Einführung in die Stochastik

Klaus Ritter

Darmstadt, SS 2008

Literatur

Insbesondere

H.-O. Georgii, *Stochastik*, 3. Auflage, de Gruyter, Berlin, 2007.

A. Irle, *Wahrscheinlichkeitstheorie und Statistik*, 2. Auflage, Teubner, Stuttgart, 2005.

U. Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 8. Auflage, Vieweg, Braunschweig, 2005.

Inhaltsverzeichnis

Kapitel I

Grundbegriffe

Ein mathematisches Modell für ein zufälliges Phänomen besteht aus einem Wahrscheinlichkeitsraum und darauf definierten Zufallsvariablen.
Themen dieser einführenden Vorlesung sind

- allgemeine Begriffsbildung und Konstruktionen,
- Analyse und Simulation stochastischer Modelle,
- Analyse empirischer Daten.

1 Wahrscheinlichkeitsräume

Die möglichen *Ergebnisse* ω eines Zufallsexperimentes bilden den *Ergebnisraum* Ω. Im allgemeinen ist Ω eine nicht-leere Menge ohne weitere Struktur.

Beispiel 1.

(i) Würfeln, $\Omega := \{1, \ldots, 6\}$.

(ii) Anzahl Anrufe in Callcenter, $\Omega := \mathbb{N}_0$.

(iii) Wartezeit bei Anruf, $\Omega := \mathbb{R}_+ := [0, \infty[$.

(iv) Verlauf eines Aktienkurses über die Zeit von 0 bis T, $\Omega := C_+([0, T])$ Menge der nicht-negativen stetigen Funktionen auf $[0, T]$.

Für gewisse Teilmengen $A \subseteq \Omega$, genannt *Ereignisse*, definiert man die Wahrscheinlichkeit ihres Eintretens ($\omega \in A$). Die Menge $\mathfrak{A}$ aller Ereignisse in Ω bilden den *Ereignisraum*.

Beispiel 2.

(i) Würfeln, Ergebnis ist gerade Zahl, $A := \{2, 4, 6\}$.

(ii) Anzahl Anrufe in Callcenter, Kapazitätsgrenze K wird nicht überschritten, $A := \{0, \dots, K\}$.

(iii) Wartezeit bei Anruf, Wartezeit liegt zwischen 1 und 2 (Minuten), $A := [1, 2]$.

(iv) Verlauf eines Aktienkurses, Kurs weicht von Anfangswert um nicht mehr als 10 (Euro) ab,

$$A := \{\omega \in C_+([0, T]) : \sup_{0 \le t \le T} |\omega(0) - \omega(t)| \le 10\}.$$

Nun: mengentheoretische Operationen mit Ereignissen.

Beispiel 3.

(i) Ereignis A oder Ereignis B tritt ein, $A \cup B$.

(ii) Ereignis A und Ereignis B treten ein, $A \cap B$.

(iii) Ereignis A tritt nicht ein, $A^{\mathrm{c}} := \Omega \setminus A$.

(iv) (mindestens) eines der Ereignisse $A_1, A_2, \dots$ tritt ein, $\bigcup_{i=1}^{\infty} A_i$.

(v) alle Ereignisse $A_1, A_2, \dots$ treten ein, $\bigcap_{i=1}^{\infty} A_i$.

Forderung: obige Operationen liefern wieder Ereignisse (Abschlußeigenschaft). Dazu führt man den Begriff der σ-Algebra ein.

Bezeichnung. Die *Potenzmenge* von Ω (Menge aller Teilmengen von Ω) wird mit $\mathfrak{P}(\Omega)$ und die *Mächtigkeit* (Anzahl der Elemente) einer endlichen Menge U wird mit $|U|$ bezeichnet.

Beispiel 4. Münzwurf, $\Omega := \{\mathrm{Z}, \mathrm{K}\}$,

$$\mathfrak{P}(\Omega) = \{\emptyset, \{\mathrm{Z}\}, \{\mathrm{K}\}, \{\mathrm{Z}, \mathrm{K}\}\}.$$

Beachte $\mathrm{Z} \notin \mathfrak{P}(\Omega)$, aber $\{\mathrm{Z}\} \in \mathfrak{P}(\Omega)$.

Satz 5. Für endliche Mengen Ω gilt

$$|\mathfrak{P}(\Omega)| = 2^{|\Omega|}.$$

Beweis. Durch Induktion über $n := |\Omega|$.
Verankerung: Gelte $|\Omega| = 0$. Dann $\Omega = \emptyset$ und $|\mathfrak{P}(\Omega)| = |\{\emptyset\}| = 1$.
Induktionsschritt: Gelte $|\Omega| = n + 1 \geq 1$ und sei die Behauptung für Mengen der Mächtigkeit n bereits bewiesen. Fixiere $\omega^* \in \Omega$. Dann

$$\begin{aligned}|\mathfrak{P}(\Omega)| &= |\{A \subseteq \Omega : \omega^* \in A\}| + |\{A \subseteq \Omega : \omega^* \notin A\}| \\ &= 2^n + 2^n = 2^{n+1}.\end{aligned}$$

□

Definition 6. $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$ heißt *σ-Algebra* (in Ω), falls

(i) $\Omega \in \mathfrak{A}$,

(ii) $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$,

(iii) $A_1, A_2, \ldots \in \mathfrak{A} \Rightarrow \bigcup_{i=1}^\infty A_i \in \mathfrak{A}$.

Forderung: der Ereignisraum ist eine σ-Algebra.

Bemerkung 7. In der Regel betrachtet man $\mathfrak{A} := \mathfrak{P}(\Omega)$, falls Ω abzählbar ist. Dies ist im allgemeinen nicht mehr möglich, falls Ω überabzählbar ist, siehe Kapitel IV.

Lemma 8. Für jede σ-Algebra $\mathfrak{A}$ gilt

(i) $\emptyset \in \mathfrak{A}$,

(ii) $A, B \in \mathfrak{A} \Rightarrow A \cup B,\ A \cap B, A \setminus B \in \mathfrak{A}$,

(iii) $A_1, A_2, \ldots \in \mathfrak{A} \Rightarrow \bigcap_{i=1}^\infty A_i \in \mathfrak{A}$.

Beweis. Ad (i): Es gilt $\emptyset = \Omega^c \in \mathfrak{A}$ nach Definition 6.(i), 6.(ii).
Ad (ii): Seien $A, B \in \mathfrak{A}$. Es gilt

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \ldots \in \mathfrak{A}$$

nach (i) und Definition 6.(iii). Es gilt $A^c \cup B^c \in \mathfrak{A}$ nach (ii) und Definition 6.(ii) und somit auch

$$A \cap B = (A^c \cup B^c)^c \in \mathfrak{A}.$$

Man verwende $A \setminus B = A \cap B^c$ um $A \setminus B \in \mathfrak{A}$ zu zeigen.
Ad (iii): Für $A_1, A_2, \ldots \in \mathfrak{A}$ gilt

$$\bigcap_{i=1}^\infty A_i = \left(\bigcup_{i=1}^\infty A_i^c\right)^c \in \mathfrak{A}$$

nach Definition 6.(ii), 6.(iii). □

Bezeichnung. Mengen $A_1, A_2, \ldots$ heißen *paarweise disjunkt* (p.d.), falls $A_i \cap A_j = \emptyset$ für $i \neq j$.

Im folgenden sei $\mathfrak{A}$ eine σ-Algebra in einer nicht-leeren Menge Ω.

Nun: Zuordung von Wahrscheinlichkeiten $P(A)$ zu den einzelnen Ereignissen $A \in \mathfrak{A}$. Dabei leitet uns folgende Vorstellung: Bei einer „großen“ Anzahl von „unabhängigen“ Wiederholungen des Zufallsexperimentes liegt die relative Häufigkeit des Eintretens von Ereignis A „nahe“ bei $P(A)$.

Definition 9. $P : \mathfrak{A} \to [0,1]$ heißt *Wahrscheinlichkeitsmaß* oder *Wahrscheinlichkeitsverteilung* (auf $\mathfrak{A}$), falls

(i) $P(\Omega) = 1$,

(ii) $A_1, A_2, \ldots \in \mathfrak{A}$ p.d. $\Rightarrow P\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty P(A_i)$ (σ-*Additivität*).
Genauer: $\ldots \Rightarrow \sum_{i=1}^\infty P(A_i)$ (absolut) konvergent und $\ldots$

Beispiel 10. Die *Gleichverteilung* auf einer endlichen Menge Ω wird definiert durch (*Laplace-Annahme*)

$$P(A) := |A|/|\Omega|, \qquad A \subseteq \Omega.$$

Speziell gilt für alle $\omega \in \Omega$

$$P(\{\omega\}) = 1/|\Omega|.$$

Beh.: P ist Wahrscheinlichkeitsmaß auf $\mathfrak{A} := \mathfrak{P}(\Omega)$.

Beweis. Offensichtlich gilt $0 \le P(A) \le 1$ und $P(\Omega) = 1$.
Für $A_1, A_2, \ldots \subseteq \Omega$ p.d. (notwendig: $A_i = \emptyset$ bis auf endlich viele i) gilt

$$\left|\bigcup_{i=1}^\infty A_i\right| = \sum_{i=1}^\infty |A_i|\,.$$

Dies zeigt die σ-Additivität. □

Definition 11. $(\Omega, \mathfrak{A}, P)$ heißt *Wahrscheinlichkeitsraum*, falls Ω eine nicht-leere Menge, $\mathfrak{A}$ eine σ-Algebra in Ω und P Wahrscheinlichkeitsmaß auf $\mathfrak{A}$ ist.

Beispiel 12. Ein stochastisches Modell für einmaliges Würfeln wird definiert durch

(i) $\Omega := \{1, \ldots, 6\}$,

(ii) $\mathfrak{A} := \mathfrak{P}(\Omega)$,

(iii) P Gleichverteilung auf Ω.

Beispiel 13. Ein stochastisches Modell für Geschlecht eines Neugeborenen wird definiert durch

(i) $\Omega := \{\mathrm{W}, \mathrm{M}\}$,

(ii) $\mathfrak{A} := \mathfrak{P}(\Omega)$,

(iii) P definiert durch $P(\{\mathrm{W}\}) := 0.4863$.

Letzteres wurde empirisch ermittelt als relative Häufigkeit unter den 25 171 123 Lebendgeburten in D in den Jahren 1970–1999. Siehe Hesse (2003, p. 23).

Beispiel 14. Hard core model der Physik, Gleichverteilung auf einer sehr großen Menge „unbekannter" Mächtigkeit.
Betrachte ein Gitter

$$G = \{1, \dots, m\}^d$$

mit $d = 3$. Mit Ω_0 bezeichnen wir die Menge aller Abbildungen (Konfigurationen)

$$\varphi : G \to \{0, 1\}.$$

Interpretation: $\varphi(x) = 1$ gdw. der Gitterpunkt $x \in G$ besetzt ist. Es gilt $|\Omega_0| = 2^{(m^d)}$.
Zwei Gitterpunkte $x, y \in G$ heißen benachbart, falls sie sich in genau einer Koordinate unterscheiden, d.h., falls $\sum_{i=1}^d |x_i - y_i| = 1$. Eine Konfiguration $\varphi \in \Omega_0$ heißt zulässig, falls benachbarte Gitterpunkte $x, y \in G$ nicht zugleich besetzt sind, also $\varphi(x) \cdot \varphi(y) = 0$ erfüllen.
Man studiert die Gleichverteilung auf der Menge Ω der zulässigen Konfigurationen. Siehe Häggström (2002).

Beispiel 15. Ein (fragwürdiges) stochastisches Modell für Pfeiltreffer auf Dartscheibe mit Radius $r > 0$ wird definiert durch

(i) $\Omega := \{\underbrace{(x, y)}_{=:\omega} \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$,

(ii) eine „σ-Algebra $\mathfrak{A}$ in Ω, deren Elementen ein 'Flächeninhalt' $\lambda(A)$ zugeordnet werden kann", siehe Kapitel IV.1 und Kapitel IV.2,

(iii) $P(A) := \lambda(A)/(\pi r^2)$.

Beachte: Es gilt $P(\{\omega\}) = 0$ für alle $\omega \in \Omega$.

Hilfreiche Intuition zu Wahrscheinlichkeiten: relative Anzahlen, siehe Beispiel 10, oder normierte Flächeninhalte, siehe Beispiel 15.
Im folgenden bezeichne $(\Omega, \mathfrak{A}, P)$ stets einen Wahrscheinlichkeitsraum.
Nun: Eigenschaften von Wahrscheinlichkeitsmaßen.

Satz 16. Für $A, B \in \mathfrak{A}$ gilt

(i) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ (*Additivität*),

(ii) $A \subseteq B \Rightarrow P(B) = P(A) + P(B \setminus A)$,

(iii) $A \subseteq B \Rightarrow P(A) \leq P(B)$ (*Monotonie*),

(iv) $P(A^c) = 1 - P(A)$,

(v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Beweis. Seien $A, B \in \mathfrak{A}$.
Ad(i): Im Fall $A \cap B = \emptyset$ gilt

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \ldots$$

mit p.d. Mengen und somit

$$P(A \cup B) = P(A) + P(B) + \sum_{i=1}^{\infty} P(\emptyset).$$

Da $P(A \cup B) < \infty$, folgt $P(\emptyset) = 0$ und weiter die Behauptung.
Ad (ii): Im Fall $A \subseteq B$ gilt

$$B = A \cup (B \setminus A)\,.$$

mit disjunkten Mengen A und $B \setminus A$. Verwende (i).
Ad (iii): Verwende (ii) und $P(B \setminus A) \geq 0$.
Ad (iv): Verwende (ii) mit $B = \Omega$, also $P(B) = 1$.
Ad (v): Verwende

$$A \cup B = A \cup (B \setminus A) = A \cup \big(B \setminus (A \cap B)\big)$$

zusammen mit (i) und (ii). □

Satz 17. Seien $A_1, A_2, \ldots \in \mathfrak{A}$. Dann gilt

(i) $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) \in [0, \infty]$ (*σ-Subadditivität*),

(ii) aus $A_1 \subseteq A_2 \subseteq \ldots$ folgt $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n)$ (*σ-Stetigkeit von unten*),

(iii) aus $A_1 \supseteq A_2 \supseteq \ldots$ folgt $P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n)$ (*σ-Stetigkeit von oben*).

Beweis. Seien $A_1, A_2, \ldots \in \mathfrak{A}$. Setze $A_0 := \emptyset$ und $B_i = A_i \setminus \left(\bigcup_{j=0}^{i-1} A_j\right)$ für $i \geq 1$. Lemma 8 sichert $B_i \in \mathfrak{A}$. Beachte, daß $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ mit p.d. Mengen B_i.
Ad (i): Mit der σ-Additivität und der Monotonie von P folgt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) \leq \sum_{i=1}^{\infty} P(A_i). \tag{1}$$

Ad (ii): Aus $A_1 \subseteq A_2 \subseteq \ldots$ folgt $P(B_i) = P(A_i) - P(A_{i-1})$, siehe Satz 16.(ii), und weiter

$$\sum_{i=1}^{n} P(B_i) = P(A_n).$$

Verwende (1) um

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty}\left(\sum_{i=1}^{n} P(B_i)\right) = \lim_{n\to\infty} P(A_n)$$

zu erhalten.
Ad (iii): Verwende (ii) und Komplementbildung. □

2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Gegeben sei ein Wahrscheinlichkeitsmaß P auf einer σ-Algebra $\mathfrak{A}$. Falls bekannt ist, daß ein Ereignis $B \in \mathfrak{A}$ eingetreten ist, kann man zu einem neuen Wahrscheinlichkeitsmaß auf $\mathfrak{A}$ übergehen. Idee: „Restriktion“ auf B und Normierung.

Definition 1. Für $A, B \in \mathfrak{A}$ mit $P(B) > 0$ heißt

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$

bedingte Wahrscheinlichkeit von A gegeben B.

Bemerkung 2. $P(\cdot \mid B)$ ist ein Wahrscheinlichkeitsmaß auf $\mathfrak{A}$ mit $P(B \mid B) = 1$.

Beispiel 3. Für die Gleichverteilung P auf einer endlichen Menge Ω und $\emptyset \neq B \subseteq \Omega$ sowie $A \subseteq \Omega$ gilt

$$P(A \mid B) = \frac{|A \cap B|}{|B|},$$

d.h. $P(A \mid B)$ ist der relative Anteil von Elementen aus A in B. Somit ist $P(\cdot \mid B)$ die „Gleichverteilung“ auf B. Analog in Beispiel 1.15.

Speziell: einmaliges Würfeln (wie üblich modelliert) und $B := \{1, 5, 6\}$. Dann

$$P(\{\omega\} \mid B) = \begin{cases} 1/3, & \text{falls } \omega \in B \\ 0, & \text{falls } \omega \notin B. \end{cases}$$

Beispiel 4. Betrachte zwei weiße (1, 2) und drei schwarze (3, 4, 5) Kugeln, ziehe zwei Kugeln ohne Zurücklegen. Gesucht ist die Wahrscheinlichkeit dafür, daß die 2. Kugel schwarz ist, falls die 1. Kugel weiß war.
Modell: Gleichverteilung auf

$$\Omega := \{(\omega_1, \omega_2) \in \{1, \ldots, 5\}^2 : \omega_1 \neq \omega_2\}.$$

Für

$$A := \{(\omega_1, \omega_2) \in \Omega : \omega_2 \geq 3\},$$
$$B := \{(\omega_1, \omega_2) \in \Omega : \omega_1 \leq 2\}$$

gilt (wie erwartet)

$$P(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{6}{8} = \frac{3}{4}.$$

Satz 5. Für p.d. Mengen $B_1, \ldots, B_n \in \mathfrak{A}$ gelte

$$\forall i: \; P(B_i) > 0$$

und

$$\bigcup_{i=1}^{n} B_i = \Omega.$$

Dann folgt für jedes $A \in \mathfrak{A}$ (Formel von der totalen Wahrscheinlichkeit)

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i).$$

Falls zusätzlich $P(A) > 0$ gilt, so folgt (Formel von Bayes)

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j) \cdot P(B_j)}.$$

Analoge Aussagen gelten für abzählbar viele Mengen B_i.

Beweis. Zur totalen Wahrscheinlichkeit: Es gilt

$$A = \bigcup_{i=1}^{n} (A \cap B_i)$$

mit p.d. Mengen $A \cap B_i$. Somit folgt

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i).$$

Zur Formel von Bayes: Es gilt für jedes $i \in \{1, \dots, n\}$

$$P(B_i \mid A) = \frac{P(B_i \cap A)}{P(A)} \cdot \frac{P(B_i)}{P(B_i)} = \frac{P(A \mid B_i) \cdot P(B_i)}{P(A)}.$$

Die Behauptung ergibt sich nun mit Hilfe der Formel von der totalen Wahrscheinlichkeit. □

Beispiel 6. Situation

- 3 Maschinen, $i = 1, 2, 3$,
- Anteil an Tagesproduktion, $r_i = 60\%, 30\%, 10\%$,
- Anteil defekter Produkte pro Maschine, $d_i = 1\%, 2\%, 3\%$.

Fragen

- Mit welcher Wahrscheinlichkeit ist ein zufällig gewähltes Produkt defekt?
- Mit welcher Wahrscheinlichkeit stammt ein defektes Produkt von Maschine 1?

Modell

(i) $\Omega := \{(1,+), (1,-), (2,+), (2,-), (3,+), (3,-)\}$,

(ii) $\mathfrak{A} := \mathfrak{P}(\Omega)$,

(iii) P definiert durch

$$P(\{i,-\}) := r_i \cdot d_i, \qquad P(\{i,+\}) := r_i \cdot (1 - d_i).$$

Für $D := \{(1,-), (2,-), (3,-)\}$ und $M_i := \{(i,+), (i,-)\}$ folgt

$$P(M_i) = r_i, \qquad P(D \mid M_i) = \frac{r_i \cdot d_i}{r_i} = d_i.$$

Man erhält

$$P(D) = d_1 r_1 + d_2 r_2 + d_3 r_3 = \frac{3}{200}$$

und

$$P(M_1 \mid D) = \frac{P(D \mid M_1) \cdot P(M_1)}{P(D)} = \frac{200}{3} \cdot d_1 r_1 = \frac{2}{5}.$$

Häufig wie in diesem Beispiel: Modellierung durch Vorgabe bedingter Wahrscheinlichkeiten, etwa bei Markov-Ketten. Siehe Übung 5:G7 und 5:H19.

Nun: ein zentraler Begriff der Wahrscheinlichkeitstheorie.

Definition 7. Zwei Ereignisse $A, B \in \mathfrak{A}$ heißen *unabhängig*, falls

$$P(A \cap B) = P(A) \cdot P(B).$$

Bemerkung 8. Im Falle $P(B) > 0$ gilt

$$A, B \text{ unabhängig} \Leftrightarrow P(A \mid B) = P(A).$$

Beispiel 9. Einmaliges Würfeln (wie üblich modelliert) und

$$B := \{1, 2, 3, 4\}, \qquad A_1 := \{2, 4, 6\}, \qquad A_2 := \{1\}.$$

Dann gilt

$$P(A_1 \mid B) = \frac{1}{2}, \qquad P(A_1) = \frac{1}{2},$$

d.h. A_1, B unabhängig. Ferner gilt

$$P(A_2 \mid B) = \frac{1}{4}, \qquad P(A_2) = \frac{1}{6},$$

d.h. A_2, B nicht unabhängig.

Beispiel 10. Zweimaliger Wurf einer fairen Münze,

$$\Omega := \{(\mathrm{Z}, \mathrm{Z}), (\mathrm{Z}, \mathrm{K}), (\mathrm{K}, \mathrm{Z}), (\mathrm{K}, \mathrm{K})\},$$

$\mathfrak{A} := \mathfrak{P}(\Omega)$ und P Gleichverteilung auf Ω. Betrachte

$$\begin{aligned} A_1 &:= \{(\mathrm{Z}, \mathrm{Z}), (\mathrm{Z}, \mathrm{K})\} && \text{1. Wurf Z,} \\ A_2 &:= \{(\mathrm{Z}, \mathrm{K}), (\mathrm{K}, \mathrm{K})\} && \text{2. Wurf K,} \\ A_3 &:= \{(\mathrm{Z}, \mathrm{K}), (\mathrm{K}, \mathrm{Z})\} && \text{Würfe verschieden.} \end{aligned}$$

Es gilt $|A_i| = 2$ und $|A_i \cap A_j| = 1$ für $i \neq j$. Also

$$A_1,\ A_2 \text{ unabh.}, \qquad A_1,\ A_3 \text{ unabh.}, \qquad A_2,\ A_3 \text{ unabh.}$$

Im folgenden sei $I = \{1, \ldots, n\}$ oder $I = \mathbb{N}$.

Definition 11. Eine Folge $(A_i)_{i \in I}$ von Ereignissen heißt *unabhängig*, falls für jede endliche Menge $\emptyset \neq J \subseteq I$ gilt

$$P\Bigl(\bigcap_{j \in J} A_j\Bigr) = \prod_{j \in J} P(A_j).$$

Vgl. Definition 7 im Spezialfall $|I| = |J| = 2$.

Bemerkung 12. Falls $(A_i)_{i\in I}$ unabhängig, so folgt die *paarweise Unabhängigkeit*

$$\forall j_1, j_2 \in I,\ j_1 \neq j_2 :\ A_{j_1}, A_{j_2} \text{ unabhängig.}$$

Die Umkehrung ist falsch, wie das folgende Beispiel zeigt.

Beispiel 13. Zweimaliger Wurf einer fairen Münze, siehe Beispiel 10. Die Folge (A_1, A_2, A_3) ist nicht unabhängig, da

$$|A_1 \cap A_2 \cap A_3| = 1.$$

Alternative Argumentation: es gilt$P(A_3|A_1 \cap A_2) = 1$ aber $P(A_3) = 1/2$, siehe Bemerkung 14.

Bemerkung 14. $(A_i)_{i\in I}$ ist genau dann unabhängig, wenn für alle endlichen Mengen $\emptyset \neq J_1, J_2 \subseteq I$ mit $J_1 \cap J_2 = \emptyset$ und $P\left(\bigcap_{j_2\in J_2} A_{j_2}\right) \neq 0$ gilt

$$P\left(\bigcap_{j_1\in J_1} A_{j_1} | \bigcap_{j_2\in J_2} A_{j_2}\right) = P\left(\bigcap_{j_1\in J_1} A_{j_1}\right).$$

Beweis. Übung 3:G3. □

Gängige Sprechweise: $A_1, A_2, \dots$ sind unabhängig statt $(A_1, A_2, \dots)$ ist unabhängig.

3 Reellwertige Zufallsvariablen

Oft interessiert man sich (nur) für spezielle Aspekte eines Zufallsexperimentes. Dazu betrachtet man Abbildungen $\Omega \to \mathbb{R}$.

Bezeichnung. Die *Indikatorfunktion* $1_U : V \to \mathbb{R}$ einer Teilmenge $U \subseteq V$ ist definiert durch

$$1_U(x) := \begin{cases} 1, & \text{falls } x \in U \\ 0, & \text{falls } x \notin U. \end{cases}$$

Beispiel 1. Anzahl Anrufe in Callcenter an Tagen $1, \dots, n$. Dazu sei

$$\Omega := \mathbb{N}_0^n := \{\underbrace{(\omega_1, \dots, \omega_n)}_{=:\omega} : \omega_i \in \mathbb{N}_0 \text{ für } i = 1, \dots, n\}$$

und $\mathfrak{A} := \mathfrak{P}(\Omega)$. Die Anzahl Anrufe an Tag i ist gegeben durch

$$X_i(\omega) := \omega_i$$

und die Gesamtanzahl der Anrufe ist gegeben durch

$$X(\omega) := \sum_{i=1}^{n} \omega_i$$

Mittels

$$Y_i(\omega) := 1_{\{K+1,\dots\}}(\omega_i)$$

wird beschrieben, ob an Tag i die Kapazitätsgrenze K überschritten wurde. Die Anzahl der Tage an, an denen die Kapazitätsgrenze K überschritten wurde, ist gegeben durch

$$Y(\omega) := \sum_{i=1}^{n} 1_{\{K+1,\dots\}}(\omega_i).$$

Definition 2. Eine Abbildung $X : \Omega \to \mathbb{R}$ heißt (reellwertige) *Zufallsvariable* (auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$), falls

$$\forall x \in \mathbb{R} : \ \{\omega \in \Omega : X(\omega) \leq x\} \in \mathfrak{A}.$$

Manchmal zugelassen: Funktionswerte $\pm\infty$. Siehe auch Lemma 8 und Lemma IV.3.1.

Bemerkung 3.

(i) Zufallsvariablen sind Abbildungen!

(ii) Für Zufallsvariablen sind die Wahrscheinlichkeiten $P(\{\omega \in \Omega : X(\omega) \leq x\})$ wohldefiniert.

(iii) Im Falle $\mathfrak{A} = \mathfrak{P}(\Omega)$ ist jede Abbildung $\Omega \to \mathbb{R}$ eine Zufallsvariable.

Bezeichnung. Suggestive Kurzschreibweise

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\}$$

für $X : \Omega \to \mathbb{R}$ und $B \subseteq \mathbb{R}$ sowie

$$\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} = \{X \in \,]-\infty, x]\}$$

für $x \in \mathbb{R}$. Analog mit „=“ usw. Also: Urbilder von Mengen, $\{X \in B\} = X^{-1}(B)$.

Beispiel 4. Callcenter

$$\Omega := \mathbb{N}_0^n, \qquad \mathfrak{A} := \mathfrak{P}(\Omega), \qquad X_i(\omega) := \omega_i.$$

Es gilt (Verknüpfung von Zufallsvariablen)

$$X = \sum_{i=1}^{n} X_i, \qquad Y_i = 1_{\{K+1,\dots\}} \circ X_i, \qquad Y = \sum_{i=1}^{n} Y_i.$$

Gängige Schreibweise $1_A(X_i)$ für $1_A \circ X_i$.

Spezielle Ereignisse: Gesamtanzahl der Anrufe liegt zwischen 1000 und 2000,

$$\{1000 \leq X \leq 2000\},$$

Kapazitätsgrenze K wurde nie überschritten,

$$\bigcap_{i=1}^{n} \{X_i \leq K\} = \bigcap_{i=1}^{n} \{Y_i = 0\} = \{Y = 0\}.$$

Definition 5. Die *Verteilungsfunktion* $F_X : \mathbb{R} \to [0,1]$ einer Zufallsvariable X auf $(\Omega, \mathfrak{A}, P)$ ist definiert durch

$$F_X(x) := P(\{X \leq x\}).$$

Wir werden sehen, daß die wahrscheinlichkeitstheoretische Struktur einer Zufallsvariable vollständig durch ihre Verteilungsfunktion beschrieben wird.

Gleichheit von Zufallsvariablen in folgendem Sinn (und damit schwächer als die Gleichheit von Abbildungen).

Definition 6. Zwei Zufallsvariablen X auf $(\Omega, \mathfrak{A}, P)$ und X' auf $(\Omega', \mathfrak{A}', P')$ heißen *identisch verteilt*, falls $F_X = F_{X'}$.

Beispiel 7. Betrachte die Gleichverteilung P auf $\Omega := \{1, \ldots, n\}$ und setze

$$X(\omega) := \omega.$$

Dann gilt

$$P(\{X = x\}) = \begin{cases} 1/n, & \text{falls } x \in \{1, \ldots, n\} \\ 0, & \text{sonst} \end{cases}$$

Betrachte den Wahrscheinlichkeitsraum $(\Omega', \mathfrak{A}', P')$ zur Modellierung des Pfeiltreffers auf einer Dartscheibe, siehe Beispiel 1.15. Definiere $A'_0 = \{(0,0)\}$ und für $x = 1, \ldots, n$ die Sektoren

$$A'_x := \{\rho(\cos\alpha, \sin\alpha) : \rho \in \,]0, r]\,,\ \alpha \in \,](x-1)/n \cdot 2\pi, x/n \cdot 2\pi]\}.$$

Sei $X'(\omega')$ der getroffene Sektor, d.h.

$$X'(\omega') := x,$$

falls $\omega' \in A'_x$. Dann gilt für $x = 1, \ldots, n$

$$P'(\{X' = x\}) = P'(A'_x) = \lambda(A'_x)/\lambda(\Omega') = 1/n,$$

sowie $P'(\{X' = x\}) = 0$ für $x \in \mathbb{R} \setminus \{1, \ldots, n\}$. Also

$$\forall x \in \mathbb{R} : \ P(\{X = x\}) = P'(\{X' = x\}). \tag{1}$$

Beh.: X und X' sind identisch verteilt.

Beweis. Für $x \in \mathbb{R}$, $M := \,]-\infty, x]$ und $D := \{1, \ldots, n\}$ gilt

$$\begin{aligned} P(\{X \le x\}) &= P(\{X \in M\} \cap \{X \in D\}) \\ &= \sum_{y \in M \cap D} \underbrace{P(\{X = y\})}_{=P'(\{X'=y\})} = P'(\{X' \le x\}). \end{aligned}$$

□

Siehe auch Lemma III.4.4.

Warnung: aus (1) folgt im allgemeinen nicht, daß X und X' identisch verteilt sind.

Bezeichnung. $\mathfrak{M} := \{M \subseteq \mathbb{R} : M \text{ oder } M^c \text{ Intervall}\}$.

Lemma 8. Für jede Zufallsvariable X auf $(\Omega, \mathfrak{A}, P)$ gilt

$$\forall M \in \mathfrak{M} : \; \{X \in M\} \in \mathfrak{A}.$$

Beweis. Im folgenden seien $a, b \in \mathbb{R}$ mit $a \le b$. Wir beweisen die Aussage in Teilschritten.

0. Für $M = \mathbb{R}$ gilt $\{X \in M\} = \Omega \in \mathfrak{A}$.

1. Für $M = \,]-\infty, b]$ gilt $\{X \in M\}$ nach Definition einer Zufallsvariable.

2. Für $M = \,]a, \infty[$ gilt $M = \mathbb{R} \setminus \,]-\infty, a]$, also

$$\{X \in M\} = \{X \in \mathbb{R}\} \setminus \{X \in \,]-\infty, a]\} \in \mathfrak{A}.$$

3. Für $M = \,]-\infty, b[$ gilt $M = \bigcup_{n=1}^{\infty} \,]-\infty, b - 1/n]$, also

$$\{X \in M\} = \bigcup_{n=1}^{\infty} \{X \in \,]-\infty, b - 1/n]\} \in \mathfrak{A}.$$

4. Für $M = [a, \infty[$ gilt $M = \mathbb{R} \setminus \,]-\infty, a[$, also

$$\{X \in M\} = \{X \in \mathbb{R}\} \setminus \{X \in \,]-\infty, a[\} \in \mathfrak{A}.$$

5. Beschränkte Intervalle sind Durchschnitte der unter 1.–4. betrachteten Intervalle, und es gilt

$$\{X \in I_1 \cap I_2\} = \{X \in I_1\} \cap \{X \in I_2\}.$$

Demnach gilt die Aussage für alle Intervalle, unter Benutzung der Definition der σ-Algebra auch für deren Komplemente. □

Satz 9. Zwei Zufallsvariablen X auf $(\Omega, \mathfrak{A}, P)$ und X' auf $(\Omega', \mathfrak{A}', P')$ sind genau dann identisch verteilt, wenn

$$\forall M \in \mathfrak{M} : \; P(\{X \in M\}) = P'(\{X' \in M\}).$$

Beweis (und Rechentechnik). Zu zeigen ist nur „$\Rightarrow$“. Wir gehen wieder in Teilschritten vor.

1. Für $M :=]a, b]$ gilt

$$\begin{aligned} P(\{X \in M\}) &= P(\{X \le b\} \setminus \{X \le a\}) \\ &= P(\{X \le b\}) - P(\{X \le a\}) \\ &= P'(\{X' \le b\}) - P'(\{X' \le a\}) = P'(\{X' \in M\}). \end{aligned}$$

2. Für $M :=]a, b[$ gilt $M = \bigcup_{n=1}^{\infty}]a, b - 1/n]$. Also nach Satz 1.17.(ii) und 1.

$$\begin{aligned} P(\{X \in M\}) &= P\left(\bigcup_{n=1}^{\infty} \{X \in]a, b - 1/n]\}\right) \\ &= \lim_{n\to\infty} P\left(\{X \in]a, b - 1/n]\}\right) \\ &= \lim_{n\to\infty} P'\left(\{X' \in]a, b - 1/n]\}\right) = P'(\{X' \in M\}). \end{aligned}$$

3. Für $M := \{a\}$ gilt

$$M = \bigcap_{n=1}^{\infty}]a - 1/n, a + 1/n[\,.$$

Also nach Satz 1.17.(iii) und 2.

$$\begin{aligned} P(\{X \in M\}) &= \lim_{n\to\infty} P\left(\{X \in]a - 1/n, a + 1/n[\}\right) \\ &= \lim_{n\to\infty} P'\left(\{X' \in]a - 1/n, a + 1/n[\}\right) = P'(\{X' \in M\}). \end{aligned}$$

Für alle weiteren Typen von Mengen $M \in \mathfrak{M}$ nutze man die σ-Additivität, Additivität und die Rechenregel für Komplemente. $\square$

Bemerkung 10. Gemäß Satz 9 bestimmt die Verteilungsfunktion F_X die Wahrscheinlichkeiten $P(\{X \in M\})$ für $M \in \mathfrak{M}$ eindeutig. Siehe auch Satz IV.3.6.

Im folgenden sei $I := \{1, \dots, n\}$ oder $I := \mathbb{N}$. Wir betrachten eine Folge $(X_i)_{i\in I}$ von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$.

Definition 11. $X_1, X_2 \dots$ heißen *identisch verteilt*, falls für alle $i, j \in I$ die Zufallsvariablen X_i und X_j identisch verteilt sind.

Definition 12. $(X_i)_{i\in I}$ heißt *unabhängig*, falls für jede Folge $(x_i)_{i\in I}$ in $\mathbb{R}$ die Folge $(\{X_i \le x_i\})_{i\in I}$ von Ereignissen unabhängig ist.

Gängige Sprechweise: $X_1, X_2, \dots$ sind unabhängig statt $(X_1, X_2, \dots)$ ist unabhängig.

Bemerkung 13. $(X_i)_{i\in I}$ ist genau dann unabhängig, wenn für alle endlichen Mengen $J \subseteq I$ mit $|J| \geq 2$ und Folgen $(x_j)_{j\in J}$ in $\mathbb{R}$ gilt

$$P\left(\bigcap_{j\in J}\{X_j \leq x_j\}\right) = \prod_{j\in J} P(\{X_j \leq x_j\}).$$

Beispiel 14. Zweimaliger Münzwurf, siehe Beispiel 2.10. Betrachte für $i = 1, 2$ die Zufallsvariablen $X_i(\omega) := 1_{\{\mathrm{Z}\}}(\omega_i)$.
Beh.: X_1 und X_2 sind unabhängig und identisch verteilt.

Beweis. Es gilt

$$\{X_i \leq x\} = \begin{cases} \emptyset, & \text{für } x < 0 \\ \{\omega \in \Omega : \omega_i = K\}, & \text{für } 0 \leq x < 1 \\ \Omega, & \text{für } x \geq 1. \end{cases}$$

Also

$$P(\{X_i \leq x\}) = \begin{cases} 0 & , \text{ falls } x < 0 \\ \frac{1}{2} & , \text{ falls } 0 \leq x < 1 \\ 1 & , \text{ falls } x \geq 1. \end{cases}$$

Insbesondere sind X_1 und X_2 identisch verteilt.
Für die Unabhängigkeit ist noch zu zeigen, daß

$$P(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}) = P(\{X_1 \leq x_1\}) \cdot P(\{X_2 \leq x_2\}). \tag{2}$$

Klar: (2) gilt, falls $x_1 < 0, x_2 < 0, x_1 \geq 1$ oder $x_2 \geq 1$. Für $x_i \in [0, 1[$ gilt

$$\begin{aligned} P(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}) &= P(\{(K, K)\}) = \frac{1}{4} \\ &= P(\{X_1 \leq x_1\}) \cdot P(\{X_2 \leq x_2\}). \end{aligned}$$

□

Satz 15. $(X_i)_{i\in\mathbb{N}}$ ist genau dann unabhängig, wenn

$$\forall n \in \mathbb{N}\ \forall M_1, \ldots, M_n \in \mathfrak{M} : P\left(\bigcap_{j=1}^{n}\{X_j \in M_j\}\right) = \prod_{j=1}^{n} P(\{X_j \in M_j\}).$$

Analog für $I := \{1, \ldots, n\}$. Siehe auch Satz IV.3.10.

Beweis. Übung 4:H13. □

Bemerkung 16. Stochastische Modelle beruhen sehr häufig auf einer unabhängigen Folge von identisch verteilten Zufallsvariablen. Abkürzung: *iid* für independent and identically distributed.

Kapitel II

Stochastische Simulation

Hier ein erster Einblick, siehe auch Grinstead, Snell (1997). Mehr zu diesem Themenkreis (Monte-Carlo-Algorithmen) in eigenen Vorlesungen und etwa bei Müller-Gronbach, Novak, Ritter (2008).

Problemstellung: Gegeben sei eine Zufallsvariable X auf $(\Omega, \mathfrak{A}, P)$ sowie eine Menge $M \in \mathfrak{M}$. Berechne $P(\{X \in M\})$.

1 Die Methode der direkten Simulation

Beispiel 1. Erfolgswahrscheinlichkeit einer Strategie beim Patience-Spiel[1]. Dazu sei P die Gleichverteilung auf der Menge Ω aller Permutationen von $\{1, \dots, 52\}$ (Kartenverteilungen). Ferner sei $X := 1_A$, wobei A die Menge aller Kartenverteilungen, bei denen die Strategie gewinnt, und $M := \{1\}$. Somit

$$P(\{X \in M\}) = P(A) = |A|/|\Omega|.$$

Hier $|\Omega| = 52! = 8.06 \ldots \cdot 10^{67}$, also Ω sehr groß.

Beispiel 2. Durchgang von Neutronen durch Materie. Dazu seien $(S_t^1)_{t \in \mathbb{N}}$, $(S_t^2)_{t \in \mathbb{N}}$ und $(S_t^3)_{t \in \mathbb{N}}$ Folgen von Zufallsvariablen, so daß $(S_t^1(\omega), S_t^2(\omega), S_t^3(\omega)) \in \mathbb{R}^3$ die Position eines Neutrons nach t Kollisionen beschreibt. Schließlich tritt einer dieser Fälle ein:

- Neutron wird von Abschirmung absorbiert, $X(\omega) := 0$,
- Neutron wird von Abschirmung reflektiert, $X(\omega) := 1$,
- Neutron passiert Abschirmung, $X(\omega) := 2$.

Gesucht ist $P(\{X = 2\})$. Hier X „kompliziert“.

[1] Ulam (1946), siehe Los Alamos Science, Special Issue.

Beispiel 3. Ruinwahrscheinlichkeiten als stark vereinfachte Version von Beispiel 2. Betrachte ein einfaches Spiel mit Gewinn oder Verlust ± 1 pro Runde mit Wahrscheinlichkeit p bzw. $1-p$, wobei $p \in]0,1[$. Dazu sei $(Y_j)_{j\in\mathbb{N}}$ iid mit

$$P(\{Y_1 = 1\}) = p, \qquad P(\{Y_1 = -1\}) = 1-p.$$

Dann ist

$$S_t := \sum_{j=1}^{t} Y_j$$

der akkumulierte Gewinn nach t Runden. Ziel sei die Vermehrung eines Startkapitals $a \in \mathbb{N}$ um einen Betrag $b \in \mathbb{N}$. Die Spieldauer wird dann beschrieben durch[2]

$$\tau(\omega) := \inf\{t \in \mathbb{N} : S_t(\omega) = -a \text{ oder } S_t(\omega) = b\}.$$

Gesucht ist die Ruinwahrscheinlichkeit

$$r(a,b,p) := P(\{\omega \in \Omega : \tau(\omega) < \infty \text{ und } S_{\tau(\omega)}(\omega) = -a\}),$$

d.h. $P(\{X = 1\})$ für $X := 1_{\{\tau<\infty\}\cap\{S_\tau=-a\}}$. Ausblick: Ruinwahrscheinlichkeit eines Schadenversicherers.

Wir zeigen an dieser Stelle nur $P(\{\tau < \infty\}) = 1$. Setze $d := a+b-1$. Klar: Spiel endet spätestens nach d Gewinnen in Folge. Betrachte also

$$G_k := \{(Y_{(k-1)\cdot d+1}, \dots, Y_{k\cdot d}) = (1,\dots,1)\}, \qquad k \in \mathbb{N}.$$

Dann

$$\{\tau > \ell \cdot d\} \subseteq \bigcap_{k=1}^{\ell} G_k^{\mathrm{c}},$$

und $G_1^{\mathrm{c}}, \dots, G_\ell^{\mathrm{c}}$ sind unabhängig mit

$$P(G_k^{\mathrm{c}}) = 1 - P\Bigg(\bigcap_{j=(k-1)\cdot d+1}^{k\cdot d} \{Y_j = 1\}\Bigg) = 1 - p^d < 1,$$

siehe Übung 3:H9 und Satz I.3.9. Es folgt

$$P(\{\tau > \ell \cdot d\}) \le (1-p^d)^\ell,$$

und weiter

$$0 \le P(\{\tau = \infty\}) \le \lim_{\ell\to\infty} P(\{\tau > \ell \cdot d\}) = 0.$$

Bezeichnung. Sei $(X_i)_{i\in I}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$. Für $\omega \in \Omega$ wird die Folge $(X_i(\omega))_{i\in I}$ reeller Zahlen als eine *Realisierung* von $(X_i)_{i\in I}$ bezeichnet. Spezialfall: $|I| = 1$, Realisierung einer Zufallsvariable.

[2]Konvention: $\inf \emptyset = \infty$.

Methode der direkten Simulation. Betrachte eine iid Folge von Zufallsvariablen $X_1', \dots, X_n'$ auf $(\Omega', \mathfrak{A}', P')$, wobei X und X_1' identisch verteilt sind.

1. „Erzeuge" eine Realisierung $(x_1', \dots, x_n')$ von $(X_1', \dots, X_n')$.
2. Approximiere $P(\{X \in M\}) = P'(\{X_1' \in M\})$ durch die relative Häufigkeit
$$\frac{1}{n} \cdot \sum_{i=1}^{n} 1_M(x_i').$$
Dies entspricht der frequentistischen Interpretation von Wahrscheinlichkeiten, siehe Seite 4.

Nun: Approximation der Verteilungsfunktion F_X von X.

Definition 4. Die *empirische Verteilungsfunktion* $F_n(\cdot\,; x_1, \dots, x_n) : \mathbb{R} \to [0,1]$ zu $x_1, \dots, x_n \in \mathbb{R}$ ist definiert durch
$$F_n(x; x_1, \dots, x_n) := 1/n \cdot \sum_{i=1}^{n} 1_{]-\infty,x]}(x_i).$$

Fragen: Konvergieren die relativen Häufigkeiten $\frac{1}{n} \cdot \sum_{i=1}^{n} 1_M(x_i')$ gegen $P(\{X \in M\})$? Konvergieren die empirischen Verteilungsfunktionen $F_n(\cdot\,; x_1', \dots, x_n')$ gegen F_X? Positive Antworten geben wir in Abschnitt VII.2.

Die Mathematische Statistik betrachtet bei diesen Fragen statt Simulationsdaten empirische Daten $x_1', \dots, x_n' \in \mathbb{R}$, siehe bereits Beispiel I.1.13.

2 Zufallszahlen

Die praktische Durchführung der direkten Simulation (Schritt 1.) beruht auf Zufallszahlengeneratoren, d.h. auf geeigneten Abbildungen $f : K \to K$, wobei $K := \{0, \dots, k-1\}$ und k sehr groß[3], etwa $k = 2^{128} = 3.4 \ldots \cdot 10^{38}$ oder $k = 2^{19937} = 4.3 \ldots \cdot 10^{6001}$.

Man erhält endliche Folgen $v_1, v_2, \ldots \in K$ durch die Initialisierung $v_0 \in K$ und die Iteration $v_\ell := f(v_{\ell-1})$. „Kleine" Abschnitte dieser Folgen „verhalten sich annähernd" wie Realisierungen von iid Zufallsvariablen, die jeweils gleichverteilt auf K sind.

Definition 1. Eine Zufallsvariable U heißt *gleichverteilt* auf $[0,1]$, falls
$$F_U(x) = \begin{cases} 0, & \text{falls } x < 0, \\ x, & \text{falls } 0 \le x \le 1, \\ 1, & \text{falls } x > 1. \end{cases}$$

Bez.: $U \sim \mathbf{U}([0,1])$.

[3] Zum Vergleich: die Anzahl der Atome in Universum liegt in der Größenordnung 10^{79}.

Bemerkung 2. Für $U \sim \mathbf{U}([0,1])$, $x \in \mathbb{R}$ und ein Intervall M mit Endpunkten $0 \leq a < b \leq 1$ gilt

$$P(\{U = x\}) = 0,$$
$$P(\{U \in M\}) = F_U(b) - F_U(a) = b - a.$$

Setze $u_\ell := v_\ell / k$. „Kleine Abschnitte" der Folgen $u_1, u_2, \ldots \in [0,1[$ „verhalten sich annähernd" wie Realisierungen von iid Zufallsvariablen, die jeweils gleichverteilt auf $[0,1]$ sind. Kurz: $u_1, u_2, \ldots$ (gleichverteilte) *Zufallszahlen* in $[0,1]$.

In dieser Vorlesung: Zufallszahlengenerator als black-box. Verwendung etwa in der Form

(i) Initialisierung `init(u0)`,

(ii) Iteration `u := rand()`.

Beispiel 3. Simulation des Spiels aus Beispiel 1.3. Dazu sei

$$T(u) := \begin{cases} 1, & \text{falls } u > 1 - p \\ -1, & \text{falls } u \leq 1 - p. \end{cases}$$

Ist $(U'_j)_{j \in \mathbb{N}}$ iid mit $U'_1 \sim \mathbf{U}([0,1])$, dann definiert

$$Y'_j := T(U'_j)$$

eine iid-Folge mit

$$P'(\{Y'_1 = 1\}) = p, \qquad P'(\{Y'_1 = -1\}) = 1 - p.$$

Dies entspricht einmaligen Wurf einer fairen ($p = 1/2$) bzw. unfairen ($p \neq 1/2$) Münze. Die Simulation der Münzwürfe geschieht also durch die Transformation

$$x'_i := T(u'_i)$$

von gleichverteilten Zufallszahlen $u'_i \in [0,1]$.

Die folgenden Abbildungen basieren auf der Wahl $p := 1/2$. Die Abbildungen II.1–II.3 zeigen für eine Simulation von 10 000 Münzwürfen die Entwicklung der relativen Häufigkeit eines Gewinns. Die Abbildung II.4 zeigt eine Entwicklung des akkumulierten Gewinns über 1000 Runden. Die Abbildung II.5 zeigt die Simulation von vier Spielverläufen im Fall $a = 20$ und $b = 60$. Für $a = 20$ und $b = 60$ sowie für $a = b = 100$ zeigen die Abbildungen II.6 und II.7 für eine Simulation von jeweils 10 000 Spielverläufen die Entwicklung des relativen Anteils der Spiele, die im Ruin enden.

Vermutung: für die Ruinwahrscheinlichkeiten $r(a,b,p)$ gilt

$$r(a,b,1/2) = \frac{b}{a+b}.$$

Beweis: Übung 6:H26.

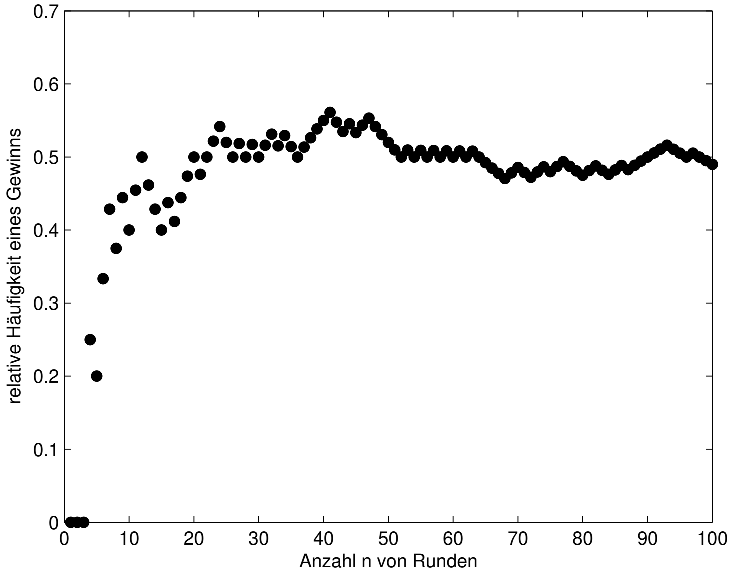

Abbildung II.1: Münzwurf, relative Häufigkeit eines Gewinns, $n \leq 100$

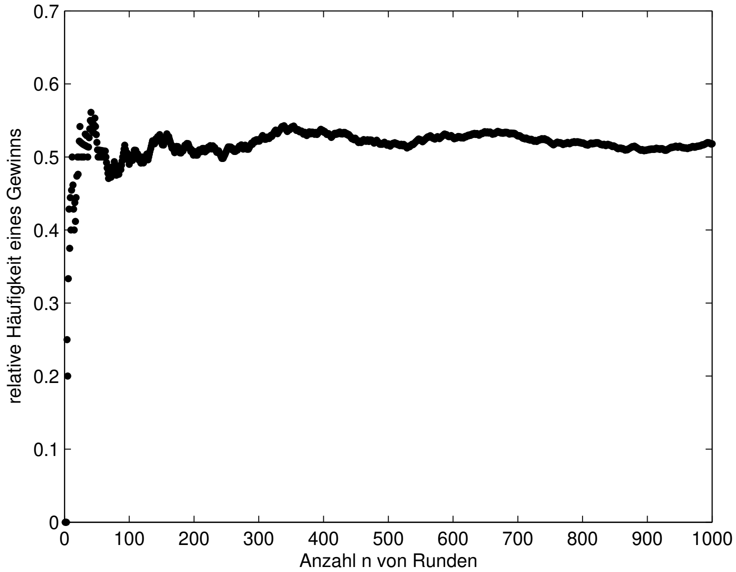

Abbildung II.2: Münzwurf, relative Häufigkeit eines Gewinns, $n \leq 1000$

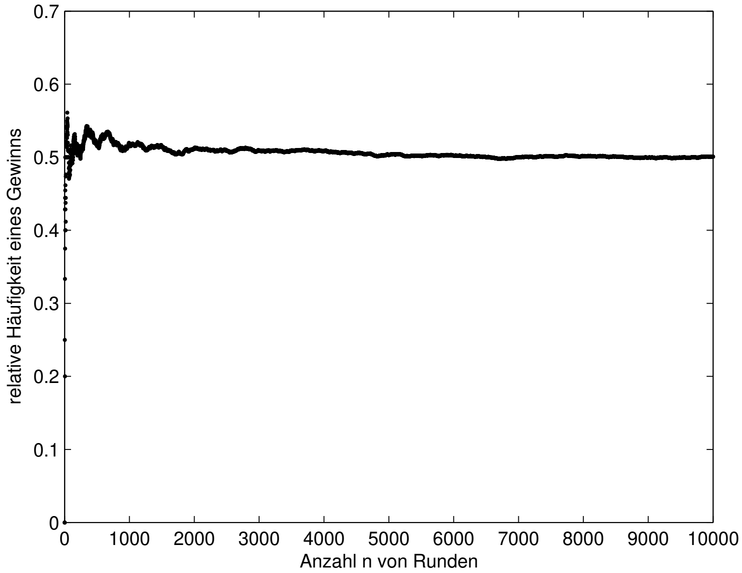

Abbildung II.3: Münzwurf, relative Häufigkeit eines Gewinns, $n \leq 10\,000$

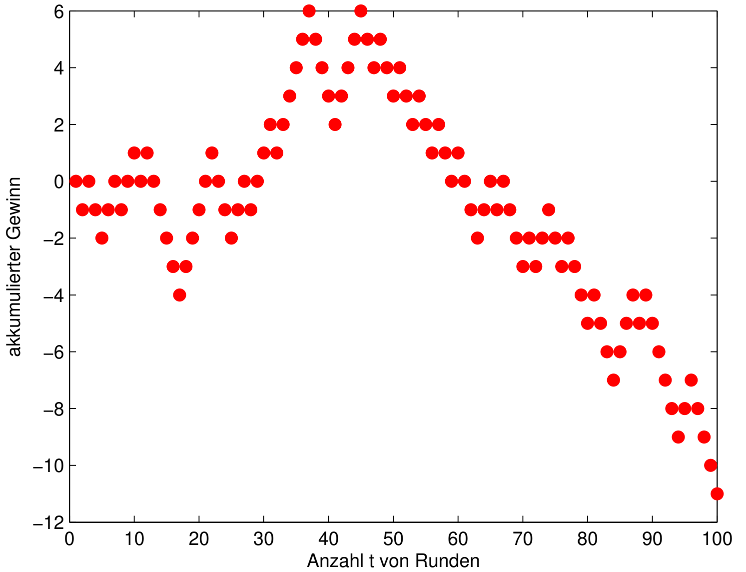

Abbildung II.4: Ruinproblem, akkumulierter Gewinn, $t \leq 100$

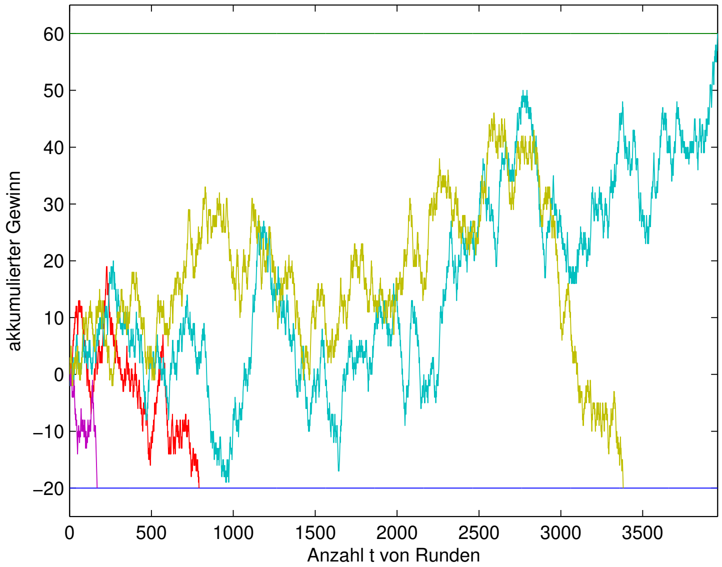

Abbildung II.5: Ruinproblem, $a = 20$, $b = 60$, vier Spielverläufe

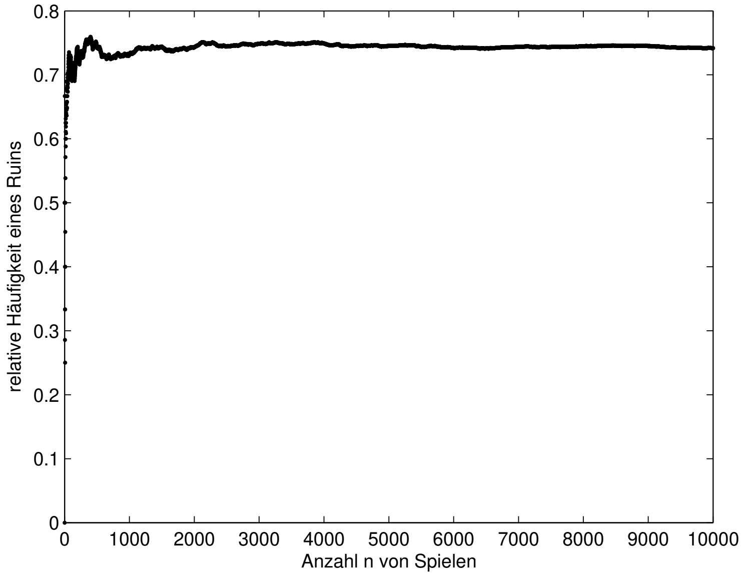

Abbildung II.6: Relative Häufigkeit eines Ruins, $a = 20$, $b = 60$, $n \leq 10\,000$

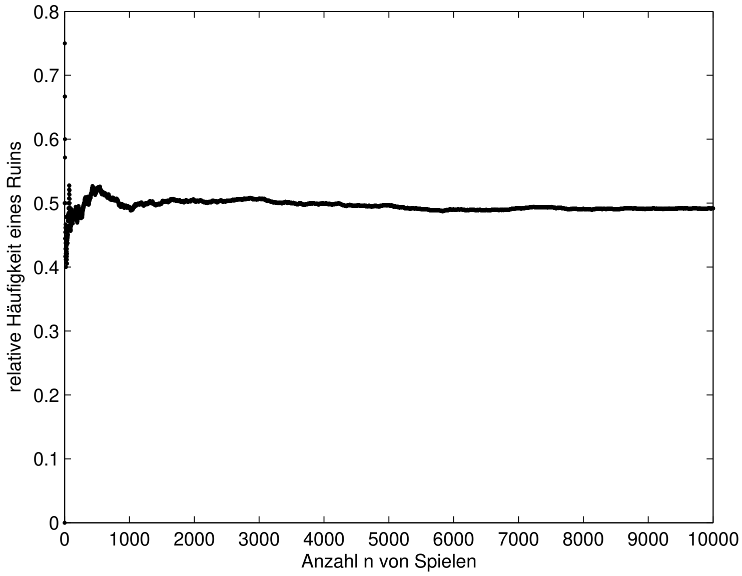

Abbildung II.7: Relative Häufigkeit eines Ruins, $a = b = 40$, $n \leq 10\,000$

3 Die Inversionsmethode

Frage: Simulation „komplizierter" Verteilungen. Eine „universelle" Möglichkeit bietet die Inversionsmethode.

Satz 1. Für jede Verteilungsfunktion F_X gilt

(i) F_X ist monoton wachsend,

(ii) F_X ist rechtsseitig stetig,

(iii) $\lim_{x\to\infty} F_X(x) = 1$ und $\lim_{x\to-\infty} F_X(x) = 0$.

Ferner gilt für alle $x \in \mathbb{R}$

$$P(\{X = x\}) = F_X(x) - \lim_{y\to x-} F_X(y)$$

und

$$P(\{X = x\}) = 0 \quad \Leftrightarrow \quad F_X \text{ stetig in } x.$$

Beweis. Übung 3:H12. □

Definition 2. Für $\alpha \in \,]0, 1[$ heißt

$$q_\alpha := \inf\{v \in \mathbb{R} : F_X(v) \geq \alpha\}$$

das α-*Quantil* der Verteilungsfunktion F_X. Speziell: $q_{1/2}$ *Median*.

Lemma 3. Für $\alpha \in \,]0,1[$ und jede Verteilungsfunktion F_X gilt

$$\{v \in \mathbb{R} : F_X(v) \geq \alpha\} = [q_\alpha, \infty[\,.$$

Beweis. Sei $I := \{v \in \mathbb{R} : F_X(v) \geq \alpha\}$. Eigenschaft (iii) aus Satz 1 sichert $\emptyset \neq I \subsetneq \mathbb{R}$, und mit Eigenschaft (i) folgt, daß I von der Form $[q, \infty[$ oder $]q, \infty[$ mit $q \in \mathbb{R}$ ist. Eigenschaft (ii) zeigt $I = [q, \infty[$. Es folgt $q = q_\alpha$. □

Gegeben sei eine Funktion $F : \mathbb{R} \to [0,1]$ mit Eigenschaften (i)–(iii) aus Satz 1 sowie eine Zufallsvariable $U \sim \mathbf{U}([0,1])$.
Definiere $T_F : \mathbb{R} \to \mathbb{R}$ durch

$$T_F(u) := \begin{cases} \inf\{v \in \mathbb{R} : F(v) \geq u\}, & \text{falls } u \in \,]0,1[, \\ 0, & \text{sonst.} \end{cases}$$

Satz 4. $X := T_F \circ U$ ist eine Zufallsvariable mit $F_X = F$.

Beweis. Für $x \in \mathbb{R}$ und $u \in \,]0,1[$ gilt

$$T_F(u) \leq x \Leftrightarrow \inf\{v \in \mathbb{R} : F(v) \geq u\} \leq x \Leftrightarrow F(x) \geq u.$$

Also

$$\{X \leq x\} = \{U \in \,]0,1[\} \cap \{F(x) \geq U\} \cup \{U \notin \,]0,1[\} \cap \underbrace{\{0 \leq x\}}_{\in \{\emptyset, \Omega\}},$$

so daß insbesondere $\{X \leq x\} \in \mathfrak{A}$. Weiter

$$P(\{X \leq x\}) = P(\{F(x) \geq U\}) = F(x).$$

□

Bemerkung 5. Achtung: die Existenz von Zufallsvariablen $U \sim \mathbf{U}([0,1])$ auf geeigneten Wahrscheinlichkeitsräumen bleibt an dieser Stelle offen und wird erst in Abschnitt IV.2 positiv geklärt.

Korollar 6. $F : \mathbb{R} \to [0,1]$ ist genau dann die Verteilungsfunktion einer Zufallsvariable, wenn F die Eigenschaften (i)–(iii) erfüllt.

Beweis. Wende die Sätze 1, 4 und IV.2.1 an. □

Also: Modellierung eines beliebigen Zufallsexperimentes mit reellwertigen Ergebnissen durch Vorgabe der Verteilungsfunktion F.

Inversionsmethode. Simulation einer Zufallsvariable X mit der Verteilungsfunktion $F := F_X$ mittels Transformation

$$T_F(u_1), T_F(u_2), \ldots$$

gleichverteilter Zufallszahlen $u_1, u_2, \ldots$ in $[0,1]$.

Beispiel 7. Das Vorgehen in Beispiel 2.3 entspricht der Inversionsmethode in einer elementaren Situation.

Kapitel III

Diskrete Modelle

In diesem Kapitel untersuchen wir stochastische Modelle für Zufallsexperimente, bei denen die Menge der möglichen Ausgänge abzählbar ist.

1 Wahrscheinlichkeitsfunktionen

Definition 1. $(\Omega, \mathfrak{A}, P)$ heißt *diskreter Wahrscheinlichkeitsraum*, falls Ω abzählbar ist und $\mathfrak{A} = \mathfrak{P}(\Omega)$ gilt.

Im folgenden seien Ω und $\mathfrak{A}$ wie oben.
Frage: Konstruktion von Wahrscheinlichkeitsmaßen auf $\mathfrak{A}$?

Definition 2. $f : \Omega \to \mathbb{R}_+$ heißt *Wahrscheinlichkeitsfunktion* (auf Ω), falls

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

Interpretation: Punktmassen $f(\omega)$. Graphische Darstellung: Stabdiagramm.

Satz 3.

(i) Jede Wahrscheinlichkeitsfunktion $f : \Omega \to \mathbb{R}_+$ definiert durch

$$P(A) := \sum_{\omega \in A} f(\omega), \qquad A \subseteq \Omega, \tag{1}$$

ein Wahrscheinlichkeitsmaß auf $\mathfrak{A}$, und es gilt

$$f(\omega) = P(\{\omega\}), \qquad \omega \in \Omega. \tag{2}$$

(ii) Für jedes Wahrscheinlichkeitsmaß P auf $\mathfrak{A}$ definiert (2) eine Wahrscheinlichkeitsfunktion f auf Ω, und es gilt (1) .

Beweis. Ad (i): Offenbar gilt $P(A) \in [0,1]$ und $P(\Omega) = 1$. Ferner gilt für $A := \bigcup_{i=1}^\infty A_i$ mit p.d. Mengen $A_i \subseteq \Omega$

$$P(A) = \sum_{\omega \in A} f(\omega) = \sum_{i=1}^\infty \sum_{\omega \in A_i} f(\omega) = \sum_{i=1}^\infty P(A_i),$$

ggf. aufgrund der absoluten Konvergenz der Reihen.
Ad (ii): klar. □

Modellierung durch Wahl von f. Siehe bereits Beispiel I.1.13 und I.2.6. Dazu

- kombinatorische Methoden, oft ausgehend von Gleichverteilungsannahmen,
- statistische Schätzung, siehe bereits Beispiel I.1.13.

Beispiel 4. Die Gleichverteilung auf einer endlichen Menge Ω entspricht der Wahrscheinlichkeitsfunktion

$$f(\omega) := \frac{1}{|\Omega|}, \qquad \omega \in \Omega.$$

2 Elementare Kombinatorik

Wir untersuchen die Abzählung von endlichen Mengen zur Berechnung von Wahrscheinlichkeiten (unter Gleichverteilungsannahmen).

Erinnerung

(i) $n! := 1 \cdot \ldots \cdot n$ für $n \in \mathbb{N}$ und $0! := 1$.

(ii) Binomialkoeffizienten: für $n \in \mathbb{N}_0$ und $k \in \{0, \ldots, n\}$ ist

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

und es gilt Rekursionsformel für $k \neq 0$

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

(iii) Binomischer Lehrsatz: für $a, b \in \mathbb{R}$ und $n \in \mathbb{N}_0$ gilt

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} \cdot a^k \cdot b^{n-k}.$$

Im folgenden seien $N, N_1, \ldots, N_k$ endliche nicht-leere Mengen und $n := |N|$.

Satz 1.
$$|N_1 \times \ldots \times N_k| = |N_1| \cdot \ldots \cdot |N_k|.$$

Beweis. Induktion über k. Hier: Induktionsschluß. Für $x \in N_{k+1}$ sei
$$A_x := \{(\omega_1, \omega_2, \ldots, \omega_{k+1}) \in N_1 \times \cdots \times N_{k+1} : x = \omega_{k+1}\}.$$
Dann $A_x \cap A_y = \emptyset$ für $x \neq y$ sowie
$$N_1 \times \ldots \times N_{k+1} = \bigcup_{x \in N_{k+1}} A_x.$$
Ferner, unter Verwendung der Induktionsannahme,
$$|A_x| = |N_1 \times \ldots N_k| = |N_1| \cdot \ldots \cdot |N_k|.$$
Fazit
$$|N_1 \times \ldots \times N_{k+1}| = \sum_{x \in N_{k+1}} |N_1 \times \ldots \times N_k| = |N_{k+1}| \cdot |N_1| \cdot \cdots \cdots \cdot |N_k|.$$
□

Bemerkung 2. Obiger Satz mit $N = N_1 = \ldots = N_k$ zeigt: Die Anzahl der Stichproben vom Umfang k aus N unter Berücksichtigung der Reihenfolge mit Zurücklegen beträgt n^k. Dies ist ebenfalls die Anzahl der Abbildungen $\{1, \ldots, k\} \to N$.

Satz 3.
$$\begin{aligned} &|\{(\omega_1, \ldots, \omega_k) \in N^k : \omega_1, \ldots, \omega_k \text{ paarw. verschieden}\}| \\ &= \frac{n!}{(n-k)!} = n \cdot (n-1) \cdots (n-k+1). \end{aligned}$$
Speziell für $k = n$: Die Anzahl der Permutationen von N beträgt $n!$

Beweis. Induktion über k. Hier: Induktionsschluß. Seien $2 \leq k+1 \leq n$ und $x \in N$. Setze
$$Q := \{(\omega_1, \ldots, \omega_{k+1}) \in N^{k+1} : \omega_1, \ldots, \omega_{k+1} \text{ paarw. verschieden}\}$$
und
$$A_x := \{\omega \in Q : \omega_{k+1} = x\}.$$
Dann
$$\begin{aligned} |A_x| &= |\{(\omega_1, \ldots, \omega_k) \in (N \backslash \{x\})^k : \omega_1, \ldots, \omega_k \text{ paarw. verschieden}\}| \\ &= (n-1) \cdot \cdots \cdot (n-1-k+1) = (n-1) \cdot \cdots \cdot (n-k). \end{aligned}$$
Also
$$|Q| = (n-1) \cdot \ldots \cdot (n-k) \cdot n = \frac{n!}{(n-(k+1))!}.$$
□

Bemerkung 4. Obiger Satz bestimmt die Anzahl der Stichproben vom Umfang k aus N unter Berücksichtigung der Reihenfolge ohne Zurücklegen. Dies ist ebenfalls die Anzahl der Injektionen $\{1, \dots, k\} \to N$.

Beispiel 5. Betrachte $\Omega := N^k$ mit der Gleichverteilung P. Gesucht sei die Wahrscheinlichkeit, daß in der Stichprobe $\omega \in \Omega$ mindestens 2 Komponenten übereinstimmen. Also

$$A := \{(\omega_1, \dots, \omega_k) \in \Omega : \exists i \neq j : \omega_i = \omega_j\}.$$

Es gilt $P(A) = 1 - P(A^c)$ und

$$\begin{aligned} P(A^c) &= \frac{|A^c|}{|\Omega|} = \frac{n!}{(n-k)! \cdot n^k} = \frac{n \cdot \ldots \cdot (n-(k-1))}{n \cdot \ldots \cdot n} \\ &= 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \ldots \cdot \left(1 - \frac{k-1}{n}\right). \end{aligned}$$

Für $n = 365$ (Geburtstagszwillinge) ergibt sich näherungsweise

k	4	16	22	23	40	64
$P(A)$	0,016	0,284	0,476	0,507	0,891	0,997

Implizit wird hier angenommen: die Geburtstage sind unabhängig und jeweils gleichverteilt auf $N = \{1, \dots, 365\}$, siehe Beispiel 3.4 und Bemerkung 3.7.

Satz 6. Für $k \in \{0, \dots, n\}$ gilt

$$|\{K \subseteq N : |K| = k\}| = \binom{n}{k}.$$

Beweis. Induktion über $n = |N|$. Hier: Induktionsschluß. Sei $|N'| = n + 1 \in \mathbb{N}$. Die Behauptung gilt offenbar für $k = 0$ und $k = n + 1$, also sei fortan $k \in \{1, \dots, n\}$. Fixiere $x \in N'$, setze $N := N' \setminus \{x\}$. Dann folgt

$$\begin{aligned} &|\{K' \subseteq N' : |K'| = k\}| \\ &\quad = |\{K' \subseteq N' : |K'| = k \wedge x \in K'\}| + |\{K' \subseteq N' : |K'| = k \wedge x \notin K'\}| \\ &\quad = |\{K \subseteq N : |K| = k-1\}| + |\{K \subseteq N : |K| = k\}| \\ &\quad = \binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}. \end{aligned}$$

$\square$

Bemerkung 7. Obiger Satz bestimmt die Anzahl der Stichproben vom Umfang k aus N ohne Berücksichtigung der Reihenfolge ohne Zurücklegen.
Vergleich der Sätze 3 und 6: es gilt

$$\binom{n}{k} \cdot k! = \frac{n!}{(n-k)!}.$$

Interpretation: Teilmenge auswählen und anordnen.

Beispiel 8. Lotto: Gleichverteilung auf

$$\Omega := \{K \subseteq \{1, \dots, 49\} : |K| = 6\}.$$

Also gilt für jedes $\omega \in \Omega$

$$P(\{\omega\}) = 1/\binom{49}{6} = 7,15\ldots \cdot 10^{-8}.$$

Die Anzahl der Stichproben vom Umfang k aus N ohne Berücksichtigung der Reihenfolge mit Zurücklegen wird in ÜBUNG 5:H18 bestimmt.

3 Produkträume

Gegeben: diskrete Wahrscheinlichkeitsräume $(\Omega_1, \mathfrak{A}_1, P_1), \dots, (\Omega_n, \mathfrak{A}_n, P_n)$ Gesucht: Modell für die „unabhängige Hintereinanderausführung“ der Einzelexperimente (Produktexperiment).

Beispiel 1. n-maliges Würfeln, n Geburten, usw. Fragwürdig bei Callcenter an n Tagen.

Definiere

$$\begin{aligned} \Omega &:= \Omega_1 \times \cdots \times \Omega_n, \\ \mathfrak{A} &:= \mathfrak{P}(\Omega), \\ f(\omega) &:= f_1(\omega_1) \cdot \ldots \cdot f_n(\omega_n), \quad \omega \in \Omega, \end{aligned}$$

wobei f_i die zu P_i gehörige Wahrscheinlichkeitsfunktion auf Ω_i bezeichnet.

Lemma 2. f ist eine Wahrscheinlichkeitsfunktion auf Ω.

Beweis. Klar: $f \geq 0$. Ferner, ggf. aufgrund der absoluten Konvergenz,

$$\begin{aligned} \sum_{\omega \in \Omega} f(\omega) &= \sum_{\omega_1 \in \Omega_1} \cdots \sum_{\omega_n \in \Omega_n} f_1(\omega_1) \cdot \ldots \cdot f_n(\omega_n) \\ &= \sum_{\omega_1 \in \Omega_1} f_1(\omega_1) \cdot \ldots \cdot \sum_{\omega_n \in \Omega_n} f_n(\omega_n) = 1. \end{aligned}$$

$\square$

Definition 3. Sei P das durch f definierte Wahrscheinlichkeitsmaß auf $\mathfrak{A}$. Dann heißt $(\Omega, \mathfrak{A}, P)$ das *Produkt der Wahrscheinlichkeitsräume* $(\Omega_i, \mathfrak{A}_i, P_i)$ und P das *Produkt der Wahrscheinlichkeitsmaße* P_i.

Beispiel 4. Für endliche Mengen Ω_i und Gleichverteilungen P_i ist das Produktmaß P die Gleichverteilung auf Ω, siehe Satz 2.1. So etwa für n-maliges Würfeln

$$\Omega = \{1, \dots, 6\}^n$$

und

$$P(\{\omega\}) = 1/6^n$$

für $\omega \in \Omega$.

Beispiel 5. Geschlecht von n Neugeborenen,

$$\Omega_i := \{\mathrm{W}, \mathrm{M}\}, \; f_i(\mathrm{W}) := p, \; f_i(\mathrm{M}) := 1 - p.$$

Also

$$f(\omega) = p^{k(\omega)} \cdot (1-p)^{n-k(\omega)}$$

mit

$$k(\omega) := |\{i \in \{1, \dots, n\} : \omega_i = \mathrm{W}\}|.$$

Im folgenden sei $(\Omega, \mathfrak{A}, P)$ das Produkt der Wahrscheinlichkeitsräume $(\Omega_i, \mathfrak{A}_i, P_i)$. Zurück zu den Einzelexperimenten gelangt man durch die Projektionen

$$X_i : \Omega \to \Omega_i,$$

d.h.

$$X_i(\omega) := \omega_i.$$

Siehe etwa Beispiel I.3.4.

Satz 6. Für $A_1 \subseteq \Omega_1, \dots, A_n \subseteq \Omega_n$ gilt

$$P\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n P(\{X_i \in A_i\}) = \prod_{i=1}^n P_i(A_i).$$

Beweis. Es gilt (vgl. Beweis Lemma 2)

$$\begin{aligned} P\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) &= P(A_1 \times \ldots \times A_n) = \sum_{\omega \in A_1 \times \ldots \times A_n} f(\omega) \\ &= \sum_{\omega_1 \in A_1} \cdots \sum_{\omega_n \in A_n} f_1(\omega_1) \cdot \ldots \cdot f_n(\omega_n) = P_1(A_1) \cdot \ldots \cdot P_n(A_n). \end{aligned}$$

Die Wahl von $A_j = \Omega_j$ für $j \neq i$ zeigt

$$P(\{X_i \in A_i\}) = P_i(A_i).$$

$\square$

Bemerkung 7. Ziel erreicht. Genauer:

(i) Das Produktraum-Modell beinhaltet die Modelle der Einzelexperimente, da

$$P(\{X_i \in A_i\}) = P_i(A_i)$$

für $i \in \{1, \ldots, n\}$ und $A_i \subseteq \Omega_i$.

(ii) Falls $\Omega_1, \ldots, \Omega_n \subseteq \mathbb{R}$ (oder bei Verwendung eines allgemeineren Begriffs von Zufallsvariablen und deren Unabhängigkeit), so sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen. Zum Beweis wähle man $A_i = \left]-\infty, x_i\right] \cap \Omega_i$ bzw. Ω_i in Satz 6.

Der Spezialfall $P_1 = \cdots = P_n$ liefert eine iid-Folge $X_1, \ldots, X_n$.

Siehe Georgii (2007, Kap. 3.2) und Krengel (2000, Kap. 2.6) zur Modellierung mehrstufiger Experimente mit Abhängigkeiten.

4 Diskrete Zufallsvariablen

Im folgenden seien $X, X_1, \ldots$ Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$.

Definition 1. X heißt *diskrete Zufallsvariable*, falls $P(\{X \in D\}) = 1$ für eine abzählbare Menge $D \subset \mathbb{R}$ gilt.

Bemerkung 2. $(\Omega, \mathfrak{A}, P)$ diskret $\Rightarrow$ $X(\Omega)$ abzählbar $\Rightarrow$ X diskret.

Beispiel 3. Pfeiltreffer auf Dartscheibe, X Nummer des getroffenen Sektors, siehe Beispiel I.3.7.

Lemma 4. Diskrete Zufallsvariablen X, X' sind genau dann identisch verteilt, wenn

$$\forall x \in \mathbb{R}: \; P(\{X = x\}) = P'(\{X' = x\}).$$

Beweis. „$\Rightarrow$": Wende Satz I.3.9 an.

„$\Leftarrow$": Betrachte eine abzählbare Menge $D \subset \mathbb{R}$ mit $P(\{X \in D\}) = P'(\{X' \in D\}) = 1$ im Beweis in Beispiel I.3.7 □

Bemerkung 5. Ist X diskret, so definiert $f(x) := P(\{X = x\})$ auf jeder abzählbaren Menge $D \subset \mathbb{R}$ mit $P(\{X \in D\}) = 1$ eine Wahrscheinlichkeitsfunktion.

Definition 6. X heißt *Bernoulli-verteilt* mit Parameter $p \in [0, 1]$, falls

$$P(\{X = 1\}) = p \qquad \text{und} \qquad P(\{X = 0\}) = 1 - p.$$

Bez.: $X \sim \mathbf{B}(1, p)$.

Beispiel 7. Betrachte n gleichartige Produkte, die voneinander unabhängig

- mit Wahrscheinlichkeit p funktionstüchtig
- mit Wahrscheinlichkeit $1-p$ defekt.

Hierbei sei $p \in [0,1]$, z.B. empirisch bestimmt als relative Häufigkeit.
Gesucht: Wahrscheinlichkeit, daß genau k Produkte funktionstüchtig sind. Daraus durch Summation: Wahrscheinlichkeit, daß mindestens k Produkte funktionstüchtig sind.
Konkretes Modell: Produktexperiment mit $\Omega_i := \{0,1\}$ und

$$f_i(\omega_i) := \begin{cases} p, & \text{falls } \omega_i = 1 \\ 1-p, & \text{falls } \omega_i = 0. \end{cases}$$

Also $\Omega := \{0,1\}^n$ Menge der Produktionsergebnisse und für $\omega \in \Omega$

$$f(\omega) := f_1(\omega_1) \cdot \ldots \cdot f_n(\omega_n).$$

Berechne bzgl. des Produktmaßes P

$$P(\{\omega \in \Omega : \sum_{i=1}^{n} \omega_i = k\}).$$

Abstraktes Modell: $X_1, \ldots, X_n$ iid mit $X_1 \sim \mathbf{B}(1,p)$ (es gilt $X_i(\omega) = \omega_i$ im konkreten Modell). Die Anzahl funktionstüchtiger Produkte ist gegeben durch

$$X := \sum_{i=1}^{n} X_i.$$

Berechne

$$P(\{X = k\}).$$

Modellierung analog bei n-fachem Münzwurf oder n Geburten, X Anzahl der geworfenen K bzw. Anzahl der weiblichen Neugeborenen.

Satz 8. Seien $X_1, \ldots, X_n$ iid mit $X_1 \sim \mathbf{B}(1,p)$. Ferner sei

$$X := \sum_{i=1}^{n} X_i.$$

Dann gilt für $k \in \{0, \ldots, n\}$

$$P(\{X = k\}) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}. \tag{1}$$

Beweis. Es gilt

$$P(\{(X_1,\ldots,X_n)\in\{0,1\}^n\}) = P\left(\bigcap_{i=1}^n \{X_i\in\{0,1\}\}\right) = \prod_{i=1}^n P(\{X_i\in\{0,1\}\}) = 1.$$

Setze

$$A_k := \{x\in\{0,1\}^n : \sum_{i=1}^n x_i = k\}.$$

Gemäß Satz 2.6 gilt

$$|A_k| = \binom{n}{k}.$$

Damit folgt für $k\in\{0,\ldots,n\}$

$$\begin{aligned} P(\{X=k\}) &= P\Big(\{X=k\}\cap \bigcup_{x\in\{0,1\}^n}\{(X_1,\ldots,X_n)=x\}\Big) \\ &= \sum_{x\in\{0,1\}^n} P(\{X=k\}\cap\{(X_1,\ldots,X_n)=x\}) \\ &= \sum_{x\in A_k}\prod_{i=1}^n P(\{X_i=x_i\}) \\ &= |A_k|\cdot p^k\cdot(1-p)^{n-k} = \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}. \end{aligned}$$

□

Definition 9. X heißt *binomialverteilt* mit Parametern $n\in\mathbb{N}$ und $p\in[0,1]$, falls (1) für alle $k\in\{0,\ldots,n\}$ gilt. Bez.: $X\sim\mathbf{B}(n,p)$.

Zu $X\sim\mathbf{B}(n,p)$ mit $n=50$ und $p=0.5,\ 0.25,\ 0.05$ stellen wir in den Abbildungen III.1–III.3. die Wahrscheinlichkeitsfunktionen auf $\{0,\ldots,n\}$ graphisch dar.

Beispiel 10. Betrachte n Produkte, unter denen sich n_0 defekte Produkte befinden. Gesucht: Wahrscheinlichkeit, daß bei Auswahl von k Produkten genau ℓ Produkte defekt sind.
Modell: Gleichverteilung P auf

$$\Omega := \{K\subseteq N : |K| = k\}.$$

Berechne $P(A_\ell)$ für

$$A_\ell := \{K\in\Omega : |K\cap N_0| = \ell\},$$

wobei $N_0\subseteq N$ fest gewählt mit $|N_0|=n_0$ sei, d.h. bestimme $|\Omega|$ und $|A_\ell|$.
Es gilt $|\Omega| = \binom{n}{k}$ und für

$$\ell\in\{\max(0,k-(n-n_0)),\ldots,\min(n_0,k)\}$$

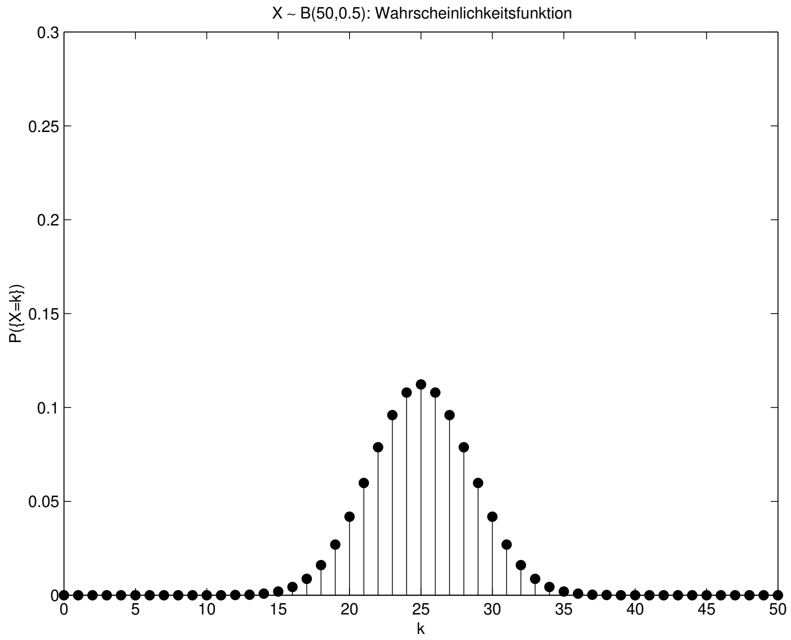

Abbildung III.1: Wahrscheinlichkeitsfunktion zu $X \sim \mathbf{B}(50, 0.5)$

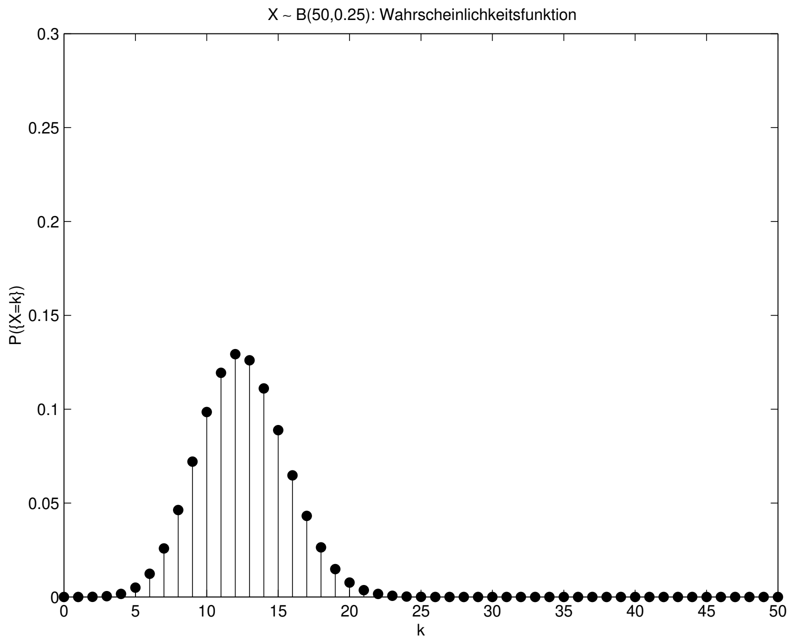

Abbildung III.2: Wahrscheinlichkeitsfunktion zu $X \sim \mathbf{B}(50, 0.25)$

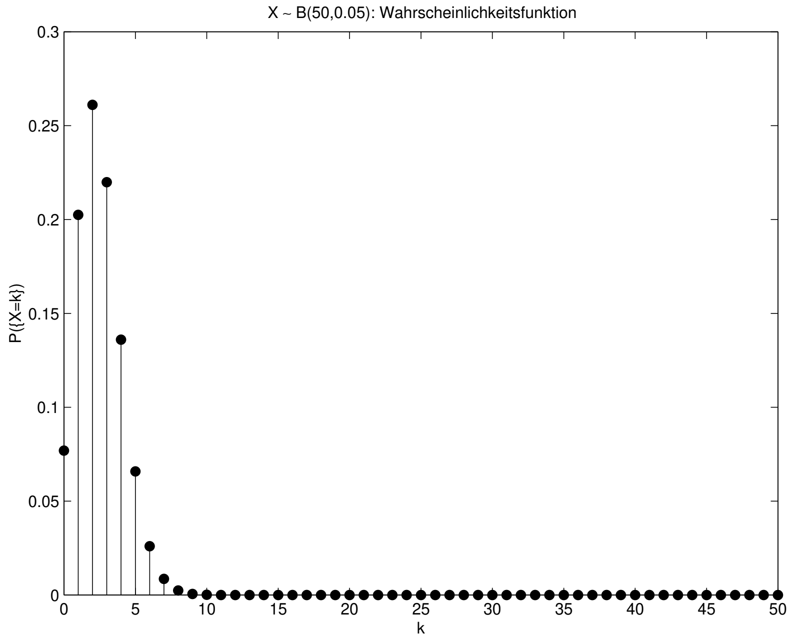

Abbildung III.3: Wahrscheinlichkeitsfunktion zu $X \sim \mathbf{B}(50, 0.05)$

ergibt sich

$$\begin{aligned}|A_\ell| &= |\{(K_0, K_1) : K_0 \subseteq N_0, |K_0| = \ell, K_1 \subseteq N \setminus N_0, |K_1| = k - \ell\}| \\ &= \binom{n_0}{\ell} \cdot \binom{n - n_0}{k - \ell}.\end{aligned}$$

Also

$$P(A_\ell) = \binom{n_0}{\ell} \cdot \binom{n - n_0}{k - \ell} / \binom{n}{k}.$$

Hiermit ergibt sich auch die Wahrscheinlichkeit, beim Skat genau 3 Asse zu erhalten, als

$$\binom{4}{3} \cdot \binom{28}{7} / \binom{32}{10} = 66/899 = 0.0734\ldots$$

Ausblick auf statistische Fragestellungen: Bekannt sei

- die Gesamtanzahl n der Produkte,
- die Stichprobengröße k,
- die Anzahl ℓ defekter Produkte in Stichprobe.

Unbekannt sei

- die Gesamtanzahl n_0 defekter Produkte.

Aufgaben:

(i) Schätze n_0.

(ii) Entscheide, ob $n_0/n \leq 0.02$.

Definition 11. X heißt *hypergeometrisch verteilt* mit Parametern $n \in \mathbb{N}$, $n_0 \in \{0, \ldots, n\}$, $k \in \{1, \ldots, n\}$, falls

$$P(\{X = \ell\}) = \binom{n_0}{\ell} \cdot \binom{n - n_0}{k - \ell} / \binom{n}{k}$$

für

$$\ell \in \{\max(0, k - (n - n_0)), \ldots, \min(n_0, k)\}.$$

Bez.: $X \sim \mathbf{H}(n, n_0, k)$.

Zu $X \sim \mathbf{H}(100, 10, 20)$ stellen wir in Abbildung III.4 die Wahrscheinlichkeitsfunktion auf $\{0, \ldots, 10\}$ graphisch dar. Wir vergleichen die Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{H}(n, n_0, 50)$ mit $n = 100$, 500, 2000 und $n_0 = n/4$ mit der Wahrscheinlichkeitsfunktion zu $Y \sim \mathbf{B}(50, 0.25)$, siehe Abbildungen III.5–III.7. Vermutung: Konvergenz. Bestätigung: Übung 6:G8.

Satz 12 (Poissonscher Grenzwertsatz)**.** Sei $X_n \sim \mathbf{B}(n, p_n)$ mit $p_n \in]0, 1[$, und gelte $\lim_{n \to \infty} n \cdot p_n = \lambda$ für $\lambda > 0$. Dann

$$\forall k \in \mathbb{N}_0 : \lim_{n \to \infty} P(\{X_n = k\}) = \exp(-\lambda) \cdot \frac{\lambda^k}{k!}.$$

Beweis. Für $n \geq k$ gilt

$$\binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} \cdot \underbrace{\left(\frac{n \cdot p_n}{\lambda}\right)^k}_{\to 1} \cdot \underbrace{\frac{(1 - \frac{n \cdot p_n}{n})^n}{(1 - p_n)^k}}_{\to \exp(-\lambda)} \cdot \underbrace{\prod_{i=0}^{k-1} \frac{n - i}{n}}_{\to 1}.$$

□

Definition 13. X heißt *Poisson-verteilt* mit Parameter $\lambda > 0$, falls

$$\forall k \in \mathbb{N}_0 : P(\{X = k\}) = \exp(-\lambda) \cdot \frac{\lambda^k}{k!}.$$

Bez.: $X \sim \mathbf{P}(\lambda)$.

Bemerkung 14. Satz 12 rechtfertigt die Approximation von $\mathbf{B}(n, p_n)$ durch $\mathbf{P}(\lambda)$, falls n „groß“ und p_n „klein“. Siehe Satz VI.1.5 zur Bedeutung von $n \cdot p_n$.

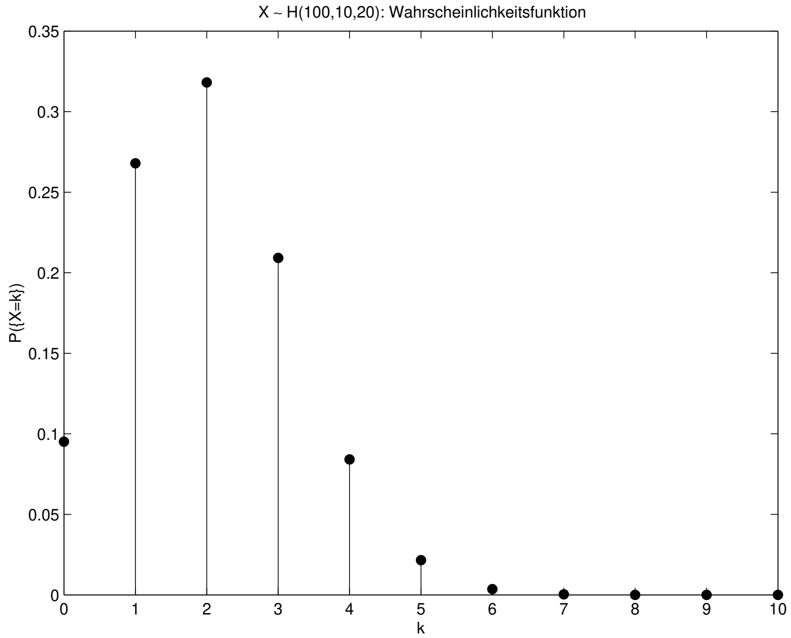

Abbildung III.4: Wahrscheinlichkeitsfunktion zu $X \sim \mathbf{H}(100, 10, 20)$

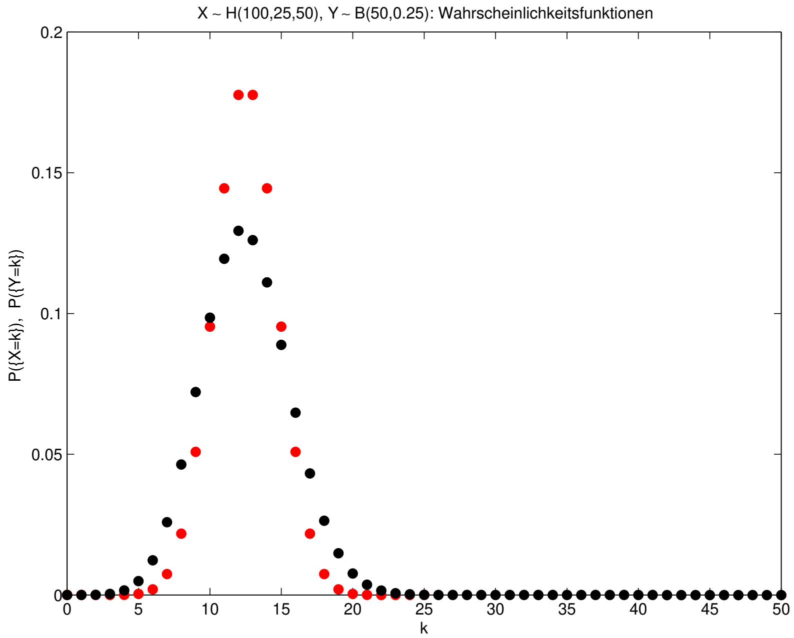

Abbildung III.5: Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{H}(100, 25, 50)$ (rot) und $Y \sim \mathbf{B}(50, 0.25)$ (schwarz)

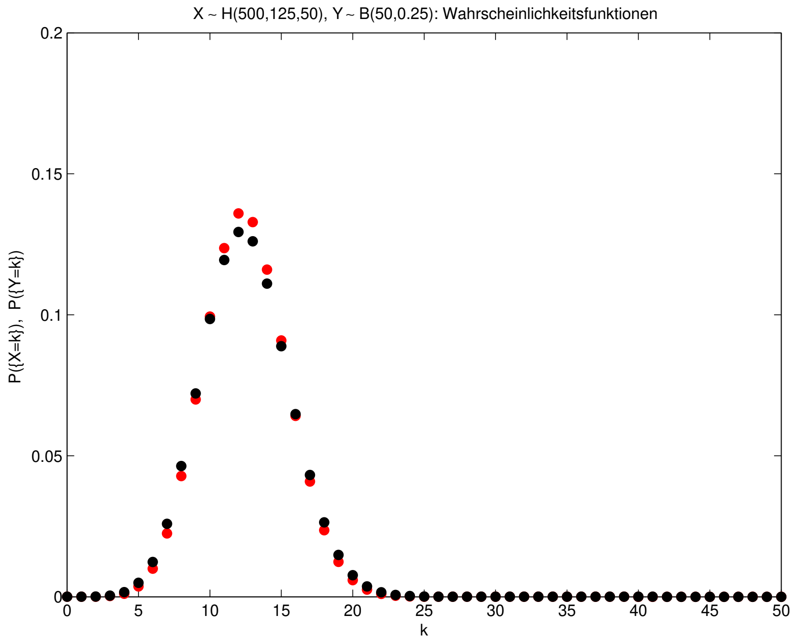

Abbildung III.6: Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{H}(500, 125, 50)$ (rot) und $Y \sim \mathbf{B}(50, 0.25)$ (schwarz)

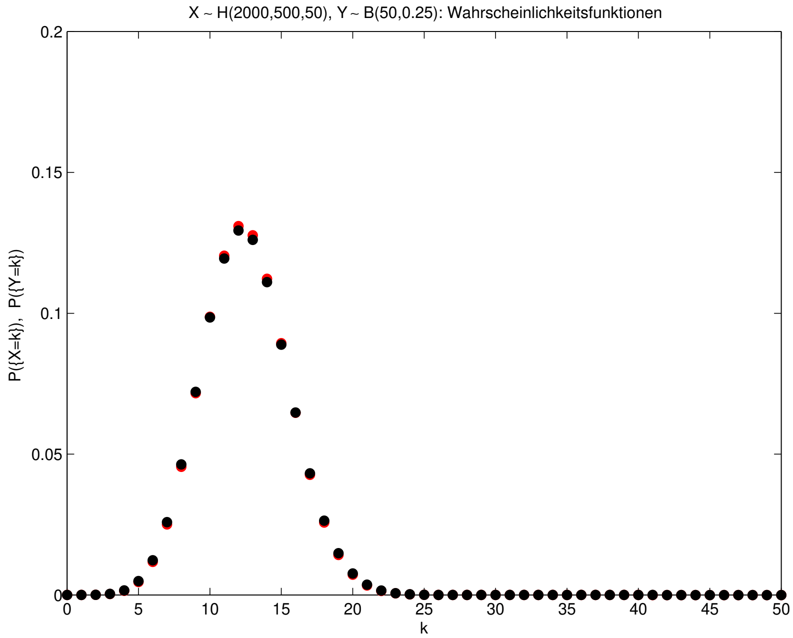

Abbildung III.7: Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{H}(2000, 500, 50)$ (rot) und $Y \sim \mathbf{B}(50, 0.25)$ (schwarz)

Zur Approximationsgüte gilt folgende Aussage. Für $X_n \sim \mathbf{B}(n,p)$, $X \sim \mathbf{P}(n \cdot p)$ und jede Menge $A \subseteq \mathbb{N}_0$ gilt

$$|P(\{X_n \in A\}) - P(\{X \in A\})| \leq 2 \cdot n \cdot p^2,$$

siehe Georgii (2007, Kap. 5.4).
Anwendung: Modellierung der Anzahl von

- Druckfehlern in Manuskript,
- Anrufen in Call-Center pro Tag,
- radioaktiven Zerfällen pro Zeiteinheit.

Wir vergleichen die Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{P}(5)$ mit mit den Wahrscheinlichkeitsfunktion zu $Y \sim \mathbf{B}(n, 5/n)$ für $n = 25, 100, 500$, siehe Abbildungen III.8–III.10.

Beispiel 15. Wir betrachten n unabhängige Würfe auf die Dartscheibe. Gesucht: Wahrscheinlichkeit, daß im k-ten Wurf erstmals das obere rechte Viertel getroffen wird.
Abstraktes Modell: $X_1, \dots, X_n$ iid mit $X_1 \sim \mathbf{B}(1,p)$, wobei $p := 1/4$.
Der Zeitpunkt des ersten Treffers im oberen rechten Viertel ist gegeben durch

$$\tau_n(\omega) := \inf\{i \in \{1, \dots, n\} : X_i(\omega) = 1\}.$$

Für $k \in \{1, \dots, n\}$ gilt

$$\{\tau_n = k\} = \bigcap_{i=1}^{k-1} \{X_i = 0\} \cap \{X_k = 1\},$$

also, unabhängig von n,

$$P(\{\tau_n = k\}) = \prod_{i=1}^{k-1} P(\{X_i = 0\}) \cdot P(\{X_k = 1\}) = (1-p)^{k-1} \cdot p.$$

Ferner gilt $P(\{\tau_n = \infty\}) = (1-p)^n$ und $\lim_{n\to\infty}(1-p)^n = 0$.

Definition 16. X heißt *geometrisch verteilt* mit Parameter $p \in]0,1]$, falls

$$\forall k \in \mathbb{N} : P(\{X = k\}) = p \cdot (1-p)^{k-1}.$$

Bez.: $X \sim \mathbf{G}(p)$.

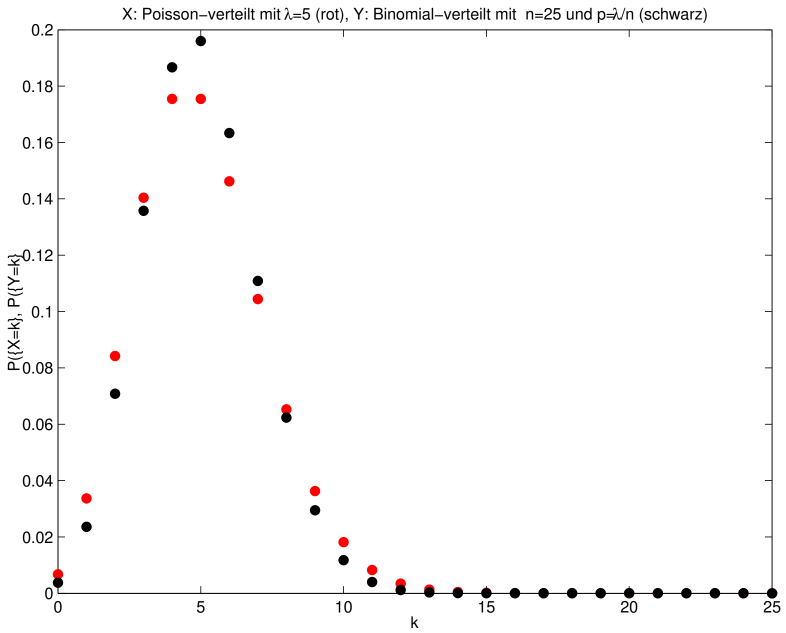

Abbildung III.8: Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{P}(5)$ (rot) und $Y \sim \mathbf{B}(n, 5/n)$ mit $n = 25$ (schwarz)

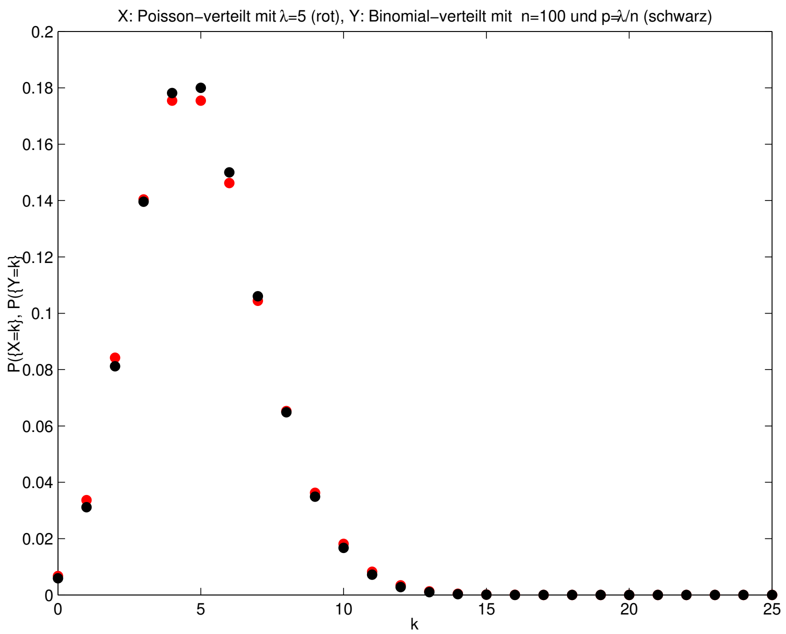

Abbildung III.9: Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{P}(5)$ (rot) und $Y \sim \mathbf{B}(n, 5/n)$ mit $n = 100$ (schwarz)

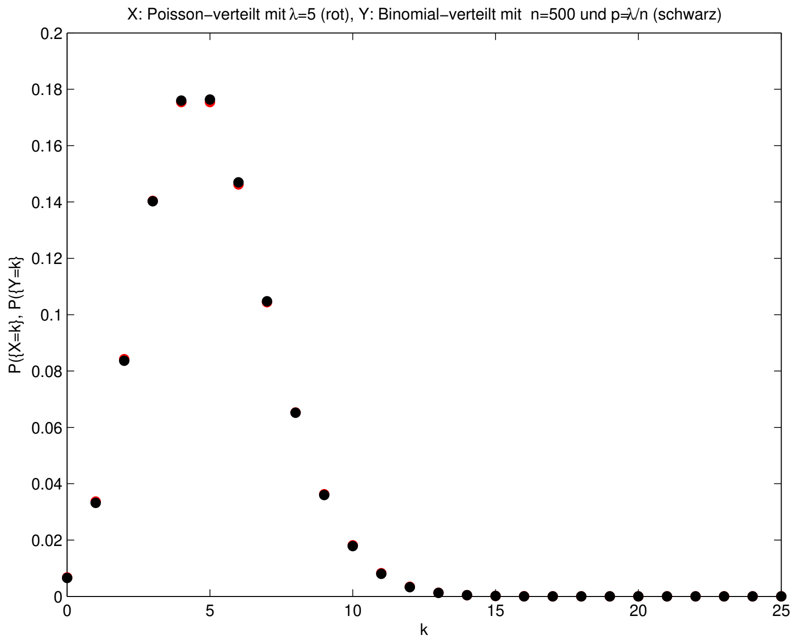

Abbildung III.10: Wahrscheinlichkeitsfunktionen zu $X \sim \mathbf{P}(5)$ (rot) und $Y \sim \mathbf{B}(n, 5/n)$ mit $n = 500$ (schwarz)

Bemerkung 17. Sei $p \in]0, 1]$. Für iid Zufallsvariablen $X_1, X_2, \dots$ mit $X_1 \sim \mathbf{B}(1, p)$ sei

$$\tau_\infty(\omega) := \inf\{i \in \mathbb{N} : X_i(\omega) = 1\}.$$

Die Rechnung aus Beispiel 15 zeigt für $k \in \mathbb{N}$

$$P(\{\tau_\infty = k\}) = p \cdot (1 - p)^{k-1},$$

so daß $\tau_\infty \sim \mathbf{G}(p)$. Beachte, daß $P(\{\tau_\infty = \infty\}) = 0$, da $\sum_{k=1}^\infty P(\{\tau_\infty = k\}) = 1$.
Vgl. Beispiel II.1.3.

Bemerkung 18. Frage: Gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ und darauf eine unendliche Folge $X_1, X_2, \dots$ von iid Zufallsvariablen mit $X_1 \sim \mathbf{B}(1, p)$? Die Antwort ist positiv, siehe Bemerkung IV.2.6, aber der zugrundeliegende Wahrscheinlichkeitsraum kann im nicht-trivialen Fall $0 < p < 1$ nicht diskret sein.
Letzteres ergibt sich wie folgt. Setze

$$S = \{\alpha = (\alpha_i)_{i \in \mathbb{N}} : \alpha_i \in \{0, 1\}\}$$

sowie

$$X(\omega) = (X_1(\omega), \dots), \qquad \omega \in \Omega.$$

Aus

$$P(\{X = \alpha\}) \le P\left(\bigcap_{i=1}^n \{X_i = \alpha_i\}\right) = \prod_{i=1}^n P(\{X_i = \alpha_i\}) \le \max(p, 1 - p)^n$$

folgt

$$P(\{X = \alpha\}) = 0.$$

Die Mengen $\{X = \alpha\}$ mit $\alpha \in S$ sind p.d. Ist Ω abzählbar, so gilt

$$\Omega = \bigcup_{\alpha \in S_0} \{X = \alpha\}$$

mit einer abzählbaren Menge $S_0 \subset S$. Damit ergibt sich der Widerspruch

$$1 = P(\Omega) = \sum_{\alpha \in S_0} P(\{X = \alpha\}) = 0.$$

5 Die symmetrische Bernoulli-Irrfahrt

Definition 1. X heißt *symmetrisch Bernoulli-verteilt*, falls

$$P(\{X = 1\}) = P(\{X = -1\}) = 1/2.$$

Bez.: $X \sim \mathbf{SB}$.

Im folgenden seien $Y_1, \dots, Y_T$ iid mit $Y_1 \sim \mathbf{SB}$. Der Einfachheit halber gelte $Y_j(\Omega) = \{-1, 1\}$ für alle $j \in \{1, \dots, T\}$. Wir setzen $S_0 := 0$ und

$$S_t := \sum_{j=1}^{t} Y_j = S_{t-1} + Y_t$$

für $t \in \{1, \dots, T\}$.

Definition 2. $(S_t)_{t \in \{0,\dots,T\}}$ heißt *symmetrische Bernoulli-Irrfahrt* (mit Zeithorizont T).

Im folgenden $(S_t)_{t \in \{0,\dots,T\}}$ wie oben.
$(S_t)_{t \in \{0,\dots,T\}}$ ist das einfachste Beispiel eines stochastischen Prozesses zur Modellierung zeitabhängiger zufälliger Phänomene. Anwendungen z. Bsp.

- Physik, S_t eine Koordinate der Position eines Teilchen nach t Kollisionen,
- faires Spiel zweier Spieler I und II, S_t Stand aus Sicht von I nach t Runden, siehe Beispiel II.1.3.

Ausblick: Irrfahrten auf Gruppen. Finanzmathematik: $]0, \infty[$ mit Multiplikation, Kartenmischen: symmetrische Gruppe mit Komposition.
In den Abbildungen III.11–III.13 zeigen wir je eine Simulation der Irrfahrt für $T = 50, 100, 1000$. Siehe Beispiel II.3.7 zur Simulation der Zufallsvariablen Y_t.
Nun: Gedächtnislosigkeit der Irrfahrt (Markov-Eigenschaft), siehe auch Georgii (2007, Kap. 6) und Krengel (2000, Kap. 15).

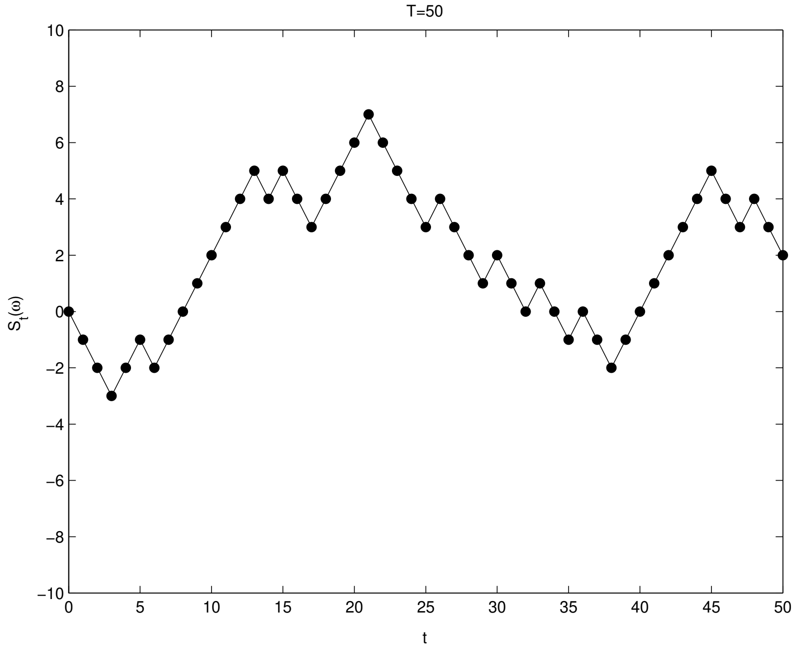

Abbildung III.11: $T = 50$, end=2, max=7, fueh=36

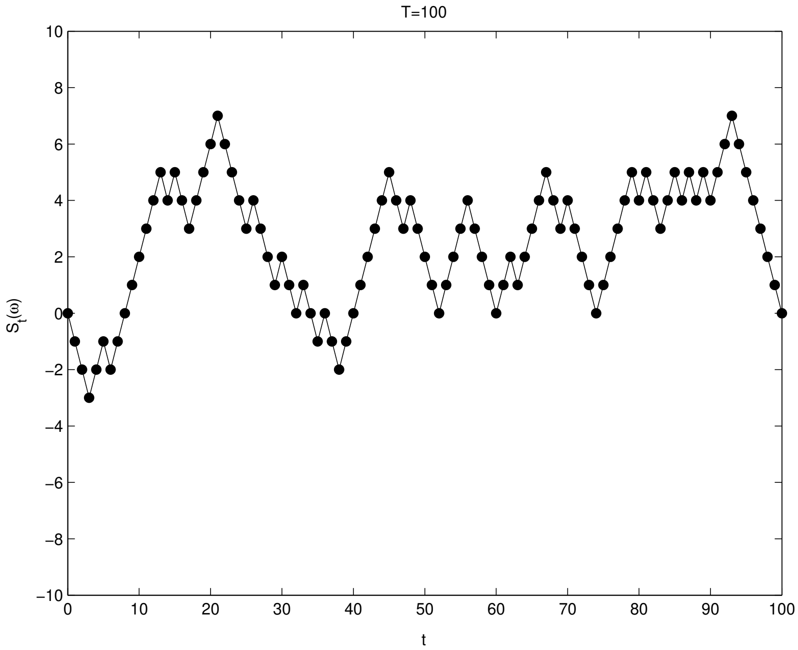

Abbildung III.12: $T = 100$, end=0, max=7, fueh=86

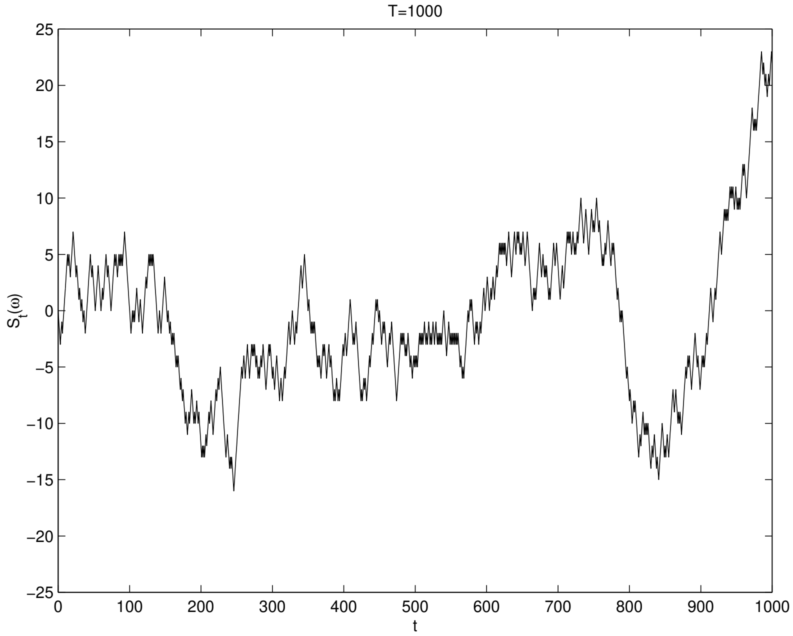

Abbildung III.13: $T = 1000$, end=22, max=23, fueh=420

Lemma 3. Für $t \in \{1, \ldots, T-1\}$, $A \subseteq \{-1, 1\}^t$ und $B \subseteq \{-1, 1\}^{T-t}$ gilt

$$\begin{aligned}
&P(\{(Y_1, \ldots, Y_t) \in A\} \cap \{(Y_{t+1}, \ldots, Y_T) \in B\}) \\
&\quad = P(\{(Y_1, \ldots, Y_t) \in A\}) \cdot P(\{(Y_{t+1}, \ldots, Y_T) \in B\}) \\
&\quad = P(\{(Y_1, \ldots, Y_t) \in A\}) \cdot P(\{(Y_1, \ldots, Y_{T-t}) \in B\}).
\end{aligned}$$

Beweis. Siehe Übung 3:H9 für eine allgemeinere Aussage im Fall von Zufallsvariablen mit Werten in $\mathbb{Z}$. Siehe auch Irle (2001, p. 163, 169). □

Satz 4. Für $t \in \{1, \ldots, T-1\}$ und $s \in \mathbb{Z}^t$ mit $P(\{(S_1, \ldots, S_t) = s\}) > 0$ gilt für alle $B \subseteq \mathbb{Z}^{T-t}$

$$\begin{aligned}
P(\{(S_{t+1}, \ldots S_T) \in B\} \mid \{(S_1, \ldots, S_t) = s\}) &= P(\{(S_{t+1}, \ldots S_T) \in B\} \mid \{S_t = s_t\}) \\
&= P(\{(S_1, \ldots S_{T-t}) \in B - (s_t, \ldots, s_t)\}).
\end{aligned}$$

Beweis. Sei t wie oben und sei $s \in \mathbb{Z}^t$. Betrachte den Spezialfall $B = \{z\}$ mit

$$z = (z_{t+1}, \ldots, z_T) \in \mathbb{Z}^{T-t}.$$

Setze $z_t := s_t$ und $s_0 := 0$. Dann gilt

$$\begin{aligned}
&P(\{(S_{t+1}, \dots S_T) = z\} \cap \{(S_1, \dots, S_t) = s\}) \\
&\quad = P\left(\bigcap_{j=1}^{t} \{Y_j = s_j - s_{j-1}\} \cap \bigcap_{j=t+1}^{T} \{Y_j = z_j - z_{j-1}\}\right) \\
&\quad = P(\{(S_1, \dots, S_t) = s\}) \cdot P\left(\bigcap_{j=1}^{T-t} \{Y_j = z_{j+t} - z_{j+t-1}\}\right) \\
&\quad = P(\{(S_1, \dots, S_t) = s\}) \cdot \underbrace{P(\{(S_1, \dots S_{T-t}) = (z_{t+1} - s_t, \dots, z_T - s_t)\})}_{=:q}.
\end{aligned}$$

Im Fall $t > 1$ erhalten wir hiermit

$$\begin{aligned}
&P(\{(S_{t+1}, \dots S_T) = z\} \cap \{S_t = s_t\}) \\
&\quad = \sum_{w \in \mathbb{Z}^{t-1}} P(\{(S_{t+1}, \dots S_T) = z\} \cap \{(S_1, \dots, S_t) = (w, s_t)\}) \\
&\quad = \sum_{w \in \mathbb{Z}^{t-1}} P(\{(S_1, \dots, S_t) = (w, s_t)\}) \cdot q \\
&\quad = P(\{S_t = s_t\}) \cdot q.
\end{aligned}$$

Hieraus folgt die Behauptung für $B = \{z\}$.

Zum Beweis des allgemeinen Falles schreibe man B als Vereinigung einelementiger Mengen. $\square$

Beispiel 5. Für t und s wie oben und $k \in \mathbb{Z}$ gilt

$$P(\{S_T = k\} \mid \{(S_1, \dots, S_t) = s\}) = P(\{S_T = k\} \mid \{S_t = s_t\}) = P(\{S_{T-t} = k - s_t\}).$$

Nun: eine zweite Sicht auf die Irrfahrt. Setze dazu $s_0 := 0$ und betrachte die Menge

$$D := \{(s_0, \dots, s_T) \in \mathbb{Z}^{T+1} : \forall 1 \leq t \leq T : |s_t - s_{t-1}| = 1\}$$

der sogenannten Pfade der Irrfahrt $(S_t)_{t \in \{0, \dots, T\}}$.

Satz 6. Es gilt $|D| = 2^T$ und

$$P(\{(S_0, \dots, S_T) = s\}) = 1/|D|$$

für alle $s \in D$.

Beweis. Der Übergang von Partialsummen zu Inkrementen zeigt

$$|D| = |\{-1, 1\}^T| = 2^T,$$

und

$$\{(S_0, \dots, S_T) = s\} = \bigcap_{t=1}^{T} \{Y_t = s_t - s_{t-1}\}.$$

Für $s \in D$ folgt $P(\{(S_0, \dots, S_T) = s\}) = 2^{-T}$. $\square$

Fazit: eine Irrfahrt entspricht der Gleichverteilung auf der Menge D ihrer Pfade. Dies ermöglicht den Einsatz kombinatorischer Methoden zum Studium der Irrfahrt.

Wir bestimmen die Verteilungsfunktionen folgender Zufallsvariablen:

(i) Spielstand nach der letzten Runde, $X := S_T$,

(ii) maximaler Spielstand, $X := \max_{t=0,\dots,T} S_t$,

(iii) Anzahl der Runden, nach denen Spieler I führt,

$$X := |\{t \in \{1,\dots,T\} : S_t \geq 0 \text{ und } S_{t-1} \geq 0\}|.$$

Beachte

$$\{S_t \geq 0\} \cap \{S_{t-1} \geq 0\} = \{S_t > 0\} \cup (\{S_t = 0\} \cap \{S_{t-1} > 0\}).$$

Definiere Abbildung $f : D \to \mathbb{Z}$

(i) $f(s_0,\dots,s_T) = s_T$,

(ii) $f(s_0,\dots,s_T) = \max_{t=0,\dots,T} s_t$,

(iii) $f(s_0,\dots,s_T) = |\{t \in \{1,\dots,T\} : s_t \geq 0 \text{ und } s_{t-1} \geq 0\}|$.

Somit

$$X = f \circ (S_0,\dots,S_T).$$

Gesucht ist

$$P(\{X = k\})$$

für $k \in \mathbb{Z}$. Siehe auch Širjaev (1988, Kap. 1.10).

Bemerkung 7. Die Simulation von n unabhängigen Wiederholungen der Irrfahrt geschieht folgendermaßen. Betrachte eine iid-Folge $Y'_1,\dots,Y'_{n\cdot T}$ mit $Y'_1 \sim \mathbf{SB}$. Definiere für $t = 0,\dots,T$ und $i = 1,\dots,n$

$$S'_{i,t} := \sum_{j=1}^{t} Y'_{(i-1)\cdot T+j}.$$

Also

$$\begin{array}{ccc}
\underbrace{Y'_1,\dots,Y'_T}_{\downarrow}, & \dots & ,\underbrace{Y'_{(n-1)\cdot T+1}, \quad \dots \quad ,Y'_{n\cdot T}}_{\downarrow} \\
\underbrace{S'_{1,0},\dots,S'_{1,T}}_{f\downarrow}, & \dots & ,\underbrace{S'_{n,0},\dots,S'_{n,T}}_{f\downarrow} \\
X'_1, & \dots & ,X'_n
\end{array}$$

Dann sind $X'_1, \ldots, X'_n$ iid und X'_1 und X sind identisch verteilt, siehe Lemma 3. Erzeuge eine Realisierung $y'_1, \ldots, y'_{n \cdot T}$ von $Y'_1, \ldots, Y'_{n \cdot T}$, setze

$$s'_{i,t} := \sum_{j=1}^{t} y'_{(i-1) \cdot T + j},$$

und approximiere die gesuchten Wahrscheinlichkeiten $P(\{X = k\})$ durch die relativen Häufigkeiten

$$\frac{1}{n} \cdot \sum_{i=1}^{n} 1_{\{k\}}(f(s'_{i,0}, \ldots, s'_{i,T}))$$

der Simulationen, bei denen f den Wert k liefert.

Im folgenden zeigen wir Simulationsergebnisse für

$$T = 50.$$

Beachte: Für die Menge D der Pfade gilt in diesem Fall

$$|D| = 2^{50} = 1.1 \ldots \cdot 10^{15}.$$

Dargestellt wird

$$k \mapsto \frac{1}{n} \cdot \sum_{i=1}^{n} 1_{\{k\}}(f(s'_{i,0}, \ldots, s'_{i,T})).$$

Die Abbildungen III.14 und III.15 zeigen Approximationen für $P(\{S_T = k\})$ auf der Basis von $n = 10^3$ und $n = 10^5$ Wiederholungen.

Satz 8. Für $t \in \{1, \ldots, T\}$ und $k \in \{-t, -t+2, \ldots, t-2, t\}$ gilt

$$P(\{S_t = k\}) = \binom{t}{(k+t)/2} \cdot 2^{-t}.$$

Beweis. Mit $U_j := (Y_j + 1)/2$ und $R := \sum_{j=1}^{t} U_j$ gilt

$$S_t = \sum_{j=1}^{t} Y_j = \sum_{j=1}^{t} (2 \cdot U_j - 1) = 2 \cdot R - t.$$

Da $U_1, \ldots, U_t$ unabhängig und $U_1 \sim \mathbf{B}(1, p)$, folgt $R \sim \mathbf{B}(t, p)$. Somit

$$P(\{S_t = k\}) = P(\{R = (k+t)/2\}).$$

$\square$

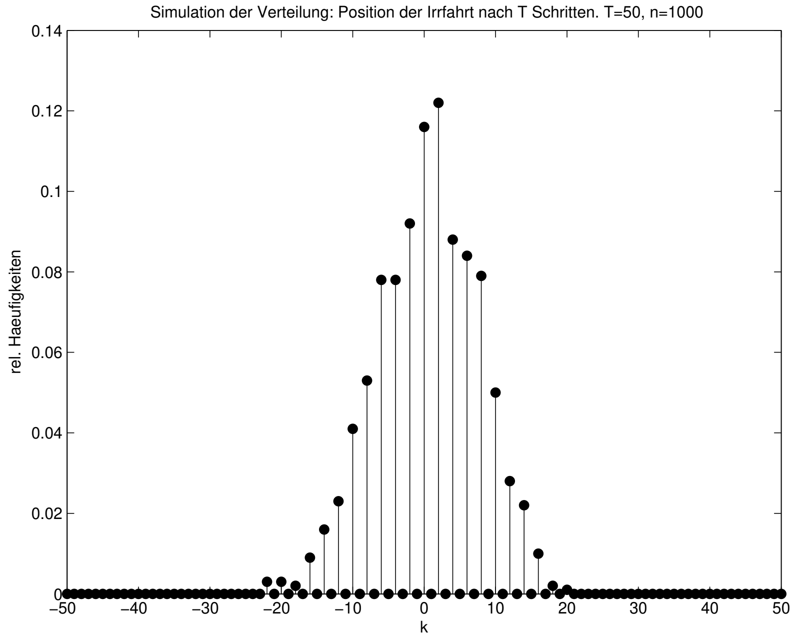

Abbildung III.14: Approximation für $P(\{S_T = k\})$, $n = 10^3$ Wdh.

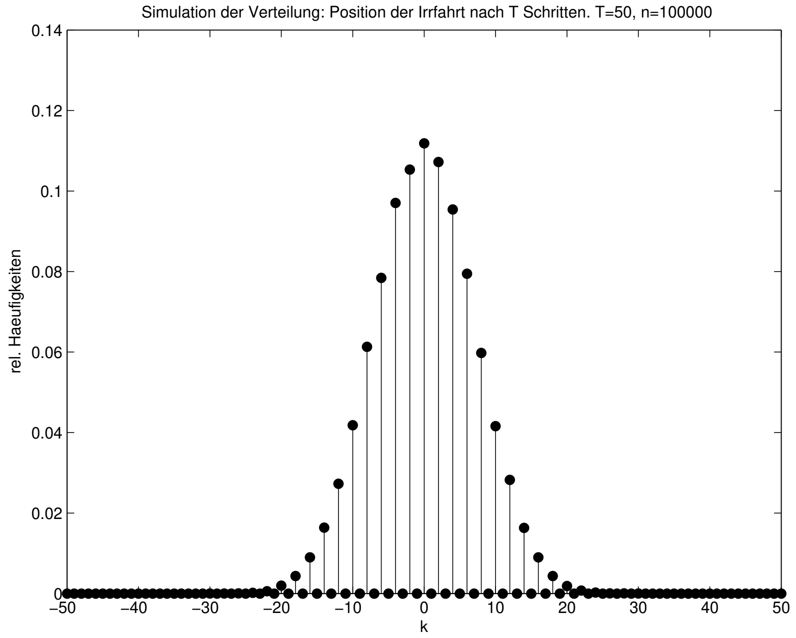

Abbildung III.15: Approximation für $P(\{S_T = k\})$, $n = 10^5$ Wdh.

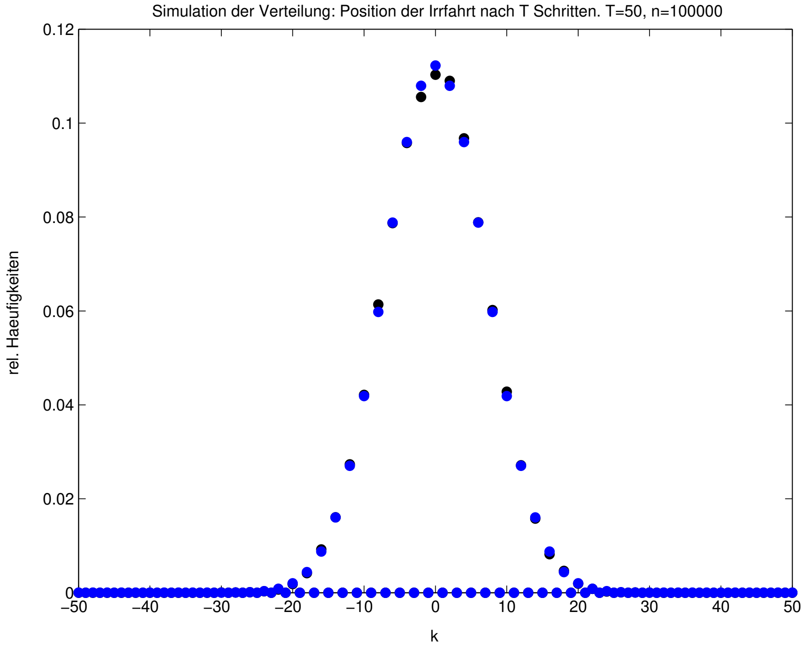

Abbildung III.16: $P(\{S_T = k\})$, exakte Werte und Approximation, $n = 10^5$ Wdh.

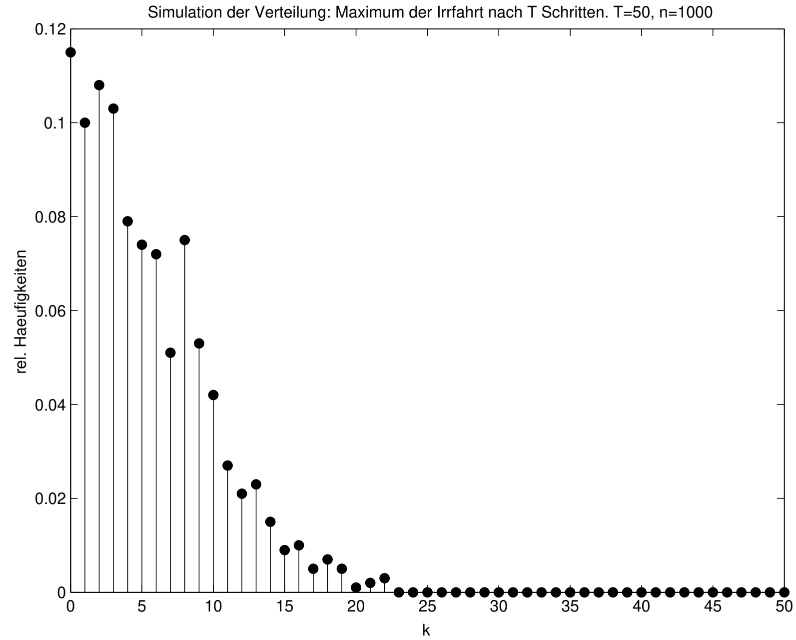

Abbildung III.17: Approximation für $P(\{\max_{t=0,\dots,T} S_t = k\})$, $n = 10^3$ Wdh.

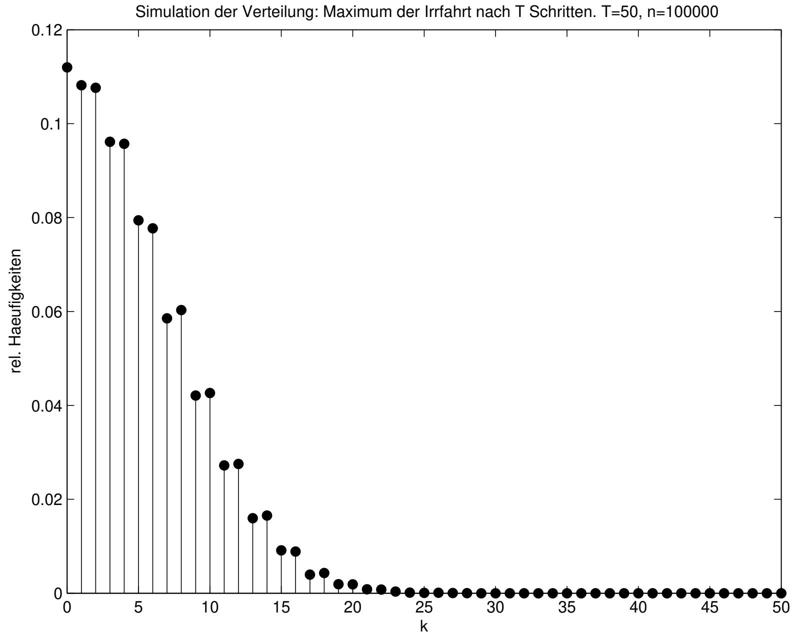

Abbildung III.18: Approximation für $P(\{\max_{t=0,\ldots,T} S_t = k\})$, $n = 10^5$ Wdh.

Abbildung III.16 zeigt für $P(\{S_T = k\})$ einen Vergleich der exakten Werten (blau) mit Simulationsergebnissen (schwarz) auf der Basis von $n = 10^5$ Wiederholungen. Die Abbildungen III.17 und III.18 zeigen Approximationen für $P(\{\max_{t=0,\ldots,T} S_t = k\})$ auf der Basis von $n = 10^3$ und $n = 10^5$ Wiederholungen.

Satz 9. Für $t \in \{1, \ldots, T\}$ und $k \in \{0, \ldots, t\}$ gilt

$$P(\{\max_{j=0,\ldots,t} S_j = k\}) = P(\{S_t = k\}) + P(\{S_t = k + 1\}).$$

Beweis. Verwende die Bezeichnung aus Satz 6. Fixiere $k \in \{0, \ldots, t\}$ und definiere

$$\tau(s) := \inf\{j \in \{0, \ldots, t\} : s_j = k\}.$$

Betrachte die Spiegelung der Pfade $s \in D$ an $(k, \ldots, k)$, definiere also $\mathcal{R} : D \to D$ durch

$$(\mathcal{R}(s))_j := \begin{cases} s_j, & \text{falls } j \leq \tau(s), \\ 2 \cdot k - s_j, & \text{falls } j > \tau(s). \end{cases}$$

Für jedes $\ell \in \mathbb{Z}$ mit $\ell \leq k$ definiert $\mathcal{R}$ eine Bijektion zwischen

$$\{s \in D : \max_{j=0,\ldots,t} s_j \geq k \wedge s_t = \ell\}$$

und

$$\{s \in D : s_t = 2 \cdot k - \ell\}.$$

Dieses sogenannte Spiegelungsprinzip liefert zusammen mit Satz 6

$$P(\{\max_{j=0,\ldots,t} S_j \geq k\} \cap \{S_t = \ell\}) = P(\{S_t = 2 \cdot k - \ell\}).$$

Es folgt

$$\begin{aligned}
&P(\{\max_{j=0,\ldots,t} S_j = k\}) \\
&\quad = \sum_{\ell=-t}^{k} P(\{\max_{j=0,\ldots,t} S_j = k\} \cap \{S_t = \ell\}) \\
&\quad = \sum_{\ell=-t}^{k} \Big(P(\{\max_{j=0,\ldots,t} S_j \geq k\} \cap \{S_t = \ell\}) - P(\{\max_{j=0,\ldots,t} S_j \geq k+1\} \cap \{S_t = \ell\}) \Big) \\
&\quad = \sum_{\ell=-t}^{k} \Big(P(\{S_t = 2 \cdot k - \ell\}) - P(\{S_t = 2 \cdot k + 2 - \ell\}) \Big) \\
&\quad = \sum_{\ell=k}^{t} P(\{S_t = \ell\}) - \sum_{\ell=k+2}^{t} P(\{S_t = \ell\}) \\
&\quad = P(\{S_t = k\}) + P(\{S_t = k+1\}).
\end{aligned}$$

$\square$

Abbildung III.19 zeigt für $P(\{\max_{t=0,\ldots,T} S_t = k\})$ einen Vergleich der exakten Werten (blau) mit Simulationsergebnissen (schwarz) auf der Basis von $n = 10^5$ Wiederholungen.

Bemerkung 10. Satz 9 zeigt die überraschende Beziehung

$$P(\{\max_{j=0,\ldots,t} S_j \leq k\}) = P(\{|S_t| \leq k\}),$$

falls $k = t \mod 2$. Die Sätze 8 und 9 liefern die explizite Formel

$$P(\{\max_{j=0,\ldots,t} S_j = k\}) = \begin{cases} \binom{t}{(k+t)/2} \cdot 2^{-t}, & \text{falls } k = t \mod 2, \\ \binom{t}{(k+1+t)/2} \cdot 2^{-t}, & \text{falls } k \neq t \mod 2 \end{cases}$$

für $k \in \{0, \ldots, t\}$.

Im folgenden sei T gerade.
Wir betrachten

$$X_T := \big|\{t \in \{1, \ldots, T\} : S_t \geq 0 \text{ und } S_{t-1} \geq 0\}\big|.$$

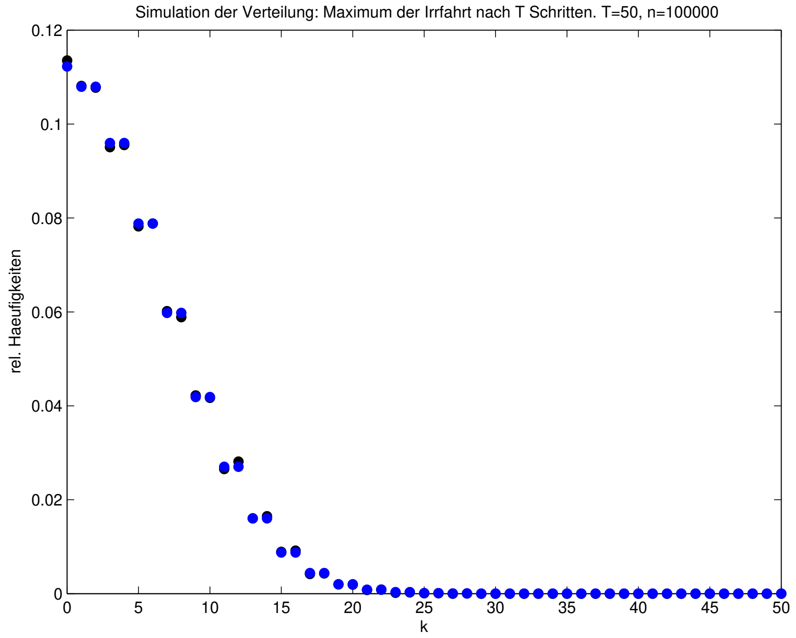

Abbildung III.19: $P(\{\max_{t=0,\dots,T} S_t = k\})$, exakte Werte und Approx., $n = 10^5$ Wdh.

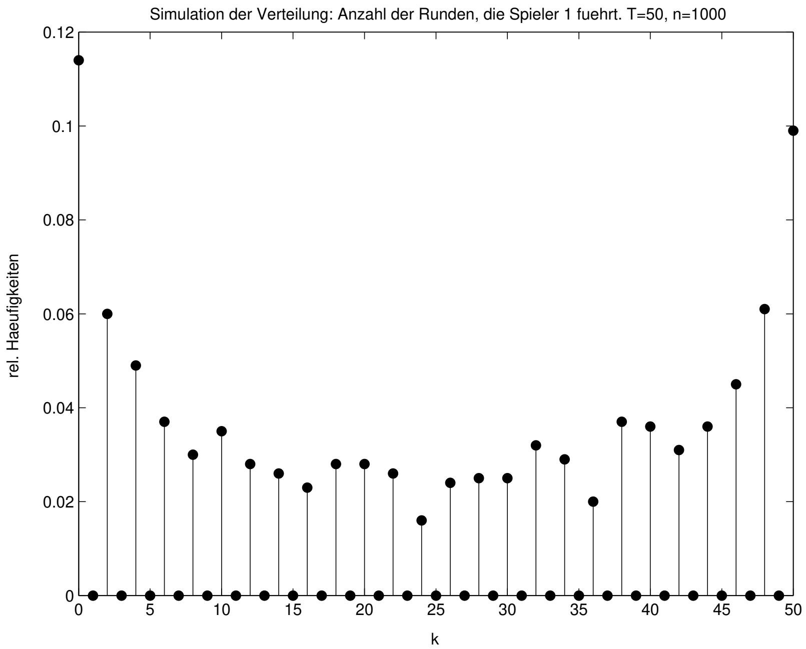

Abbildung III.20: Approximation für $P(\{X_T = k\})$, $n = 10^3$ Wdh.

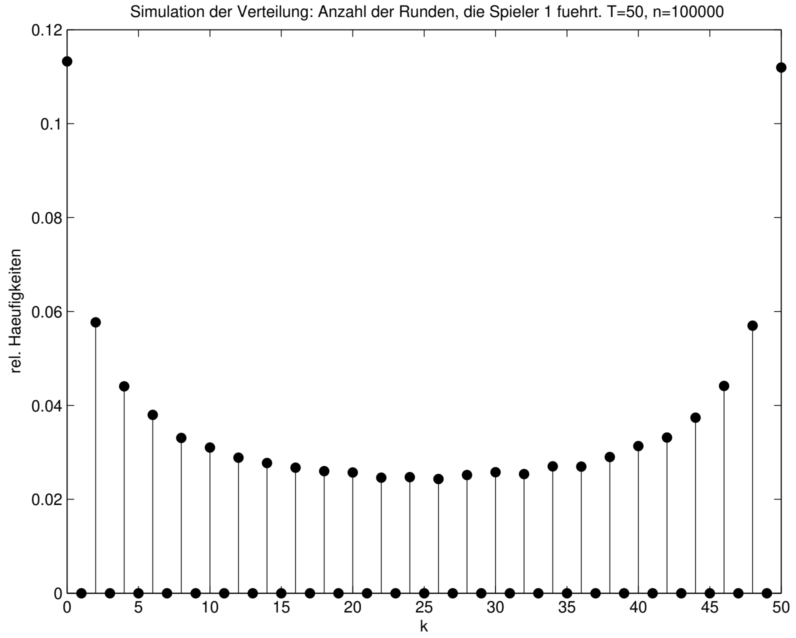

Abbildung III.21: Approximation für $P(\{X_T = k\})$, $n = 10^5$ Wdh.

Bemerkung 11. $P(\{X_T = x\}) = 0$, falls x ungerade.

Die Abbildungen III.20 und III.21 zeigen Approximationen für $P(\{X_T = k\})$ auf der Basis von $n = 10^3$ und $n = 10^5$ Wiederholungen. Wie sehen ein völlig überraschendes Monotonieverhalten.

Wir setzen

$$\tau_T := \inf\{t \in \{1, \ldots, T\} : S_t = 0\} = \inf\{t \in \{2, 4, \ldots, T\} : S_t = 0\}.$$

Falls die Irrfahrt bis zur Zeit T nicht zum Startwert zurückkehrt, liefert die Zufallsvariable τ_T den Wert ∞. Andernfalls gibt τ_T den ersten Zeitpunkt der Rückkehr an. Vgl. Beispiel 4.15.

Lemma 12. Für $k \in \{1, \ldots, T/2\}$ gilt

$$\sum_{\ell=1}^{k} P(\{\tau_T = 2\ell\}) \cdot P(\{S_{2k-2\ell} = 0\}) = P(\{S_{2k} = 0\}).$$

Beweis. Es gilt

$$\{S_{2k} = 0\} = \bigcup_{\ell=1}^{k} \{S_{2k} = 0\} \cap \{\tau_T = 2\ell\}$$

und

$$\{S_{2k} = 0\} \cap \{\tau_T = 2\ell\} = \{\tau_T = 2\ell\} \cap \left\{ \sum_{j=2\ell+1}^{2k} Y_j = 0 \right\}.$$

Mit Lemma 3 folgt

$$
\begin{aligned}
P(\{S_{2k}=0\}) &= \sum_{\ell=1}^{k} P(\{\tau_T = 2\ell\}) \cdot P\left(\left\{\sum_{j=1}^{2k-2\ell} Y_j = 0\right\}\right) \\
&= \sum_{\ell=1}^{k} P(\{\tau_T = 2\ell\}) \cdot P(\{S_{2k-2\ell}=0\}).
\end{aligned}
$$

□

Satz 13. Für $k \in \{0,\dots,T/2\}$ gilt

$$P(\{X_T = 2k\}) = P(\{S_{2k}=0\}) \cdot P(\{S_{T-2k}=0\}).$$

Beweis. Setze

$$h(y_{2\ell+1},\dots,y_{2m}) := \left|\left\{t \in \{2\ell+1,\dots 2m\} : \sum_{j=2\ell+1}^{t} y_j \geq 0 \text{ und } \sum_{j=2\ell+1}^{t-1} y_j \geq 0\right\}\right|$$

für $\ell, m \in \{0,\dots,T/2\}$ mit $\ell < m$. Per Induktion nach $m \in \{1,\dots,T/2\}$ zeigen wir, daß

$$P(\{h(Y_1,\dots,Y_{2m}) = 2k\}) = P(\{S_{2k}=0\}) \cdot P(\{S_{2m-2k}=0\})$$

für alle $k \in \{0,\dots,m\}$ gilt.

Setze

$$A_{k,m} := \{h(Y_1,\dots,Y_{2m}) = 2k\}$$

für $k \in \mathbb{Z}$. Klar: $A_{k,m} = \emptyset$, falls $k<0$ oder $k>m$. Für $k=0$ und $k=m$ erhalten wir aufgrund von Satz 9

$$P(A_{0,m}) = P(A_{m,m}) = P(\{\max_{t=0,\dots,2m} S_t = 0\}) = P(\{S_{2m}=0\}) \cdot P(\{S_0=0\}),$$

was insbesondere die Induktionsverankerung beinhaltet.

Gelte $m>1$ und sei $k \in \{1,\dots,m-1\}$. Dann

$$A_{k,m} = \bigcup_{\ell=1}^{m-1} \underbrace{A_{k,m} \cap \{\tau_T = 2\ell\}}_{=:B_\ell}.$$

Setze $A_- := \{\tau_T = 2\ell\} \cap \{S_1 = -1\}$ und $A_+ := \{\tau_T = 2\ell\} \cap \{S_1 = 1\}$. Dann

$$
\begin{aligned}
B_\ell &= A_{k,m} \cap A_- \cup A_{k,m} \cap A_+ \\
&= \{h(Y_{2\ell+1},\dots,Y_{2m}) = 2k\} \cap A_- \cup \{h(Y_{2\ell+1},\dots,Y_{2m}) = 2k-2\ell\} \cap A_+.
\end{aligned}
$$

Es gilt $P(A_-) = P(A_+) = 1/2 \cdot P(\{\tau_T = 2\ell\})$. Lemma 3 sichert

$$P(B_\ell) = 1/2 \cdot P(\{\tau_T = 2\ell\}) \cdot (P(A_{k,m-\ell}) + P(A_{k-\ell,m-\ell})).$$

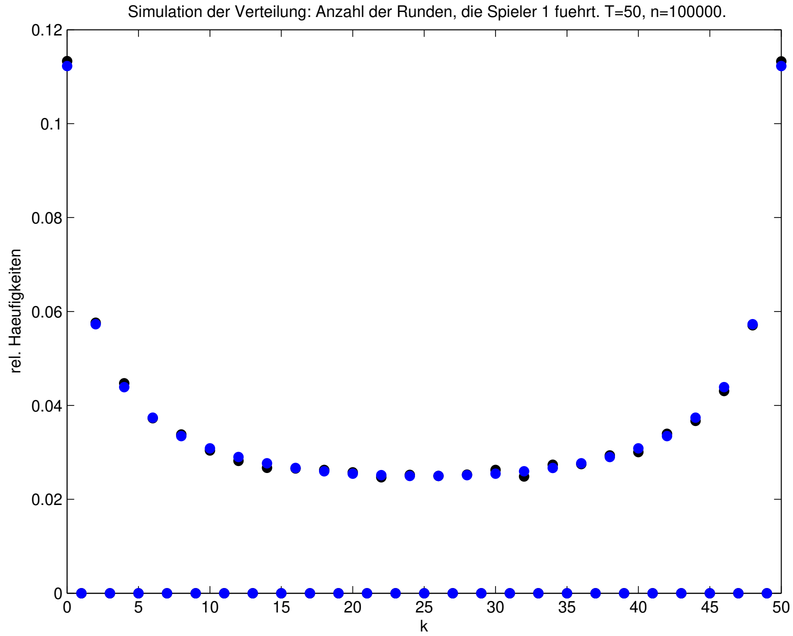

Abbildung III.22: $P(\{X_T = k\})$, exakte Werte und Approximation, $n = 10^5$ Wdh.

Somit

$$\begin{aligned} P(A_{k,m}) &= \frac{1}{2} \cdot \sum_{\ell=1}^{m-1} P(\{\tau_T = 2\ell\}) \cdot (P(A_{k,m-\ell}) + P(A_{k-\ell,m-\ell})) \\ &= \frac{1}{2} \cdot \sum_{\ell=1}^{m-k} P(\{\tau_T = 2\ell\}) \cdot P(A_{k,m-\ell}) + \frac{1}{2} \cdot \sum_{\ell=1}^{k} P(\{\tau_T = 2\ell\}) \cdot P(A_{k-\ell,m-\ell}). \end{aligned}$$

Wende die Induktionsannahme und Lemma 12 an, um

$$\begin{aligned} P(A_{k,m}) = \frac{1}{2} \cdot \sum_{\ell=1}^{m-k} & P(\{\tau_T = 2\ell\}) \cdot P(\{S_{2k} = 0\}) \cdot P(\{S_{2m-2k-2\ell} = 0\}) \\ & + \frac{1}{2} \cdot \sum_{\ell=1}^{k} P(\{\tau_T = 2\ell\}) \cdot P(\{S_{2k-2\ell} = 0\}) \cdot P(\{S_{2m-2k} = 0\}) \\ = P(\{S_{2k} = 0\}) & \cdot P(\{S_{2m-2k} = 0\}) \end{aligned}$$

zu erhalten. □

Abbildung III.22 zeigt für $P(\{X_T = k\})$ einen Vergleich der exakten Werten (blau) mit Simulationsergebnissen (schwarz) auf der Basis von $n = 10^5$ Wiederholungen.

Jetzt: asymptotische Betrachtungen.

Bezeichnung. $a_n \approx b_n$ (asymptotische Äquivalenz) für Folgen von Zahlen $a_n, b_n > 0$, falls

$$\lim_{n\to\infty} a_n/b_n = 1.$$

Lemma 14 (Stirlingsche Formel)**.** Es gilt

$$n! \approx \sqrt{2\pi\, n} \cdot (n/e)^n.$$

Korollar 15. Es gilt

$$P(\{S_{2n} = 0\}) \approx \frac{1}{\sqrt{\pi\, n}}.$$

Beweis. Satz 8 und Lemma 14 zeigen für $T = 2n$

$$\begin{aligned} P(\{S_{2n} = 0\}) &= 2^{-2n} \cdot \binom{2n}{n} \\ &\approx 2^{-2n} \cdot \frac{\sqrt{2\pi\, 2n} \cdot (2n)^{2n} \cdot \exp(-2n)}{\left(\sqrt{2\pi\, n} \cdot n^n \cdot \exp(-n)\right)^2} \\ &= \sqrt{1/(\pi\, n)}. \end{aligned}$$

□

Korollar 16. Es gilt

$$\frac{P(\{X_{4n} = 4n\})}{P(\{X_{4n} = 2n\})} \approx \sqrt{\pi\, n/2}.$$

Beweis. Satz 13 und Korollar 15 zeigen für $T = 4n$

$$\frac{P(\{X_{4n} = 4n\})}{P(\{X_{4n} = 2n\})} = \frac{P(\{S_{4n} = 0\})}{\left(P(\{S_{2n} = 0\})\right)^2} \approx \frac{\sqrt{1/(\pi\, 2n)}}{1/(\pi\, n)} = \sqrt{\pi\, n/2}.$$

□

Beispiel 17. Für $T = 1000$ gilt $\sqrt{\pi\, T/8} = 19.8166\ldots$ und

$$\frac{P(\{X_{1000} = 1000\})}{P(\{X_{1000} = 500\})} = 19.8315\ldots.$$

Nun betrachten wir den relativen Anteil

$$Z_T := \frac{1}{T} \cdot |\{t \in \{1, \ldots, T\} : S_t \geq 0 \text{ und } S_{t-1} \geq 0\}|$$

der Runden, nach denen Spieler I führt.

Wir vergleichen die Verteilungsfunktionen von Z_T für $T = 50, 100$ und 150, siehe Abbildung III.23. Vermutung: Konvergenz. Bestätigung und Bestimmung des Grenz-

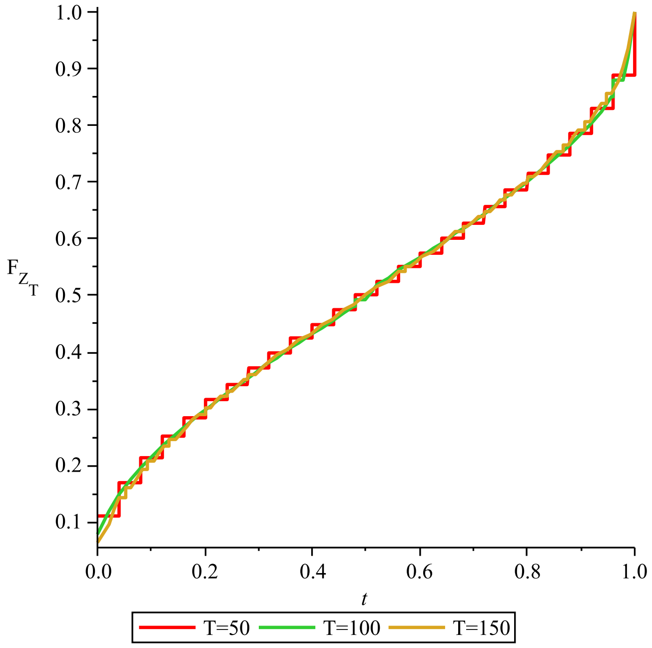

Abbildung III.23: Verteilungfunktionen von Z_T für $T = 50,\ 100,\ 150$

wertes folgt.
Dazu sei

$$f(y) := \frac{1}{\pi \cdot \sqrt{y \cdot (1-y)}}, \qquad 0 < y < 1,$$

und

$$F(x) = \int_0^x f(y)\, dy, \qquad x \in [0,1].$$

Bemerkung 18. Es gilt

$$F(x) = \frac{2}{\pi} \cdot \arcsin\left(\sqrt{x}\right)$$

und somit insbesondere $\int_0^1 f(y)\, dy = 1$.

Satz 19 (Arcussinus-Gesetz)**.** Für alle $0 < u < v < 1$ gilt

$$\lim_{n\to\infty} P(\{u \leq Z_{2n} \leq v\}) = F(v) - F(u).$$

Beweis. Fixiere $0 < u < v < 1$ und setze

$$K_n := \{k \in \{0,\dots,n\} : u \leq k/n \leq v\}$$

sowie

$$a_{k,n} := P(\{S_{2k} = 0\}) \cdot P(\{S_{2n-2k} = 0\}),$$

für $k \in K_n$. Satz 13 zeigt

$$P(\{u \leq Z_{2n} \leq v\}) = P(\{u \cdot 2n \leq X_{2n} \leq v \cdot 2n\}) = \sum_{k \in K_n} P(\{X_{2n} = 2k\}) = \sum_{k \in K_n} a_{k,n}.$$

Verwende Korollar 15 und $\lim_{n\to\infty} \inf K_n = \lim_{n\to\infty} \inf\{n-k : k \in K_n\} = \infty$ um

$$\lim_{n\to\infty} \sup_{k \in K_n} \left| \frac{n \cdot a_{k,n}}{f(k/n)} - 1 \right| = 0 \tag{1}$$

zu erhalten. Aufgrund der Stetigkeit von f auf $[u, v]$ gilt

$$\lim_{n\to\infty} \sum_{k \in K_n} \frac{f(k/n)}{n} = \int_u^v f(y)\, dy.$$

Es folgt

$$\lim_{n\to\infty} \left| \sum_{k \in K_n} a_{k,n} - \sum_{k \in K_n} \frac{f(k/n)}{n} \right| \leq \lim_{n\to\infty} \sum_{k \in K_n} \left(\frac{f(k/n)}{n} \cdot \sup_{k \in K_n} \left| \frac{n \cdot a_{k,n}}{f(k/n)} - 1 \right| \right) = 0$$

und somit

$$\lim_{n\to\infty} P(\{u \leq Z_{2n} \leq v\}) = \int_u^v f(y)\, dy.$$

□

Bemerkung 20. Gleichung (1) heißt lokales Arcussinus-Gesetz und besagt, daß die reskalierten Wahrscheinlichkeiten $a_{k,n}$ lokal gleichmäßig gegen f konvergieren.

Korollar 21. Für alle $x \in [0, 1]$ gilt

$$\lim_{n\to\infty} P(\{Z_{2n} \leq x\}) = F(x).$$

Beweis. Für $x = 0$ folgt die Aussage aus Satz 13 und Korollar 15. Für $0 < x \leq 1$ wählen wir $0 < u < \min(x, 1/2)$. Da $P(\{Z_{2n} < u\}) = P(\{Z_{2n} > 1 - u\})$, folgt

$$P(\{Z_{2n} < u\}) = 1/2 \cdot (1 - P(\{u \leq Z_{2n} \leq 1 - u\})).$$

Satz 19 sichert

$$\lim_{n\to\infty} P(\{Z_{2n} < u\}) = F(u).$$

Wende Satz 19 an. □

Definition 22. Eine Zufallsvariable X mit der Verteilungsfunktion

$$F_X(x) = \begin{cases} 0, & \text{falls } x < 0 \\ \frac{2}{\pi} \cdot \arcsin\left(\sqrt{x}\right), & \text{falls } x \in [0,1] \\ 1, & \text{falls } x > 1 \end{cases}$$

heißt *arcussinus-verteilt*.

Bemerkung 23. Mit den Sätzen 4.12 und 19 und mit Übung 6:G8 haben wir einige Konvergenzsätze für Folgen diskreter Zufallsvariablen kennengelernt. Der zugrundeliegende Konvergenzbegriff wird in Abschnitt VII.3 allgemein gefaßt. In Satz 19 zeigt sich erstmals, daß eine nicht-diskrete Verteilung als Grenzwert einer Folge diskreter Verteilungen auftreten kann.

Wir bestimmen die Wahrscheinlichkeit dafür, daß die Irrfahrt bis zur Zeit $2k$ nicht zum Startwert zurückkehrt.

Satz 24. Für $k \in \{1, \dots, T/2\}$ gilt

$$P(\{\tau_T > 2k\}) = P(\{S_{2k} = 0\}).$$

Beweis. Es gilt

$$\begin{aligned} \{\tau_T > 2k\} &= \bigcap_{j=1}^{2k} \{S_j \neq 0\} \\ &= \left(\{S_1 = -1\} \cap \bigcap_{j=2}^{2k} \{S_j < 0\}\right) \cup \left(\{S_1 = 1\} \cap \bigcap_{j=2}^{2k} \{S_j > 0\}\right) \end{aligned}$$

Es folgt

$$P(\{\tau_T > 2k\}) = 2 \cdot P\left(\{S_1 = -1\} \cap \bigcap_{j=2}^{2k} \{S_j < 0\}\right).$$

Mit den Sätzen 4 und 9 ergibt sich

$$\begin{aligned} P\left(\bigcap_{j=2}^{2k} \{S_j < 0\} \,|\, \{S_1 = -1\}\right) &= P\left(\bigcap_{j=1}^{2k-1} \{S_j \leq 0\}\right) \\ &= P(\{\max_{j=0,\dots,2k} S_j = 0\}) = P(\{S_{2k} = 0\}). \end{aligned}$$

□

Wie wahrscheinlich ist es nie zum Startwert zurückzukehren? Zur Untersuchung dieser Frage betrachten wir eine iid-Folge $(Y_t)_{t \in \mathbb{N}}$ mit $Y_1 \sim \mathbf{SB}$ und definieren die symmetriche Bernoulli-Irrfahrt mit unendlichem Zeithorizont durch $S_0 = 0$ sowie $S_t = S_{t-1} + Y_t$ für $t \in \mathbb{N}$. (Man beachte jedoch Bemerkung 4.18.)

Wir setzen

$$\tau_\infty := \inf\{t \in \mathbb{N} : S_t = 0\}.$$

Satz 25. Es gilt

$$P(\{\tau_\infty = \infty\}) = 0.$$

Beweis. Es gilt

$$\{\tau_\infty = \infty\} = \bigcap_{n=1}^{\infty} \{\tau_{2n} = \infty\}$$

und $\{\tau_{2n} = \infty\} \supset \{\tau_{2n+2} = \infty\}$, so daß

$$P(\{\tau_\infty = \infty\}) = \lim_{n\to\infty} P(\{\tau_{2n} = \infty\}) = \lim_{n\to\infty} P(\{S_{2n} = 0\})$$

mit den Sätzen I.1.17 und 24 folgt. Wende Korollar 15 an. $\square$

Ausblick: Rekurrenz und Transienz in Dimension $d \in \mathbb{N}$.

Kapitel IV

Grundlagen allgemeiner Modelle

Bisher in erster Linie studiert: diskrete Wahrscheinlichkeitsräume und diskrete Zufallsvariablen. Dieser Rahmen erlaubt beispielsweise nicht

- die Beschreibung des unendlich-oft wiederholten Münzwurf, siehe Bemerkung III.4.18,
- die Konstruktion von Wahrscheinlichkeitsräumen $(\Omega, \mathfrak{A}, P)$ und reellwertigen Zufallsvariablen X mit

  $$\forall x \in \mathbb{R}: \ P(\{X = x\}) = 0.$$

  Beispiel: Wartezeit, Koordinaten von Pfeiltreffer auf Dartscheibe, fehlerhafter Meßwert, kontinuierliches „Glücksrad“, usw.

Darüber hinaus ist die Menge der diskreten Verteilungen nicht abgeschlossen bzgl. Verteilungskonvergenz, siehe Bemerkung III.5.23 und Abschnitt VII.3.

1 Die Borelsche σ-Algebra in $\mathbb{R}^d$

Beispiel 1. Kontinuierliches „Glücksrad“. Versuch einer stochastischen Modellierung durch

(i) $\Omega := [0, 1[$ (Kreislinie der Länge 1),

(ii) $\mathfrak{A} := \mathfrak{P}(\Omega)$,

(iii) ein Wahrscheinlichkeitsmaß P auf $\mathfrak{A}$ mit folgenden Eigenschaften:

- $P([a, b]) = b - a$ für $0 \leq a < b < 1$
- $P(A) = P(B)$, falls B aus A durch „Rotation“ hervorgeht.

Definiere für $\omega, \omega' \in [0,1[$ und $A \subseteq [0,1[$

$$\begin{aligned} \omega \oplus \omega' &:= \omega + \omega' - \lfloor \omega + \omega' \rfloor, \\ \omega \oplus A &:= \{\omega \oplus a : a \in A\}. \end{aligned}$$

Frage: Existiert ein Wahrscheinlichkeitsmaß P auf $\mathfrak{P}([0,1[)$ mit

$$\forall A \subseteq [0,1[\ \forall \omega \in [0,1[: P(\omega \oplus A) = P(A)?$$

Antwort: Nein.

Beweis. Sei $Q := \mathbb{Q} \cap \Omega$. Betrachte die Äquivalenzrelation

$$\omega \sim \omega' :\Leftrightarrow \exists\, q \in Q : \omega' = \omega \oplus q$$

auf Ω und die zugehörigen Äquivalenzklassen

$$[r] = \{\omega \in \Omega : \omega \sim r\}.$$

Wähle ein Repräsentantensystem $R \subseteq \Omega$ (Auswahlaxiom), d.h.

$$\forall \omega \in \Omega \exists_1\, r \in R : \omega \in [r].$$

Es gilt für $q_1, q_2 \in Q$ mit $q_1 \neq q_2$

$$(q_1 \oplus R) \cap (q_2 \oplus R) = \emptyset.$$

Schließlich erfüllt P mit obigen Eigenschaften

$$1 = P(\Omega) = P(\bigcup_{q \in Q} q \oplus R) = \sum_{q \in Q} P(q \oplus R) = \sum_{q \in Q} P(R).$$

Widerspruch. $\square$

Folglich gibt es keine „Gleichverteilung“ auf $\mathfrak{P}([0,1[)$. Ausweg: betrachte eine kleinere σ-Algebra.

Im folgenden sei $\Omega \neq \emptyset$ und $\mathfrak{E} \subseteq \mathfrak{P}(\Omega)$ sowie

$$\begin{aligned} \mathbb{A} &:= \{\mathfrak{A} \subseteq \mathfrak{P}(\Omega) : \mathfrak{A}\, \sigma\text{-Algebra},\ \mathfrak{E} \subseteq \mathfrak{A}\}, \\ \sigma(\mathfrak{E}) &:= \bigcap_{\mathfrak{A} \in \mathbb{A}} \mathfrak{A} = \{A \subseteq \Omega : \forall \mathfrak{A} \in \mathbb{A} :\ A \in \mathfrak{A}\}. \end{aligned}$$

Beachte, daß $\mathfrak{P}(\Omega) \in \mathbb{A}$.

Lemma 2. $\sigma(\mathfrak{E})$ ist die kleinste σ-Algebra, die $\mathfrak{E}$ umfaßt, d.h.

(i) $\sigma(\mathfrak{E})$ ist σ-Algebra,

(ii) $\mathfrak{E} \subseteq \sigma(\mathfrak{E})$,

(iii) $\forall \mathfrak{A} \in \mathbb{A} : \sigma(\mathfrak{E}) \subseteq \mathfrak{A}$.

Beweis. Ad (i), exemplarisch der Nachweis einer der Eigenschaften: Gelte $A_1, A_2, \ldots \in \sigma(\mathfrak{E})$. Dann

$$\forall \mathfrak{A} \in \mathbb{A} : \ A_1, A_2, \ldots \in \mathfrak{A},$$

so daß

$$\forall \mathfrak{A} \in \mathbb{A} : \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A},$$

da $\mathfrak{A} \in \mathbb{A}$ σ-Algebra. Dies zeigt $\bigcup_{i=1}^{\infty} A_i \in \sigma(\mathfrak{E})$.
Ad (ii): Nach Definition gilt $\forall \mathfrak{A} \in \mathbb{A} : \ \mathfrak{E} \subseteq \mathfrak{A}$, d.h.

$$\forall \mathfrak{A} \in \mathbb{A} \forall E \in \mathfrak{E} : \ E \in \mathfrak{A}.$$

Somit

$$\forall E \in \mathfrak{E} : \ E \in \sigma(\mathfrak{E}).$$

Ad (iii): Klar nach Definition. □

Definition 3. $\sigma(\mathfrak{E})$ heißt die von $\mathfrak{E}$ *erzeugte σ-Algebra* (in Ω).
Vgl. erzeugter Untervektorraum.

Beispiel 4. Für $\Omega = \{1, 2, 3, 4\}$ und $\mathfrak{E} = \{\{1, 2, 3\}, \{2\}\}$ gilt

$$\begin{aligned} \sigma(\mathfrak{E}) &= \{\emptyset, \{2\}, \{4\}, \{1, 3\}, \{2, 4\}, \{1, 2, 3\}, \{1, 3, 4\}, \Omega\} \\ &= \{A \subseteq \Omega : \{1, 3\} \subseteq A \text{ oder } \{1, 3\} \cap A = \emptyset\}. \end{aligned}$$

Definition 5. Für $d \in \mathbb{N}$ und

$$\mathfrak{O}_d := \{O \subseteq \mathbb{R}^d : O \text{ offen}\}$$

heißt

$$\mathfrak{B}_d := \sigma(\mathfrak{O}_d)$$

die *Borelsche σ-Algebra* in $\mathbb{R}^d$. Die Elemente $B \in \mathfrak{B}_d$ heißen *Borel-Mengen* (in $\mathbb{R}^d$).

Bemerkung 6.

(i) $A \subseteq \mathbb{R}^d$ abgeschlossen $\Rightarrow A \in \mathfrak{B}_d$, da A^c offen,

(ii) $\mathfrak{M} \subsetneq \mathfrak{B}_1$.

Lemma 7.

$$A_1, \dots, A_d \in \mathfrak{B}_1 \Rightarrow A_1 \times \cdots \times A_d \in \mathfrak{B}_d.$$

Beweis. Siehe Irle (2001, p. 151). Stichwort: Produkt-σ-Algebra. □

Bemerkung 8. Es gilt $\mathfrak{B}_d \subsetneq \mathfrak{P}(\mathbb{R}^d)$, siehe Bemerkung 2.5. Uns werden in dieser Vorlesung jedoch (außer der Menge R aus Beispiel 1.1) keine Mengen aus $\mathfrak{P}(\mathbb{R}^d) \backslash \mathfrak{B}_d$ begegnen. Siehe dazu auch Krengel (2000, p. 130).

Satz 9.

(i) Gilt

$$\forall x_1, \dots, x_d \in \mathbb{R} : P(]-\infty, x_1] \times \cdots \times]-\infty, x_d]) = Q(]-\infty, x_1] \times \cdots \times]-\infty, x_d])$$

für Wahrscheinlichkeitsmaße P und Q auf $\mathfrak{B}_d$, so folgt

$$P = Q.$$

(ii) Sind $P_1, \dots, P_d$ Wahrscheinlichkeitsmaße auf $\mathfrak{B}_1$, so existiert ein Wahrscheinlichkeitsmaß P auf $\mathfrak{B}_d$ mit

$$\forall x_1, \dots, x_d \in \mathbb{R} : P(]-\infty, x_1] \times \cdots \times]-\infty, x_d]) = \prod_{i=1}^{d} P_i(]-\infty, x_i]).$$

Beweis. Ad (i): Siehe Irle (2001, p. 151) oder Georgii (2007, p. 16). Stichwort: durchschnittstabiler Erzeuger.
Ad (ii): Siehe Irle (2001, p. 168). Stichwort: Produkt-Maß. □

Bemerkung 10. Satz 9.(i) sichert, daß das Wahrscheinlichkeitsmaß P gemäß Satz 9.(ii) eindeutig bestimmt ist. Ferner gilt

$$P(A_1 \times \cdots \times A_d) = \prod_{i=1}^{d} P_i(A_i)$$

für dieses Wahrscheinlichkeitsmaß P und alle $A_1, \dots, A_d \in \mathfrak{B}_1$.

2 Das d-dimensionale Lebesgue-Maß

Satz 1. Es existiert genau eine σ-additive Abbildung $\lambda_d : \mathfrak{B}_d \to [0, \infty]$ mit

$$\forall a_i \le b_i : \ \lambda_d([a_1, b_1] \times \cdots \times [a_d, b_d]) = \prod_{i=1}^{d} (b_i - a_i).$$

Diese erfüllt

$$\lambda_d(a + Q(A)) = \lambda_d(A)$$

für alle $a \in \mathbb{R}^d$, $A \in \mathfrak{B}_d$ und alle orthogonalen Abbildungen $Q : \mathbb{R}^d \to \mathbb{R}^d$.

Beweis. Siehe Meintrup, Schäffler (2005, Anhang A.1). □

Definition 2. λ_d gemäß Satz 1 heißt *Lebesgue-Maß* auf $\mathfrak{B}_d$ oder d-dimensionales Lebesgue-Maß.

Bemerkung 3. Für das Lebesgue-Maß gelten die Sätze I.1.16 (i)–(iii) und I.1.17 (i), (ii) entsprechend.

Beispiel 4. Modellierung des kontinuierlichen „Glücksrad" durch

(i) $\Omega := [0,1[$,

(ii) $\mathfrak{A} := \{A \in \mathfrak{B}_1 : A \subseteq \Omega\}$,

(iii) $P(A) := \lambda_1(A)$.

Es gilt $P(\Omega) = 1$, und $\mathfrak{A}$ ist eine σ-Algebra in Ω, so daß $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum ist. Für $A \in \mathfrak{A}$ und $\omega \in \Omega$ setzen wir

$$A_1 := A \cap [0, 1-\omega[\in \mathfrak{A}, \qquad A_2 := A \cap [1-\omega, 1[\in \mathfrak{A}$$

und erhalten

$$\begin{aligned} P(\omega \oplus A) &= P(\omega \oplus A_1) + P(\omega \oplus A_2) = P(\omega + A_1) + P(\omega + A_2 - 1) \\ &= P(A_1) + P(A_2) = P(A). \end{aligned}$$

Bemerkung 5. Aus den Beispielen 1.1 und 4 ergibt sich $\mathfrak{B}_1 \neq \mathfrak{P}(\mathbb{R})$. Insbesondere gilt $R \notin \mathfrak{B}_1$.

Bemerkung 6. Auf dem Wahrscheinlichkeitsraum aus Beispiel 4 existiert eine Folge $X_1, X_2, \ldots$ von iid Zufallsvariablen mit $X_1 \sim \mathbf{B}(1,p)$, siehe auch Bemerkung III.4.18 Im Spezialfall $p = 1/2$ leistet

$$X_n(\omega) = \sum_{i=1}^{2^{n-1}} 1_{A_i}(\omega)$$

mit

$$A_i = [(2i-1)/2^n, 2i/2^n[$$

das Verlangte, siehe Übung 8:H35. Bemerkt sei, daß $X_n(\omega)$ die n-te Stelle der Binärentwicklung von ω ist.

3 Verteilungen

Im folgenden betrachten wir einen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ und Abbildungen $X_1, \ldots, X_d : \Omega \to \mathbb{R}$. Setze

$$X := (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d.$$

Lemma 1. Äquivalent sind

(i) $X_1, \ldots, X_d$ sind Zufallsvariablen,

(ii) $\forall A \in \mathfrak{B}_d : \ \{X \in A\} \in \mathfrak{A}$.

Beweis. „(ii) $\Rightarrow$ (i)“

$$\{X_i \le c_i\} = \{X_i \le c_i\} \cap \bigcap_{j \neq i} \{X_j \in \mathbb{R}\} = \{X \in A\}$$

für $A = \mathbb{R}^{i-1} \times]-\infty, c_i] \times \mathbb{R}^{d-i}$. Da A abgeschlossen, folgt $A \in \mathfrak{B}_d$ und wg. (ii) auch $\{X_i \le c_i\} \in \mathfrak{A}$.
„(i) $\Rightarrow$ (ii)“ Siehe Irle (2001, p. 151). Stichwort: Produkt-Meßbarkeit. □

Vgl. Lemma 1 mit Lemma I.3.8 im Fall $d = 1$. Vgl. Begriffe der Meßbarkeit und Stetigkeit von Abbildungen.

Definition 2. X heißt (d-dimensionaler) *Zufallsvektor*, falls eine der beiden Eigenschaften aus Lemma 1 erfüllt ist.

Ein Zufallsvektor mit $d > 1$ dient zur gemeinsamen Modellierung mehrerer Aspekte eines Zufallsexperimentes. Siehe bereits Beispiel I.3.1.
Fortan sei $d \in \mathbb{N}$ und $X = (X_1, \ldots, X_d)$ sei ein Zufallsvektor.

Satz 3.

$$P_X(A) := P(\{X \in A\}), \qquad A \in \mathfrak{B}_d,$$

definiert ein Wahrscheinlichkeitsmaß auf $\mathfrak{B}_d$.

Beweis. Klar: $P_X \ge 0$ und $P_X(\mathbb{R}^d) = P(\{X \in \mathbb{R}^d\}) = P(\Omega) = 1$.
Für $A_1, A_2, \ldots \in \mathfrak{B}_d$ p.d. und $A := \bigcup_{i=1}^\infty A_i$ gilt

$$\begin{aligned} P_X(A) &= P\Big(\Big\{X \in \bigcup_{i=1}^\infty A_i\Big\}\Big) = P\Big(\bigcup_{i=1}^\infty \underbrace{\{X \in A_i\}}_{\in \mathfrak{A} \text{ p.d.}}\Big) \\ &= \sum_{i=1}^\infty P(\{X \in A_i\}) = \sum_{i=1}^\infty P_X(A_i). \end{aligned}$$

□

Definition 4. P_X heißt *Verteilung* von X. Im Falle $d > 1$ heißt P_X auch *gemeinsame Verteilung* der Zufallsvariablen $X_1, \dots, X_d$, und $P_{X_1}, \dots, P_{X_d}$ heißen *(eindim.) Randverteilungen* von X.

Wir können somit von Binomial-Verteilungen, geometrischen Verteilungen etc. als Wahrscheinlichkeitsmaßen sprechen.

Bemerkung 5. Jedes Wahrscheinlichkeitsmaß P auf $\mathfrak{B}_d$ ist Verteilung eines Zufallsvektors: betrachte $X(\omega) := \omega$ auf dem Wahrscheinlichkeitsraum $(\mathbb{R}^d, \mathfrak{B}_d, P)$.

Satz 6. Für Zufallsvariablen X auf $(\Omega, \mathfrak{A}, P)$ und X' auf $(\Omega', \mathfrak{A}', P')$ sind äquivalent

i) X, X' identisch verteilt,

ii) $P_X = P'_{X'}$.

Beweis. „(ii) $\Rightarrow$ (i)“: klar. „(i) $\Rightarrow$ (ii)“: Satz 1.9. $\square$

Definition 7. d-dimensionale Zufallsvektoren X auf $(\Omega, \mathfrak{A}, P)$ und X' auf $(\Omega', \mathfrak{A}', P')$ heißen *identisch verteilt*, falls $P_X = P'_{X'}$.

Viele Fragestellungen der Stochastik betreffen nicht die konkrete Gestalt des zugrundeliegenden Wahrscheinlichkeitsraumes und der betrachteten Zufallsvariablen sondern nur ihre gemeinsame Verteilung. Beispiel: Unabhängigkeit, siehe Satz 10.

Beispiel 8. Sei P die Gleichverteilung auf $\Omega := \{0,1\}^2$ (zweimaliger Münzwurf) und sei $X_i(\omega) := \omega_i$. Also sind X_1 und X_2 diskret, und es gilt $P(\{X = x\}) = 1/4$ für $x \in D := \Omega$.

Bestimmung der gemeinsamen Verteilung von X_1 und X_2: Für $A \in \mathfrak{B}_2$ gilt

$$P_X(A) = P(\{X \in A\} \cap \{X \in D\}) = \sum_{x \in A \cap D} P(\{X = x\}) = |A \cap D|/4.$$

Bestimmung der Randverteilungen: Für $B \in \mathfrak{B}_1$ gilt

$$P_{X_1}(B) = P(\{X \in B \times \mathbb{R}\}) = |(B \times \mathbb{R}) \cap D|/4 = |B \cap \{0,1\}|/2.$$

Analog zeigt man $P_{X_2}(B) = |B \cap \{0,1\}|/2$.

Insbesondere gilt $P_{X_1} = P_{X_2}$, obwohl $X_1 \neq X_2$.

Jetzt betrachten wir $X'_1(\omega) := \omega_1$ und $X'_2(\omega) := \omega_1$. Wiederum sind X'_1 und X'_2 diskret, und es gilt $P(\{X' = x\}) = P(\{X_1 = x_1\}) = 1/2$ für $x \in D' := \{(0,0),(1,1)\}$.

Bestimmung der gemeinsamen Verteilung: Für $A \in \mathfrak{B}_2$ gilt

$$P_{X'}(A) = P(\{X' \in A\} \cap \{X' \in D'\}) = \sum_{x \in A \cap D'} P(\{X = x\}) = |A \cap D'|/2.$$

Also gilt $P_X \neq P_{X'}$.
Für die Randverteilungen gilt

$$P_{X_1'} = P_{X_2'} = P_{X_1} = P_{X_2}.$$

Dies zeigt, daß die gemeinsame Verteilung durch die eindimensionalen Randverteilungen im allgemeinen nicht eindeutig bestimmt ist. Siehe jedoch Bemerkung 11.(ii).

Beispiel 9. Sei P die Gleichverteilung auf $\Omega := \{1, \dots, 6\}^2$ (zweimaliges Würfeln). Setze

$$X_1(\omega) := \omega_1, \qquad X_2(\omega) := \omega_1 + \omega_2.$$

Für

$$D := \{x \in \{1, \dots, 6\} \times \{2, \dots, 12\} : 1 \leq x_2 - x_1 \leq 6\}$$

gilt $P(\{X \in D\}) = 1$ sowie $P(\{X = x\}) = 1/36$ für alle $x \in D$. Also

$$P_X(A) = \sum_{x \in A \cap D} P(\{X = x\}) = \frac{1}{36} \cdot |A \cap D|.$$

Satz 10. $(X_1, \dots, X_d)$ genau dann unabhängig, wenn

$$\forall A_1, \dots, A_d \in \mathfrak{B}_1 : \quad \underbrace{P\Big(\bigcap_{i=1}^{d} \{X_i \in A_i\}\Big)}_{=P_X(A_1 \times \dots \times A_d)} = \prod_{i=1}^{d} \underbrace{P(\{X_i \in A_i\})}_{=P_{X_i}(A_i)}.$$

Kurz: die gemeinsame Verteilung ist das Produkt der Randverteilungen.

Beweis. „$\Leftarrow$“ Klar. „$\Rightarrow$“ Gemäß Satz 1.9.(ii) wird durch

$$Q(]-\infty, x_1] \times \dots \times]-\infty, x_d]) := \prod_{i=1}^{d} P_{X_i}(]-\infty, x_i])$$

ein Wahrscheinlichkeitsmaß Q auf $\mathfrak{B}_d$ definiert. Satz 1.9.(i) und die Voraussetzung sichern $Q = P_X$. Wende Bemerkung 1.10 an.

Wir geben einen elementaren Beweis unter der zusätzlichen Annahme, daß $X_1, \dots, X_d$ diskret. Wähle abzählbare Menge $D \subset \mathbb{R}$ mit

$$P(\{X \in D^d\}) = P\Big(\bigcap_{i=1}^{d} \{X_i \in D\}\Big) = 1.$$

Für $A = A_1 \times \dots \times A_d$ folgt mit Satz I.3.15

$$\begin{aligned} P\Big(\bigcap_{i=1}^{d} \{X_i \in A_i\}\Big) &= P(\{X \in A\}) = \sum_{x \in A \cap D^d} P(\{X = x\}) \\ &= \sum_{x_1 \in A_1 \cap D} \cdots \sum_{x_d \in A_d \cap D} \prod_{i=1}^{d} P(\{X_i = x_i\}) = \prod_{i=1}^{d} P(\{X_i \in A_i\}). \end{aligned}$$

□

Bemerkung 11.

(i) Satz I.3.15 enthält eine Teilaussage von Satz 10.

(ii) Sind $X_1, \ldots, X_d$ unabhängig, so ist die gemeinsame Verteilung P_X eindeutig durch die Randverteilungen $P_{X_1}, \ldots, P_{X_d}$ bestimmt, siehe Beweis von Satz 10. Vgl. Modellierung in Kapitel III.

(iii) Diskrete Zufallsvariablen $X_1, \ldots, X_d$ sind genau dann unabhängig, wenn

$$\forall x \in \mathbb{R}^d : \; P(\{X = x\}) = \prod_{i=1}^{d} P(\{X_i = x_i\}).$$

Beweis. „$\Rightarrow$“ Siehe Satz I.3.15 bzw. Satz 10. „$\Leftarrow$“ Siehe Beweis von Satz 10.

Kapitel V

Absolutstetige Modelle

Ist X eine diskrete Zufallsvariable auf $(\Omega, \mathfrak{A}, P)$, so gilt $P(\{X \in D\}) = 1$ für eine abzählbare Menge $D \subset \mathbb{R}$, und für alle $A \in \mathfrak{B}_1$ folgt

$$P(\{X \in A\}) = P(\{X \in A\} \cap \{X \in D\}) = \sum_{x \in A \cap D} P(\{X = x\}).$$

In diesem Kapitel untersuchen wir Zufallsvariablen, bei denen die Summation der Wahrscheinlichkeitsfunktion $x \mapsto P(\{X = x\})$ über $A \cap D$ durch die Integration einer geeigneten Funktion $x \mapsto f_X(x)$ über A ersetzt wird. Also

$$P(\{X \in A\}) = \int_A f_X(x)\, dx.$$

1 Wahrscheinlichkeitsdichten

Bemerkung 1. Integralbegriffe für Funktionen $f : \mathbb{R}^d \to \mathbb{R}$:

(i) Lebesgue-Integral (Vorlesung Analysis IV), siehe auch Georgii (2007, p. 17).

(ii) Spezialfall: Uneigentliches Riemann-Integral (Walther, Analysis II, Springer-Verlag, 1990, §7.20).

(iii) Spezialfall: Für abgeschlossene Intervalle $B_i \subseteq \mathbb{R}$ und $B := B_1 \times \cdots \times B_d \subseteq \mathbb{R}^d$ sei $f|_B$ stetig. Setze $B^{(K)} := B \cap [-K, K]^d$. Falls $\sup_{K \in \mathbb{N}} \int_{B^{(K)}} |f(x)|\, dx < \infty$, so gilt

$$\int_B f(x)\, dx = \lim_{K \to \infty} \int_{B^{(K)}} f(x)\, dx.$$

Berechnung von $\int_{B^{(K)}} f(x)\, dx$ mit $B^{(K)} = [a_1, b_1] \times \cdots \times [a_d, b_d]$ als iteriertes Integral

$$\int_{B^{(K)}} f(x)\, dx = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(x_1, \dots, x_d)\, dx_d \dots dx_1.$$

Definition 2. $f : \mathbb{R}^d \to \mathbb{R}_+$ heißt *Wahrscheinlichkeitsdichte*, kurz *Dichte*, falls f (Lebesgue)-integrierbar mit

$$\int_{\mathbb{R}^d} f(x)\,dx = 1.$$

Satz 3. Jede Dichte f definiert durch

$$P(A) := \int_A f(x)\,dx, \qquad A \in \mathfrak{B}_d,$$

ein Wahrscheinlichkeitsmaß auf $\mathfrak{B}_d$.

Beweis. Klar: $P \geq 0$ und $P(\mathbb{R}^d) = 1$.
Für $A_1, A_2, \ldots \in \mathfrak{B}_d$ p.d. und $A := \bigcup_{i=1}^\infty A_i$ gilt

$$\begin{aligned} P(A) &= \int_{\mathbb{R}^d} 1_A(x) \cdot f(x)\,dx = \int_{\mathbb{R}^d} \sum_{i=1}^\infty 1_{A_i}(x) \cdot f(x)\,dx \\ &= \sum_{i=1}^\infty \int_{\mathbb{R}^d} 1_{A_i}(x) \cdot f(x)\,dx = \sum_{i=1}^\infty P(A_i) \end{aligned}$$

aufgrund des Satzes von der monotonen Konvergenz. □

Vgl. Satz 3 mit Satz III.1.3 über Wahrscheinlichkeitsfunktionen.
Nun zur Eindeutigkeit von Dichten.

Lemma 4. Seien $f, g : \mathbb{R}^d \to \mathbb{R}$ integrierbar. Dann sind äquivalent:

(i) $\forall A \in \mathfrak{B}_d : \displaystyle\int_A f(x)\,dx = \int_A g(x)\,dx,$

(ii) $\lambda_d(\{x \in \mathbb{R}^d : f(x) \neq g(x)\}) = 0.$

Beweis. Folgt aus Meintrup, Schäffler (2005, Satz 2.15). □

Ausblick: singuläre Verteilungen.

2 Absolutstetig verteilte Zufallsvariablen

Satz 1.3 und Bemerkung IV.3.5 erlauben die Modellierung von Verteilungen durch Vorgabe ihrer Dichten. Ein erstes Beispiel haben wir mit Definition III.5.22 kennengelernt.

Im folgenden sei $d \in \mathbb{N}$ und $X = (X_1, \ldots, X_d)$ bezeichne einen d-dimensionaler Zufallsvektor auf $(\Omega, \mathfrak{A}, P)$.

Definition 1. X heißt *absolutstetig verteilt*, falls P_X eine Dichte besitzt. Diese wird ggf. mit f_X bezeichnet.

Definition 2. Sei $B \in \mathfrak{B}_d$ mit Lebesgue-Maß (Länge, Flächeninhalt, Volumen) $\lambda_d(B) \in]0, \infty[$. Ein Zufallsvektor X mit Dichte

$$f_X(x) = \frac{1}{\lambda_d(B)} \cdot 1_B(x)$$

heißt *gleichverteilt* auf B. Bez.: $X \sim \mathbf{U}(B)$.

Bemerkung 3. Für $X \sim \mathbf{U}(B)$ und $A \in \mathfrak{B}_d$ gilt

$$P_X(A) = \frac{1}{\lambda_d(B)} \cdot \int_A 1_B(x)\, dx = \frac{\lambda_d(A \cap B)}{\lambda_d(B)}.$$

Beispiel 4. Dichte und Verteilungsfunktion von $X \sim \mathbf{U}([a, b])$ mit $a < b$ sind gegeben durch

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } x \in [a, b] \\ 0, & \text{sonst} \end{cases}$$

und

$$F_X(x) = \begin{cases} 0, & \text{falls } x < a \\ \frac{x-a}{b-a}, & \text{falls } x \in [a, b] \\ 1, & \text{sonst.} \end{cases}$$

Beispiel 5. $X \sim \mathbf{U}(B)$ zur Modellierung von Pfeiltreffer auf Dartscheibe, Glücksrad. Wichtige Anwendung der Gleichverteilung: Zufallszahlen und stochastische Simulation, siehe Kapitel II.

Definition 6. Eine Zufallsvariable X mit Dichte

$$f_X(x) = \begin{cases} \lambda \cdot \exp(-\lambda x), & \text{falls } x \geq 0 \\ 0, & \text{sonst} \end{cases}$$

für $\lambda > 0$ heißt *exponentialverteilt* mit Parameter λ. Bez.: $X \sim \mathbf{Exp}(\lambda)$.

Bemerkung 7. Für $X \sim \mathbf{Exp}(\lambda)$ und $x > 0$ gilt

$$F_X(x) = \lambda \cdot \int_0^x \exp(-\lambda y)\; dy = 1 - \exp(-\lambda x)$$

Klar: $F_X(x) = 0$, falls $x \leq 0$.

Beispiel 8. Dichten und Verteilungsfunktionen exponentialverteilter Zufallsvariablen sind in den Abbildungenen V.1 und V.2 dargestellt.

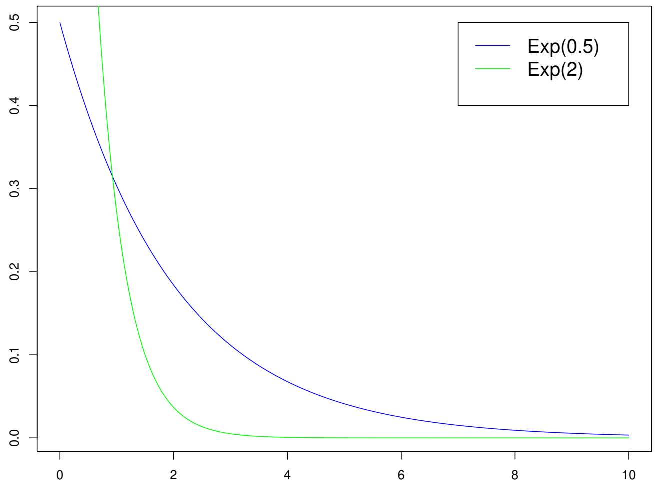

Abbildung V.1: Dichten exponentialverteilter Zufallsvariablen

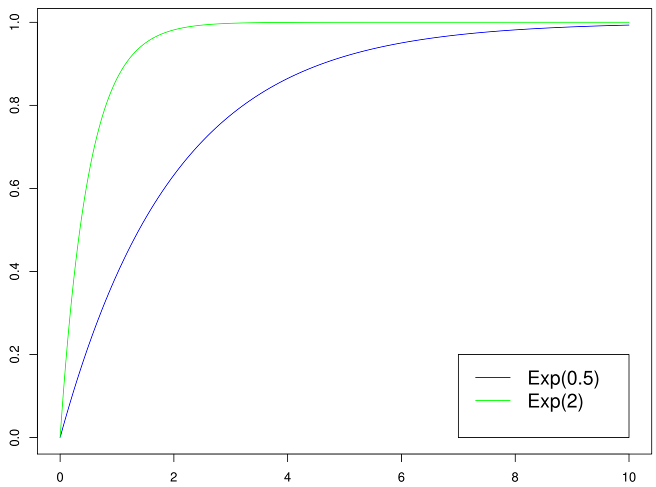

Abbildung V.2: Verteilungsfunktionen exponentialverteilter Zufallsvariablen

Nun: Charakterisierung der Exponentialverteilung durch ihre Gedächtnislosigkeit.

Satz 9. Für eine Zufallsvariable X mit

- $P(\{X > 0\}) = 1$ und
- $\forall t > 0 : P(\{X > t\}) > 0$

sind äquivalent

(i) $\exists \lambda > 0 : \; X \sim \mathbf{Exp}(\lambda)$,

(ii) $\forall s, t > 0 : P(\{X > t + s\} \mid \{X > t\}) = P(\{X > s\})$.

Beweis. Siehe Übung 9:H36 und vgl. Übung 6:H24. □

Beispiel 10. $X \sim \mathbf{Exp}(\lambda)$ zur Modellierung von Lebensdauern, Wartezeiten, radioaktivem Zerfall. Beim radioaktiven Zerfall bezeichne X den Zerfallszeitpunkt. Die Halbwertszeit $h > 0$ ist definiert durch

$$P(\{X \leq h\}) = \frac{1}{2}.$$

Man erhält

$$h = \ln(2)/\lambda.$$

Siehe auch Beispiel VII.3.8.

Lemma 11. Für $\mu \in \mathbb{R}$ und $\sigma > 0$ gilt

$$\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}.$$

Beweis. OBdA $\mu = 0$ und $\sigma = 1$ (Substitutionsregel). Es gilt

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx\right)^2 &= \int_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2}\right) d(x,y) \\
&= \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{r^2}{2}\right) \cdot r \, dr \, d\varphi \\
&= 2\pi \cdot \left(-\exp\left(-\frac{r^2}{2}\right)\right) \Big|_0^{\infty} = 2\pi.
\end{aligned}$$

□

Definition 12. Eine Zufallsvariable X mit Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

für $\mu \in \mathbb{R}$ und $\sigma > 0$ heißt *normalverteilt* mit Parametern μ und σ^2. Bez.: $X \sim \mathbf{N}(\mu, \sigma^2)$. *Standard-Normalverteilung* als Spezialfall: $\mu = 0$ und $\sigma = 1$.

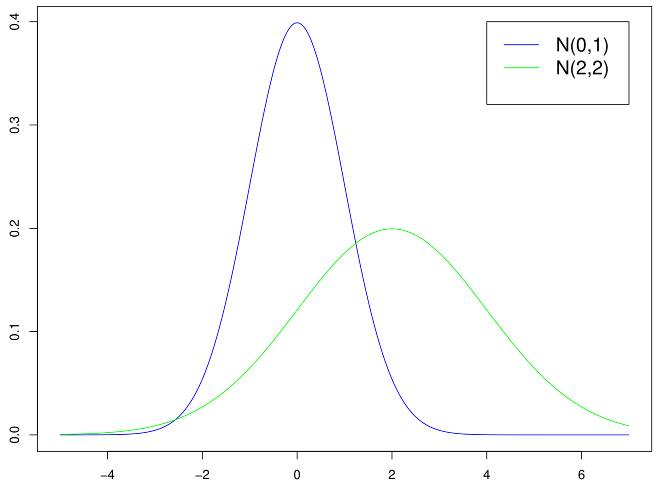

Abbildung V.3: Dichten normalverteilter Zufallsvariablen

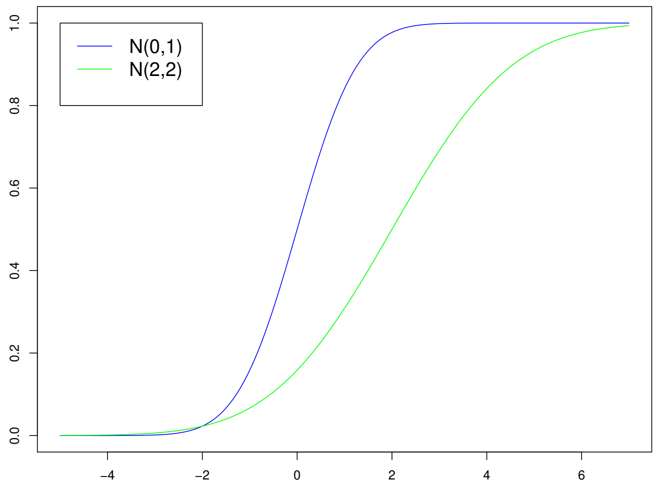

Abbildung V.4: Verteilungsfunktionen normalverteilter Zufallsvariablen

Beispiel 13. Dichten und Verteilungsfunktionen normalverteilter Zufallsvariablen sind in den Abbildungenen V.3 und V.4 dargestellt.

Beispiel 14. $X \sim \mathbf{N}(\mu, \sigma^2)$ zur Modellierung von (Meß)Fehlern. Siehe Abschnitt VII.3 zur zentralen Rolle der Normalverteilung in der Stochastik.

Bemerkung 15. Es gibt keine explizite Formel für die Verteilungsfunktion F_X, falls $X \sim \mathbf{N}(\mu, \sigma^2)$. Bez.: $\Phi = F_X$, falls $X \sim \mathbf{N}(0, 1)$, also

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2}\right) dy.$$

Zur Berechnung von Φ und entsprechender Quantile: Numerik, Tabellen, Plots.

Nun: der Zusammenhang zwischen Dichten und Verteilungsfunktionen.

Satz 16. Für eine Zufallsvariable X sind äquivalent

(i) F_X stetig differenzierbar,

(ii) X absolutstetig verteilt mit stetiger Dichte f_X.

Gilt (i) oder (ii), so folgt $F_X' = f_X$.

Beweis. "(i) $\Rightarrow$ (ii)" Nach dem Hauptsatz der Differential- und Integralrechnung gilt

$$F_X(x) = \int_{-\infty}^{x} F_X'(u)\, du.$$

Gem. Satz IV.1.9 stimmt P_X mit dem Wahrscheinlichkeitsmaß mit der Dichte F_X' überein.
"(ii) $\Rightarrow$ (i)" Hauptsatz der Differential- und Integralrechnung. $\square$

Ausblick: absolutstetige Funktionen.
Nun speziell: mehrdimensionale Dichten. Analytisches Hilfsmittel: der Satz von Fubini.

Lemma 17. Falls f_X Dichte von P_X ist, so besitzt P_{X_i} die Dichte

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{d-1}} f_X(\overline{x}_1, x_i, \overline{x}_2)\, d(\overline{x}_1, \overline{x}_2)$$

mit

$$\overline{x}_1 = (x_1, \dots, x_{i-1}), \qquad \overline{x}_2 = (x_{i+1}, \dots, x_d).$$

Beweis. Für $A_i \in \mathfrak{B}_1$ sei $A := \mathbb{R}^{i-1} \times A_i \times \mathbb{R}^{d-i}$. Dann

$$\begin{aligned} P(\{X_i \in A_i\}) = P(\{X \in A\}) &= \int_A f_X(x)\,dx \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{A_i} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(x_1, \ldots, x_d)\,dx_d \ldots dx_1 \\ &= \int_{A_i} \underbrace{\int_{\mathbb{R}^{d-1}} f_X(\overline{x}_1, x_i, \overline{x}_2)\,d(\overline{x}_1, \overline{x}_2)}_{=:g(x_i)}\,dx_i, \end{aligned}$$

und g ist eine Dichte. □

Beispiel 18. Pfeiltreffer auf Dartscheibe. Hier

$$f_X(x_1, x_2) := \frac{1}{\pi r^2} \cdot 1_K(x_1, x_2)$$

mit

$$K := \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq r^2\}.$$

Also für $x_1 \in [-r, r]$

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_X(x_1, x_2)\,dx_2 = \frac{1}{\pi r^2} \cdot \int_{-\sqrt{r^2 - x_1^2}}^{\sqrt{r^2 - x_1^2}} 1\,dx_2 = \frac{2}{\pi r^2} \cdot \sqrt{r^2 - x_1^2}$$

sowie $f_{X_1}(x_1) = 0$, falls $|x_1| > r$. Klar

$$f_{X_1} = f_{X_2}.$$

Definition 19. Das *Tensorprodukt* $f_1 \otimes \ldots \otimes f_d : \mathbb{R}^d \to \mathbb{R}$ von Abbildungen $f_i : \mathbb{R} \to \mathbb{R}$ ist definiert durch

$$f_1 \otimes \ldots \otimes f_d(x) := f_1(x_1) \cdot \ldots \cdot f_d(x_d).$$

Vgl. Abschnitt III.3.

Lemma 20. Falls $f_1, \ldots, f_d$ Dichten auf $\mathbb{R}$, so ist $f_1 \otimes \ldots \otimes f_d$ Dichte auf $\mathbb{R}^d$.

Beweis. Klar. Vgl. Lemma III.3.2. □

Satz 21.

(i) Falls $X_1, \ldots, X_d$ unabhängig mit Dichten f_{X_i}, so besitzt X die Dichte

$$f_X = f_{X_1} \otimes \ldots \otimes f_{X_d}.$$

(ii) Falls X die Dichte

$$f_X = f_1 \otimes \ldots \otimes f_d$$

mit eindimensionalen Dichten f_i besitzt, so sind $X_1, \ldots, X_d$ unabhängig mit Dichten $f_{X_i} = f_i$.

Beweis. Ad (i): Gemäß Satz 1.3 und Lemma 20 definiert

$$Q(A) := \int_A f_{X_1} \otimes \ldots \otimes f_{X_d}(x)\, dx, \qquad A \in \mathfrak{B}_d,$$

ein Wahrscheinlichkeitsmaß auf $\mathfrak{B}_d$. Speziell für $A := A_1 \times \cdots \times A_d$ mit $A_i :=]-\infty, b_i]$

$$\begin{aligned} P(\{X \in A\}) &= \prod_{i=1}^{d} P(\{X_i \in A_i\}) = \prod_{i=1}^{d} \int_{A_i} f_{X_i}(x_i)\, dx_i \\ &= \int_{A_1} \cdots \int_{A_d} \prod_{i=1}^{d} f_{X_i}(x_i)\, dx_d \ldots dx_1 = Q(A). \end{aligned}$$

Satz IV.1.9.(i) zeigt $P_X = Q$.
Ad (ii): Für $A_1, \ldots, A_d \in \mathfrak{B}_1$ und $A := A_1 \times \cdots \times A_d$ gilt

$$P(\{X \in A\}) = \int_A f_X(x)\, dx = \int_{A_1} f_1(x_1)\, dx_1 \cdots \int_{A_d} f_d(x_d)\, dx_d.$$

Insbesondere

$$P(\{X_i \in A_i\}) = \int_{A_i} f_i(x_i)\, dx_i,$$

d.h. f_i ist Dichte von X_i, und weiter

$$P\left(\bigcap_{i=1}^{d} \{X_i \in A_i\}\right) = \prod_{i=1}^{d} P(\{X_i \in A_i\}).$$

□

Bemerkung 22. Satz 21 ermöglicht die Modellierung der „unabhängigen Hintereinanderausführung“ von Einzelexperimenten, deren Verteilungen Dichten besitzen.

Beispiel 23. Pfeiltreffer auf Dartscheibe, siehe Beispiel 18. Satz 21 zeigt, daß X_1, X_2 nicht unabhängig sind.

Definition 24. Ein d-dimensionaler Zufallsvektor X mit Dichte

$$f_X(x) = (2\pi)^{-d/2} \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^{d} x_i^2\right)$$

heißt *standard-normalverteilt* (in $\mathbb{R}^d$).

Beispiel 25. Die Dichte eines 2-dim. standard-normalverteilten Zufallsvektors ist in Abbildung V.5 dargestellt.

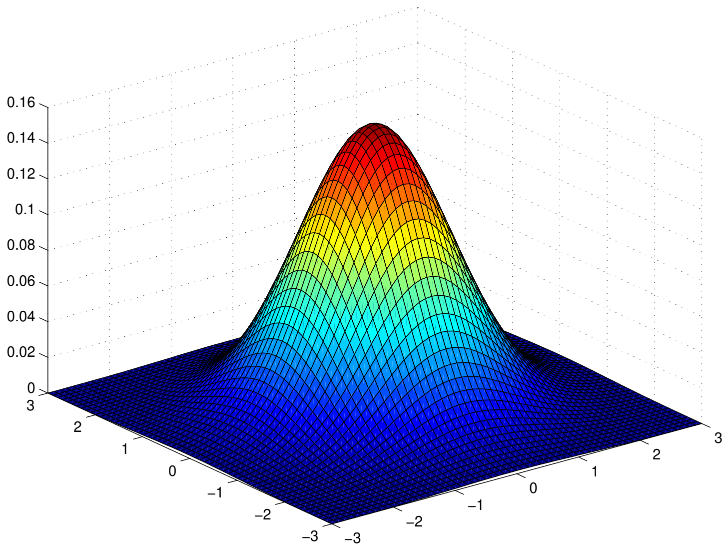

Abbildung V.5: Dichte eines 2-dim. standard-normalverteilten Zufallsvektors

.

Kapitel VI

Erwartungswert und Varianz

Erwartungswert und Varianz sind fundamentale Kenngrößen der Verteilung einer reellwertigen Zufallsvariable.

1 Der Erwartungswert

Erwartungswert: „mittlerer Wert“ einer Zufallsvariablen, „Schwerpunkt“ ihrer Verteilung. Die allgemeine Definition basiert auf dem abstraktem Lebesgue-Integral, siehe Georgii (2007, Abschn. 4.1.2).

Bemerkung 1. Die Menge der Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ bildet einen Vektorraum, der auch abgeschlossen unter Multiplikation ist. Auf dem Untervektorraum $\mathfrak{L}_1 = \mathfrak{L}_1(\Omega, \mathfrak{A}, P)$ der *integrierbaren Zufallsvariablen* erklärt man das Integral

$$\mathrm{E}(X) = \int_\Omega X(\omega)\, dP(\omega),$$

genannt *Erwartungswert* von X. Für $A \in \mathfrak{A}$ gilt $1_A \in \mathfrak{L}_1$ und

$$\mathrm{E}(1_A) = P(A).$$

Die Abbildung $\mathrm{E} : \mathfrak{L}_1 \to \mathbb{R}$ ist linear und monoton, d.h. für $X, Y \in \mathfrak{L}_1$ und $\alpha, \beta \in \mathbb{R}$ gilt

- $\mathrm{E}(\alpha X + \beta Y) = \alpha\, \mathrm{E}(X) + \beta\, \mathrm{E}(Y)$,
- $X \leq Y \Rightarrow \mathrm{E}(X) \leq \mathrm{E}(Y)$.

Ferner gilt für Zufallsvariablen X und Y auf $(\Omega, \mathfrak{A}, P)$

- $|X| \leq Y \wedge Y \in \mathfrak{L}_1 \Rightarrow X \in \mathfrak{L}_1$,

- falls $X \geq 0$: $X \in \mathfrak{L}_1 \wedge \mathrm{E}(X) = 0 \Leftrightarrow P(\{X > 0\}) = 0$.

Wir betrachten die Spezialfälle

(D) X diskret, also $P(\{X \in D\}) = 1$ für eine abzählbare Menge, $D \subset \mathbb{R}$

(A) X absolutstetig mit Dichte f_X.

Die folgenden Sätze dienen uns als Definitionen, siehe Irle (2001, Kap. 8) oder Georgii (2007, Abschn. 4.1).

Satz 2. Im Fall (D) gilt $X \in \mathfrak{L}_1$ genau dann, wenn $\sum_{x \in D} |x| \cdot P(\{X = x\}) < \infty$. Gegebenenfalls folgt

$$\mathrm{E}(X) = \sum_{x \in D} x \cdot P(\{X = x\}).$$

Satz 3. Im Fall (A) gilt $X \in \mathfrak{L}_1$ genau dann, wenn $\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx < \infty$. Gegebenenfalls folgt

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

Bemerkung 4. Die Integrierbarkeit von X und gegebenenfalls $\mathrm{E}(X)$ hängt nur von der Verteilung P_X ab.

Satz 5.

$$\begin{aligned}
X \sim \mathbf{B}(n,p) \quad &\Rightarrow \quad \mathrm{E}(X) = n \cdot p \\
X \sim \mathbf{G}(p) \quad &\Rightarrow \quad \mathrm{E}(X) = 1/p \\
X \sim \mathbf{P}(\lambda) \quad &\Rightarrow \quad \mathrm{E}(X) = \lambda \\
X \sim \mathbf{U}([a,b]) \quad &\Rightarrow \quad \mathrm{E}(X) = (a+b)/2 \\
X \sim \mathbf{Exp}(\lambda) \quad &\Rightarrow \quad \mathrm{E}(X) = 1/\lambda \\
X \sim \mathbf{N}(\mu, \sigma^2) \quad &\Rightarrow \quad \mathrm{E}(X) = \mu
\end{aligned}$$

Beweis. Für $X \sim \mathbf{B}(1,p)$ gilt

$$\mathrm{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p.$$

Für $X \sim \mathbf{B}(n,p)$ könne wir wegen Bemerkung 4 und Satz III.4.8 oBdA annehmen, daß

$$X = \sum_{i=1}^{n} X_i$$

mit $X_1, \dots, X_n$ iid und $X_1 \sim \mathbf{B}(1,p)$. Also

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = n \cdot \mathrm{E}(X_1) = n \cdot p.$$

Beachte, daß hier die Unabhängigkeit nicht verwendet wurde.

Sei f_X die Dichte von $X \sim \mathbf{Exp}(\lambda)$. Dann gilt

$$\int_{-\infty}^{\infty} |x| \cdot f_X(x)\ dx = \int_0^{\infty} x \cdot \lambda \cdot \exp(-\lambda x)\ dx = -x \cdot \exp(-\lambda x)\Big|_0^{\infty} + \int_0^{\infty} \exp(-\lambda x)\, dx$$
$$= \frac{-1}{\lambda} \exp(-\lambda x)\Big|_0^{\infty} = \frac{1}{\lambda}.$$

Siehe ÜBUNG 10:H40 für die restlichen Verteilungen. □

Beispiel 6. Betrachte Rückkehrzeit

$$\tau_\infty := \inf\{t \in \mathbb{N} : S_t = 0\}$$

der symmetrischen Bernoulli-Irrfahrt $(S_t)_{t\in\mathbb{N}_0}$ mit unendlichem Zeithorizont. Es gilt $P(\{\tau_\infty < \infty\}) = 1$, siehe Satz III.5.25, aber $\tau_\infty \notin \mathfrak{L}_1$, siehe Grinstead, Snell (1997, Thm. 12.3) und Korollar III.5.15.

Nun: Hilfsmittel zur Berechnung von Erwartungswerten.

Bezeichnung. $(\Omega_j)_{j\in J}$ *abzählbare Partition* von Ω, falls

(i) J abzählbar,

(ii) $\Omega_j \in \mathfrak{A}$ für alle $j \in J$,

(iii) $\Omega_j \cap \Omega_\ell = \emptyset$ für $j \neq \ell$,

(iv) $\Omega = \bigcup_{j\in J} \Omega_j$.

Lemma 7. Sei X eine Zufallsvariable auf $(\Omega, \mathfrak{A}, P)$, und sei $(\Omega_j)_{j\in J}$ eine abzählbare Partition von Ω. Gelte

$$\forall j \in J\ \exists c_j \in \mathbb{R}\ \forall \omega \in \Omega_j :\ X(\omega) = c_j.$$

Dann gilt $X \in \mathfrak{L}_1$ genau dann, wenn

$$\sum_{j\in J} |c_j| \cdot P(\Omega_j) < \infty.$$

Gegebenenfalls folgt

$$\mathrm{E}(X) = \sum_{j\in J} c_j \cdot P(\Omega_j).$$

Beweis. Für

$$D := \{c_j : j \in J\}$$

gilt

$$P(\{X \in D\}) = 1.$$

Setze $J_x := \{j \in J : c_j = x\}$ für $x \in D$. Dann

$$\begin{aligned}\sum_{x \in D} |x| \cdot P(\{X = x\}) &= \sum_{x \in D} |x| \cdot \sum_{j \in J_x} P(\Omega_j) = \sum_{x \in D} \sum_{j \in J_x} |c_j| \cdot P(\Omega_j) \\ &= \sum_{j \in J} |c_j| \cdot P(\Omega_j).\end{aligned}$$

Berechnung des Erwartungswertes analog ohne Beträge. □

Bemerkung 8. Satz 2 beruht auf der abzählbaren Partition $(\{X = x\})_{x \in D}$.

Korollar 9. Falls $(\Omega, \mathfrak{A}, P)$ diskret, gilt $X \in \mathfrak{L}_1$ genau dann, wenn

$$\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\{\omega\}) < \infty.$$

Gegebenenfalls folgt

$$\mathrm{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}).$$

Beweis. Wähle $J := \Omega$, $\Omega_j := \{j\}$ und $c_j := X(j)$ in Lemma 7. □

Definition 10. $g : \mathbb{R}^d \to \mathbb{R}$ heißt *Borel-meßbar*, falls

$$\forall A \in \mathfrak{B}_1 : \; g^{-1}(A) \in \mathfrak{B}_d.$$

Bemerkung 11.

(i) Die Menge der Borel-meßbaren Funktionen $\mathbb{R}^d \to \mathbb{R}$ bildet einen Vektorraum, der auch abgeschlossen unter Multiplikation ist.

(ii) Jede stetige Funktion $\mathbb{R}^d \to \mathbb{R}$ ist Borel-meßbar.

(iii) $g = 1_B$ mit $B \in \mathfrak{B}_d$ ist Borel-meßbar.

Siehe Irle (2001, Kap. 7).

Nun: ein Transformationssatz für den Erwartungswert.

Lemma 12. Sei X ein d-dimensionaler Zufallsvektor mit Dichte f_X und $h : \mathbb{R}^d \to \mathbb{R}$ Borel-meßbar. Genau dann gilt $h(X) \in \mathfrak{L}_1$, wenn $\int_{\mathbb{R}^d} |h(x)| \cdot f_X(x)\, dx < \infty$. Gegebenenfalls folgt

$$\mathrm{E}(h(X)) = \int_{\mathbb{R}^d} h(x) \cdot f_X(x)\, dx < \infty.$$

Beweis. Siehe Irle (2001, Satz 8.25). □

Beachte: Lemma 12 gilt ohne die Annahme, daß $h(X)$ absolutstetig verteilt ist.

Satz 13. Sind $X, Y \in \mathfrak{L}_1$ unabhängig, so folgt $X \cdot Y \in \mathfrak{L}_1$ und

$$\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

Beweis. Siehe Irle (2001, Satz 10.16). Hier Beweis unter der zusätzlichen Annahme, daß X, Y diskrete Zufallsvariablen sind.

Zunächst nehmen wir sogar an, daß $X(\Omega)$ und $Y(\Omega)$ abzählbar sind. Dann wird durch

$$\Omega_{(x,y)} = \{(X,Y) = (x,y)\}, \qquad (x,y) \in X(\Omega) \times Y(\Omega),$$

eine abzählbare Partition von Ω definiert. Nach Voraussetzung gilt

$$\sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |x \cdot y| \cdot P(\Omega_{(x,y)}) = \sum_{x \in X(\Omega)} |x| \cdot P(\{X = x\}) \cdot \sum_{y \in Y(\Omega)} |y| \cdot P(\{Y = y\}) < \infty,$$

und Lemma 7 zeigt $X \cdot Y \in \mathfrak{L}_1$. Berechnung des Erwartungswertes analog ohne Beträge.

Etwas allgemeiner gilt $P(\{X \in D\}) = P(\{Y \in D\}) = 1$ für eine abzählbare Menge $D \subset \mathbb{R}$. Setze

$$X' := 1_D(X) \cdot X \quad \text{und} \quad Y' := 1_D(Y) \cdot Y.$$

Dann

$$P(\{(X,Y) \neq (X',Y')\}) \leq P(\{X \notin D\} \cup \{Y \notin D\}) = 0.$$

Also folgt $P_{(X,Y)} = P_{(X',Y')}$ und somit

$$X \cdot Y \in \mathfrak{L}_1 \Leftrightarrow X' \cdot Y' \in \mathfrak{L}_1$$

sowie ggf.

$$\mathrm{E}(X \cdot Y) = \mathrm{E}(X' \cdot Y').$$

Ebenso folgt $\mathrm{E}(X) = \mathrm{E}(X')$ und $\mathrm{E}(Y) = \mathrm{E}(Y')$. Satz IV.3.10 zeigt, daß X' und Y' unabhängig sind. Nach Konstruktion sind $X'(\Omega)$ und $Y'(\Omega)$ abzählbar. Wende den ersten Teil des Beweises an.

Im Spezialfall absolutstetiger Zufallsvariablen X, Y verwendet man den Satz von Fubini, Satz V.2.21.(i) und Lemma 12. □

2 Varianz und Kovarianz

Varianz: „Streuungsmaß" für die Verteilung einer Zufallsvariablen.

Definition 1. X *quadratisch integrierbar*, falls $X^2 \in \mathfrak{L}_1$. Bez.: $\mathfrak{L}_2 = \mathfrak{L}_2(\Omega, \mathfrak{A}, P)$ Menge der quadratisch integrierbaren Zufallsvariablen.

Satz 2. $\mathfrak{L}_2$ ist Untervektorraum von $\mathfrak{L}_1$.

Beweis. Verwende $|X| \leq 1 + X^2$ und $(X+Y)^2 \leq 2\,X^2 + 2\,Y^2$. □

Bemerkung 3. Die quadratische Integrierbarkeit von X und gegebenenfalls $\mathrm{E}(X^2)$ hängt nur von der Verteilung P_X ab.

Satz 4. Im Fall (D) gilt $X \in \mathfrak{L}_2$ genau dann, wenn $\sum_{x \in D} x^2 \cdot P(\{X = x\}) < \infty$. Gegebenenfalls folgt

$$\mathrm{E}(X^2) = \sum_{x \in D} x^2 \cdot P(\{X = x\}).$$

Beweis. Schließe wie im Beweis von Satz 1.13. Wende dabei Lemma 1.7 mit $J := D$, $\Omega_j := \{X = j\}$ und $c_j := j^2$ an. □

Satz 5. Im Fall (A) gilt $X \in \mathfrak{L}_2$ genau dann, wenn $\int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx < \infty$. Gegebenenfalls folgt

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx.$$

Beweis. Wende Lemma 1.12 mit $h(x) := x^2$ an. □

Definition 6. Für $X \in \mathfrak{L}_2$ heißt

$$\mathrm{Var}(X) := \mathrm{E}(X - \mathrm{E}(X))^2$$

die *Varianz* und $\sqrt{\mathrm{Var}(X)}$ die *Standardabweichung* von X.

Nun: Abschätzung für die Konzentration einer Zufallsvariable um ihren Erwartungswert.

Bemerkung 7. Für $X \in \mathfrak{L}_2$ gilt

$$\mathrm{Var}(X) = 0 \Leftrightarrow P(\{X = \mathrm{E}(X)\}) = 1.$$

Satz 8 (Tschebyschev-Ungleichung)**.** Für $X \in \mathfrak{L}_2$ und $\varepsilon > 0$ gilt

$$P(\{|X - \mathrm{E}(X)| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}(X).$$

Beweis. Für $A := \{|X - \mathrm{E}(X)| \geq \varepsilon\} \in \mathfrak{A}$ gilt

$$\varepsilon^2 \cdot 1_A \leq (X - \mathrm{E}(X))^2 \cdot 1_A \leq (X - \mathrm{E}(X))^2.$$

Es folgt

$$\varepsilon^2 \cdot P(A) = \varepsilon^2 \cdot \mathrm{E}(1_A) \leq \mathrm{E}(X - \mathrm{E}(X))^2.$$

□

Satz 9. Für $X \in \mathfrak{L}_2$ und $\alpha, \beta \in \mathbb{R}$ gilt

(i) $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$,

(ii) $\mathrm{Var}(\alpha \cdot X + \beta) = \alpha^2 \cdot \mathrm{Var}(X)$.

Beweis. Ad (i): Es gilt

$$(X - \mathrm{E}(X))^2 = X^2 - 2 \cdot X \cdot \mathrm{E}(X) + (\mathrm{E}(X))^2.$$

Es folgt

$$\mathrm{E}(X - \mathrm{E}(X))^2 = \mathrm{E}(X^2) - 2 \cdot (\mathrm{E}(X))^2 + (\mathrm{E}(X))^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.$$

Ad (ii): Es gilt

$$\alpha \cdot X + \beta - \mathrm{E}(\alpha \cdot X + \beta) = \alpha \cdot (X - \mathrm{E}(X)).$$

Es folgt

$$\mathrm{Var}(\alpha \cdot X + \beta) = \mathrm{E}(\alpha^2 \cdot (X - \mathrm{E}(X))^2) = \alpha^2 \cdot \mathrm{Var}(X).$$

□

Bemerkung 10. Für $X, Y \in \mathfrak{L}_2$ gilt $X \cdot Y \in \mathfrak{L}_1$. Zum Beweis verwende man $|X \cdot Y| \leq X^2 + Y^2$.

Definition 11. Betrachte $X, Y \in \mathfrak{L}_2$.

(i) Die *Kovarianz* von X und Y ist definiert durch

$$\mathrm{Cov}(X, Y) = \mathrm{E}((X - \mathrm{E}(X)) \cdot (Y - \mathrm{E}(Y))).$$

(ii) X, Y heißen *unkorreliert*, falls $\mathrm{Cov}(X, Y) = 0$.

(iii) Falls $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$, so ist der *Korrelationskoeffizient* von X und Y definiert durch

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Bemerkung 12.

(i) Für $X, Y \in \mathfrak{L}_2$ gilt

$$\operatorname{Cov}(X, Y) = \mathrm{E}(X \cdot Y) - \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

(ii) $X, Y \in \mathfrak{L}_2$ unabhängig $\Rightarrow X, Y$ unkorreliert, siehe Satz 1.13. Die Umkehrung ist falsch, siehe ÜBUNG 10:G16.

Satz 13 (Formel von Bienaymé)**.** Falls $X_1, \dots, X_n \in \mathfrak{L}_2$ paarweise unkorreliert,

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i).$$

Beweis. Setze $Y_i := X_i - \mathrm{E}(X_i)$ („zentrieren"). Für $i \neq j$ gilt $\mathrm{E}(Y_i \cdot Y_j) = 0$. Also

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^n X_i\right) &= \operatorname{Var}\left(\sum_{i=1}^n Y_i\right) = \mathrm{E}\left(\sum_{i=1}^n Y_i\right)^2 \\ &= \sum_{i,j=1}^n \mathrm{E}(Y_i \cdot Y_j) = \sum_{i=1}^n \mathrm{E}(Y_i^2) = \sum_{i=1}^n \operatorname{Var}(X_i).\end{aligned}$$

□

Beispiel 14. Für $X_1 \sim \mathbf{B}(1, 1/2)$ und $X_2 = -X_1$ gilt $\operatorname{Var}(X_1 + X_2) = 0$ und $\operatorname{Var}(X_1) = \operatorname{Var}(X_2) = 1/4$.

Satz 15.

$$\begin{aligned}X \sim \mathbf{B}(n,p) \quad &\Rightarrow \quad \operatorname{Var}(X) = n \cdot p \cdot (1-p) \\ X \sim \mathbf{G}(p) \quad &\Rightarrow \quad \operatorname{Var}(X) = (1-p)/p^2 \\ X \sim \mathbf{P}(\lambda) \quad &\Rightarrow \quad \operatorname{Var}(X) = \lambda \\ X \sim \mathbf{U}([a,b]) \quad &\Rightarrow \quad \operatorname{Var}(X) = (b-a)^2/12 \\ X \sim \mathbf{Exp}(\lambda) \quad &\Rightarrow \quad \operatorname{Var}(X) = 1/\lambda^2 \\ X \sim \mathbf{N}(\mu, \sigma^2) \quad &\Rightarrow \quad \operatorname{Var}(X) = \sigma^2\end{aligned}$$

Beweis. Für $X \sim \mathbf{B}(1, p)$ gilt

$$\operatorname{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = p - p^2 = p \cdot (1-p).$$

Für $X \sim \mathbf{B}(n, p)$ können wir wegen der Bemerkungen 1.4 und 3 und Satz III.4.8 oBdA annehmen, daß

$$X = \sum_{i=1}^n X_i$$

mit $X_1, \ldots, X_n$ iid und $X_1 \sim \mathbf{B}(1, p)$. Mit Satz 13 folgt

$$\mathrm{Var}(X) = n \cdot p \cdot (1 - p).$$

Für $X \sim \mathbf{Exp}(\lambda)$ gilt

$$\begin{aligned}
&\int_0^\infty x^2 \cdot \lambda \cdot \exp(-\lambda x)\, dx \\
&\qquad = -x^2 \cdot \exp(-\lambda x)\Big|_0^\infty + \int_0^\infty 2x \cdot \exp(-\lambda x)\, dx \\
&\qquad = \frac{2}{\lambda} \cdot \mathrm{E}(X) = \frac{2}{\lambda^2}.
\end{aligned}$$

Demnach gilt $\mathrm{E}(X^2) = \dfrac{2}{\lambda^2}$ und

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Siehe Übung 10:H40 für die restlichen Verteilungen. □

Satz 16 (Cauchy-Schwarzsche-Ungleichung)**.** Für $X, Y \in \mathfrak{L}_2$ gilt

$$|\,\mathrm{E}(X \cdot Y)| \leq \sqrt{\mathrm{E}(X^2) \cdot \mathrm{E}(Y^2)}.$$

Beweis. Übung 10:H42. □

Bemerkung 17. Für $X, Y \in \mathfrak{L}_2$ mit $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$ gilt

$$-1 \leq \rho(X, Y) \leq 1,$$

Satz 18. Für $X, Y \in \mathfrak{L}_2$ mit $\mathrm{Var}(X) > 0$ seien

$$b^* = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}, \qquad a^* = \mathrm{E}(Y) - b^* \cdot \mathrm{E}(X).$$

Dann

$$\mathrm{E}(Y - (a^* + b^* \cdot X))^2 = \mathrm{Var}(Y) \cdot (1 - \rho^2(X, Y)) \tag{1}$$

und

$$\mathrm{E}(Y - (a^* + b^* \cdot X))^2 \leq \mathrm{E}(Y - (a + b \cdot X))^2 \tag{2}$$

für alle $a, b \in \mathbb{R}$.

Beweis. Übung 10:H43. □

Bemerkung 19.

(i) Interpretation von (2): $a^* + b^* \cdot X$ ist die beste lineare Vorhersage von Y bei Beobachtung von X bzgl. des $\mathfrak{L}_2$-Abstandes.

(ii) Interpretation von (1): $\rho^2(X, Y)$ und $\mathrm{sgn}(\rho(X, Y))$ geben den Grad und die Richtung des linearen Zusammenhang von X und Y an. Extremfälle:

$$|\rho(X, Y)| = 1 \Leftrightarrow P(\{Y = a^* + b^* \cdot X\}) = 1,$$

und $\rho(X, Y) = 0$ gilt genau dann, wenn X nicht in die beste lineare Vorhersage eingeht.

(iii) Interpretation von Satz 18: orthogonale Projektion von Y auf den von 1 und X erzeugten Unterraum des Hilbertraumes L_2.

Kapitel VII

Grenzwertsätze

In diesem Kapitel sei $(X_i)_{i\in\mathbb{N}}$ ein Folge von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$. Wir studieren das Verhalten der Partialsummen

$$S_n := \sum_{i=1}^{n} X_i$$

für große Werte n. Untersucht wird dazu die Konvergenz geeignet normierter Partialsummenfolgen. Als Standard-Annahme sei $(X_i)_{i\in\mathbb{N}}$ iid und $X_1 \in \mathfrak{L}_2$ mit $\mathrm{Var}(X_1) > 0$. Schwächere Annahmen werden ggf. explizit genannt.

Beispiel 1. Das einfachste Beispiel ist durch $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \sim \mathbf{B}(1,p)$ für $p \in \,]0,1[$ gegeben. Die Transformation $\widetilde{X}_i = 2 \cdot X_i - 1$ liefert wiederum eine iid-Folge und führt auf die *einfache Irrfahrt*

$$\widetilde{S}_n = \sum_{i=1}^{n} \widetilde{X}_i = 2 \cdot S_n - n$$

mit Parameter p. Hierdurch Modellierung der akkumulierten Gewinne bei einfachem Spiel, speziell $p = 19/37 = 0.5135\ldots$: Roulette aus Sicht des Kasinos. In Abschnitt III.5 wurde der Fall $p = 1/2$ analysiert.

1 Schwaches Gesetz der großen Zahlen

Wir studieren zunächst die Konvergenz des arithmetischen Mittels S_n/n gegen $\mathrm{E}(X_1)$.

Bemerkung 1. Sei $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \sim \mathbf{B}(1, 1/2)$. Dann

$$P(\{S_{2n}/(2n) = \mathrm{E}(X_1)\}) = 2^{-2n} \cdot \binom{2n}{2} \approx 1/\sqrt{\pi\, n},$$

vgl. Korollar III.5.15, so daß insbesondere $\lim_{n\to\infty} P(\{S_{2n}/(2n) = \mathrm{E}(X_1)\}) = 0$. Man untersucht deshalb das asymptotische Verhalten von $P(\{|S_n/n - \mathrm{E}(X_1)| \geq \varepsilon\})$ im Fall $\varepsilon > 0$.

Satz 2 (Schwaches Gesetz der großen Zahlen[1])**.** Sei $(X_i)_{i\in\mathbb{N}}$ paarweise unkorreliert und identisch verteilt. Dann gilt für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P(\{|S_n/n - \mathrm{E}(X_1)| \geq \varepsilon\}) = 0.$$

Beweis. Es gilt $X_i \in \mathfrak{L}_2$ sowie $\mathrm{E}(S_n/n) = \mathrm{E}(X_1)$ und gemäß Satz VI.2.13

$$\mathrm{Var}(S_n) = n \cdot \mathrm{Var}(X_1).$$

Also sichert Satz VI.2.8

$$P(\{|S_n/n - \mathrm{E}(X_1)| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}(S_n/n) = \frac{1}{\varepsilon^2} \cdot \frac{\mathrm{Var}(X_1)}{n}. \qquad (1)$$

□

Bemerkung 3. Was besagt Satz 2 für das Verhalten von S_n? Es gilt

$$\{|S_n/n - \mathrm{E}(X_1)| < \varepsilon\} = \{(\mathrm{E}(X_1) - \varepsilon) \cdot n \leq S_n \leq (\mathrm{E}(X_1) + \varepsilon) \cdot n\}.$$

Also folgt für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P(\{(\mathrm{E}(X_1) - \varepsilon) \cdot n \leq S_n \leq (\mathrm{E}(X_1) + \varepsilon) \cdot n\}) = 1.$$

Beispiel 4. Sei $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \sim \mathbf{B}(1, p)$. Gemäß (1) gilt

$$P(\{|S_n/n - p| \geq \varepsilon\}) \leq \frac{p \cdot (1-p)}{\varepsilon^2 \cdot n} \leq \frac{1}{4 \cdot \varepsilon^2 \cdot n}. \qquad (2)$$

Für großes n kennt man wesentlich bessere Abschätzungen, da die Folge $(X_i)_{i\in\mathbb{N}}$ gleichmäßig beschränkt ist.

Satz 5 (Hoeffdingsche Ungleichung)**.** Sei $(X_i)_{i\in\mathbb{N}}$ iid mit $0 \leq X_i \leq 1$ für alle $i \in \mathbb{N}$. Dann gilt für alle $\varepsilon > 0$ und $n \in \mathbb{N}$

$$P(\{|S_n/n - \mathrm{E}(X_1)| \geq \varepsilon\}) \leq 2 \cdot \exp(-2 \cdot \varepsilon^2 \cdot n).$$

Beweis. Siehe Müller-Gronbach, Novak, Ritter (2008, Kap. 3.4). □

Beispiel 6. Vergleich der Hoeffding-Ungleichung mit (2). Für $\varepsilon = 10^{-2}$ und $n = 10^4$ gilt

$$\frac{1}{4 \cdot \varepsilon^2 \cdot n} = 0.25 \quad \text{und} \quad 2 \cdot \exp(-2 \cdot \varepsilon^2 \cdot n) = 0.270\ldots,$$

und für $\varepsilon = 10^{-2}$ und $n = 10^5$ gilt

$$\frac{1}{4 \cdot \varepsilon^2 \cdot n} = 0.025 \quad \text{und} \quad 2 \cdot \exp(-2 \cdot \varepsilon^2 \cdot n) = 4.12 \cdots 10^{-9}.$$

[1] Jakob Bernoulli (1713), Khintchine (1928)

2 Starkes Gesetz der großen Zahlen

Satz 1.2 macht (unmittelbar) keine Aussage über einzelne Realisierungen der Folge $(S_n/n)_{n\in\mathbb{N}}$. Wir betrachten deshalb nun einen stärkeren Konvergenzbegriff.

Definition 1. Für $A_1, A_2, \dots \in \mathfrak{A}$ heißt

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

der *Limes superior* der Folge $(A_n)_{n\in\mathbb{N}}$.

Bemerkung 2. Es gilt

$$\limsup_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ für unendlich viele } n\},$$

siehe Übung 1:H4.

Lemma 3 (Borel-Cantelli-Lemma)**.**

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 0.$$

Beweis. Setze $B_n := \bigcup_{m=n}^{\infty} A_m$. Satz I.1.17 und die Voraussetzung implizieren

$$P(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} P(B_n) \leq \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) = 0$$

□

Bezeichnung. Eine Eigenschaft gilt für *fast alle* ω bzw. *fast sicher*, falls sie für alle ω aus einer Menge $A \in \mathfrak{A}$ mit $P(A) = 1$ gilt.

Satz 4 (Starkes Gesetz der großen Zahlen[2])**.** Sei $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \in \mathfrak{L}_1$. Dann gilt fast sicher

$$\lim_{n\to\infty} S_n/n = \mathrm{E}(X_1).$$

Beweis. Siehe Irle (2001, Kap. 11) und Vorlesung „Probability Theory“.

Hier unter der zusätzlichen Annahme, daß $X_1 \in \mathfrak{L}_2$. Es genügt den Fall $\mathrm{E}(X_1) = 0$ zu betrachten. Wir zeigen zunächst die Konvergenz der Teilfolge $(S_{n^2}/n^2)_{n\in\mathbb{N}}$. Setze

$$A_{n,\varepsilon} := \{|S_{n^2}/n^2| > \varepsilon\}, \qquad B_\varepsilon := \{\limsup_{n\to\infty} |S_{n^2}/n^2| > \varepsilon\}$$

[2] Borel (1909), Kolmogorov (1930), Etemadi (1981)

für $n \in \mathbb{N}$ und $\varepsilon > 0$. Dann

$$B_\varepsilon \subseteq \limsup_{n\to\infty} A_{n,\varepsilon},$$

und mit (1.1) folgt

$$P(A_{n,\varepsilon}) \leq \frac{\mathrm{Var}(X_1)}{\varepsilon^2} \cdot \frac{1}{n^2}.$$

Lemma 3 sichert $P(B_\varepsilon) = 0$ und mit Satz I.1.17.(ii) ergibt sich

$$P(\{\limsup_{n\to\infty} |S_{n^2}/n^2| > 0\}) = \lim_{\ell\to\infty} P(B_{1/\ell}) = 0,$$

d.h. fast sicher gilt

$$\lim_{n\to\infty} S_{n^2}/n^2 = 0. \tag{1}$$

Für $n \in \mathbb{N}$ definieren wir $k_n \in \mathbb{N}$ durch $k_n^2 \leq n < (k_n + 1)^2$. Wie oben zeigt man

$$P(\{|S_n - S_{k_n^2}| > \varepsilon \cdot k_n^2\}) \leq \frac{\mathrm{Var}(X_1)}{\varepsilon^2} \cdot \frac{n - k_n^2}{k_n^4} \leq \frac{c}{k_n^3}$$

mit einer Konstanten $c > 0$. Wie oben folgt, daß

$$\lim_{n\to\infty} (S_n/k_n^2 - S_{k_n^2}/k_n^2) = 0$$

fast sicher gilt, und zusammen mit (1) ergibt sich hieraus, daß

$$\lim_{n\to\infty} S_n/k_n^2 = 0$$

fast sicher gilt. Schließlich verwende man $\lim_{n\to\infty} n/k_n^2 = 1$. □

Die Abbildungen VII.1–VII.3 zeigen je eine Realisierung der Folge $(S_1, \dots, S_n/n)$ von arithmetischen Mitteln im Fall $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \sim \mathbf{B}(1, 0.7)$.

Bemerkung 5. Was besagt Satz 4 für Folgen von Realisierungen der Zufallsvariablen S_n? Für fast alle $\omega \in \Omega$ gilt

$$\forall \varepsilon > 0 \ \exists n_0(\varepsilon, \omega) \ \forall n \geq n_0(\varepsilon, \omega) : \ (\mathrm{E}(X_1) - \varepsilon) \cdot n \leq S_n(\omega) \leq (\mathrm{E}(X_1) + \varepsilon) \cdot n.$$

Korollar 6. Sei $(X_i)_{i\in\mathbb{N}}$ iid und $B \in \mathfrak{B}_1$. Dann gilt fast sicher

$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{i=1}^{n} 1_B(X_i) = P(\{X_1 \in B\}).$$

Beweis. Für $Z_i := 1_B(X_i)$ gilt $(Z_i)_{i\in\mathbb{N}}$ iid mit $Z_1 \sim \mathbf{B}(1, p)$, siehe ÜBUNG 3:H11. Beachte, daß $\mathrm{E}(Z_1) = P(\{X_1 \in B\})$, und wende Satz 4 an. □

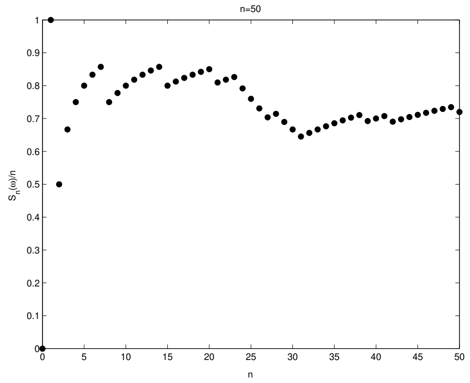

Abbildung VII.1: Realisierung arithmetischer Mittel für $n \leq 50$, $X_1 \sim \mathbf{B}(1, 0.7)$

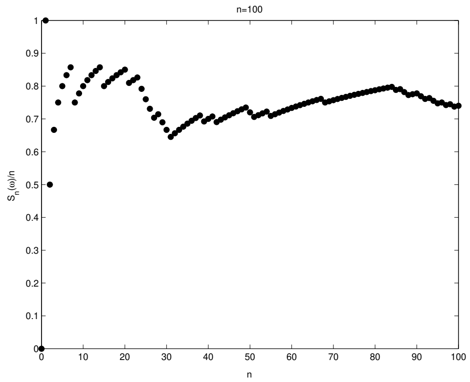

Abbildung VII.2: Realisierung arithmetischer Mittel für $n \leq 100$, $X_1 \sim \mathbf{B}(1, 0.7)$

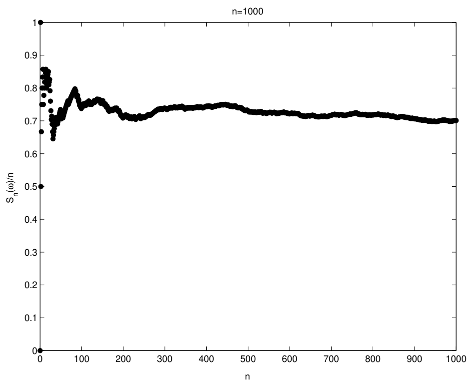

Abbildung VII.3: Realisierung arithmetischer Mittel für $n \leq 1000$, $X_1 \sim \mathbf{B}(1, 0.7)$

Bemerkung 7. Korollar 6 liefert eine theoretische Rechtfertigung der Methode der direkten Simulation, siehe Abschitt II.1. Siehe auch Seite 4 zur frequentistischen Interpretation von Wahrscheinlichkeiten.

Wie Satz 4 zeigt, kann die direkte Simulation ebenso für die approximative Berechnung von Erwartungswerten eingesetzt werden. Gegeben sei eine Zufallsvariable X auf $(\Omega, \mathfrak{A}, P)$. Betrachte eine iid Folge von Zufallsvariablen $X_1', \dots, X_n'$ auf $(\Omega', \mathfrak{A}', P')$, wobei X und X_1' identisch verteilt sind.

1. „Erzeuge" eine Realisierung $(x_1', \dots, x_n')$ von $(X_1', \dots, X_n')$.
2. Approximiere $\mathrm{E}(X)$ durch die das arithmetische Mittel
$$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i'.$$

Beispiel 8. Zur Interpretation der Halbwertszeit: Sei $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \sim \mathbf{Exp}(\lambda)$. Die zugehörige Halbwertszeit ist der Median $h = \ln(2)/\lambda$, siehe Beispiel V.2.10. Gemäß Korollar 6 gilt fast sicher
$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{i=1}^{n} 1_{]-\infty,h]}(X_i) = P(\{X_1 \leq h\}) = \frac{1}{2}.$$

Beispiel 9. Betrachte die Zufallsvariable[3]
$$X := \max\{k \in \mathbb{N}_0 : \sum_{i=1}^{k} Y_i \leq 1\},$$
wobei $(Y_i)_{i\in\mathbb{N}}$ iid mit $Y_1 \sim \mathbf{Exp}(\lambda)$ für $\lambda := 5$. Berechne $\mathrm{E}(X)$. Vgl. ÜBUNG 11:H47. Anwendung: Warteschlangentheorie, Stichwort Poisson-Prozeß. Siehe Abbildungen VII.4–VII.7 zur Approximation von $\mathrm{E}(X)$; die ersten 10 Simulationen lieferten die Werte 7, 1, 4, 6, 7, 6, 8, 5, 11, 5.

Im folgenden sei $(X_i)_{i\in\mathbb{N}}$ iid, und $F := F_{X_1}$ bezeichne die Verteilungsfunktion von X_1. Wir betrachten empirische Verteilungsfunktionen
$$\widehat{F}_n(x;\omega) := F_n(x; X_1(\omega), \dots, X_n(\omega)), \qquad x \in \mathbb{R}, \omega \in \Omega,$$
auf der Basis von Realisierungen von $(X_1, \dots, X_n)$ und untersuchen die Konvergenz von $\widehat{F}_n$ gegen F.

Bemerkung 10. Korollar 6 zeigt für alle $x \in \mathbb{R}$, daß für fast alle ω gilt
$$\lim_{n\to\infty} \widehat{F}_n(x;\omega) = F(x).$$

Verschärfung im folgenden Satz von Glivenko-Cantelli (Hauptsatz der Mathematischen Statistik): fast sicher gleichmäßige Konvergenz.

[3]Konvention: $\max M = \infty$ für unbeschränkte Mengen $M \subseteq \mathbb{N}$.

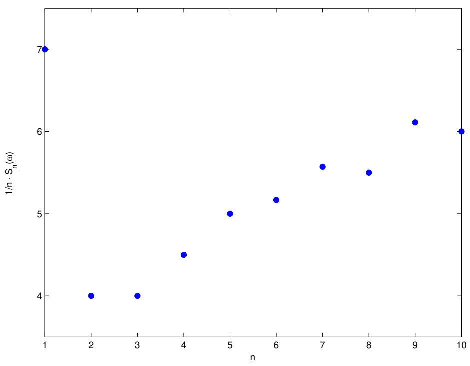

Abbildung VII.4: Realisierung arithmetischer Mittel für $n \leq 10$

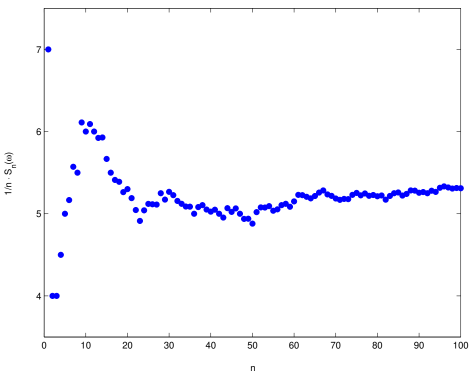

Abbildung VII.5: Realisierung arithmetischer Mittel für $n \leq 100$

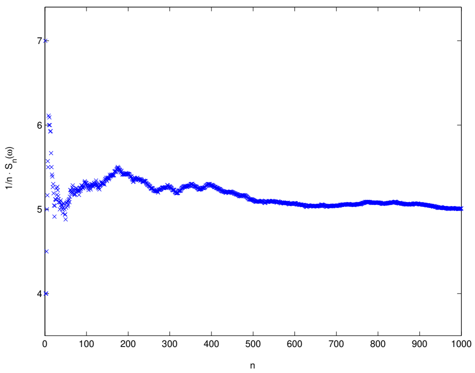

Abbildung VII.6: Realisierung arithmetischer Mittel für $n \leq 1000$

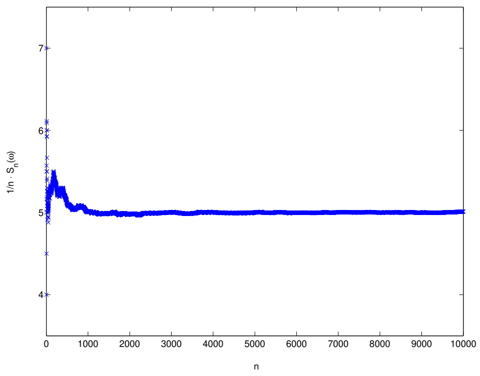

Abbildung VII.7: Realisierung arithmetischer Mittel für $n \leq 10000$

Satz 11 (Satz von Glivenko-Cantelli)**.** Für fast alle ω gilt

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\widehat{F}_n(x;\omega) - F(x)| = 0.$$

Beweis. Man betrachte vorab den Spezialfall einer stetigen und streng monoton wachsenden Verteilungsfunktion F.

Zur Behandlung des allgemeinen Falls schreiben wir

$$G^-(x) := \lim_{y\to x-} G(y)$$

für Funktionen $G : \mathbb{R} \to \mathbb{R}$ mit existierenden linksseitigen Grenzwerten, und wir setzen

$$\mathfrak{E} := \{A \in \mathfrak{A} : P(A) = 1\}.$$

Fixiere $k \in \mathbb{N}$. Betrachte für $\ell = 1, \ldots, k-1$ die ℓ/k-Quantile $q_{\ell,k}$ von F, setze ferner $q_{0,k} := -\infty$ und $q_{k,k} := \infty$. Korollar 6 zeigt

$$\forall \ell \in \{1, \ldots, k-1\}\ \exists A \in \mathfrak{E}\ \forall \omega \in A : \quad \lim_{n\to\infty} \widehat{F}_n^-(q_{\ell,k};\omega) = F^-(q_{\ell,k})$$

und

$$\forall \ell \in \{1, \ldots, k-1\}\ \exists A \in \mathfrak{E}\ \forall \omega \in A : \quad \lim_{n\to\infty} \widehat{F}_n(q_{\ell,k};\omega) = F(q_{\ell,k}).$$

Für

$$\Delta_{n,k}(\omega) = \max_{\ell\in\{0,\ldots,k\}} \max\Big(\big|\widehat{F}_n(q_{\ell,k};\omega) - F(q_{\ell,k})\big|, \big|\widehat{F}_n^-(q_{\ell,k};\omega) - F^-(q_{\ell,k})\big|\Big)$$

folgt

$$\exists A_k \in \mathfrak{E}\ \forall \omega \in A_k : \ \lim_{n\to\infty} \Delta_{n,k}(\omega) = 0. \tag{2}$$

Lemma II.3.3 zeigt

$$F^-(q_{\ell,k}) - F(q_{\ell-1,k}) \leq \ell/k - (\ell-1)/k = 1/k.$$

Hiermit folgt für $x \in]q_{\ell-1,k}, q_{\ell,k}[$ und $\omega \in \Omega$

$$\begin{aligned}\widehat{F}_n(x;\omega) &\leq \widehat{F}_n^-(q_{\ell,k};\omega) \leq F^-(q_{\ell,k}) + \Delta_{n,k}(\omega) \leq F(q_{\ell-1,k}) + 1/k + \Delta_{n,k}(\omega)\\ &\leq F(x) + 1/k + \Delta_{n,k}(\omega)\end{aligned}$$

sowie

$$\begin{aligned}\widehat{F}_n(x;\omega) &\geq \widehat{F}_n(q_{\ell-1,k};\omega) \geq F(q_{\ell-1,k}) - \Delta_{n,k}(\omega) \geq F^-(q_{\ell,k}) - 1/k - \Delta_{n,k}(\omega)\\ &\geq F(x) - 1/k - \Delta_{n,k}(\omega).\end{aligned}$$

Für $\omega \in A_k$ ergibt sich gemäß (2)

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} |\widehat{F}_n(x;\omega) - F(x)| \leq 1/k.$$

Fazit: für

$$A := \bigcap_{k\in\mathbb{N}} A_k \in \mathfrak{A}$$

gilt

$$P(A) = 1$$

und

$$\forall \omega \in A: \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\widehat{F}_n(x;\omega) - F(x)| = 0.$$

$\square$

Die Abbildungen VII.8–VII.16 illustrieren den Satz von Glivenko-Cantelli, wobei jeweils schwarz eine empirische Verteilungsfunktion und blau die Verteilungsfunktion dargestellt sind. Verwendet werden $n = 50$, 250, 1000 Simulationen. Die zugrundeliegende Verteilungen sind gegeben durch

$$P(\{X_1 = 1\}) = 0.3, \quad P(\{X_1 = 2\}) = 0.2, \quad P(\{X_1 = 3\}) = 0.5$$

in den Abbildungen VII.8–VII.10,

$$X_1 \sim \mathbf{U}([0,1])$$

in den Abbildungen VII.11–VII.13 und

$$X_1 \sim \mathbf{Exp}(1/2)$$

in den Abbildungen VII.14–VII.16.

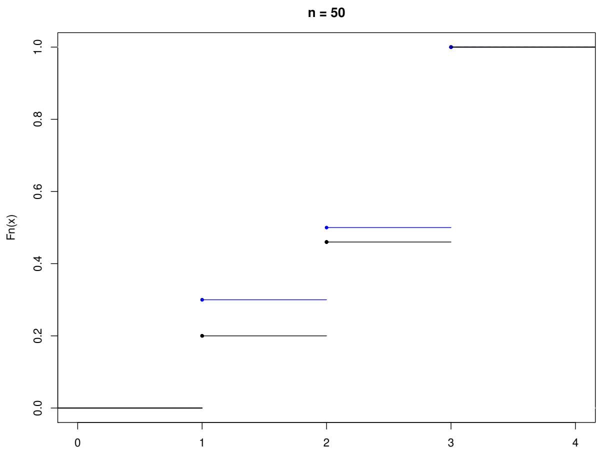

Abbildung VII.8: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 50$

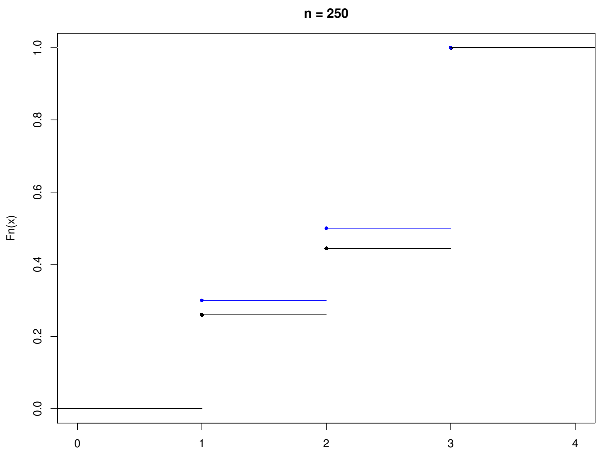

Abbildung VII.9: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 250$

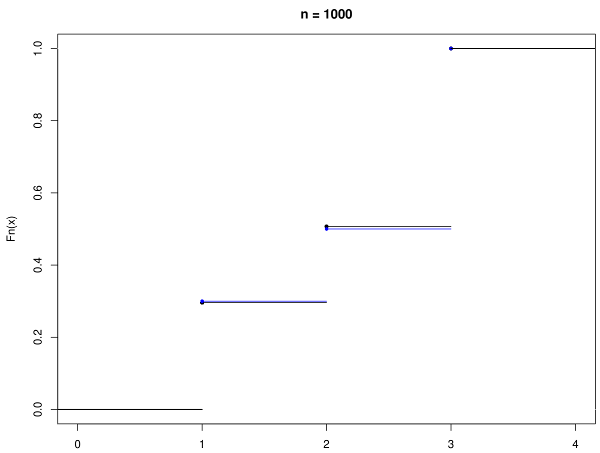

Abbildung VII.10: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 1000$

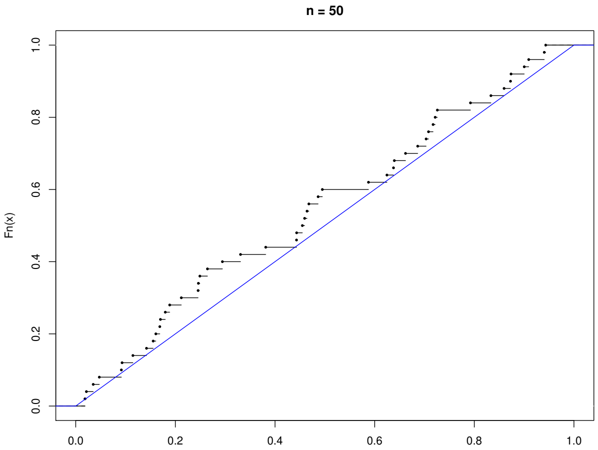

Abbildung VII.11: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 50$

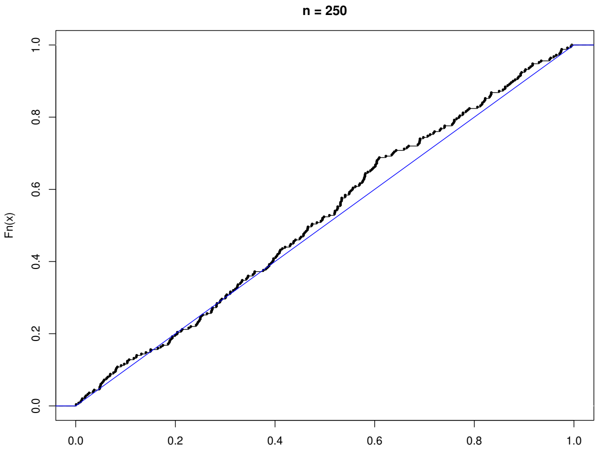

Abbildung VII.12: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 250$

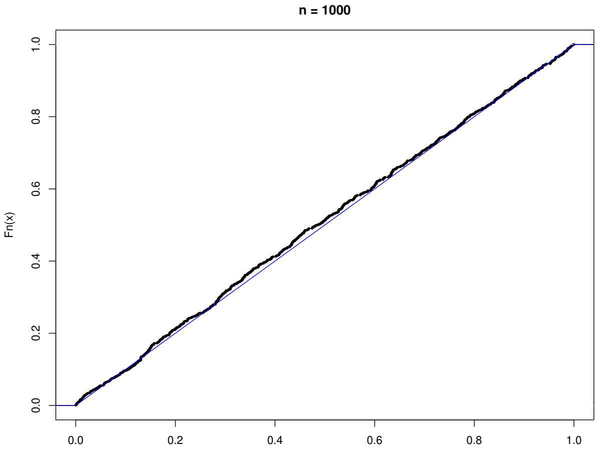

Abbildung VII.13: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 1000$

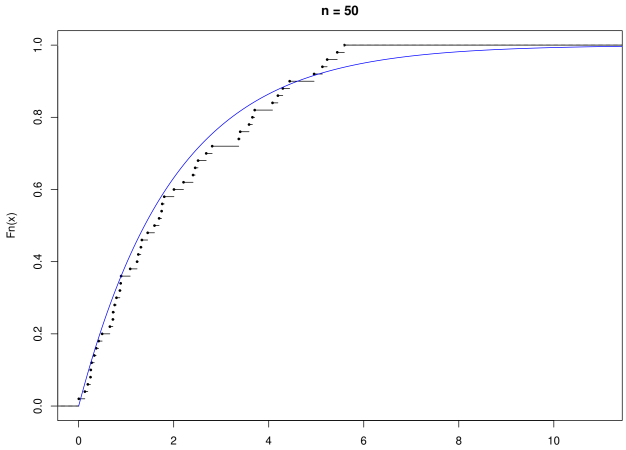

Abbildung VII.14: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 50$

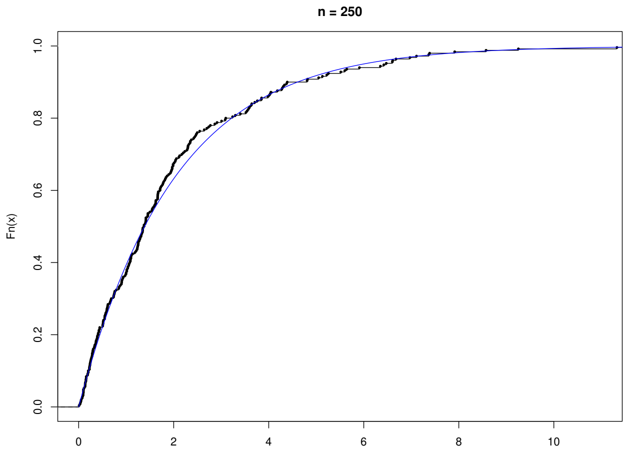

Abbildung VII.15: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 250$

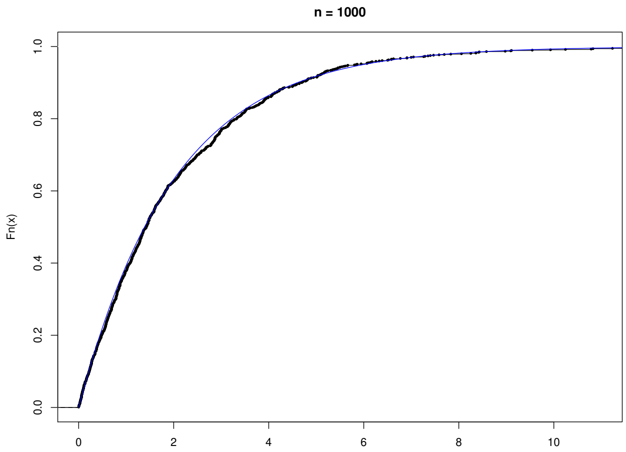

Abbildung VII.16: Verteilungsfunktion und empirische Verteilungsfunktion, $n = 1000$

Abschließend: Vergleich der Konvergenzbegriffe im Starken und Schwachen Gesetz der großen Zahlen. Anwendung des folgenden Resultates mit $Y_n := S_n/n - \mathrm{E}(X_1)$.

Satz 12. Gilt fast sicher

$$\lim_{n\to\infty} Y_n = 0$$

für eine Folge von Zufallsvariablen $(Y_n)_{n\in\mathbb{N}}$ auf $(\Omega, \mathfrak{A}, P)$, so folgt für alle $\varepsilon > 0$

$$\lim_{n\to\infty} P(\{|Y_n| \geq \varepsilon\}) = 0.$$

Beweis. Übung 11:H45. Siehe auch Übung 11:G18. □

3 Zentraler Grenzwertsatz

Beispiel 1. Sei $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \sim \mathbf{B}(1,p)$ für $p \in]0,1[$. Bestimme ein Intervall um $\mathrm{E}(S_n)$, das S_n mit vorgegebener Wahrscheinlichkeit $\alpha \in]0,1[$ enthält.
Es gilt

$$P(\{|S_n - \mathrm{E}(S_n)| \leq \delta\}) = \sum_{k\in K_n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

mit

$$K_n := \{k \in \{0,\dots,n\} : |k - n \cdot p| \leq \delta\}.$$

Für jede Folge von Zahlen $\delta_n > 0$ gilt

$$\lim_{n\to\infty} \delta_n/\sqrt{n} = 0 \Rightarrow \lim_{n\to\infty} P(\{|S_n - \mathrm{E}(S_n)| \leq \delta_n\}) = 0,$$

siehe Übung 11:H46, und, unter den allgemeinen Voraussetzungen von Satz 1.2,

$$\lim_{n\to\infty} \delta_n/\sqrt{n} = \infty \Rightarrow \lim_{n\to\infty} P(\{|S_n - \mathrm{E}(S_n)| \leq \delta_n\}) = 1.$$

Dies motiviert die Untersuchung des Falls $\delta_n = c \cdot \sqrt{n}$ mit einer Konstanten $c > 0$.
Analog für einfache Irrfahrten, siehe Abbildungen VII.17–VII.21.

Im folgenden bezeichne φ die Dichte der Standard-Normalverteilung und Φ die zugehörige Verteilungsfunktion.

Satz 2 (Grenzwertsatz von de Moivre-Laplace)**.** Sei $(X_i)_{i\in\mathbb{N}}$ iid mit $X_1 \sim \mathbf{B}(1,p)$ für $p \in]0,1[$. Setze

$$\sigma_n := \sqrt{n \cdot p \cdot (1-p)}.$$

Dann gilt für alle $u < v$

$$\lim_{n\to\infty} P(\{u \cdot \sigma_n \leq S_n - \mathrm{E}(S_n) \leq v \cdot \sigma_n\}) = \Phi(v) - \Phi(u).$$

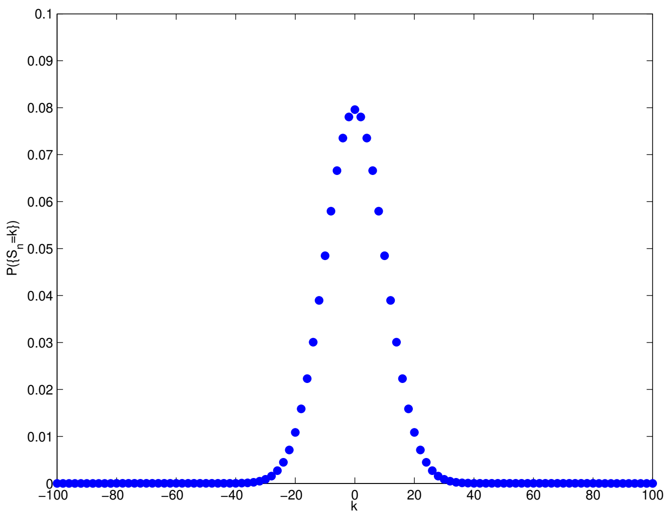

Abbildung VII.17: Symmetrische Bernoulli-Irrfahrt, Wahrscheinlichkeitsfunktion von S_n für $n = 100$

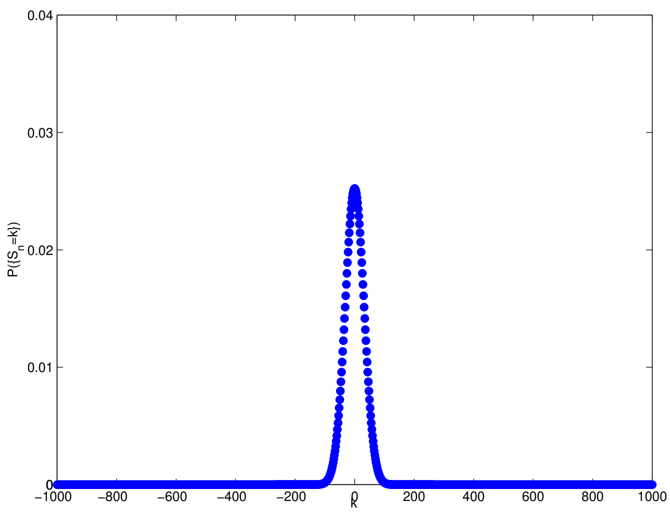

Abbildung VII.18: Symmetrische Bernoulli-Irrfahrt, Wahrscheinlichkeitsfunktion von S_n für $n = 1000$

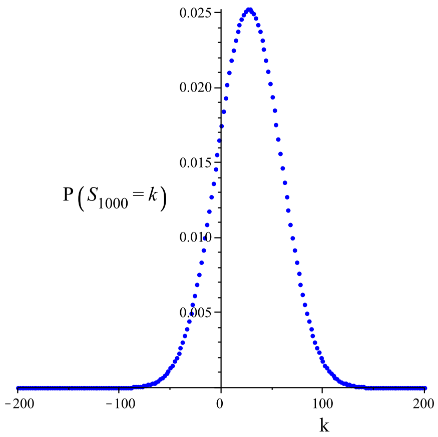

Abbildung VII.19: Einfache Irrfahrt, $p = 19/37$, Wahrscheinlichkeitsfunktion von S_n für $n = 1000$

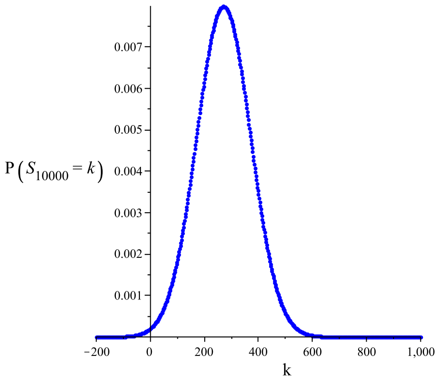

Abbildung VII.20: Einfache Irrfahrt, $p = 19/37$, Wahrscheinlichkeitsfunktion von S_n für $n = 10\,000$

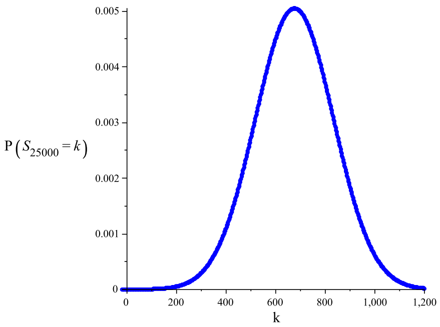

Abbildung VII.21: Einfache Irrfahrt, $p = 19/37$, Wahrscheinlichkeitsfunktion von S_n für $n = 25\,000$

Beweis. Vgl. Beweis von Satz III.5.19. Fixiere $u < v$ und setze

$$b_{n,k} := \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

und

$$x_{n,k} := \frac{k - n \cdot p}{\sigma_n}$$

für $k \in \{0, \ldots, n\}$ sowie

$$K_n := \{k \in \{0, \ldots, n\} : u \leq x_{n,k} \leq v\}.$$

Klar

$$P(\{u \cdot \sigma_n \leq S_n - \mathrm{E}(S_n) \leq v \cdot \sigma_n\}) = \sum_{k \in K_n} b_{n,k}.$$

Bestimme die Asymptotik von $b_{n,k}$ gleichmäßig für $k \in K_n$. Lemma III.5.14 sichert

$$\lim_{n\to\infty} \max_{k \in K_n} \left| \frac{\sigma_n \cdot b_{n,k}}{c_{n,k}} - 1 \right| = 0$$

mit

$$c_{n,k} := \frac{1}{\sqrt{2\pi}} \cdot \left(\frac{n \cdot p}{k}\right)^k \cdot \left(\frac{n \cdot (1-p)}{n-k}\right)^{n-k}.$$

Setze

$$h(y) := y \cdot \ln(y/p) + (1-y) \cdot \ln((1-y)/(1-p)), \qquad y \in \left]0,1\right[.$$

Dann

$$\sqrt{2\pi} \cdot c_{n,k} = \exp(-n \cdot h(k/n)).$$

Nun: Taylor-Entwicklung von h um p. Man verifiziert $h(p) = h'(p) = 0$ und $h''(p) = 1/(p \cdot (1-p))$, so daß

$$\max_{k \in K_n} \left| h(k/n) - \frac{x_{n,k}^2}{2 \cdot n} \right| \leq M \cdot n^{-3/2}$$

mit einer Konstanten $M > 0$ gilt. Es folgt

$$\lim_{n\to\infty} \max_{k \in K_n} \left| \frac{c_{n,k}}{\varphi(x_{n,k})} - 1 \right| = 0.$$

Fazit:

$$\lim_{n\to\infty} \max_{k \in K_n} \left| \frac{\sigma_n \cdot b_{n,k}}{\varphi(x_{n,k})} - 1 \right| = 0. \tag{1}$$

Ferner gilt

$$\lim_{n\to\infty} \sum_{k \in K_n} \frac{1}{\sigma_n} \cdot \varphi(x_{n,k}) = \Phi(v) - \Phi(u) \tag{2}$$

aufgrund der Stetigkeit von φ.

Aus (1) und (2) folgt die Behauptung. □

Bemerkung 3. Gleichung (1) besagt, daß reskalierte Gewichte von Binomialverteilungen lokal gleichmäßig gegen die Normalverteilungsdichte konvergieren.

Beispiel 4. Fortsetzung von Beispiel 1. Für $\delta_n := q \cdot \sigma_n$ gilt aufgrund von Satz 2

$$\lim_{n\to\infty} P(\{|S_n - \mathrm{E}(S_n)| \leq \delta_n\}) = \Phi(q) - \Phi(-q) = 2 \cdot \Phi(q) - 1.$$

Wähle

$$q := \Phi^{-1}((1+\alpha)/2)$$

als $(1+\alpha)/2$-Quantil der Standard-Normalverteilung. Approximativ leistet dann

$$\delta_n := q \cdot \sqrt{p \cdot (1-p)} \cdot \sqrt{n}$$

das Verlangte, d.h.

$$\lim_{n\to\infty} P(\{|S_n - \mathrm{E}(S_n)| \leq \delta_n\}) = \alpha. \tag{3}$$

Beispielsweise gilt $q = 1.645\ldots$ für $\alpha = 0.9$, $q = 2.576\ldots$ für $\alpha = 0.99$ und $q = 3.291\ldots$ für $\alpha = 0.999$.

Ebenso gilt (3) für die einfache Irrfahrt $(S_n)_{n\in\mathbb{N}}$ mit der Wahl

$$\delta_n := q \cdot 2\sqrt{p \cdot (1-p)} \cdot \sqrt{n}.$$

Für $n = 10^6$ Runden des einfachen Spiels beim Roulette, d.h. für $p = 19/37$, erhalten wir somit approximativ, daß

$$23\,737 \leq S_n \leq 30\,316$$

mit der Wahrscheinlichkeit 0.999 gilt.

Den Sätzen III.4.12, III.5.19, 2 sowie Übung 6:G8 liegt ein gemeinsamer Konvergenzbegriff zugrunde.

Bezeichnung. $\mathfrak{C}_Z = \{x \in \mathbb{R} : F_Z \text{ stetig in } x\}$ *Menge der Stetigkeitspunkte* der Verteilungsfunktion F_Z von Z.

Beispiel 5. Für $Z \sim \mathbf{U}([a,b])$ und $Z \sim \mathbf{N}(\mu, \sigma^2)$ gilt $\mathfrak{C}_Z = \mathbb{R}$. Für $Z \sim \mathbf{B}(n,p)$ mit $p \in \,]0,1[$ gilt $\mathfrak{C}_Z = \mathbb{R} \setminus \{0,\ldots,n\}$.

Definition 6. Eine Folge $(Z_n)_{n\in\mathbb{N}}$ von Zufallsvariablen *konvergiert in Verteilung* gegen eine Zufallsvariable Z, falls

$$\forall x \in \mathfrak{C}_Z : \lim_{n\to\infty} F_{Z_n}(x) = F_Z(x)$$

Bez.: $Z_n \xrightarrow{\mathrm{d}} Z$

Beispiel 7. Für $Z_n = a_n$ und $Z = a$ mit $a, a_n \in \mathbb{R}$ gilt

$$Z_n \xrightarrow{\mathrm{d}} Z \Leftrightarrow \lim_{n\to\infty} a_n = a.$$

Beachte: Aus $a_n > a$ folgt $F_{Z_n}(a) = 0$, während $F_Z(a) = 1$.

Bemerkung 8. Es existiert eine Metrik ρ auf der Menge der Wahrscheinlichkeitsmaße auf $\mathfrak{B}_1$, so daß

$$Z_n \xrightarrow{\mathrm{d}} Z \Leftrightarrow \lim_{n\to\infty} \rho(P_{Z_n}, P_Z) = 0,$$

siehe Vorlesung „Probability Theory". In diesem Sinn beschreibt die Verteilungskonvergenz die Approximation stochastischer Modelle.

Lemma 9. Gelte $P(\{Z \in \mathbb{Z}\}) = P(\{Z_n \in \mathbb{Z}\}) = 1$ für alle $n \in \mathbb{N}$. Dann

$$Z_n \xrightarrow{\mathrm{d}} Z \Leftrightarrow \forall k \in \mathbb{Z}: \lim_{n\to\infty} P(\{Z_n = k\}) = P(\{Z = k\}).$$

Beweis. Siehe Übung 8:H33. □

Beispiel 10. Zum rein diskreten Fall:

(i) Gelte $Z_n \sim \mathbf{B}(n, p_n)$ mit $\lim_{n\to\infty} n \cdot p_n = \lambda > 0$. Dann $Z_n \xrightarrow{\mathrm{d}} Z$, falls $Z \sim \mathbf{P}(\lambda)$. Siehe Satz III.4.12.

(ii) Gelte $Z_n \sim \mathbf{H}(n, n_0(n), k)$ mit $n_0(n) \in \{1, \ldots, n\}$ und $\lim_{n\to\infty} n_0(n)/n = p \in]0,1[$. Dann $Z_n \xrightarrow{\mathrm{d}} Z$, falls $Z \sim \mathbf{B}(k, p)$. Siehe Übung 6:G8

Beispiel 11. Zum Übergang diskret/kontinuierlich:

(i) Sei X_{2n} die Anzahl der Führungszeitpunkte bei einer symmetrischen Bernoulli-Irrfahrt der Länge $2n$ und sei

$$Z_{2n} = \frac{X_{2n}}{2n}.$$

Dann $Z_n \xrightarrow{\mathrm{d}} Z$, falls Z arcussinus-verteilt. Siehe Satz III.5.19.

(ii) Sei $S_n \sim \mathbf{B}(n, p)$ mit $p \in]0,1[$ und

$$Z_n := \frac{S_n - n \cdot p}{\sigma_n}.$$

Dann $Z_n \xrightarrow{\mathrm{d}} Z$, falls $Z \sim \mathbf{N}(0,1)$. Siehe Satz 2.

Lemma 12. Sei F_Z stetig und gelte $Z_n \xrightarrow{\mathrm{d}} Z$. Dann

(i) $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_{Z_n}(x) - F_Z(x)| = 0$,

(ii) $\forall A \in \mathfrak{M} : \lim_{n\to\infty} P(\{Z_n \in A\}) = P(\{Z \in A\})$.

Beweis. Ad (i): $\boxed{\text{Übung}}$.
Ad (ii): Gilt nach Definition für $A =]-\infty, x]$. Für $A = \{x\}$ und $\varepsilon > 0$

$$P(\{Z_n \in A\}) \leq P(\{x - \varepsilon < Z_n \leq x\}) = P(\{Z_n \leq x\}) - P(\{Z_n \leq x - \varepsilon\}).$$

Somit

$$0 \leq \limsup_{n\to\infty} P(\{Z_n \in A\}) \leq F_Z(x) - F_Z(x - \varepsilon).$$

Aufgrund der Stetigkeit von F_Z in x folgt

$$\lim_{n\to\infty} P(\{Z_n \in A\}) = 0 = P(\{Z \in A\}).$$

$\square$

Bemerkung 13. Die Aussage von Lemma 12.(ii) gilt im allgemeinen nicht für alle $A \in \mathfrak{B}_1$, wie Beispiel 11 zeigt.

Satz 2 erweist sich als Spezialfall einer Konvergenzaussage, die unter sehr allgemeinen Voraussetzungen gilt.

Satz 14 (Zentraler Grenzwertsatz)**.** Sei $(X_i)_{i\in\mathbb{N}}$ iid und $X_1 \in \mathfrak{L}_2$ mit

$$\sigma := \sqrt{\mathrm{Var}(X_1)} > 0.$$

Setze $\mu := \mathrm{E}(X_1)$ sowie

$$S_n^* := \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}}.$$

Dann gilt $S_n^* \xrightarrow{\mathrm{d}} Z$, falls $Z \sim \mathbf{N}(0, 1)$.

Beweis. Irle (2001, Kap. 12) und Vorlesung „Probability Theory“. $\square$

Bezeichnung. Die Zufallsvariablen S_n^* heißen *standardisierte Summenvariablen*.

Bemerkung 15. Es gilt $\mathrm{E}(S_n^*) = 0$ und $\mathrm{Var}(S_n^*) = 1$ sowie

$$S_n^* = \sum_{i=1}^n \underbrace{\frac{X_i - \mathrm{E}(X_i)}{\sigma \cdot \sqrt{n}}}_{=:X_i^*}$$

mit $\mathrm{E}(X_i^*) = 0$ und $\mathrm{Var}(X_i^*) = 1/n$. Der zentrale Grenzwertsatz besagt also grob: „Ein Gesamteffekt, der Summe vieler kleiner zentrierter unabhängiger Einzeleffekte ist, ist näherungsweise normalverteilt.“

Die Abbildungen VII.22 und VII.23 illustrieren den Zentralen Grenzwertsatz für Summen exponentialverteilter Zufallsvariablen.

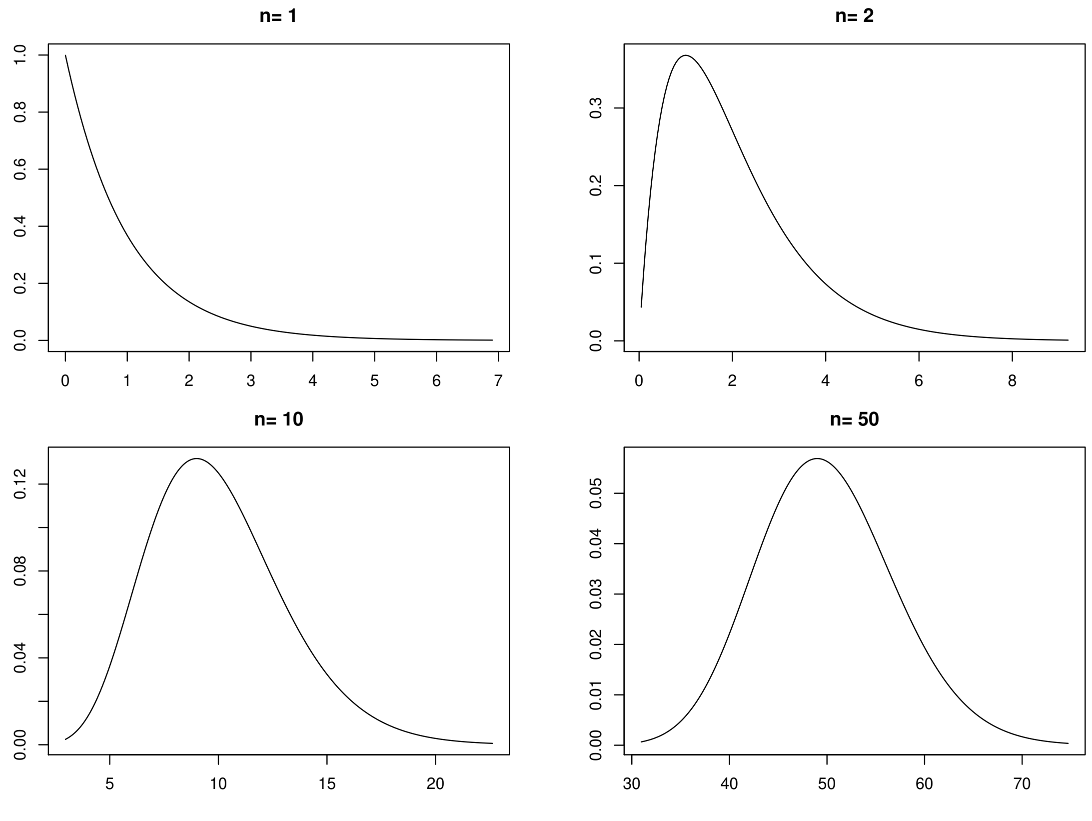

Abbildung VII.22: Dichten von Summen n unabhängiger **Exp**(1)-verteilter Zufallsvariablen

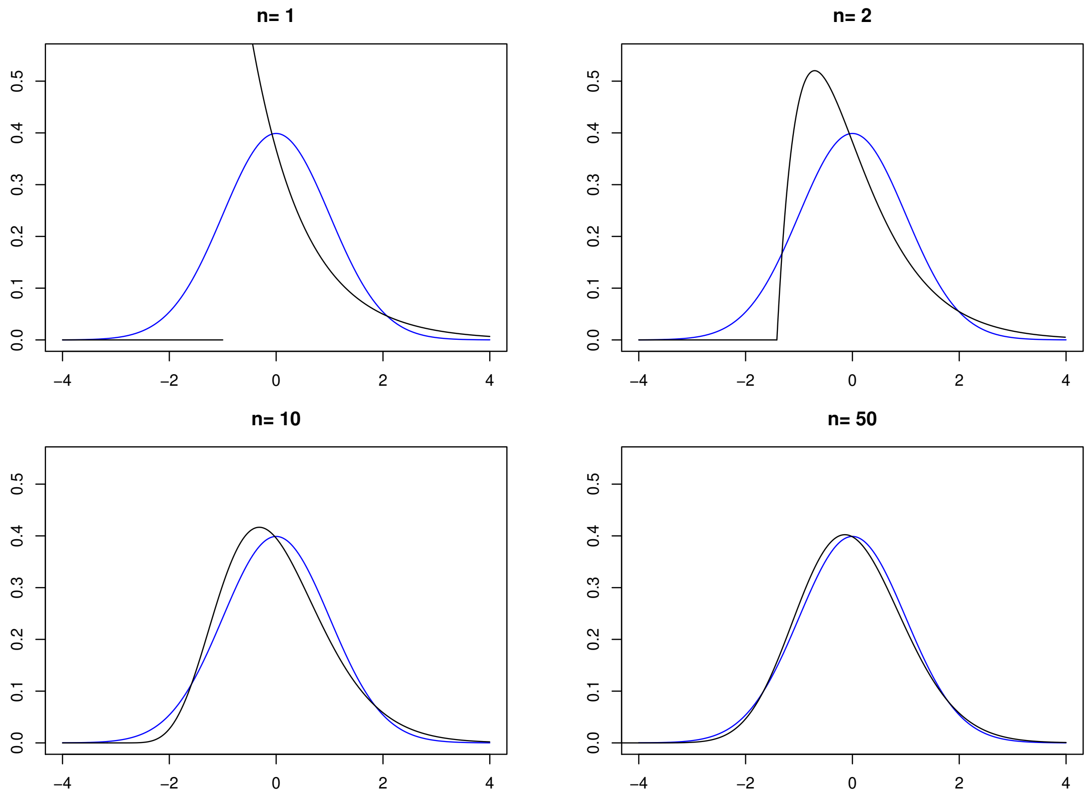

Abbildung VII.23: Dichten von standardisierten Summen n unabhängiger **Exp**(1)-verteilter Zufallsvariablen

Beispiel 16. Überbuchung von Flugverbindungen:

- Kapazität $K \in \mathbb{N}$
- Buchungsanzahl $n \in \mathbb{N}$, wobei die Passagiere unabhängig jeweils mit Wahrscheinlichkeit $p \in \,]0,1[$ erscheinen

Gegeben $\alpha \in \,]0,1[$. Bestimme $n \in \mathbb{N}$, so daß eine Überbuchung ungefähr mit Wahrscheinlichkeit α auftritt.
Modell: $X_1, \ldots, X_n$ iid, $X_1 \sim \mathbf{B}(1,p)$.
Es gilt $\Big\{\sum_{i=1}^n X_i > K\Big\} = \{S_n^* > c_n\}$ mit

$$S_n^* := \sum_{i=1}^n (X_i - p)/\sqrt{n \cdot p \cdot (1-p)},$$
$$c_n := (K - n \cdot p)/\sqrt{n \cdot p \cdot (1-p)}.$$

Somit ist c_n näherungsweise durch $c_n = \Phi^{-1}(1-\alpha)$ gegeben.
Für $K := 1000$, $p := 0.9$ und $\alpha := 0.01$ ergibt sich näherungsweise $c_n = 2.33$ und

$$n = 1086.$$

Hiermit gilt für die erwartete Anzahl nicht beförderter Passagiere

$$\mathrm{E}\Big(\max(\sum_{i=1}^n X_i - K, 0)\Big) \leq 86 \cdot P\Big(\Big\{\sum_{i=1}^n X_i > K\Big\}\Big) \doteq 0.86.$$

Zum Vergleich die exakten Werte

$$\mathrm{E}\Big(\max(\sum_{i=1}^n X_i - K, 0)\Big) = 0.0287\ldots,$$
$$P\Big(\Big\{\sum_{i=1}^n X_i > K\Big\}\Big) = 0.00811\ldots,$$
$$86 \cdot P\Big(\Big\{\sum_{i=1}^n X_i > K\Big\}\Big) = 0.697\ldots$$

Kapitel VIII

Schließende Statistik

1 Statistische Modellbildung und statistisches Entscheiden

Grundform wichtiger statistischer Fragestellungen:

(i) Zufallsexperiment mit unbekannter Verteilung Q.

(ii) Verteilungsannahme: $Q \in \mathfrak{P}$ für eine Menge $\mathfrak{P}$ von Wahrscheinlichkeitsmaßen auf $\mathfrak{B}_d$.

(iii) (a) Schätzproblem. Gegeben: Abbildung $\eta : \mathfrak{P} \to \mathbb{R}$. Bestimme $\eta(Q)$, einen „Aspekt" der Verteilung Q.

(b) Testproblem. Gegeben $\emptyset \neq \mathfrak{P}_0 \subsetneq \mathfrak{P}$. Entscheide, ob $Q \in \mathfrak{P}_0$.

(iv) Verfügbar: Stichprobe $(x_1, \dots, x_n) \in \mathbb{R}^{n \cdot d}$ aus n-maliger unabhängiger Wiederholung des Zufallsexperimentes.

Bezeichnung. Fortan bezeichnen wir mit $\mathbf{B}(n,p)$, $\mathbf{N}(\mu,\sigma^2)$, ... auch die entsprechenden Wahrscheinlichkeitsmaße auf $\mathfrak{B}_1$.

Beispiel 1. Geschlecht eines Neugeborenen ($1 \triangleq$ W, $0 \triangleq$ M), siehe Beispiel I.1.13. Hier $d = 1$ und

(i) $Q = \mathbf{B}(1,p)$, wobei p die Wahrscheinlichkeit, daß Neugeborenes weiblich.

(ii) $\mathfrak{P} := \{\mathbf{B}(1,p) : p \in]0,1[\}$.

(iii) (a) $\eta(\mathbf{B}(1,p)) := p$.

(b) $\mathfrak{P}_0 := \{\mathbf{B}(1,1/2)\}$ oder $\mathfrak{P}_0 := \{\mathbf{B}(1,p) : p < 1/2\}$.

(iv) Geschlecht bei n Lebendgeburten.

Bei einem Stichprobenumfang von $n = 25\,171\,123$ scheint eine „verläßliche“ Bestimmung von $\eta(Q)$ und eine Entscheidung, ob $\eta(Q) \in \mathfrak{P}_0$, möglich.
Das empirische Mittel

$$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{12\,241\,392}{25\,171\,123} = 0.4863\,\ldots$$

legt nahe, daß $\eta(Q)$ ungefähr 0.48 beträgt und daß $Q \neq \mathbf{B}(1, 1/2)$ gilt. Siehe Hesse (2003, p. 23).

Bemerkung 2. Studiert werden auch Varianten dieser Grundform, z. Bsp. abhängige Beobachtungen oder $\mathbb{R}^k$-wertige Abbildungen η. Die Modellierung und Analyse solcher Fragestellungen ist Gegenstand der Mathematischen Statistik (Gütekriterien, Optimalitätsaussagen, Quantifizierung von Risiken).

Bemerkung 3. Oft ist $\mathfrak{P}$ in natürlicher Weise parametrisiert, siehe Beispiel 1.

Fortan betrachten wir der Einfachheit halber den Fall $d = 1$.

Definition 4. Ein *statistisches Experiment* mit *Parameterraum* Θ und *Stichprobenraum* $\mathbb{R}^n$ ist gegeben durch

(i) eine Familie $(\Omega, \mathfrak{A}, P^\vartheta)_{\vartheta \in \Theta}$ von Wahrscheinlichkeitsräumen,

(ii) einen Zufallsvektor $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ mit folgenden Eigenschaften. Für alle $\vartheta \in \Theta$ ist

$$(X_1, \ldots, X_n) \text{ unabhängig bzgl. } P^\vartheta$$

und

$$P^\vartheta_{X_1} = \ldots = P^\vartheta_{X_n}.$$

Ferner ist $P^\vartheta_{X_1} \neq P^{\vartheta'}_{X_1}$ für alle $\vartheta, \vartheta' \in \Theta$ mit $\vartheta \neq \vartheta'$.

Bemerkung 5. In vorliegenden Kontext beschreibt die Menge

$$\mathfrak{P} = \{P^\vartheta_{X_1} : \vartheta \in \Theta\}$$

die sogenannte *Verteilungsannahme*. Die konkrete Gestalt von Ω, $\mathfrak{A}$ und den Wahrscheinlichkeitsmaßen P^ϑ ist im folgenden irrelevant. Die in Definition 4.(ii) geforderte Injektivität dient nur der mathematischen Bequemlichkeit, da sie die Identifikation der Verteilung $P^\vartheta_{X_1} \in \mathfrak{P}$ mit dem Parameter $\vartheta \in \Theta$ erlaubt.

Wir nehmen an, daß die vorliegenden Daten $(x_1, \ldots, x_n) \in \mathbb{R}^n$ eine Realisierung von $(X_1, \ldots, X_n)$ sind, d.h.

$$(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega)) \in \mathbb{R}^n$$

für ein $\omega \in \Omega$.

Definition 6. Gegeben sei ein statistisches Experiment mit Parameterraum Θ. Ein *Schätzproblem* ist definiert durch eine Abbildung

$$\gamma : \Theta \to \mathbb{R}.$$

Definition 7. Gegeben sei ein statistisches Experiment mit Stichprobenraum $\mathbb{R}^n$. Eine *Schätzfunktion* ist eine Borel-meßbare Abbildung

$$g_n : \mathbb{R}^n \to \mathbb{R}.$$

Bemerkung 8. Die Abbildung γ und ebenso η mit $\eta(P^{\vartheta}_{X_1}) = \gamma(\vartheta)$ beschreibt einen „Aspekt“ der Verteilungen $P^{\vartheta}_{X_1}$. Wichtige Beispiele sind der Erwartungswert und die Varianz.

Zur Lösung des Schätzproblems wählt man eine Schätzfunktion g_n und schätzt $\gamma(\vartheta)$ durch

$$g_n(x_1, \dots, x_n) = g_n(X_1(\omega), \dots, X_n(\omega)).$$

Ziel: Für jedes $\vartheta \in \Theta$ liegen die Werte der Zufallsvariable

$$g_n(X_1, \dots, X_n)$$

auf $(\Omega, \mathfrak{A}, P^{\vartheta})$ „nahe “ bei $\gamma(\vartheta)$.

Beispiel 9. Geschlecht eines Neugeborenen, siehe Beispiel 1. Modellierung durch

$$\Theta :=]0, 1[$$

und

$$P^{p}_{X_1} := \mathbf{B}(1, p)$$

für $p \in \Theta$. Das durch

$$\gamma(p) := p$$

definierte Schätzproblem beschreibt die Bestimmung der zugrundeliegenden Verteilung. Als Schätzfunktion für γ haben wir bereits

$$g_n(x_1, \dots, x_n) := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{1}$$

betrachtet.

Definition 10. Gegeben sei ein statistisches Experiment mit Parameterraum Θ. Ein *Testproblem* ist definiert durch eine Menge

$$\emptyset \neq \Theta_0 \subsetneq \Theta,$$

genannt *Hypothese*.

Definition 11. Gegeben sei ein statistisches Experiment mit Stichprobemraum $\mathbb{R}^n$. Ein *Verwerfungsbereich* ist eine Borel-Menge

$$R_n \in \mathfrak{B}_n.$$

Bemerkung 12. Die Hypothese Θ_0 definiert eine Teilmenge

$$\emptyset \neq \mathfrak{P}_0 = \{P_{X_1}^{\vartheta} : \vartheta \in \Theta_0\} \subsetneq \mathfrak{P}.$$

Zur Lösung des Testproblems wählt man einen Verwerfungsbereich R_n und lehnt die Hypothese Θ_0 (bzw. „$\vartheta \in \Theta_0$“ oder „$P_{X_1}^{\vartheta} \in \mathfrak{P}_0$“) genau dann ab, wenn

$$(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega)) \in R_n$$

gilt.

Ziel: Für jedes $\vartheta \in \Theta_0$ ist die Wahrscheinlichkeit

$$P^{\vartheta}(\{(X_1, \ldots, X_n) \in R_n\})$$

des *Fehlers 1. Art* „klein“, und für jedes $\vartheta \in \Theta \setminus \Theta_0$ ist die Wahrscheinlichkeit

$$P^{\vartheta}(\{(X_1, \ldots, X_n) \notin R_n\})$$

des *Fehlers 2. Art* „klein“. Siehe jedoch Bemerkung 3.9.(ii).

Die in den Bemerkungen 8 und 12 genannten Ziele bei Schätz- und Testproblem entsprechen jeweils einer worst case-Analyse über alle $\vartheta \in \Theta$.

Beispiel 13. Geschlecht eines Neugeborenen, siehe Beispiel 9. Die durch

$$\Theta_0 := \{1/2\}$$

oder

$$\Theta_0 :=]0, 1/2[$$

definierten Testprobleme betreffen die Hypothesen, daß Neugeborene mit gleicher Wahrscheinlichkeit weiblich wie männlich sind bzw. daß weibliche Lebendgeburten mit kleinerer Wahrscheinlichkeit auftreten.

Im Fall $\Theta_0 := \{1/2\}$ ist die Wahl eines Verwerfungsbereiche von der Form

$$R_n := \{x \in \mathbb{R}^n : |g_n(x) - 1/2| \geq k_n\}$$

mit g_n gemäß (1) und $k_n > 0$ naheliegend.

Beispiel 14. Analog: Halbwertszeit, siehe Beispiel VII.2.8. Hier gilt

$$\Theta :=]0, \infty[$$

und

$$P_{X_1}^{\lambda} := \mathbf{Exp}(\lambda)$$

für $\lambda \in \Theta$. Die Bestimmung der Halbwertszeit entspricht einem Schätzproblem mit

$$\gamma(\lambda) := \ln(2)/\lambda.$$

Die Entscheidung, ob die Halbwertszeit h_0 vorliegt, entspricht einem Testproblem mit

$$\Theta_0 := \{\ln(2)/h_0\}.$$

Ausblick: nicht-parametrische Statistik.

2 Schätzprobleme

Wir betrachten für ein statistisches Experiment, gegeben durch $(\Omega, \mathfrak{A}, P^{\vartheta})_{\vartheta \in \Theta}$ und $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$, ein Schätzproblem, gegeben durch

$$\gamma : \Theta \to \mathbb{R}.$$

Bezeichnung. Wir setzen

$$\mathfrak{L}_1^{\vartheta} = \mathfrak{L}_1(\Omega, \mathfrak{A}, P^{\vartheta}), \qquad \mathfrak{L}_2^{\vartheta} = \mathfrak{L}_2(\Omega, \mathfrak{A}, P^{\vartheta})$$

und bezeichnen mit E^{ϑ} und Var^{ϑ} den Erwartungswert bzw. die Varianz bzgl. P^{ϑ}. Eine Eigenschaft gilt P^{ϑ}-fast sicher (f.s.), falls sie für alle ω aus einer Menge $A \in \mathfrak{A}$ mit $P^{\vartheta}(A) = 1$ gilt.

Definition 1. Die *Schätzvariable* zu einer Schätzfunktion $g_n : \mathbb{R}^n \to \mathbb{R}$ ist die Zufallsvariable $g_n(X)$.

Definition 2. Eine Schätzfunktion $g_n : \mathbb{R}^n \to \mathbb{R}$ heißt *erwartungstreu* für γ, falls

$$\forall \vartheta \in \Theta : \mathrm{E}^{\vartheta}(g_n(X)) = \gamma(\vartheta).$$

Eine Folge von Schätzfunktionen $g_n : \mathbb{R}^n \to \mathbb{R}$ heißt

(i) *schwach konsistent* für γ, falls

$$\forall \vartheta \in \Theta \; \forall \varepsilon > 0 : \lim_{n \to \infty} P^{\vartheta}(\{|g_n(X) - \gamma(\vartheta)| > \varepsilon\}) = 0.$$

(ii) *stark konsistent* für γ, falls

$$\forall \vartheta \in \Theta : \lim_{n\to\infty} g_n(X) = \gamma(\vartheta) \qquad P^\vartheta\text{-f.s.}$$

Bemerkung 3. Die starke Konsistenz impliziert die schwache Konsistenz, siehe Satz VII.2.12.

Wir untersuchen die Schätzung eines Erwartungswertes, siehe Beispiele 1.9 und 1.14.

Definition 4. Sei $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Dann heißt

$$\overline{x}_n := \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

empirisches Mittel der Stichprobe x.

Die zugehörige Schätzvariable ist

$$\overline{X}_n := \frac{1}{n} \cdot \sum_{i=1}^n X_i.$$

Satz 5. Gelte $X_1 \in \mathfrak{L}_1^\vartheta$ für alle $\vartheta \in \Theta$, und sei

$$\gamma(\vartheta) := \mathrm{E}^\vartheta(X_1).$$

Dann definieren die empirischen Mittel $g_n(x) := \overline{x}_n$ eine stark konsistente Folge von erwartungstreuen Schätzfunktionen für γ.

Beweis. Für jedes $\vartheta \in \Theta$ gilt

$$\mathrm{E}^\vartheta(g_n(X)) = \mathrm{E}^\vartheta(\overline{X}_n) = \frac{1}{n} \cdot \sum_{i=1}^n \mathrm{E}^\vartheta(X_i) = \gamma(\vartheta).$$

Die starke Konsistenz ist genau das starke Gesetz der großen Zahlen. $\square$

Beispiel 6. In Beispiel 1.9 mit den Daten aus Beispiel 1.1 gilt für jedes $p \in \,]0,1[$ aufgrund der Tschebyschev-Ungleichung

$$P^p(\{|\overline{X}_n - p| \geq 10^{-2}\}) \leq 9.9 \cdot 10^{-5},$$
$$P^p(\{|\overline{X}_n - p| \geq 10^{-3}\}) \leq 9.9 \cdot 10^{-3}$$

und aufgrund der Hoeffdingschen Ungleichung, siehe Satz VII.1.5,

$$P^p(\{|\overline{X}_n - p| \geq 10^{-2}\}) \leq 9.2 \cdot 10^{-2187},$$
$$P^p(\{|\overline{X}_n - p| \geq 10^{-3}\}) \leq 2.7 \cdot 10^{-22}.$$

Wir untersuchen nun die Schätzung der Varianz.

Definition 7. Sei $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Dann heißt

$$s_n^2 := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x}_n)^2$$

empirische Varianz der Stichprobe x.

Satz 8. Gelte $X_1 \in \mathfrak{L}_2^\vartheta$ für alle $\vartheta \in \Theta$, und sei

$$\gamma(\vartheta) := \mathrm{Var}^\vartheta(X_1).$$

Ferner sei $n > 1$. Dann definieren die empirischen Varianzen $v_n(x) := s_n^2$ eine stark konsistente Folge von erwartungstreuen Schätzern für γ.

Beweis. Setze $\mu(\vartheta) := \mathrm{E}^\vartheta(X_1)$. Es gilt

$$\begin{aligned} v_n(x) &= \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left((x_i - \mu(\vartheta)) + (\mu(\vartheta) - \overline{x}_n)\right)^2 \\ &= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \mu(\vartheta))^2 - \frac{n}{n-1} \cdot (\mu(\vartheta) - \overline{x}_n)^2 \end{aligned}$$

Mit Satz VI.2.13 folgt

$$\begin{aligned} \mathrm{E}^\vartheta(v_n(X)) &= \frac{n}{n-1} \cdot \gamma(\vartheta) - \frac{n}{n-1} \cdot \mathrm{Var}^\vartheta(\overline{X}_n) \\ &= \frac{n}{n-1} \cdot \gamma(\vartheta) - \frac{1}{n-1} \cdot \gamma(\vartheta) = \gamma(\vartheta). \end{aligned}$$

Nach dem starken Gesetz der großen Zahlen gilt P^ϑ-f.s.

$$\lim_{n\to\infty} \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \mu(\vartheta))^2 = \mathrm{E}^\vartheta (X_1 - \mu(\vartheta))^2 = \mathrm{Var}^\vartheta(X_1)$$

und

$$\lim_{n\to\infty} \overline{X}_n = \mu(\vartheta).$$

Es folgt P^ϑ-f.s.

$$\lim_{n\to\infty} v_n(X) = \gamma(\vartheta).$$

□

Bemerkung 9. Für die empirische Varianz gilt

$$s_n^2 = \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}_n^2 \right),$$

wie der erste Teil des Beweises von Satz 8 mit $\mu(\vartheta) := 0$ zeigt.

Beispiel 10. In Beispiel 1.1 beträgt die empirische Varianz 0.2498 . . . .

Beispiel 11. Sei $\Theta := [0,1]$ und für $p \in \Theta$ gelte $P^p_{X_1} = \mathbf{B}(1,p)$. Ferner sei

$$\gamma(p) := \sqrt{p \cdot (1-p)}$$

die entsprechende Standardabweichung. Für jede Schätzfunktion g_1 gilt

$$\mathrm{E}^p(g_1(X)) = p \cdot g_1(1) + (1-p) \cdot g_1(0).$$

Also existiert keine erwartungstreue Schätzfunktion für γ. Analog für $n > 1$. Stichwort: asymptotische Erwartungstreue.

Wir studieren nun folgende Fragen:

(i) Wie definiert man die „Güte“ einer Schätzfunktion?

(ii) Kennt man „optimale“ Schätzfunktionen?

(iii) Wie „verläßlich“ ist ein Schätzwert?

Zunächst zur Frage (i).

Definition 12. Der *Quadratmittel-Fehler* einer Schätzfunktion g_n für γ ist definiert als

$$R^\vartheta(g_n) := \mathrm{E}^\vartheta(g_n(X) - \gamma(\vartheta))^2,$$

falls $g_n(X) \in \mathfrak{L}^\vartheta_2$.

Definition 13. Der *Bias* einer Schätzfunktion g_n für γ ist definiert als

$$B^\vartheta(g_n) := \mathrm{E}^\vartheta(g_n(X)) - \gamma(\vartheta),$$

falls $g_n(X) \in \mathfrak{L}^\vartheta_1$.

Bemerkung 14.

(i) Die Erwartungstreue ist äquivalent zu

$$\forall \vartheta \in \Theta : B^\vartheta(g_n) = 0.$$

(ii) Der Quadratmittel-Fehler wird auch als Risiko bei quadratischer Verlustfunktion bezeichnet. Es gilt

$$R^\vartheta(g_n) = \mathrm{Var}^\vartheta(g_n(X)) + (B^\vartheta(g_n))^2,$$

siehe Übung 13:G22.

Im folgenden sei
$$g_n^*(x) := \overline{x}_n.$$

Beispiel 15. Für die Schätzung des Erwartungswertes gilt $B^\vartheta(g_n^*) = 0$ und
$$R^\vartheta(g_n^*) = \frac{1}{n} \cdot \mathrm{Var}^\vartheta(X_1),$$
falls $X_1 \in \mathfrak{L}_2^\vartheta$. In der Situation von Beispiel 1.9 folgt
$$R^p(g_n^*) = \frac{p \cdot (1-p)}{n}$$
und weiter
$$\sup_{p \in \Theta} R^p(g_n^*) = \frac{1}{4 \cdot n}.$$
In der Situation von Beispiel 1.14 gilt $\gamma(\lambda) = \ln(2) \cdot \mathrm{E}^\lambda(X_1)$, und es folgt
$$R^\lambda(\ln(2) \cdot g_n^*) = \frac{(\ln(2))^2}{\lambda^2 \cdot n},$$
so daß der Quadratmittelfehler zwar für jeden Parameter $\lambda > 0$ mit $n \to \infty$ gegen null strebt, aber für jedes $n \in \mathbb{N}$ unbeschränkt auf $\Theta =]0, \infty[$ ist.

Wir untersuchen die Frage (ii) der Optimalität exemplarisch für die Problemstellung aus Beispiel 1.9. Im folgenden sei
$$\gamma(p) := \mathrm{E}^p(X_1) = p.$$
Für $p \in]0, 1[$ und $x \in D := \{0, 1\}^n$ sei
$$L_x(p) := P^p(\{X = x\}) = p^{k(x)} \cdot (1-p)^{n-k(x)}$$
mit
$$k(x) := |\{i \in \{1, \ldots, n\} : x_i = 1\}|$$
die sogenannte Likelihood-Funktion, vgl. Übung 6:G9, sowie
$$\ell_x(p) := \ln(L_x(p))$$
die sogenannte Log-Likelihood-Funktion.

Satz 16 (Ungleichung von Fréchet, Cramér, Rao)**.** Jede erwartungstreue Schätzfunktion $g_n : \mathbb{R}^n \to \mathbb{R}$ für γ erfüllt
$$\forall p \in]0, 1[: \ \ R^p(g_n) \geq \frac{1}{\mathrm{E}^p\big((\ell'_X(p))^2\big)}.$$

Beweis. Für g_n wie oben gilt

$$p = \mathrm{E}^p(g_n(X)) = \sum_{x \in D} g_n(x) \cdot L_x(p)$$

und somit

$$1 = \sum_{x \in D} g_n(x) \cdot L'_x(p) = \sum_{x \in D} g_n(x) \cdot \ell'_x(p) \cdot L_x(p) = \mathrm{E}^p(g_n(X) \cdot \ell'_X(p)).$$

Aus $\sum_{x \in D} L_x(p) = 1$ folgt

$$0 = \sum_{x \in D} L'_x(p) = \sum_{x \in D} \ell'_x(p) \cdot L_x(p) = \mathrm{E}^p(\ell'_X(p)).$$

Also liefert die Cauchy-Schwarzsche Ungleichung

$$1 = \Big(\mathrm{E}^p((g_n(X) - \gamma(p)) \cdot \ell'_X(p))\Big)^2 \leq \mathrm{Var}^p(g_n(X)) \cdot \mathrm{E}^p\big((\ell'_X(p))^2\big).$$

□

Bemerkung 17. Im Beweis von Satz 16 wurde die konkrete Verteilungsannahme nicht wesentlich genutzt. Die untere Schranke des Satzes gilt deshalb unter viel allgemeineren Voraussetzungen. Siehe Krengel (2000, §4.5) und Irle (2001, p. 308).

Wir erhalten die Optimalität des empirischen Mittels in einem sehr starken Sinn.

Satz 18. Es gilt

$$\forall p \in \,]0,1[\,: \;\; R^p(g_n^*) = \inf \{R^p(g_n) : g_n \text{ erwartungstreue Schätzfunktion für } \gamma\}.$$

Beweis. Man verifiziert

$$\mathrm{E}^p\big((\ell'_X(p))^2\big) = \frac{n}{p \cdot (1-p)},$$

siehe Krengel (2000, p. 68), und wendet Satz 16 und Beispiel 15 an. □

Satz 18 besagt, daß g_n^* eine gleichmäßig beste erwartungstreue Schätzfunktion in der vorliegenden Situation ist. Letzteres ist wesentlich, wie das folgende Beispiel zeigt.

Beispiel 19. Sei $\Theta := \mathbb{R}$ und

$$P_{X_1}^{\vartheta} := \mathbf{U}([\vartheta - 1/2, \vartheta + 1/2])$$

für $\vartheta \in \Theta$. Zu schätzen ist

$$\gamma(\vartheta) := \mathrm{E}^{\vartheta}(X_1) = \vartheta.$$

Setze

$$g_3(x_1, x_2, x_3) := \frac{\max(x_1, x_2, x_3) + \min(x_1, x_2, x_3)}{2}.$$

Dann ist g_3 erwartungstreu für γ mit

$$\forall \vartheta \in \Theta : \;\; R^{\vartheta}(g_3) < R^{\vartheta}(g_3^*).$$

Siehe Übung 13:H51.

Schließlich behandeln wir Frage (iii).

Definition 20. Sei $\alpha \in \,]0,1[$. Zwei Borel-meßbare Abbildungen $\ell_n, r_n : \mathbb{R}^n \to \mathbb{R}$ bilden ein *Konfidenzintervall* für γ zum Niveau $1-\alpha$, falls

$$\forall \vartheta \in \Theta : \; P^\vartheta(\{\gamma(\vartheta) \in [\ell_n(X), r_n(X)]\}) \geq 1-\alpha.$$

Bemerkung 21. Gesucht sind „möglichst kleine“ Konfidenzintervalle. Man beachte, daß nicht $\gamma(\vartheta)$ sondern $[\ell_n(X), r_n(X)]$ zufällig ist.

Im folgenden behandeln wir den Spezialfall

$$\gamma(\vartheta) := \mathrm{E}^\vartheta(X_1)$$

und Intervalle der Form

$$[\overline{x}_n - b_n(x), \overline{x}_n + b_n(x)]$$

mit Borel-meßbaren Abbildungen

$$b_n : \mathbb{R}^n \to [0,\infty[\,.$$

Bemerkung 22. Die Abbildung b_n definiert genau dann Konfidenzintervall für den Erwartungswert zum Niveau $1-\alpha$, wenn

$$\forall \vartheta \in \Theta : \; P^\vartheta(\{|\overline{X}_n - \mathrm{E}^\vartheta(X_1)| \leq b_n(X)\}) \geq 1-\alpha.$$

Zur Festlegung von b_n ist deshalb die Tschebyschev-Ungleichung anwendbar, falls (eine Schranke für) $\sup_{\vartheta\in\Theta} \mathrm{Var}^\vartheta(X_1)$ bekannt ist. Auf diese Weise erhält man jedoch oft zu große Konfidenzintervalle.

Beispiel 23. Fortsetzung von Beispiel 6. Konfidenzintervalle deterministischer Breite $2 \cdot b_n$ mittels der

(i) Tschebyschev-Ungleichung: Es gilt

$$\frac{1}{4 \cdot b_n^2 \cdot n} = \alpha \Leftrightarrow b_n = \frac{1}{\sqrt{4 \cdot \alpha \cdot n}}.$$

(ii) Hoeffdingschen Ungleichung: Es gilt

$$2 \cdot \exp(-2 \cdot b_n^2 \cdot n) = \alpha \Leftrightarrow b_n = \sqrt{\frac{\ln(2/\alpha)}{2 \cdot n}}.$$

Auf diese Weise erhält man

α	b_n per T-Ungl.	b_n per H-Ungl.
0.05	$4.5 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$
0.01	$1.0 \cdot 10^{-3}$	$3.3 \cdot 10^{-4}$
0.001	$3.2 \cdot 10^{-3}$	$3.8 \cdot 10^{-4}$

Nun bestimmen wir Konfidenzintervalle für den Erwartungswert unter Normalverteilungsannahmen. Wir unterscheiden dabei zwei Fälle:

(i) Die Varianz $\sigma^2 > 0$ ist bekannt. Also $\Theta := \mathbb{R}$ und $P^{\mu}_{X_1} := \mathbf{N}(\mu, \sigma^2)$ für $\mu \in \Theta$.

(ii) Die Varianz ist unbekannt. Also $\Theta := \mathbb{R} \times]0, \infty[$ und $P^{(\mu,\sigma)}_{X_1} := \mathbf{N}(\mu, \sigma^2)$ für $(\mu, \sigma) \in \Theta$.

Zuächst stellen wir einige Eigenschaften der Normalverteilung zusammen.

Satz 24. Seien $a, b, \mu, \mu_i \in \mathbb{R}$ mit $a \neq 0$ und $\sigma, \sigma_i \in]0, \infty[$.

(i) Falls $X \sim \mathbf{N}(\mu, \sigma^2)$, dann $a \cdot X + b \sim \mathbf{N}(a \cdot \mu + b, a^2 \cdot \sigma^2)$.

(ii) Falls $X_1, \dots, X_n$ unabhängig und $X_i \sim \mathbf{N}(\mu_i, \sigma_i^2)$, dann $\sum_{i=1}^n X_i \sim \mathbf{N}(\mu, \sigma^2)$ mit $\mu := \sum_{i=1}^n \mu_i$ und $\sigma^2 := \sum_{i=1}^n \sigma_i^2$.

Beweis. Übung 12:H51. □

Satz 25. Unter der Normalverteilungsannahme definiert

$$b_n := \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1 - \alpha/2)$$

bei bekannter Varianz σ^2 ein Konfidenzintervall für den Erwartungswert zum Niveau $1 - \alpha$.

Beweis. Satz 24 zeigt

$$Z := \frac{\sqrt{n}}{\sigma} \cdot \big(\overline{X}_n - \mathrm{E}^{\mu}(X_1)\big) \sim \mathbf{N}(0, 1)$$

bzgl. P^{μ}. Für alle $b_n > 0$ folgt

$$\begin{aligned} P^{\mu}(\{|\overline{X}_n - \mathrm{E}^{\mu}(X_1)| \leq b_n\}) &= P^{\mu}(\{|Z| \leq \sqrt{n}/\sigma \cdot b_n\}) \\ &= 2 \cdot \Phi(\sqrt{n}/\sigma \cdot b_n) - 1. \end{aligned}$$

Schließlich gilt

$$2 \cdot \Phi(\sqrt{n}/\sigma \cdot b_n) - 1 = 1 - \alpha \Leftrightarrow b_n = \sigma/\sqrt{n} \cdot \Phi^{-1}(1 - \alpha/2).$$

□

Beispiel 26. Die Abbildungen VIII.1 und VIII.2 zeigen 5 bzw. 50 Konfidenzintervalle nach der Konstruktion aus Satz 25 mit

$$\alpha := 0.05, \qquad \sigma := 2, \qquad n := 15$$

(und $\mu := 3$).

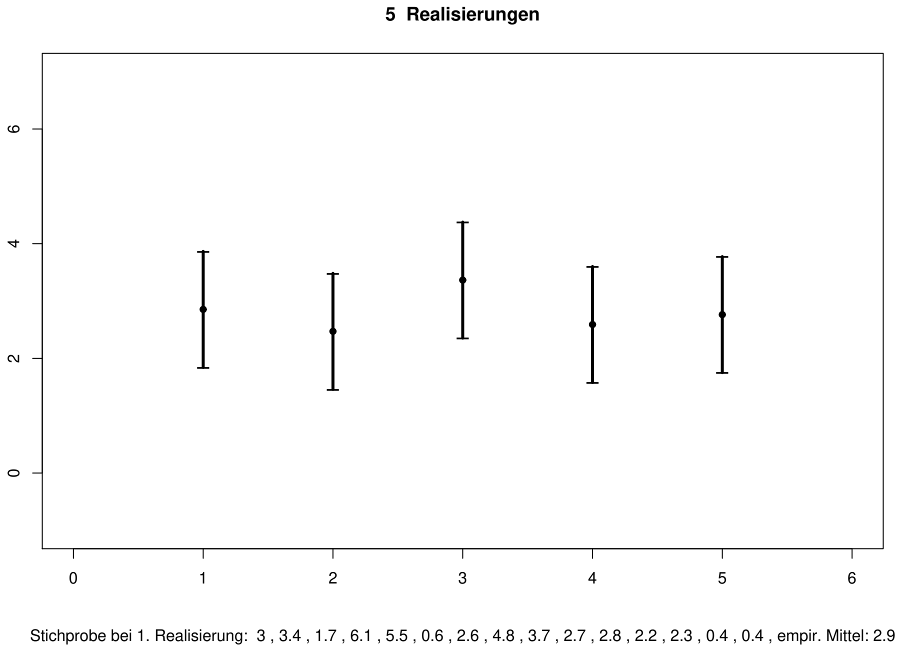

Abbildung VIII.1: Konfidenzintervalle unter Normalverteilungsannahme bei bekannter Varianz

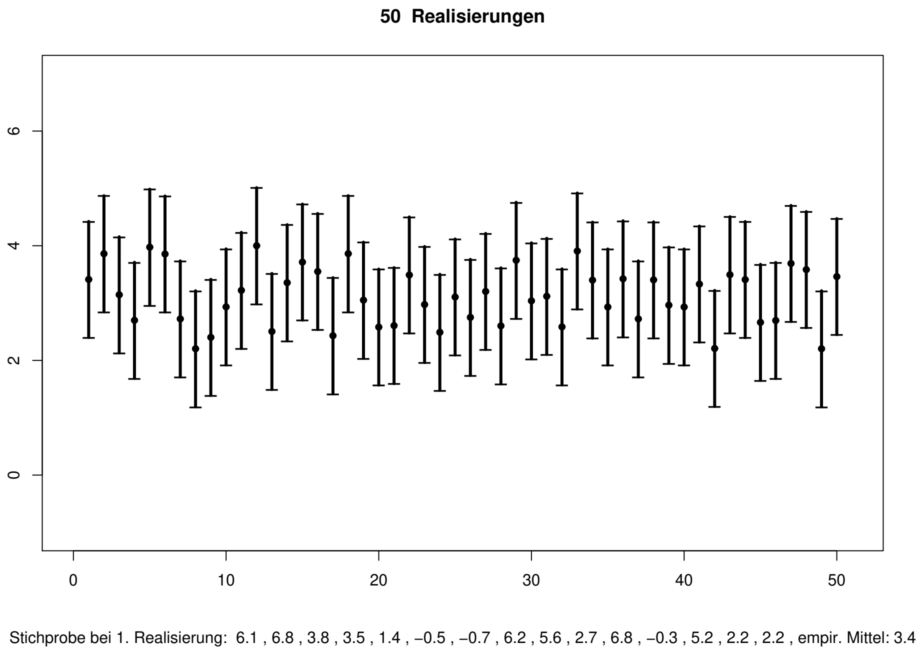

Abbildung VIII.2: Konfidenzintervalle unter Normalverteilungsannahme bei bekannter Varianz

Bei unbekannter Varianz ist es naheliegend σ^2 durch die empirische Varianz $v_n(x) := s_n^2$ zu ersetzen. Im folgenden sei $n > 1$ und $X'_1, \ldots, X'_n$ iid mit $X'_1 \sim \mathbf{N}(0,1)$. Setze

$$\overline{X'}_n := \sum_{i=1}^{n} X'_i/n,$$
$$X' := (X'_1, \ldots, X'_n)$$

und

$$T'_n := \frac{\overline{X'}_n}{\sqrt{v_n(X')/n}}.$$

Beachte: in der Definition von T'_n ist der Nenner fast sicher ungleich Null.

Lemma 27. Die Zufallsvariable T'_n besitzt die Dichte

$$f_n(x) := \frac{\Gamma(n/2)}{\Gamma((n-1)/2) \cdot \sqrt{\pi \cdot (n-1)}} \cdot \left(1 + x^2/(n-1)\right)^{-n/2}.$$

Beweis. Siehe Irle (2001, Kapitel 20). □

Bemerkung 28. Die Dichte f_n ist symmetrisch, und für alle $x \in \mathbb{R}$ gilt

$$\lim_{n\to\infty} f_n(x) = 1/\sqrt{2\pi} \cdot \exp(-x^2/2).$$

Definition 29. Die Verteilung der Zufallsvariable T'_n heißt *t-Verteilung* mit $n-1$ Freiheitsgraden. Bez.: $\mathbf{t}_{n-1}$.

Die Abbildungen VIII.3 und VIII.4 zeigen die Dichten und Verteilungsfunktion von $\mathbf{t}_5$ bzw. $\mathbf{t}_{20}$. Zum Vergleich sind auch die Dichte und die Verteilungsfunktion von $\mathbf{N}(0,1)$ angegeben.

Bemerkung 30. Zur Berechnung der Verteilungsfunktion von $\mathbf{t}_n$ und entsprechender Quantile: Numerik, Tabellen, Plots.

Lemma 31. Bzgl. $P^{(\mu,\sigma)}$ gilt

$$\frac{\overline{X}_n - \mu}{\sqrt{v_n(X)/n}} \sim \mathbf{t}_{n-1}.$$

Beweis. Setze

$$X'_i := \frac{X_i - \mu}{\sigma}.$$

Bzgl. $P^{(\mu,\sigma)}$ gilt $X'_1, \ldots, X'_n$ iid und $X'_1 \sim \mathbf{N}(0,1)$. Ferner gilt

$$\overline{X'_n} = \sum_{i=1}^{n} \frac{X'_i}{n} = \frac{\overline{X_n} - \mu}{\sigma},$$

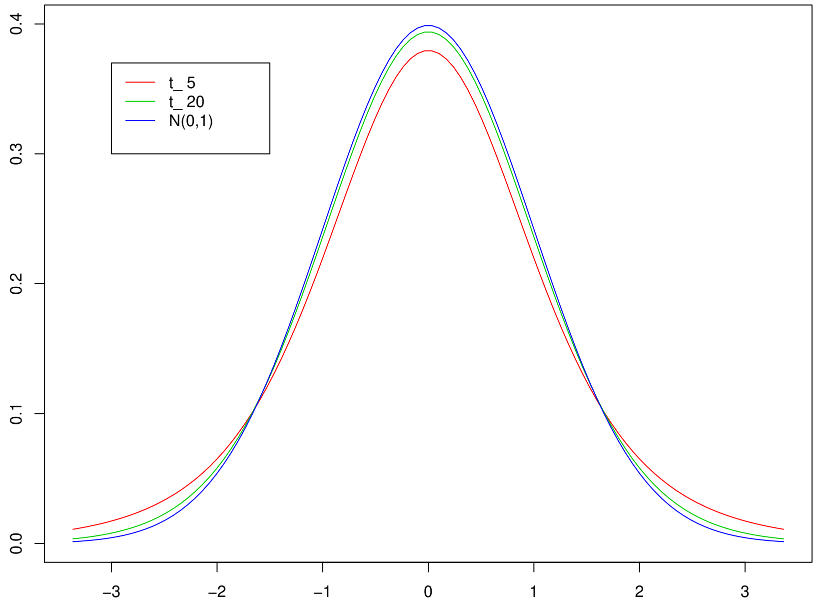

Abbildung VIII.3: Dichten von t-Verteilungen

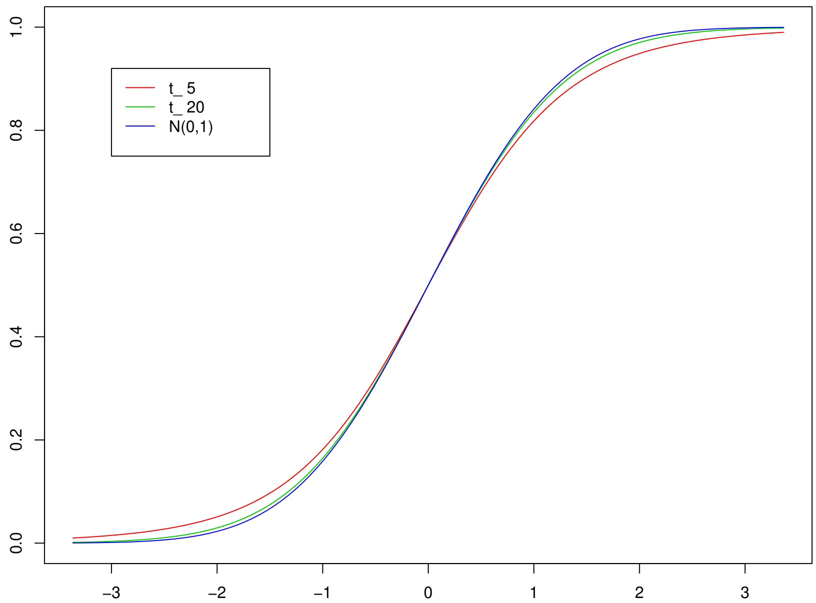

Abbildung VIII.4: Verteilungsfunktionen von t-Verteilungen

$$X_i' - \overline{X_n'} = \frac{X_i - \overline{X_n}}{\sigma}$$

und

$$(n-1)\cdot v_n(X') = \sum_{i=1}^{n}(X_i' - \overline{X_n'})^2 = \frac{1}{\sigma^2}\cdot\sum_{i=1}^{n}(X_i - \overline{X_n})^2 = \frac{(n-1)\cdot v_n(X)}{\sigma^2}.$$

Fazit

$$\frac{\overline{X_n'}}{\sqrt{v_n(X')}/\sqrt{n}} = \frac{\overline{X_n}-\mu}{\sigma}\cdot\frac{\sigma}{\sqrt{v_n(X)}/\sqrt{n}} = \frac{\overline{X_n}-\mu}{\sqrt{v_n(X)}/\sqrt{n}}.$$

□

Satz 32. Sei $t_{n-1;1-\alpha/2}$ das $(1-\alpha/2)$-Quantil von $\mathbf{t}_{n-1}$. Unter der Normalverteilungsannahme definiert

$$b_n(x) := \sqrt{\frac{v_n(x)}{n}}\cdot t_{n-1;1-\alpha/2}$$

bei unbekannter Varianz ein Konfidenzintervall für den Erwartungswert zum Niveau $1-\alpha$.

Beweis. Vgl. Beweis von Satz 25. Sei $F_n := F_{Z_n}$ die Verteilungsfunktion von

$$Z_n := \frac{\overline{X}_n - \mu}{\sqrt{v_n(X)/n}}.$$

Für alle $\mu \in \mathbb{R}$ und $\sigma > 0$ gilt gemäß Lemma 31

$$\begin{aligned} P^{(\mu,\sigma)}(\{|\overline{X_n}-\mu| \le b_n(X)\}) &= P^{(\mu,\sigma)}(\{|Z_n| \le t_{n-1;1-\alpha/2}\}) \\ &= F(t_{n-1;1-\alpha/2}) - F(-t_{n-1;1-\alpha/2}) \\ &= 2\cdot F(t_{n-1;1-\alpha/2}) - 1 \\ &= 2(1-\alpha/2) - 1 = 1-\alpha. \end{aligned}$$

□

Beispiel 33. Für $\alpha := 0.05$ ergibt sich

n	21	51	101
$t_{n-1;1-\alpha/2}$	$2.09\ldots$	$2.01\ldots$	$1.98\ldots$

Zum Vergleich: $\Phi^{-1}(1-\alpha/2) = 1.96\ldots$.

Beispiel 34. Die Abbildung VIII.5 zeigt 50 Konfidenzintervalle nach der Konstruktion aus Satz 32 mit

$$\alpha := 0.05, \qquad n := 15$$

(und $\mu := 3$ sowie $\sigma := 2$).

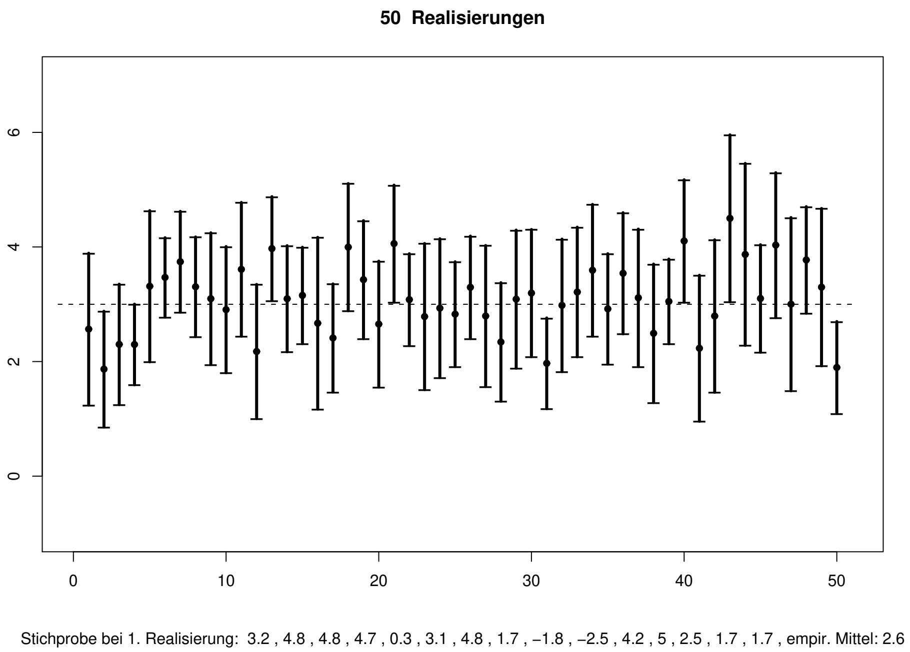

Abbildung VIII.5: Konfidenzintervalle unter Normalverteilungsannahme bei unbekannter Varianz

Ausblick: asymptotische Konfidenzintervalle. Gelte

$$\forall \vartheta \in \Theta : X_1 \in \mathfrak{L}_2^\vartheta \wedge \mathrm{Var}^\vartheta(X_1) > 0.$$

Setze

$$b_n(x) := \sqrt{\frac{v_n(x)}{n}} \cdot \Phi^{-1}(1 - \alpha/2).$$

Satz 35. Für jedes $\alpha \in]0,1[$ gilt

$$\forall \vartheta \in \Theta : \lim_{n\to\infty} P^\vartheta(\{|\overline{X}_n - \mathrm{E}^\vartheta(X_1)| \leq b_n(X)\}) = 1 - \alpha.$$

Beweis. Beruht auf dem Zentralen Grenzwertsatz. Siehe $\boxed{\text{MC-Buch}}$. □

Beispiel 36. Wir ergänzen die Konfidenzintervalle aus Beispiel 23 um asymptotische Konfidenzintervalle nach der Konstruktion aus Satz 35. Siehe Beispiel 10 zum entsprechenden Wert der empirischen Varianz.

α	b_n per T-Ungl.	b_n per H-Ungl.	b_n asymp. per ZGS
0.05	$4.5 \cdot 10^{-4}$	$2.8 \cdot 10^{-4}$	$1.95 \cdot 10^{-4}$
0.01	$1.0 \cdot 10^{-3}$	$3.3 \cdot 10^{-4}$	$2.57 \cdot 10^{-4}$
0.001	$3.2 \cdot 10^{-3}$	$3.8 \cdot 10^{-4}$	$3.27 \cdot 10^{-4}$

3 Testprobleme

Wir betrachten für ein statistisches Experiment, gegeben durch $(\Omega, \mathfrak{A}, P^\vartheta)_{\vartheta\in\Theta}$ und $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$, ein Testproblem, gegeben durch die Hypothese

$$\emptyset \neq \Theta_0 \subsetneq \Theta$$

(„der unbekannte Parameter liegt in Θ_0“, „$\vartheta \in \Theta_0$“).

Bezeichnung. Die Menge $\Theta \setminus \Theta_0$ heißt *Alternative*.

Beispiel 1. Bei einer Maschine zur Produktion von Gewinderingen stellt sich die Frage: Wird in der Produktion ein Nennmaß μ_0 für den Innendurchmesser eingehalten? Die vorliegenden Daten sind die Innendurchmesser $x_1, \ldots, x_n$ von n Gewinderingen.

Wir nehmen an, daß diese Innendurchmesser eine Realisierung von unabhängigen normalverteilten Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 sind. Zu unterscheiden sind zwei Fälle:

(i) Die Präzision der Maschine, gegeben durch σ^2, ist bekannt.

(ii) Die Präzision der Maschine ist unbekannt.

Definition 2. Ein Verwerfungsbereich $R_n \in \mathfrak{B}_n$ definiert einen *Signifikanztest* zum Niveau $\alpha \in]0, 1[$ (für die Hypothese Θ_0), falls

$$\forall \vartheta \in \Theta_0 : P^\vartheta(\{X \in R_n\}) \leq \alpha.$$

Bemerkung 3. Ein Verwerfungsbereich R_n legt folgende Entscheidung fest: man lehnt die Hypothese genau dann ab, wenn $x \in R_n$ gilt. Der Verwerfungsbereich definiert genau dann einen Signifikanztest zum Niveau α, wenn die Wahrscheinlichkeit des Fehlers 1. Art für jedes $\vartheta \in \Theta_0$ durch α beschränkt ist, siehe Bemerkung 1.12.

Um kleine Wahrscheinlichkeiten für Fehler 2. Art zu erreichen, sucht man eine Signifikanztest zu gegebenem Niveau α, dessen Verwerfungsbereich R_n „möglichst groß“ ist. In dieser Vorlesung werden keine Optimalitätsaussagen der Testtheorie behandelt.

Bemerkung 4. Die Formulierung eines Testproblems ist symmetrisch in der Hypothese Θ_0 und der Alternative $\Theta \setminus \Theta_0$. Diese Symmetrie gilt nicht bei Signifikanztests. Signifikanztest zum Niveau α für

- „Θ_0 = schuldig“: Freispruch Schuldiger mit Wahrscheinlichkeit höchstens α.
- „Θ_0 = unschuldig“: Verurteilung Unschuldiger mit Wahrscheinlichkeit höchstens α.

Bemerkung 5. In der Regel (und oBdA) wählt man einen Verwerfungsbereich der Form
$$R_n := \{g_n \in K_n\}$$
mit einer Borel-meßbaren Abbildung $g_n : \mathbb{R}^n \to \mathbb{R}$ und einer Menge $K_n \in \mathfrak{M}$. Die Zufallsvariable $g_n(X)$ heißt in diesem Kontext *Teststatistik* und K_n heißt *kritischer Bereich*. Hierdurch ist folgende Entscheidung festgelegt: Lehne die Hypothese genau dann ab, wenn $g_n(x) \in K_n$ gilt.

Man verwendet „plausible" Teststatistiken $g_n(X)$, deren Verteilungen $P^{\vartheta}_{g_n(X)}$ für $\vartheta \in \Theta_0$ (approximativ) bekannt sind.

Beispiel 6. Normalverteilungsannahme mit bekannter Varianz $\sigma^2 > 0$. Fortsetzung von Fall (i) in Beispiel 1. Gegeben sei $\mu_0 \in \mathbb{R}$. Entscheide, ob der unbekannte Erwartungswert gleich μ_0 ist. Formal wird dieses Testproblem definiert durch
$$\Theta := \mathbb{R}, \qquad P^{\mu}_{X_1} := \mathbf{N}(\mu, \sigma^2), \qquad \Theta_0 := \{\mu_0\}.$$
Wir verwenden
$$g_n(x) := \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}$$
für $x \in \mathbb{R}^n$. Falls die Hypothese korrekt ist, ist $g_n(X)$ $\mathbf{N}(0,1)$-verteilt. Also gilt für jedes $k > 0$
$$P^{\mu_0}(\{|g_n(X)| \geq k\}) = 1 - \Phi(k) + \Phi(-k) = 2(1 - \Phi(k)).$$
Somit definiert
$$R_n := \{|g_n| \geq k\}$$
genau dann einen Signifikanztest zum Niveau α, wenn $2(1-\Phi(k)) \leq \alpha$ gilt, d.h. wenn
$$k \geq \Phi^{-1}(1 - \alpha/2).$$
Klar: die Fehlerwahrscheinlichkeiten 2. Art sind monoton wachsend in k. Man wählt also
$$k = \Phi^{-1}(1 - \alpha/2).$$
Speziell für $\alpha = 0.05$ ergibt sich $k = 1.96\ldots$.

Dies entspricht folgender Entscheidung: Lehne die Hypothese Θ_0 genau dann ab, wenn
$$\left|\overline{x}_n - \mu_0\right| \geq \sigma/\sqrt{n} \cdot \Phi^{-1}(1 - \alpha/2).$$
Vgl. Satz 2.25.

Bezeichnung. Die Vorgehensweise aus Beispiel 6 wird als *zweiseitiger Gauß-Test* bezeichnet.

Beispiel 7. Das entsprechende einseitige Testproblem ist gegeben durch

$$\Theta := \mathbb{R}, \qquad P^{\mu}_{X_1} := \mathbf{N}(\mu, \sigma^2), \qquad \Theta_0 := \left]-\infty, \mu_0\right],$$

siehe ÜBUNG 14:G25.

Beispiel 8. Fortsetzung von Beispiel 1.13, Geschlecht eines Neugeborenen. Man widerlege die Annahme, daß Jungen und Mädchen mit gleicher Wahrscheinlichkeit geboren werden. Formal ist dieses Testproblem definiert durch

$$\Theta := \left]0, 1\right[, \qquad P^{p}_{X_1} = \mathbf{B}(1, p), \qquad \Theta_0 = \{1/2\}.$$

Als Teststatistik verwenden wir

$$g_n(X) := \frac{\overline{X}_n - 1/2}{1/(2\sqrt{n})} = \frac{\sum_{i=1}^{n} X_i - n \cdot 1/2}{1/2 \cdot \sqrt{n}}.$$

Beachte, daß $n = 25\,171\,123$. Also „ist" $g_n(X)$ bzgl. $P^{1/2}$ standard-normalverteilt. Deshalb führen wir einen Gauß-Test als sogenannten asymptotischen α-Niveau-Test durch: Lehne die Hypothese genau dann ab, wenn

$$\left|\overline{x}_n - 1/2\right| \geq 1/(2\sqrt{n}) \cdot \Phi^{-1}(1 - \alpha/2).$$

Für $n = 25\,171\,123$ gilt $1/(2\sqrt{n}) = 9.9 \ldots 10^{-5}$ und man erhält

α	$1/(2\sqrt{n}) \cdot \Phi^{-1}(1 - \alpha/2)$
10^{-2}	$2.56 \cdots \cdot 10^{-4}$
10^{-3}	$3.27 \cdots \cdot 10^{-4}$
10^{-20}	$9.30 \cdots \cdot 10^{-4}$

Für das empirische Mittel $\overline{x}_n = 0.4863\ldots$ der Daten gilt

$$|\overline{x}_n - 1/2| = 1.37 \cdots \cdot 10^{-2},$$

so daß die Hypothese für alle diese Werte α verworfen wird.

Bemerkung 9. Wir untersuchen die Wahrscheinlichkeiten für Fehler 2. Art bei einem zweiseitigen Gauß-Test, siehe Beispiel 6. Sei $\mu \neq \mu_0$ und

$$\mu_n := \sqrt{n} \cdot (\mu - \mu_0)/\sigma.$$

Dann ist $g_n(X)$ $\mathbf{N}(\mu_n, 1)$-verteilt bzgl. P^{μ} und

$$\begin{aligned} P^{\mu}(\{|g_n(X)| < k\}) &= P^{\mu}(\{-k - \mu_n < g_n(X) - \mu_n < k - \mu_n\}) \\ &= \Phi(k - \mu_n) - \Phi(-k - \mu_n) =: a_n(\mu). \end{aligned}$$

Es gilt:

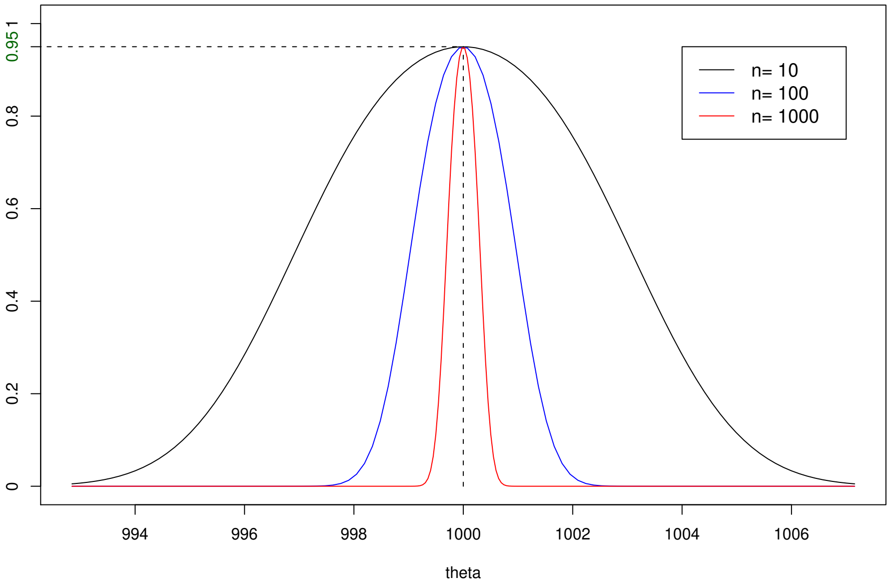

Abbildung VIII.6: Operationscharakteristiken des zweiseitigen Gauß-Tests für $\alpha := 0.05$

(i) Für jedes $\mu \neq \mu_0$ konvergiert $(a_n(\mu))_{n\in\mathbb{N}}$ sehr schnell gegen Null, siehe $\boxed{\text{Übung}}$.

(ii) Für jedes $n \in \mathbb{N}$ gilt

$$\sup_{\mu\in\Theta\setminus\Theta_0} a_n(\mu) = 1 - \alpha.$$

Vgl. Bemerkung 1.12.

Definition 10. Die *Operationscharakteristik* f eines durch Verwerfungsbereich $R_n \in \mathfrak{B}_n$ gegebenen Tests ist die durch

$$f(\vartheta) := P^{\vartheta}(\{X \notin R_n\})$$

definierte Abbildung $f : \Theta \to [0,1]$.

Bemerkung 11.

(i) Die Abbildung $f|_{\Theta\setminus\Theta_0}$ liefert die Wahrscheinlichkeiten für Fehler 2. Art.

(ii) Ein Signifikanztest zum Niveau α ist durch $f|_{\Theta_0} \geq 1 - \alpha$ charakterisiert.

Beispiel 12. Abbildung VIII.6 zeigt für $n = 10$, 100, 1000 Operationscharakteristiken des zweiseitigen Gauß-Tests für $\Theta_0 := \{1000\}$, $\sigma^2 := 25$ und $\alpha := 0.05$.

Beispiel 13. Normalverteilungsannahme mit unbekannter Varianz. Fortsetzung von Fall (ii) in Beispiel 1. Gegeben sei $\mu_0 \in \mathbb{R}$. Entscheide, ob der unbekannte Erwartungswert gleich μ_0 ist. Formal wird dieses Testproblem definiert durch

$$\Theta := \mathbb{R} \times]0, \infty[\,, \qquad P_{X_1}^{(\mu,\sigma)} := \mathbf{N}(\mu, \sigma^2), \qquad \Theta_0 := \{\mu_0\} \times]0, \infty[\,.$$

Wir betrachten die durch

$$g_n(x) := \frac{\overline{x}_n - \mu_0}{\sqrt{s_n^2/n}}.$$

definierte Teststatistik

$$g_n(X) := \frac{\overline{X}_n - \mu_0}{\sqrt{v_n(X)/n}}.$$

Vgl. Satz 2.32.

Satz 14. Sei $t_{n-1;1-\alpha/2}$ das $(1-\alpha/2)$-Quantil von $\mathbf{t}_{n-1}$. Für das Testproblem aus Beispiel 13 definiert der Verwerfungsbereich

$$R_n := \{|g_n| \geq t_{n-1;1-\alpha/2}\}.$$

einen Signifikanztest zum Niveau α.

Beweis. Unter Verwendung von Lemma 2.31 schließt man wie in Beispiel 6. □

Bezeichnung. Die Vorgehensweise aus Satz 14 wird als *zweiseitiger t-Test* bezeichnet.

Wir betrachten nun für Zufallsexperimente mit Werten in einer endlichen Menge folgende Probleme. Teste,

(i) ob ein Zufallsexperiment einer gegebenen Verteilung genügt,

(ii) ob die Komponenten eines vektorwertigen Zufallsexperimentes unabhängig sind.

Beispiel 15.

(i) Teste, ob ein Würfel fair ist.

(ii) Teste, ob Einkommen und politische Präferenz unabhängig sind.

Im folgenden sei

$$\emptyset \neq M \subseteq \mathbb{R}^d$$

endlich und

$$\Theta = \Big\{(p_k)_{k\in M} \in [0,1]^M : \sum_{k\in M} p_k = 1\Big\}$$

die Menge aller Wahrscheinlichkeitsfunktionen auf M. Für

$$\mathbf{p} = (p_k)_{k \in M} \in \Theta$$

sei die Verteilung von X_1 unter $P^{\mathbf{p}}$ gegeben durch

$$P^{\mathbf{p}}_{X_1}(\{k\}) = p_k \qquad k \in M.$$

Der Einfachheit halber setzen wir voraus, daß $X_1, \ldots, X_n$ nur Werte aus M annehmen.
Zunächst untersuchen wir Fragestellung (i), wobei die Hypothese durch

$$\Theta_0 = \{\mathbf{p}^0\}$$

mit einer Wahrscheinlichkeitsfunktion $\mathbf{p}^0 \in \Theta$ mit $p^0_k > 0$ für alle $k \in M$ definiert sei.

Beispiel 16. Man wählt $M := \{1, \ldots, 6\}$ und $p^0_k := 1/6$ für $k \in M$ bei der Frage, ob ein Würfel fair ist. Siehe auch Beispiel 8.

Im folgenden gelte oBdA

$$M = \{1, \ldots, m\}$$

mit $m \geq 2$. Wir betrachten die absoluten Häufigkeiten der Werte $1, \ldots, m$ in einer Stichprobe und definiere dazu

$$H : M^n \to \mathbb{N}^m_0$$

durch

$$H_k(x_1, \ldots, x_n) := \sum_{i=1}^{n} 1_{\{k\}}(x_i)$$

für $k = 1, \ldots, m$. Naheliegend ist folgendes Vorgehen: Verwerfung der Hypothese Θ_0, falls der „Abstand“ von $1/n \cdot H(x_1, \ldots, x_n)$ und $\mathbf{p}^0$ „groß“ ist.
Dazu: Bestimmung der Verteilung des Zufallsvektors $H(X)$.

Beispiel 17. Für $M := \{1, 2\}$ gilt

$$H_1(X) = \sum_{i=1}^{n} \underbrace{1_{\{1\}}(X_i)}_{=:X'_i}.$$

Aus $P^{\mathbf{p}}_{X_1} = \mathbf{B}(1, p_1)$ folgt $H_1(X) \sim \mathbf{B}(n, p_1)$ bzgl. $P^{\mathbf{p}}$. Klar: $H_2(X) = n - H_1(X)$. Also folgt für $k \in \{0, \ldots, n\}$

$$P^{\mathbf{p}}(\{H(X) = (k, n-k)\}) = \binom{n}{k} \cdot p_1^k \cdot p_2^{n-k}.$$

Bemerkung 18. Analog folgt allgemein, daß $H_k(X) \sim \mathbf{B}(n, p_k)$ bzgl. $P^{\mathbf{p}}$ gilt. Man beachte, daß die Komponenten $H_k(X)$ des Zufallsvektors $H(X)$ nicht unabhängig sind.

Satz 19. Für $\mathbf{p} \in \Theta$ und $h \in \mathbb{N}_0^m$ gilt

$$P^{\mathbf{p}}(\{H(X) = h\}) = n! \cdot \prod_{k=1}^{m} \frac{p_k^{h_k}}{h_k!},$$

falls $\sum_{k=1}^m h_k = n$. Andernfalls gilt $P^{\mathbf{p}}(\{H(X) = h\}) = 0$.

Beweis. Offenbar gilt $P^{\mathbf{p}}(\{H(X) = h\}) = 0$, falls $\sum_{k=1}^m h_k \neq n$. Im folgenden gelte $\sum_{k=1}^m h_k = n$. Für $x \in M^n$ mit $H(x) = h$ ergibt sich

$$P^{\mathbf{p}}(\{X = x\}) = p_{x_1} \cdots p_{x_n} = \prod_{k=1}^{m} p_k^{h_k}.$$

Die Anzahl der Stichproben mit absoluten Häufigkeiten h_k ist gegeben durch

$$\begin{aligned} |\{x \in M^n : H(x) = h\}| &= \binom{n}{h_1} \cdot \binom{n - h_1}{h_2} \cdots \binom{n - (h_1 + \cdots + h_{m-1})}{h_m} \\ &= \frac{n!}{h_1! \cdots h_m!}. \end{aligned}$$

Fazit:

$$P^{\mathbf{p}}(\{H(X) = h\}) = \sum_{x \in M^n,\ H(x) = h} P^{\mathbf{p}}(\{X = x\}) = n! \cdot \prod_{k=1}^{m} \frac{p_k^{h_k}}{h_k!}.$$

□

Definition 20. Ein m-dimensionaler Zufallsvektor Y auf $(\Omega, \mathfrak{A}, P)$ ist *multinomialverteilt* mit Parametern $n, m \in \mathbb{N}$ und $\mathbf{p} \in [0,1]^m$, wobei $\sum_{k=1}^m p_k = 1$, falls

$$P(\{Y = y\}) = \begin{cases} n! \cdot \prod_{k=1}^m \frac{p_k^{y_k}}{y_k!}, & \text{falls } y \in \mathbb{N}_0^m \text{ mit } \sum_{k=1}^m y_k = n \\ 0, & \text{andernfalls.} \end{cases}$$

Bez.: $Y \sim \mathbf{M}(n, m, \mathbf{p})$.

Bemerkung 21. Satz 19 zeigt, daß $H(X) \sim \mathbf{M}(n, m, \mathbf{p})$ bzgl. $P^{\mathbf{p}}$ gilt.

Beispiel 22. Wir betrachten 12 Würfe eines fairen Würfels. Für

$$n := 12, \qquad m := 6, \qquad \mathbf{p} := (1/6, \ldots, 1/6)$$

gilt

$$\mathbf{M}(12, 6, \mathbf{p})(\{(3, 0, 0, 2, 6, 1)\}) = \frac{12!}{3! \cdot 2! \cdot 6!} \cdot 6^{-12} = 2.54 \ldots 10^{-5}$$

und

$$\mathbf{M}(12, 6, \mathbf{p})(\{(2, 2, 2, 2, 2, 2)\}) = \frac{12!}{2^6} \cdot 6^{-12} = 3.43 \ldots 10^{-3}.$$

Bemerkung 23. Satz 19 ermöglicht prinzipiell die Konstruktion eines Signifikanztests zum Niveau α für $\Theta_0 = \{\mathbf{p}^0\}$: Wähle eine (möglichst kleine endliche) Menge $A \subset \mathbb{N}_0^m$ mit

$$P^{\mathbf{p}^0}(\{H(X) \in A\}) \geq 1 - \alpha,$$

und verwende den Verwerfungsbereich

$$R_n := \{H \notin A\}.$$

Dieses Prinzip ist allgemeine bei diskreten Teststatistiken anwendbar.

Beispiel 24. Wir betrachten Multinomialverteilungen mit

$$m := 3.$$

In den Abbildungen VIII.7 und VIII.9 ist

$$(h_1, h_2) \mapsto M(n, m, \mathbf{p})(\{(h_1, h_2, n - (h_1 + h_2))\}) \tag{1}$$

für $\mathbf{p} := (1/4, 1/4, 1/2)$ und $n := 10$ bzw. $n := 50$ (mit Interpolation) dargestellt. Optimale Wahlen der Verwerfungsbereiche für $\alpha = 0.05$ sind in den Abbildung VIII.8 und VIII.10 dargestellt: verwirf die Hypothese genau dann, wenn (h_1, h_2) „schwarz markiert“ ist. Auf diese Weise erhält man

$$P^{\mathbf{p}^0}(\{X \in R_n\}) = 0.0488\ldots$$

für $n = 10$ und

$$P^{\mathbf{p}^0}(\{X \in R_n\}) = 0.0497\ldots$$

für $n = 50$.

Abbildung VIII.11 zeigt (1) für $n = 10$ und $\mathbf{p} = (1/8, 3/8, 1/2)$.

Nachteile der Vorgehensweise gemäß Bemerkung 23: Abhängigkeit von n, m und $\mathbf{p}^0$ und hoher Rechenaufwand, falls n groß. Deshalb: asymptotischer α-Niveau-Test, siehe auch Beispiel 8. Beachte: $H = H^{(n)}$ und $X = X^{(n)}$ hängen von n ab.

Wir definieren $Q = Q^{(n)} : \mathbb{N}_0^m \to \mathbb{R}$ durch

$$Q^{(n)}(h) = \sum_{k=1}^{m} \frac{(h_k - n \cdot p_k^0)^2}{n \cdot p_k^0}$$

und betrachten die Teststatistik $Q(H(X)) = Q^{(n)}(H^{(n)}(X^{(n)}))$. Als partielle Motivation halten wir fest, daß das starke Gesetz der großen Zahlen zeigt, daß

$$\lim_{n\to\infty} H_k(X)/n = P^{\mathbf{p}^0}(\{X_1 = k\}) = p_k^0$$

$P^{\mathbf{p}^0}$-f.s. gilt.

Wir untersuchen die Konvergenz der Verteilung von $Q(H(X))$ bzgl. $P^{\mathbf{p}^0}$ für $n \to \infty$.

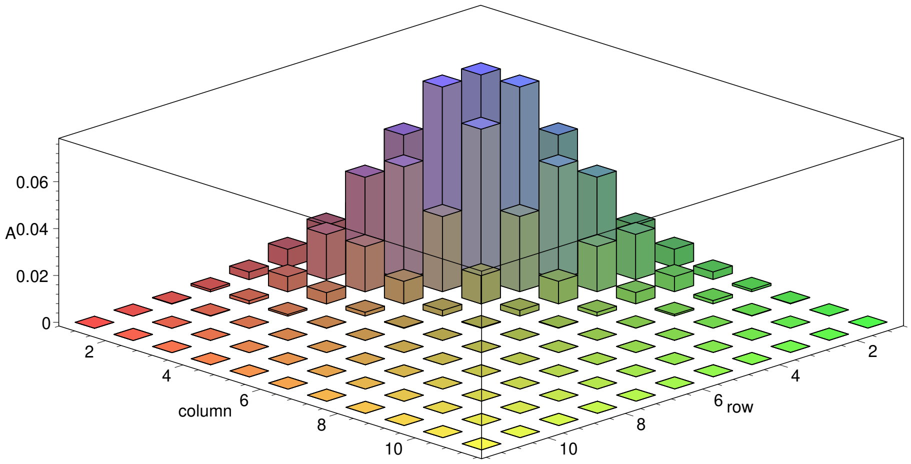

Abbildung VIII.7: Multinomialverteilung mit $n := 10$, $m := 3$, $\mathbf{p} := (1/4, 1/4, 1/2)$

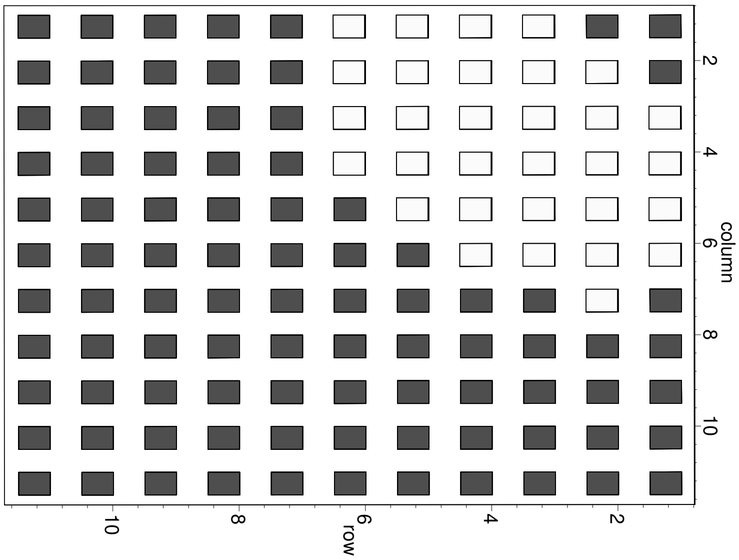

Abbildung VIII.8: Verwerfungsbereich in der Situation von Abb. VIII.7 für $\alpha := 0.05$

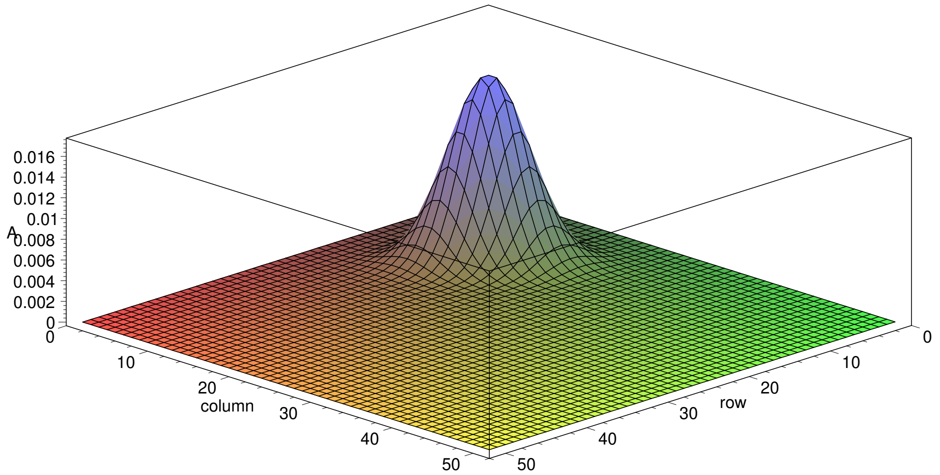

Abbildung VIII.9: Multinomialverteilung mit $n := 50$, $m := 3$, $\mathbf{p} := (1/4, 1/4, 1/2)$

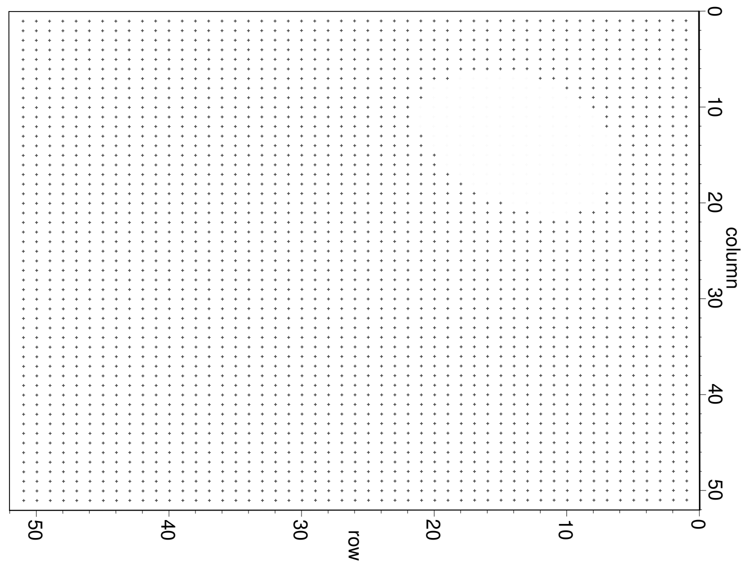

Abbildung VIII.10: Verwerfungsbereich in der Situation von Abb. VIII.9 für $\alpha := 0.05$

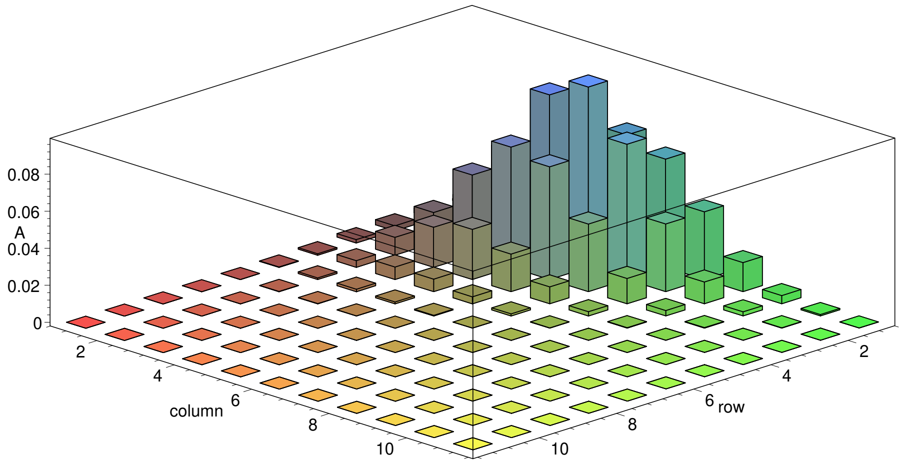

Abbildung VIII.11: Multinomialverteilung mit $n := 10$, $m := 3$, $\mathbf{p} := (1/8, 3/8, 1/2)$

Definition 25. Sei $d \in \mathbb{N}$. Eine Zufallsvariable Y mit Dichte

$$f_Y(x) = \frac{1}{2^{d/2} \cdot \Gamma(d/2)} \cdot x^{d/2-1} \cdot \exp(-x/2), \qquad x > 0,$$

und $f_Y(x) = 0$ für $x \leq 0$ heißt χ^2-*verteilt* mit d Freiheitsgraden. Bez.: $Y \sim \chi^2_d$.

Bemerkung 26. Für iid Zufallsvariablen $Y_1, \ldots, Y_d$ mit $Y_1 \sim \mathbf{N}(0,1)$ und $Y := \sum_{i=1}^d Y_i^2$ gilt $Y \sim \chi^2_d$. Siehe Krengel (2000, §14).

Abbildung VIII.12 zeigt die Dichten von χ^2_d für $d = 1, 2, 4, 6$.

Satz 27. Gelte

$$Z_n \sim \mathbf{M}(n, m, \mathbf{p})$$

für $m \geq 2$ und $\mathbf{p} \in [0,1]^m$, wobei $\sum_{k=1}^m p_k = 1$, sowie

$$Y \sim \chi^2_{m-1}.$$

Dann folgt

$$Q^{(n)}(Z_n) \xrightarrow{\mathrm{d}} Y.$$

Beweis. Siehe Krengel (2000, §14). □

Satz 28. Sei $\chi^2_{m-1;1-\alpha}$ das $(1-\alpha)$-Quantil der χ^2-Verteilung mit $m-1$ Freiheitsgraden. Dann definiert der Verwerfungsbereich

$$R_n := \{Q^{(n)} \circ H^{(n)} \geq \chi^2_{m-1;1-\alpha}\}$$

einen Signifikanztest zum asymptotischen Niveau α für $\Theta_0 = \{\mathbf{p}^0\}$, d.h.

$$\lim_{n\to\infty} P^{\mathbf{p}^0}(\{X^{(n)} \in R_n\}) = \alpha.$$

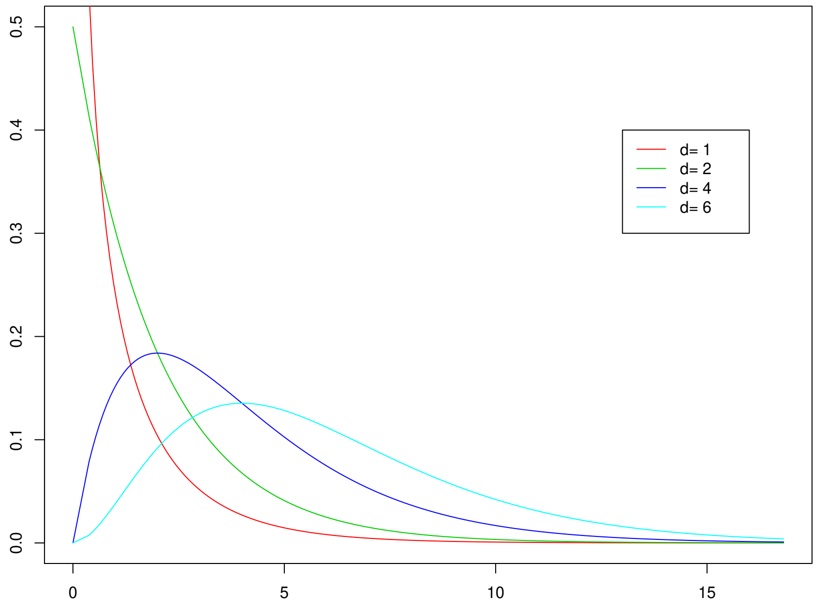

Abbildung VIII.12: Dichten von χ^2-Verteilungen

Beweis. Sei F die Verteilungsfunktion von χ^2_{m-1}. Es gilt

$$\{X^{(n)} \in R_n\} = \{Q^{(n)}(H^{(n)}(X^{(n)})) \geq \chi^2_{m-1;1-\alpha}\}.$$

Da F stetig ist, folgt mit den Sätzen 19 und 27

$$\lim_{n\to\infty} P^{\mathbf{p}^0}(\{X^{(n)} \in R_n\}) = 1 - F(\chi^2_{m-1,1-\alpha}) = \alpha.$$

$\square$

Bezeichnung. Die Vorgehensweise aus Satz 28 wird als χ^2-*Anpassungstest* bezeichnet.

Beispiel 29. Wir betrachten die Situation aus Beispiel 24 mit der Hypothese $\Theta_0 = \{\mathbf{p}^0\}$ für $m := 3$ und für $\mathbf{p}^0 := (1/4, 1/4, 1/2)$.
Wir wählen $\alpha := 0.05$ und erhalten $\chi^2_{m-1;1-\alpha} = 5.99\ldots$. Für $n := 10$ und

$$H(x_1, \ldots, x_n) := (0, 1, 9)$$

ergibt sich

$$Q(H(x_1, \ldots, x_n)) = 33/5 = 6.6 \geq \chi^2_{m-1;1-\alpha},$$

so daß der χ^2-Anpassungstest die Hypothese Θ_0 verwirft. Für $n := 50$ und

$$H(x_1, \ldots, x_n) := (9, 15, 26)$$

ergibt sich
$$Q(H(x_1, \ldots, x_n)) = 38/25 = 1.52 < \chi^2_{m-1;1-\alpha},$$
so daß der χ^2-Anpassungstest die Hypothese Θ_0 nicht verwirft.

Bemerkung 30. Der χ^2-Anpassungstest läßt sich nach einer Diskretisierung auch im Falle absolutstetiger Verteilungen auf $\mathbb{R}^d$ anwenden. Zum Testen der Hypothese, daß die unbekannte Verteilung von X_1 gleich P_0 ist, wählen wir $m \in \mathbb{N}$ und p.d. Mengen $B_k \in \mathfrak{B}_d$ mit $\bigcup_{k=1}^m B_k = \mathbb{R}^d$. Wir setzen
$$X_i' := \sum_{k=1}^m k \cdot 1_{B_k}(X_i)$$
und betrachten als neue und schwächere Hypothes $\Theta_0 = \{\mathbf{p}^0\}$ mit
$$p_k^0 := P_0(B_k).$$

Beispiel 31. Wir illustrieren dieses Vorgehen anhand folgender Frage: Liefert ein Zufallszahlengenerator auf $[0,1]$ gleichverteilte Zufallszahlen? Hier gilt $P_0 := \mathbf{U}([0,1])$. Wir wählen
$$B_1 := \left]-\infty, 1/m\right], \qquad B_m := \left]1-1/m, \infty\right[$$
und für $k = 2, \ldots, m-1$
$$B_k := \left](k-1)/m, k/m\right].$$
Somit gilt für $k = 1, \ldots, m$
$$p_k^0 := P_0(B_k) = 1/m.$$

Literatur

H.-O. Georgii, *Stochastik*, 3. Auflage, de Gruyter, Berlin, 2007.

Ch. M. Grinstead, J. L. Snell, *Introduction to Probability*, 2nd Edition[1], AMS, Providence, 1997.

O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge Univ. Press, Cambridge, 2002.

Ch. Hesse, *Angewandte Wahrscheinlichkeitstheorie*, Vieweg, Braunschweig, 2003.

A. Irle, *Wahrscheinlichkeitstheorie und Statistik*, 2. Auflage, Teubner, Stuttgart, 2005.

U. Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 8. Auflage, Vieweg, Braunschweig, 2005.

J. Lehn, H. Wegmann, *Einführung in die Statistik*, 5. Auflage, Teubner, Stuttgart, 2004.

E. Lesigne, *Heads or Tails: An Introduction to Limit Theorems in Probability*, AMS, Providence, 2005.

D. Meintrup, St. Schäffler, *Stochastik*, Springer, Berlin, 2005.

Th. Müller-Gronbach, E. Novak, K. Ritter, *Monte-Carlo-Algorithmen*, erscheint bei Springer, Berlin, 2008.

A. N. Širjaev, *Wahrscheinlichkeit*, Deutscher Verlag der Wissenschaften, Berlin, 1988.
A. N. Shiryayev, *Probability*, Springer, New York, 1984.

[1] siehe auch:
http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html

Definitionen und Bezeichnungen