n_{k+1}$, sodass $|X_{n_k} - X| \xrightarrow[k\to\infty]{\mathbb{P}\text{-f.s.}} 0$ gilt.

Stimmt.

Beweis. Seien $\varepsilon > 0$. Nach Voraussetzung existiert eine Teilfolge $(n_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ sodass

$$\mathbb{P}(|X_{n_k} - X| > 2^{-k}) \leqslant 2^{-k}.$$

TODO: induktiv konstruierbar.

Nun gilt

$$\sum_{k\in\mathbb{N}} \mathbb{P}(|X_{n_k} - X| > 2^{-k}) \leqslant \sum_{k\in\mathbb{N}} 2^{-k} = 1 < \infty.$$

Das Lemma von BOREL-CANTELLI impliziert

$$\mathbb{P}\left(\limsup_{k\to\infty} |X_{n_k} - X| > 2^{-k}\right) = \mathbb{P}\left(\bigcap_{k\in\mathbb{N}} \bigcup_{\ell\geqslant k} \{|X_{n_\ell} - X| > 2^{-\ell}\}\right) = 0$$

und somit

$$\mathbb{P}\left(\bigcup_{k\in\mathbb{N}} \bigcap_{\ell\geqslant k} \{|X_{n_\ell} - X| \geqslant 2^{-\ell}\}\right) = 1$$

und (vgl. Beweis)

$$\lim_{k\to\infty} \mathbb{P}\left(\bigcup_{\ell\geqslant k} \{|X_{n_\ell} - X| > 2^{-\ell}\}\right) = \lim_{k\to\infty} \mathbb{P}\left(|X_{n_k} - X| > 0\right) = 0$$

□

3 $X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-f.s.}} X$ impliziert $\mathbb{E}[|X|] \leqslant \liminf_{n\to\infty} \mathbb{E}[|X_n|]$.

Stimmt. Da $X_n \to X$ fast überall gilt, gilt auch $|X_n| \to |X|$ fast überall. Nach dem Lemma von FATOU folgt

Lemma von FATOU

$$\mathbb{E}[\liminf_{n\to\infty} |X_n|] \leqslant \liminf_{n\to\infty} \mathbb{E}[|X_n|].$$

Da aufgrund der fast überall Konvergenz der lim mit dem lim inf fast überall übereinstimmt, und das Integral Nullmenge ignoriert, folgt die Behauptung.

3 Für unkorrelierte reellwertige Zufallsvariablen $(X_k)_{k\in\mathbb{N}}$ auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $|X_k| \leqslant 52$ für alle $k \in \mathbb{N}$ gilt für $S_n := \sum_{k=1}^n X_k$

1 $\frac{1}{n}\sum_{k\in\mathbb{N}} X_k - \mathbb{E}[X_k] \xrightarrow[n\to\infty]{\text{in W.keit}} 0.$

Stimmt.

$$\frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}[X_k] \overset{\text{Def.}}{=} \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}[(X_k - \mathbb{E}[X_k])^2] \leqslant \frac{1}{n^2}\sum_{k=1}^n (2\cdot 52)^2$$
$$\leqslant \frac{1}{n^2} 2n \cdot 104^2 = \frac{104^2}{n} \xrightarrow{n\to\infty} 0.$$

Nach der Verallgemeinerung des schwachen Gesetzes der großen Zahlen folgt die Aussage.

2 $\frac{S_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}\text{-f.s.}} c \in \mathbb{R}.$

Falsch.

3 $\sqrt{n}\operatorname{Var}\left(\frac{S_n}{n}\right) \xrightarrow{n\to\infty} 0.$

Stimmt. Nach den Rechenregeln für die Varianz und der Identität von B... gilt

$$\left|\sqrt{n}\operatorname{Var}\left(\frac{S_n}{n}\right)\right| = \left|\frac{\sqrt{n}}{n^2}\sum_{k=1}^n \operatorname{Var}(X_k)\right| \leqslant \frac{104n}{n^{\frac{3}{2}}} = \frac{104}{\sqrt{n}} \xrightarrow{n\to\infty} 0.$$

4 Seien $(X_k)_{k\in\mathbb{N}}$ u.i.v. auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathbb{E}[X_1] =: \mu$ und $\operatorname{Var}[X_1] =: \sigma^2 \in (0,\infty)$ sowie $S_n := \sum_{k=1}^n X_k$. Dann gilt

1 $\frac{S_n - n\mu}{n\sigma^2} \xrightarrow[n\to\infty]{\text{d}} Z \sim \mathcal{N}(0,1)$

Falsch. Sieht fast aus wie zentraler Grenzwertsatz, es fehlt aber ein Faktor.

2 $\mathbb{P}\left(\frac{S_n}{n} \leqslant \mu\right) \xrightarrow{n\to\infty} \frac{1}{2}.$

Stimmt. Nach zentralem Grenzwertsatz konvergiert S_n/n gegen eine normalverteilte Zufallsvariable, die um ihren Erwartungswert symmetrisch ist.

3 $(X_k)_{k\in\mathbb{N}}$ genügt dem schwachen schwachen Gesetz der großen Zahlen.

Stimmt. Da die Zufallsvariablen unabhängig sind, sind die insbesondere unkorreliert. Da ihre Varianz endlich ist, sind die Voraussetzungen des schwachen Gesetzes der großen Zahlen erfüllt.

5 Sei X eine reellwertige Zufallsvariable. Dann gilt für alle $c > 0$

1 $\mathbb{P}(|X| \geqslant c) \leqslant \frac{\mathbb{E}[X^2]}{c^2},$

Stimmt. Markovsche Ungleichung mit $\Phi(t) = t^2$, da $|X|^2 = X^2$ und $|X|$ nichtnegativ.

2 $\mathbb{P}(|X| \geqslant c) \leqslant \mathbb{E}[e^{X-c}]$,

Stimmt. Markovsche Ungleichung mit $\Phi(t) = e^t$.

3 $\mathbb{P}(|X| \geqslant c) \leqslant \frac{\mathrm{Var}[X]}{c^2}$.

Falsch. Betrachte eine degenerierte (= konstante) Zufallsvariable, z.B. $X \sim \mathrm{Ber}(1)$, also $X \equiv 1$. Dann gilt $\mathrm{Var}[X] = 1 \cdot (1-1) = 0$ aber

$$\mathbb{P}(X \geqslant 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = 0) = 1 + 0 > 0.$$

Stannat 2013 Seien $(X_k)_{k\in\mathbb{N}}$ unabhängige auf $[0,1]$ gleichverteilte Zufallsvariablen und $f : [0,1] \to \mathbb{R}$ eine stetige beschränkte Funktion Zeigen Sie, dass dann

1 gilt:

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \int_0^1 f(x)\,\mathrm{d}x\right) = 1$$

Beweis. Dies ist das starke Gesetz der großen Zahlen für $Y_k := f(X_k)$. Die Folge $(Y_k)_{k\in\mathbb{N}}$ is unabhängig, weil f messbar ist. Ferner gilt für alle $k \in \mathbb{N}$

$$\mathbb{E}[Y_k] = \mathbb{E}[f(X_k)] = \int_{\mathbb{R}} f(x) \cdot f_{X_k}(x) dx = \int_0^1 f(x) dx,$$

wobei $f_{X_k} := \mathbb{1}_{[0,1]}$ die Dichte von X_k ist. Das starkes Gesetz der großen Zahlen impliziert $\frac{1}{n}\sum_{k=1}^n Y_k \xrightarrow[n\to\infty]{\mathbb{P}\text{-f.s.}} \mathbb{E}[Y_1]$ und somit insbesondere stochastische Konvergenz. □

2 für alle $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \int_0^1 f(x)\,\mathrm{d}x\right| \leqslant \frac{\varepsilon}{\sqrt{n}}\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2(f)}} \int_{-\varepsilon}^{\varepsilon} \exp\left(-\frac{y^2}{2\sigma^2(f)}\right) \mathrm{d}y$$

gilt, wobei $\sigma^2(f) := \int_0^1 f^2(x)\,\mathrm{d}x - \left(\int_0^1 f(x)\,\mathrm{d}x\right)^2$ ist.

Beweis. Es gilt

$$\mathrm{Var}(Y_k) = \mathbb{E}[Y_k^2] - \mathbb{E}[Y_k]^2 = \sigma^2(f).$$

Nach dem zentralen Grenzwertsatz (ZG) folgt mit $m := \mathbb{E}[Y_1]$

und $S_n := \sum_{k=1}^n Y_k$

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(-\varepsilon \leqslant \sqrt{\frac{n}{\sigma^2(f)}}\left(\frac{S_n}{n} - m\right) \leqslant \varepsilon\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{\sigma^2(f)}}\left|\left(\frac{S_n}{n} - m\right)\right| \leqslant \frac{\varepsilon}{\sqrt{n}}\right) \\
&\overset{\text{(ZG)}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{y^2}{2}}\,\mathrm{d}y = \frac{1}{\sqrt{2\pi\sigma^2(f)}} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{y^2}{2\sigma^2(f)}}\,\mathrm{d}y,
\end{aligned}$$

aber die letzte Gleichheilt gilt nur für große ε... **TODO** □