



ELEMENTARE WAHRSCHEINLICHKEITSRECHNUNG

SKRIPT
JUN.-PROF. DR. ZAKHAR KABLUCHKO

UNIVERSITÄT ULM
INSTITUT FÜR STOCHASTIK

Inhaltsverzeichnis

Vorwort	1
Literatur	1
Kapitel 1. Grundbegriffe der Wahrscheinlichkeitstheorie	3
1.1. Zufallsexperimente, Ausgänge, Grundmenge	3
1.2. Ereignisse	4
1.3. Wahrscheinlichkeiten	6
Kapitel 2. Kombinatorik	11
2.1. Geburtstagsproblem	11
2.2. Urnenmodelle	12
2.3. Hypergeometrische Verteilung	16
2.4. Binomialverteilung und Multinomialverteilung	18
Kapitel 3. Zufallsvariablen	21
Kapitel 4. Unabhängigkeit	23
4.1. Unabhängigkeit von Ereignissen	23
4.2. Produkträume	26
4.3. Bedingte Wahrscheinlichkeiten	27
4.4. Unabhängigkeit von Zufallsvariablen	29
Kapitel 5. Erwartungswert	31
Kapitel 6. Diskrete Verteilungen	37
6.1. Gleichverteilung	37
6.2. Bernoulli-Experimente und die Binomialverteilung	37
6.3. Poisson-Verteilung	39
6.4. Geometrische Verteilung	41
6.5. Negative Binomialverteilung	43
Kapitel 7. Wahrscheinlichkeitstheorie und Maßtheorie	47
7.1. Vorüberlegungen	47
7.2. Geometrische Wahrscheinlichkeiten	48
7.3. Algebren	49
7.4. σ -Algebren	51
7.5. Limes superior und Limes inferior für Folgen von Mengen	52
7.6. Borel- σ -Algebra	53
7.7. Maße	54
7.8. Wahrscheinlichkeitsmaße	55

7.9.	Das Lemma von Borel–Cantelli	57
Kapitel 8.	Zufallsvariablen: Die allgemeine Definition	61
8.1.	Zufallsvariablen	61
8.2.	Zufallsvektoren	63
8.3.	Verteilungsfunktion einer Zufallsvariable	64
8.4.	Definition und Eigenschaften des Erwartungswerts	69
8.5.	Diskrete und absolut stetige Verteilungen	71
8.6.	Beispiele von absolut stetigen Verteilungen	72
8.7.	Singuläre Verteilungen	78
8.8.	Zerlegungssatz von Lebesgue	80
8.9.	Verteilungsfunktion eines Zufallsvektors	81
8.10.	Diskrete und absolut stetige Zufallsvektoren	82
8.11.	Randverteilungen eines Zufallsvektors	82
8.12.	Unabhängigkeit und Produktformeln	84
8.13.	Transformationsformel für die Dichte	85
8.14.	Faltungsformeln	86
8.15.	Transformationsformel für den Erwartungswert	91
8.16.	Multiplikativität des Erwartungswerts	93
Kapitel 9.	Varianz und Kovarianz	95
9.1.	Varianz	95
9.2.	Kovarianz und Korrelationskoeffizient	98
Kapitel 10.	Gesetz der großen Zahlen	105
10.1.	Zwei Beispiele	105
10.2.	Konvergenz in Wahrscheinlichkeit und L^2 -Konvergenz	105
10.3.	Ungleichungen von Markow und Tschebyschew	106
10.4.	Schwaches Gesetz der großen Zahlen	107
10.5.	Fast sichere Konvergenz	108
10.6.	Starkes Gesetz der großen Zahlen: Erste Version	111
10.7.	Starkes Gesetz der großen Zahlen: Zweite Version	114
10.8.	Der Fall eines unendlichen Erwartungswerts	120
10.9.	Anwendungen des Gesetzes der großen Zahlen	121
Kapitel 11.	Ungleichungen	129
11.1.	Jensen-Ungleichung	129
11.2.	Ljapunow-Ungleichung	130
11.3.	Young-Ungleichung	131
11.4.	Hölder-Ungleichung	131
11.5.	Minkowski-Ungleichung	132
11.6.	L^p -Räume und L^p -Konvergenz	133
Kapitel 12.	Analytische Methoden	135
12.1.	Erzeugende Funktion	135
12.2.	Summen mit einer zufälligen Anzahl von Summanden	138
12.3.	Verzweigungsprozesse	139

12.4.	Momenterzeugende Funktion (Laplace-Transformierte)	142
12.5.	Charakteristische Funktion (Fourier-Transformierte)	144
Kapitel 13.	Der zentrale Grenzwertsatz	153
13.1.	Konvergenz in Verteilung	153
13.2.	Eine Charakterisierung der Konvergenz in Verteilung	157
13.3.	Satz von Helly	159
13.4.	Stetigkeitssatz von Lévy	161
13.5.	Der zentrale Grenzwertsatz	163
13.6.	Beweis des zentralen Grenzwertsatzes	165
13.7.	Sätze von Lindeberg und Ljapunow	166
Kapitel 14.	Irrfahrt	175
14.1.	Berechnung einer Ruinwahrscheinlichkeit	175
14.2.	Rückkehr der Irrfahrt zum Ursprung	177
14.3.	Verteilung des Maximums der Irrfahrt	182
14.4.	Arcussinus-Gesetz	183
14.5.	Gesetz vom iterierten Logarithmus	185

Vorwort

Dies ist ein Skript zur Vorlesung “Elementare Wahrscheinlichkeitsrechnung und Statistik”, die an der Universität Ulm im Wintersemester 2012/13 gehalten wurde. Für die Erstellung der L^AT_EX-Version des Skripts bedanke ich mich bei Herrn Uli Armbruster, Frau Linda Bolay und Frau Melanie Herz.

Zakhar Kabluchko

Literatur

Es gibt sehr viele Lehrbücher über Wahrscheinlichkeitstheorie. Folgende Lehrbücher benutzen keine oder wenig Maßtheorie:

1. H. Dehling und B. Haupt. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Springer-Verlag.
2. U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg-Verlag.
3. K. Bosch. *Elementare Einführung in die Wahrscheinlichkeitsrechnung: Mit 82 Beispielen und 73 Übungsaufgaben mit vollständigem Lösungsweg*. Vieweg-Verlag.
4. N. Henze. *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls. Mit über 220 Übungsaufgaben und Lösungen*. Vieweg-Verlag.
5. A. Wakolbinger und G. Kersting. *Elementare Stochastik*. Springer-Verlag.
6. O. Häggström. *Streifzüge durch die Wahrscheinlichkeitstheorie*. Springer-Verlag.

Hier ist eine Liste von Büchern, die die Maßtheorie benutzen:

1. H.-O. Georgii. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. De Gruyter.
2. H. Bauer. *Wahrscheinlichkeitstheorie*. De Gruyter.
3. R. Durrett. *Probability: Theory and Examples*. Cambridge University Press.
4. A. Klenke. *Wahrscheinlichkeitstheorie*. Springer-Verlag.
5. G. Grimmett and R. Stirzaker. *Probability and Random Processes*. Oxford University Press.
6. A. Gut. *Probability: A graduate course*. Springer-Verlag.

Folgendes Buch von Feller ist ein Klassiker:

1. W. Feller. *An introduction to probability theory and its applications*. Vol. I/II. Wiley and Sons.

Grundbegriffe der Wahrscheinlichkeitstheorie

1.1. Zufallsexperimente, Ausgänge, Grundmenge

In der Stochastik betrachten wir Zufallsexperimente. Die Ausgänge eines Zufallsexperiments fassen wir zu einer Menge zusammen. Diese Menge bezeichnen wir mit Ω und nennen sie die Grundmenge des Experiments.

BEISPIEL 1.1.1. Das einfachste Beispiel eines Zufallsexperiments ist das Werfen einer Münze. Die Münze hat zwei Seiten, die wir “Kopf” und “Zahl” nennen und mit K bzw. Z abkürzen. Es gibt also zwei Ausgänge: K und Z . Die Grundmenge besteht aus zwei Elementen:

$$\Omega = \{K, Z\}.$$

BEISPIEL 1.1.2. Ein anderes Zufallsexperiment ist das Werfen eines Würfels. Der Würfel hat 6 Seiten, die mit den Zahlen $1, \dots, 6$ beschriftet sind. Das Experiment hat also 6 Ausgänge und die Grundmenge ist

$$\Omega = \{1, \dots, 6\}.$$

BEISPIEL 1.1.3. Nun erweitern wir das Experiment. Werden zwei Münzen geworfen, so erhalten wir eine aus 4 Elementen bestehende Grundmenge

$$\Omega = \{(K, K), (K, Z), (Z, K), (Z, Z)\}.$$

Werden nun drei Münzen geworfen, so besteht die Grundmenge aus 8 Elementen:

$$\Omega = \{(K, K, K), (K, K, Z), (K, Z, K), (Z, K, K), (K, Z, Z), (Z, K, Z), (Z, Z, K), (Z, Z, Z)\}.$$

Wenn wir nun allgemeiner n Münzen werfen, so ergibt sich für die Grundmenge

$$\Omega = \{K, Z\}^n \stackrel{\text{def}}{=} \{(a_1, \dots, a_n) : a_1, \dots, a_n \in \{K, Z\}\}.$$

Diese Grundmenge besteht aus 2^n Ausgängen, also $\#\Omega = 2^n$.

Wir können das obige Beispiel verallgemeinern.

BEISPIEL 1.1.4. Wir betrachten ein beliebiges Experiment mit Grundmenge E . Dieses Experiment soll n -mal durchgeführt werden. Die Grundmenge Ω ergibt sich dann zu

$$\Omega = E^n \stackrel{\text{def}}{=} \{(e_1, \dots, e_n) : e_i \in E\}.$$

Hier ist die Anzahl Ausgänge $\#\Omega = (\#E)^n$.

Noch allgemeiner können wir auch verschiedene Experimente durchführen.

BEISPIEL 1.1.5 (Produktexperiment). Wir führen n Experimente mit Grundmengen E_1, \dots, E_n *unabhängig* voneinander aus. Die Grundmenge ist dann ein sogenanntes kartesisches Produkt

$$\Omega = E_1 \times \dots \times E_n \stackrel{\text{def}}{=} \{(e_1, \dots, e_n) : e_1 \in E_1, \dots, e_n \in E_n\}.$$

Die Anzahl der Ausgänge ist $\#\Omega = (\#E_1) \cdot \dots \cdot (\#E_n)$.

BEISPIEL 1.1.6. Werfen wir eine Münze und einen Würfel, so haben wir $E_1 = \{K, Z\}$, $E_2 = \{1, \dots, 6\}$ und das kartesische Produkt $\Omega = E_1 \times E_2$ besteht aus $2 \cdot 6 = 12$ Elementen:

(K,1)	(K,2)	(K,3)	(K,4)	(K,5)	(K,6)
(Z,1)	(Z,2)	(Z,3)	(Z,4)	(Z,5)	(Z,6)

BEISPIEL 1.1.7. Ein weiteres, einfacheres Beispiel für ein Zufallsexperiment ist das Geschlecht eines Kindes bei der Geburt, also

$$\Omega = \{\text{Junge, Mädchen}\}.$$

BEISPIEL 1.1.8. Wir stellen uns eine Versicherung vor, bei welcher n Personen versichert sind. Jede dieser Personen wird einen Schaden melden, oder eben nicht. Daher ist dies vergleichbar mit einem n -maligen Münzwurf.

In den obigen Beispielen ist die Grundmenge endlich. Man kann sich auch Experimente mit einer unendlichen Grundmenge vorstellen.

BEISPIEL 1.1.9. Ein Spieler hat einen Würfel und würfelt so lange, bis er die erste 6 würfelt. Prinzipiell könnte dies unendlich lange dauern. Als Ausgang des Experiments betrachten wir die Anzahl der Würfe. Daher ist hier die Grundmenge

$$\Omega = \mathbb{N} \cup \{\infty\}.$$

1.2. Ereignisse

DEFINITION 1.2.1. Ein *Ereignis* ist eine Teilmenge der Grundmenge Ω .

BEISPIEL 1.2.2. Wir betrachten wieder das einfache Würfeln. Die Grundmenge ist $\Omega = \{1, \dots, 6\}$. Dann gibt es beispielsweise folgende Ereignisse:

$$A = \text{“eine gerade Zahl wird gewürfelt”} = \{2, 4, 6\},$$
$$B = \text{“eine ungerade Zahl wird gewürfelt”} = \{1, 3, 5\}.$$

BEISPIEL 1.2.3. Nun würfeln wir zweimal. Die Grundmenge ist $\Omega = \{1, \dots, 6\}^2$ mit $\#\Omega = 36$. Nun wollen wir als die Summe der Augenzahlen beispielsweise 10 haben. Dieses Ereignis kann sich durch 3 Wurfkombinationen ergeben, nämlich

$$A = \{(6, 4), (5, 5), (4, 6)\}.$$

Hier ist zu beachten, dass es sich bei $(6, 4)$ und $(4, 6)$ um verschiedene Ausgänge handelt.

BEISPIEL 1.2.4. Als Spezialfälle existieren:

- (1) unmögliches Ereignis, welches nie eintritt, $A = \emptyset$.
- (2) sicheres Ereignis, welches immer eintritt, $A = \Omega$.

DEFINITION 1.2.5. Ein *Elementarereignis* ist ein aus nur einem Element bestehendes Ereignis, also $A = \{\omega\}$ mit $\omega \in \Omega$. Jedes Ereignis setzt sich somit aus Elementarereignissen zusammen.

BEMERKUNG 1.2.6. Die Anzahl der möglichen Ereignisse errechnet sich durch $2^{\#\Omega}$.

Seien $A, B \subset \Omega$ Ereignisse. Mit mengentheoretischen Operationen lassen sich weitere Ereignisse konstruieren, nämlich

- $A \cup B = \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\}$: “ A tritt ein *oder* B tritt ein”.
- $A \cap B = \{\omega \in \Omega : \omega \in A \text{ und } \omega \in B\}$: “ A tritt ein *und* B tritt ein”.
- $A^c = \{\omega \in \Omega | \omega \notin A\}$: “ A tritt *nicht* ein” (Komplement von A).
- $B^c = \{\omega \in \Omega | \omega \notin B\}$: “ B tritt *nicht* ein” (Komplement von B).
- $A \setminus B = \{\omega \in \Omega : \omega \in A, \omega \notin B\}$: “ A tritt ein, aber B tritt nicht ein”.
- $B \setminus A = \{\omega \in \Omega : \omega \in B, \omega \notin A\}$: “ B tritt ein, aber A tritt nicht ein”.
- $A \Delta B = (A \setminus B) \cup (B \setminus A)$: “ A tritt ein *oder* B tritt ein, *aber nicht beide*” (symmetrische Differenz).

BEMERKUNG 1.2.7. Die Mengendifferenz \setminus ist nicht kommutativ: $A \setminus B \neq B \setminus A$. Es gilt $A \setminus B = A \cap B^c$ und $A^c = \Omega \setminus A$.

DEFINITION 1.2.8. Zwei Ereignisse A und B heißen *disjunkt*, falls $A \cap B = \emptyset$.

BEISPIEL 1.2.9. Folgende Paare von Ereignissen sind disjunkt:

- $A \setminus B$ und $B \setminus A$.
- A und A^c .
- A und \emptyset .

DEFINITION 1.2.10. Wir schreiben $A \subset B$, falls alle Elemente von A auch in B enthalten sind.

BEISPIEL 1.2.11. Wir betrachten ein Experiment, bei dem zwei Münzen geworfen werden. Man kann auch eine Münze zweimal werfen. Wir betrachten folgende Ereignisse:

$$A = \text{“erste Münze zeigt Kopf”} = \{KK, KZ\},$$

$$B = \text{“zweite Münze zeigt Kopf”} = \{KK, ZK\}.$$

Nun können wir diese beiden Ereignisse verknüpfen:

$$A \cap B = \text{“beide Münzen zeigen Kopf”} = \{KK\},$$

$$A \cup B = \text{“mindestens eine Münze zeigt Kopf”} = \{KK, KZ, ZK\},$$

$$A \Delta B = \text{“genau eine Münze zeigt Kopf”} = \{KZ, ZK\}.$$

Beachte, dass $KK \notin A \Delta B$. Man kann weitere Ereignisse definieren:

$$\text{“beide Münzen zeigen Zahl”} = A^c \cap B^c,$$

$$\text{“keine Münze zeigt Kopf”} = (A \cup B)^c.$$

Diese Ereignisse sind gleich. Analog sind die folgenden Ereignisse gleich:

$$\text{“nicht beide Münzen zeigen Kopf”} = (A \cap B)^c,$$

$$\text{“mindestens eine Münze zeigt Zahl”} = A^c \cup B^c.$$

SATZ 1.2.12 (De Morgan Regeln). Für beliebige Ereignisse $A, B \subset \Omega$ gilt

- (1) $(A \cup B)^c = A^c \cap B^c$.
- (2) $(A \cap B)^c = A^c \cup B^c$.

BEWEIS. Zu (1): $\omega \in (A \cup B)^c \Leftrightarrow \omega \notin A \cup B \Leftrightarrow \omega \notin A, \omega \notin B \Leftrightarrow \omega \in A^c \cap B^c$.
Beweis von (2) ist analog. □

BEMERKUNG 1.2.13. Man kann die Regeln auf beliebige Anzahl von Ereignissen verallgemeinern: Für beliebige Ereignisse $A_1, \dots, A_n \subset \Omega$ gilt

- (1) $(A_1 \cup \dots \cup A_n)^c = A_1^c \cap \dots \cap A_n^c$.
- (2) $(A_1 \cap \dots \cap A_n)^c = A_1^c \cup \dots \cup A_n^c$.

SATZ 1.2.14. Für beliebige Ereignisse $A, B, C \subset \Omega$ gelten folgende Gesetze:

- (1) Gesetz der doppelten Negation: $(A^c)^c = A$.
- (2) Erstes Distributivgesetz: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- (3) Zweites Distributivgesetz: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

BEWEIS. Zu (2): $x \in (A \cap B) \cup (A \cap C) \Leftrightarrow x \in A \cap B$ oder $x \in A \cap C \Leftrightarrow x \in A$ und ($x \in B$ oder $x \in C$) $\Leftrightarrow x \in A \cap (B \cup C)$. \square

1.3. Wahrscheinlichkeiten

BEISPIEL 1.3.1. Buffon und Pearson haben mit der Münze experimentiert:

- Buffon: 4040 Münzwürfe, davon 2048 Kopf.
- Pearson: 24000 Münzwürfe, davon 12012 Kopf.

Also zeigte die Münze in beiden Fällen ungefähr in 50% aller Fälle Kopf. Deshalb sagt man, dass die Wahrscheinlichkeit von "Kopf" gleich $1/2$ ist, jedenfalls dann, wenn die Münze fair (symmetrisch) ist.

BEHAUPTUNG 1.3.2 (Empirisches Gesetz der großen Zahlen). *Betrachte ein Experiment mit der Grundmenge Ω und sei $A \subset \Omega$ ein Ereignis. Wir wiederholen das Experiment n -mal unabhängig voneinander. Sei $N_n(A)$ eine Variable, die zählt, wie oft das Ereignis A eingetreten ist. Dann existiert der Grenzwert*

$$\mathbb{P}[A] := \lim_{n \rightarrow \infty} \frac{N_n(A)}{n} \in [0, 1].$$

BEMERKUNG 1.3.3. Die Zahl $N_n(A)/n$ ist die *relative Häufigkeit* des Eintretens von A in n Experimenten. Die Zahl $\mathbb{P}[A]$ heißt die *Wahrscheinlichkeit* von A .

DEFINITION 1.3.4. Sei Ω eine endliche oder abzählbare Menge. Eine Funktion $p : \Omega \rightarrow [0, 1]$ heißt *Wahrscheinlichkeitsfunktion*, falls

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

Interpretation: $p(\omega)$ ist die Wahrscheinlichkeit des Ausgangs $\omega \in \Omega$.

BEMERKUNG 1.3.5. Die Wahrscheinlichkeit eines Ereignisses $A \subset \Omega$ ist definiert durch

$$\mathbb{P}[A] \stackrel{\text{def}}{=} \sum_{\omega \in A} p(\omega).$$

DEFINITION 1.3.6. Bei einem *Laplace-Experiment* nehmen wir an, dass alle Ausgänge die gleiche Wahrscheinlichkeit haben. Sei $\#\Omega = n$ endlich, dann gilt

$$p(\omega) = \frac{1}{n} \text{ für alle } \omega \in \Omega.$$

Somit gilt für jedes Ereignis $A \subset \Omega$:

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega}.$$

BEISPIEL 1.3.7. Wir würfeln mit einem fairen (=symmetrischen) Würfel zweimal. Die Grundmenge ergibt sich dann zu $\Omega = \{1, \dots, 6\}^2$ mit $\#\Omega = 6^2 = 36$.

- Für das Ereignis $A = \text{“Augensumme} = 10\text{”} = \{(4, 6), (5, 5), (6, 4)\}$ ergibt sich eine Wahrscheinlichkeit von

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega} = \frac{3}{36} = \frac{1}{12}.$$

- Für das Ereignis $B = \text{“Augensumme} = 11\text{”} = \{(6, 5), (5, 6)\}$ ergibt sich eine Wahrscheinlichkeit von

$$\mathbb{P}[B] = \frac{\#B}{\#\Omega} = \frac{2}{36} = \frac{1}{18}.$$

- Für das Ereignis $C = \text{“Augensumme} = 12\text{”} = \{(6, 6)\}$ ergibt sich eine Wahrscheinlichkeit von

$$\mathbb{P}[C] = \frac{\#C}{\#\Omega} = \frac{1}{36}.$$

BEMERKUNG 1.3.8. Nicht jedes Experiment ist ein Laplace-Experiment. Beispiele:

- (1) Das Werfen einer Reißzwecke ist kein Laplace-Experiment, da

$$p(\text{“Landung auf dem Kopf”}) \neq p(\text{“seitliche Landung”}).$$

- (2) Die Bestimmung der Blutgruppe ist kein Laplace-Experiment, da nicht alle Blutgruppen die gleiche Wahrscheinlichkeit haben.
- (3) Unfaire (=unsymmetrische) Würfel oder Münzen.

BEISPIEL 1.3.9 (Falsches Modell). Wir werfen zwei Münzen gleichzeitig. Es gibt drei mögliche Ausgänge:

$$\omega_1 = \text{“beide Kopf”}, \quad \omega_2 = \text{“beide Zahl”}, \quad \omega_3 = \text{“verschiedene Symbole”}.$$

Hieraus ergibt sich die Grundmenge $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Allerdings ist aus Erfahrung bekannt, dass die Laplace-Annahme $p(\omega_1) = p(\omega_2) = p(\omega_3) = \frac{1}{3}$ falsch ist. Das oben beschriebene Modell ist falsch.

Im richtigen Modell sind die Münzen unterscheidbar (man stelle sich vor, dass sie mit verschiedenen Farben, etwa rot und gelb, markiert sind). Das richtige Modell hat 4 mögliche Ausgänge:

$$\begin{aligned} \omega_1 &= \text{“beide Münzen zeigen Kopf”}, \\ \omega_2 &= \text{“beide Münzen zeigen Zahl”}, \\ \omega_3 &= \text{“rote Münze zeigt Kopf, gelbe Münze zeigt Zahl”}, \\ \omega_4 &= \text{“rote Münze zeigt Zahl, gelbe Münze zeigt Kopf”}. \end{aligned}$$

Beachte, dass ω_3 und ω_4 zwei verschiedene Ausgänge sind. Die Grundmenge ist

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\} = \{KK, ZZ, KZ, ZK\}$$

und $p(\omega) = \frac{1}{4}$ für alle $\omega \in \Omega$. Somit gilt

$$\mathbb{P}[\text{“verschiedene Symbole”}] = \frac{\#\{ZK, KZ\}}{4} = \frac{1}{2} \neq \frac{1}{3}.$$

DEFINITION 1.3.10. Sei Ω eine Grundmenge. Die Menge aller Ereignisse in Ω heißt die *Potenzmenge* von Ω und wird mit 2^Ω bezeichnet. Die Elemente der Potenzmenge sind also alle möglichen Ereignisse $A \subset \Omega$. Es gilt $\#2^\Omega = 2^{\#\Omega}$.

DEFINITION 1.3.11. Sei Ω eine endliche oder abzählbare Menge. Eine Funktion $\mathbb{P} : 2^\Omega \rightarrow [0, 1]$ heißt ein *Wahrscheinlichkeitsmaß* auf Ω , falls folgende zwei Bedingungen gelten:

- (1) $\mathbb{P}[\Omega] = 1$.
- (2) Für beliebige paarweise disjunkte Ereignisse $A_1, A_2, \dots \subset \Omega$ gilt

$$(1.3.1) \quad \mathbb{P}[\cup_{k=1}^{\infty} A_k] = \sum_{k=1}^{\infty} \mathbb{P}[A_k].$$

Dabei heißen Ereignisse $A_1, A_2, \dots \subset \Omega$ *paarweise disjunkt* (oder einfach *disjunkt*), falls $A_i \cap A_j = \emptyset$ für alle $i \neq j$. Eigenschaft (1.3.1) heißt *σ -Additivität*.

BEMERKUNG 1.3.12. Die Funktion \mathbb{P} ordnet jedem Ereignis $A \subset \Omega$ eine Zahl $\mathbb{P}[A]$ zu. Die Zahl $\mathbb{P}[A]$ heißt die *Wahrscheinlichkeit* des Ereignisses A . Die Wahrscheinlichkeit kann nur Werte im Intervall $[0, 1]$ annehmen.

DEFINITION 1.3.13. Ein Paar (Ω, \mathbb{P}) mit den oben aufgelisteten Eigenschaften heißt *diskreter Wahrscheinlichkeitsraum*.

BEMERKUNG 1.3.14. Das Wort “diskret” bezieht sich dabei auf die Forderung, dass Ω endlich oder abzählbar sein soll. Später werden wir auch allgemeinere (überabzählbare) Wahrscheinlichkeitsräume betrachten.

BEMERKUNG 1.3.15. Ist eine Wahrscheinlichkeitsfunktion $p : \Omega \rightarrow [0, 1]$ gegeben, so definiert $\mathbb{P}[A] := \sum_{\omega \in A} p(\omega)$ ein Wahrscheinlichkeitsmaß auf Ω . Umgekehrt, ist $\mathbb{P} : 2^\Omega \rightarrow [0, 1]$ ein Wahrscheinlichkeitsmaß auf Ω , so ist $p(\omega) := \mathbb{P}[\{\omega\}]$ eine Wahrscheinlichkeitsfunktion. Beide Begriffe sind somit äquivalent.

Wir leiten nun einige Eigenschaften von \mathbb{P} her.

LEMMA 1.3.16. *Unmögliches Ereignis hat Wahrscheinlichkeit 0. Das heißt, $\mathbb{P}[\emptyset] = 0$.*

BEWEIS. Setze $A_1 = A_2 = \dots = \emptyset$ in (1.3.1). Es ergibt sich $\mathbb{P}[\emptyset] = \mathbb{P}[\emptyset] + \mathbb{P}[\emptyset] + \dots$. Das kann nur für $\mathbb{P}[\emptyset] = 0$ gelten. \square

LEMMA 1.3.17 (Additivität). *Für beliebige paarweise disjunkte Ereignisse $A_1, \dots, A_n \subset \Omega$ gilt*

$$\mathbb{P}[\cup_{k=1}^n A_k] = \sum_{k=1}^n \mathbb{P}[A_k].$$

BEWEIS. Setze $A_{n+1} = A_{n+2} = \dots = \emptyset$ in (1.3.1) und benutze $\mathbb{P}[\emptyset] = 0$. \square

LEMMA 1.3.18. Für jedes Ereignis $A \subset \Omega$ gilt:

$$\mathbb{P}[A^c] = 1 - \mathbb{P}[A].$$

BEWEIS. Ereignisse A und A^c sind disjunkt und es gilt $A \cup A^c = \Omega$. Mit der Additivität ergibt sich $1 = \mathbb{P}[\Omega] = \mathbb{P}[A \cup A^c] = \mathbb{P}[A] + \mathbb{P}[A^c]$. \square

LEMMA 1.3.19. Für beliebige Ereignisse $A, B \subset \Omega$ gilt

$$\mathbb{P}[A \setminus B] = \mathbb{P}[A] - \mathbb{P}[A \cap B].$$

BEWEIS. Ereignisse $A \cap B$ und $A \setminus B$ sind disjunkt und es gilt $(A \cap B) \cup (A \setminus B) = A$. Mit der Additivität folgt $\mathbb{P}[A \cap B] + \mathbb{P}[A \setminus B] = \mathbb{P}[A]$. \square

LEMMA 1.3.20. Für beliebige Ereignisse $A, B \subset \Omega$ (nicht unbedingt disjunkt) gilt:

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

BEWEIS. Ereignisse $A \setminus B$ und B sind disjunkt und $(A \setminus B) \cup B = A \cup B$. Mit der Additivität folgt $\mathbb{P}[A \setminus B] + \mathbb{P}[B] = \mathbb{P}[A \cup B]$. Mit Lemma 1.3.19 folgt $\mathbb{P}[A] - \mathbb{P}[A \cap B] + \mathbb{P}[B] = \mathbb{P}[A \cup B]$.

LEMMA 1.3.21 (Siebformel). Für beliebige Ereignisse $A_1, A_2, \dots, A_n \subset \Omega$ (nicht unbedingt disjunkt) gilt

$$\mathbb{P}[A_1 \cup \dots \cup A_n] = \sum_i \mathbb{P}[A_i] - \sum_{i < j} \mathbb{P}[A_i \cap A_j] + \sum_{i < j < k} \mathbb{P}[A_i \cap A_j \cap A_k] - \dots + (-1)^{n+1} \mathbb{P}[A_1 \cap \dots \cap A_n].$$

LEMMA 1.3.22 (Monotonie). Für beliebige Ereignisse $A \subset B$ gilt $\mathbb{P}[A] \leq \mathbb{P}[B]$.

BEWEIS. Ereignisse A und $B \setminus A$ sind disjunkt und es gilt $A \cup (B \setminus A) = B$. Mit der Additivität folgt $\mathbb{P}[A] + \mathbb{P}[B \setminus A] = \mathbb{P}[B]$. Das Lemma folgt, denn $\mathbb{P}[B \setminus A] \geq 0$. \square

LEMMA 1.3.23 (Subadditivität). Für beliebige Ereignisse $A, B \subset \Omega$ (nicht unbedingt disjunkt) gilt

$$\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B].$$

BEWEIS. Dies folgt aus Lemma 1.3.20, denn $\mathbb{P}[A \cap B] \geq 0$. \square

LEMMA 1.3.24 (Subadditivität). Für beliebige Ereignisse $A_1, \dots, A_n \subset \Omega$ gilt:

$$\mathbb{P}[A_1 \cup \dots \cup A_n] \leq \mathbb{P}[A_1] + \dots + \mathbb{P}[A_n].$$

BEWEIS. Dies folgt aus Lemma 1.3.23 mit Induktion. \square

BEISPIEL 1.3.25 (Gegenereignis betrachten). Wir werfen 10 faire Münzen. Berechnen möchten wir die Wahrscheinlichkeit des Ereignisses A , welches wie folgt definiert ist:

$A =$ "mindestens eine Münze zeigt Kopf".

LÖSUNG. Die Grundmenge ist hier $\Omega = \{K, Z\}^{10}$ mit $\#\Omega = 2^{10}$. Das Komplement des Ereignisses A ist

$$A^c = \text{“keine Münze zeigt Kopf”} = \{Z \dots Z\}.$$

Somit besteht A^c aus nur einem Ausgang: $\#A^c = 1$. Daraus ergibt sich die Wahrscheinlichkeit für das Komplement von A zu $\mathbb{P}[A^c] = \frac{1}{2^{10}}$. Also errechnet sich die Wahrscheinlichkeit von A durch

$$\mathbb{P}[A] = 1 - \mathbb{P}[A^c] = 1 - \frac{1}{2^{10}}.$$

2.2. Urnenmodelle

Es sei eine Urne mit n Bällen gegeben. Die Bälle seien mit $1, \dots, n$ beschriftet. Wir betrachten folgendes Zufallsexperiment: es wird k Mal ein Ball aus der Urne gezogen und seine Nummer notiert. Es gibt nun 4 Möglichkeiten:

- die Bälle werden mit/ohne Zurücklegen gezogen;
- die Nummern werden mit/ohne Berücksichtigung der Reihenfolge notiert.

Modell 1: Ziehen mit Reihenfolge und mit Zurücklegen. “Ziehen mit Zurücklegen” heißt, dass nach jeder Ziehung der gezogene Ball zurück in die Urne gelegt wird. Insbesondere kann ein Ball mehrmals aus der Urne gezogen werden. “Ziehen mit Reihenfolge” heißt, dass zwei Ausgänge des Experiments auch dann als unterschiedlich angesehen werden, wenn sie sich nur durch die Reihenfolge der gezogenen Bälle unterscheiden. Zum Beispiel gibt es für $n = 4$ Bälle und $k = 2$ Ziehungen folgende Ausgänge:

(1,1)	(1,2)	(1,3)	(1,4)
(2,1)	(2,2)	(2,3)	(2,4)
(3,1)	(3,2)	(3,3)	(3,4)
(4,1)	(4,2)	(4,3)	(4,4)

Beachte: Elemente $(1, 1), \dots, (4, 4)$ sind präsent (da Ziehen mit Zurücklegen). Elemente $(1, 2)$ und $(2, 1)$ gelten als verschiedene Elemente (da Ziehen mit Reihenfolge).

Beim Ziehen mit Reihenfolge und mit Zurücklegen handelt sich um ein Produktexperiment: das Ziehen eines Balls aus einer Urne mit n Bällen wird k Mal unter gleichen Bedingungen wiederholt. Die Grundmenge kann somit wie folgt dargestellt werden:

$$\Omega = \{(a_1, \dots, a_k) : a_i \in \{1, \dots, n\}\}.$$

Es gilt $\#\Omega = n^k$.

BEISPIEL 2.2.1. k -maliges Würfeln. Man stelle sich eine Urne mit 6 Bällen $1, \dots, 6$ vor. Anstatt einmal zu würfeln kann man auch einen Ball aus dieser Urne ziehen. Anstatt k -mal zu würfeln, kann man das Ziehen k -mal wiederholen. Da eine Augenzahl mehrmals gewürfelt werden kann, müssen die Bälle zurück in die Urne gelegt werden. Beim Würfeln müssen die Ausgänge “zuerst 1 gewürfelt, dann 2” und “zuerst 2 gewürfelt, dann 1” als unterschiedlich angesehen werden. Also wird die Reihenfolge berücksichtigt.

BEISPIEL 2.2.2. Geburtstage von k Studenten. Die möglichen Geburtstage können als $n = 365$ Bälle dargestellt werden.

Modell 2: Ziehen mit Reihenfolge und ohne Zurücklegen. “Ziehen ohne Zurücklegen” heißt, dass ein aus der Urne gezogener Ball nicht mehr in die Urne gelegt wird. Insbesondere kann jeder Ball höchstens einmal gezogen werden. Zum Beispiel ergeben sich für $n = 4$ Bälle und $k = 2$ Ziehungen folgende Möglichkeiten:

	(1,2)	(1,3)	(1,4)
(2,1)		(2,3)	(2,4)
(3,1)	(3,2)		(3,4)
(4,1)	(4,2)	(4,3)	

Beachte: Elemente $(1, 1), \dots, (4, 4)$ sind nicht präsent, da wir ohne Zurücklegen ziehen. Elemente $(1, 2)$ und $(2, 1)$ gelten als unterschiedlich, da die Reihenfolge berücksichtigt wird.

Die Grundmenge kann wie folgt dargestellt werden:

$$\Omega = \{(a_1, \dots, a_k) : a_i \in \{1, \dots, n\}, a_i \neq a_j \text{ für } i \neq j\}.$$

Elemente von Ω können als geordnete k -elementige Teilmengen von $\{1, \dots, n\}$ angesehen werden. Die Anzahl der Ausgänge in Ω kann so bestimmt werden: für die erste Ziehung gibt es n Möglichkeiten, für die zweite $n - 1$, usw. Für die letzte Ziehung gibt es $n - k + 1$ Möglichkeiten. Somit gilt

$$\#\Omega = n(n - 1) \cdot \dots \cdot (n - k + 1) \stackrel{\text{def}}{=} (n)_k.$$

BEMERKUNG 2.2.3. Für $k > n$ hat das Experiment keinen Sinn: wir können nicht mehr Bälle ziehen, als es in der Urne gibt. Es gilt dann logischerweise $(n)_k = 0$ (das Experiment hat keine Ausgänge).

BEMERKUNG 2.2.4. Ein wichtiger Spezialfall ist der Fall $k = n$. In diesem Fall wird jeder Ball aus der Urne genau einmal gezogen, es geht nur darum, in welcher Reihenfolge das geschieht. Die Ausgänge sind somit Mögliche Permutationen von n Bällen. Zum Beispiel gibt es für $n = 3$ Bälle folgende 6 Möglichkeiten:

$$(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1).$$

Die Anzahl der Permutationen von n unterscheidbaren Objekten ist $n! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \dots \cdot n$. Das kann man wie folgt sehen: an die erste Stelle kann man n mögliche Objekte Stellen, für die zweite Stelle kann man aus $(n - 1)$ möglichen Objekten auswählen, usw.

Modell 3: Ziehen ohne Reihenfolge und ohne Zurücklegen. “Ziehen ohne Reihenfolge” heißt, dass 2 Ausgänge, die sich nur durch die Reihenfolge der gezogenen Bälle unterscheiden, als gleich gelten. “Ziehen ohne Zurücklegen” heißt, dass ein Ball höchstens einmal gezogen werden kann. Für $n = 4$ Bälle und $k = 2$ Ziehungen ergeben sich folgende Möglichkeiten:

	(1,2)	(1,3)	(1,4)
		(2,3)	(2,4)
			(3,4)

Beachte: Ausgänge $(1, 1), \dots, (4, 4)$ sind nicht präsent (da Ziehen ohne Zurücklegen). Die Ausgänge $(1, 2)$ und $(2, 1)$ gelten als gleich (da die Reihenfolge nicht berücksichtigt wird). Deshalb haben wir in der obigen Tabelle nur einen dieser beiden Ausgänge aufgeführt.

Die Grundmenge kann so dargestellt werden:

$$\Omega = \{(a_1, \dots, a_k) : a_1 < \dots < a_k, a_i \in \{1, \dots, n\}\}$$

Alternativ kann man sich Ω als die Menge aller (ungeordneten) k -elementigen Teilmengen von $\{1, \dots, n\}$ vorstellen:

$$\Omega = \{\{a_1, \dots, a_k\} : a_i \in \{1, \dots, n\}, a_i \neq a_j \text{ für } i \neq j\}.$$

Zur Erinnerung: in einer Menge sind die Elemente nach Definition nicht geordnet, so dass z.B. $\{1, 2\} = \{2, 1\}$. Die Anzahl der Elemente von Ω kann wie folgt bestimmt werden. Zuerst können wir *mit* Reihenfolge und mit Zurücklegen ziehen. Es gibt $(n)_k$ Ausgänge. Nun müssen wir aber die Reihenfolge vergessen. Das heißt, wir müssen Ausgänge, die sich nur

durch Permutationen unterscheiden, identifizieren (z.B. müssen $(1, 2)$ und $(2, 1)$ identifiziert werden). Da man $k!$ Permutationen von k Elementen hat, werden jeweils $k!$ Ausgänge zu einem Ausgang zusammengefasst. Es gilt also

$$\#\Omega = \frac{\binom{n}{k}}{k!} = \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{k!} \stackrel{\text{def}}{=} \binom{n}{k}.$$

DEFINITION 2.2.5. Der *Binomialkoeffizient* $\binom{n}{k}$ ist die Anzahl der k -elementigen Teilmengen einer n -elementigen Menge:

$$\binom{n}{k} = \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

BEISPIEL 2.2.6 (Lotto). Aus einer Urne mit 49 Kugeln mit den Nummern $1, 2, \dots, 49$ werden 6 Kugeln ohne Zurücklegen gezogen. Um zu gewinnen, muss man die Nummern der gezogenen Kugeln erraten. Man tippt auf eine Kombination, etwa auf $(1, 2, \dots, 6)$. Wie wahrscheinlich ist das Ereignis

$$A = \text{“man hat die richtige Kombination getippt”}.$$

Die Reihenfolge, in der die Kugeln gezogen werden, muss beim Lotto nicht erraten werden.

LÖSUNG 1. Stellen wir uns vor, dass alle 6 Kugeln *gleichzeitig*, mit einem Griff, aus der Urne gezogen werden. Es wird also eine 6-elementige Teilmenge von $\{1, 2, \dots, 49\}$ zufällig ausgewählt.

$$\Omega = \text{Menge aller 6-elementigen Teilmengen von } \{1, \dots, 49\}.$$

Es gilt somit $\#\Omega = \binom{49}{6}$. Nur eine Kombination (nämlich, $\{1, 2, \dots, 6\}$) führt dazu, dass man gewinnt. Somit gilt $\#A = 1$ und

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega} = \frac{1}{\binom{49}{6}} = \frac{1}{13983816} = 7,15 \cdot 10^{-8}.$$

LÖSUNG 2. Stellen wir uns vor, dass die Kugeln nacheinander gezogen werden und die Nummern der Kugeln mit Berücksichtigung der Reihenfolge notiert werden. Es gilt dann

$$\Omega = \text{Menge aller geordneten 6-elementigen Teilmengen von } \{1, \dots, 49\}.$$

Es gilt $\#\Omega = (49)_6 = 49 \cdot 48 \cdot \dots \cdot 44$. Nun führt aber nicht nur die Kombination $(1, 2, 3, 4, 5, 6)$ zum Gewinn, sondern zum Beispiel auch die Kombination $(2, 1, 3, 4, 5, 6)$, genauso wie jede andere Permutation von $(1, \dots, 6)$. Es gibt $6!$ solche Permutationen, also $\#A = 6!$. Wir erhalten

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega} = \frac{6!}{49 \cdot 48 \cdot \dots \cdot 44} = \frac{1}{\binom{49}{6}} = \frac{1}{13983816} = 7,15 \cdot 10^{-8}.$$

SATZ 2.2.7 (Eigenschaften der Binomialkoeffizienten). *Es gelten folgende Formeln:*

- (1) $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$.
- (2) $\binom{n}{k} = \binom{n}{n-k}$.
- (3) $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$.

KOROLLAR 2.2.8. *Es gilt $\sum_{k=0}^n \binom{n}{k} = 2^n$ und $\sum_{k=0}^n (-1)^k \binom{n}{k} = 0$.*

BEISPIEL 2.2.9. In einem Raum gibt es n Plätze. Der Raum wird von k Studenten betreten, die die Plätze besetzen. Auf einem Platz kann maximal 1 Student sitzen. Wieviele Möglichkeiten gibt es?

LÖSUNG. Das Problem ist nicht eindeutig gestellt. Sind die Studenten *unterscheidbar*, so handelt es sich um das Ziehen ohne Zurücklegen und *mit* Reihenfolge. Jeder Student “zieht” sich einen Platz. Ohne Zurücklegen, denn kein Platz kann zweimal gezogen werden. Mit Reihenfolge, denn die Studenten sind unterscheidbar: “Student A setzt sich auf Platz 1 und Student B auf Platz 2” ist ein anderer Ausgang, als “Student A setzt sich auf Platz 2 und Student B auf Platz 1”. Es gibt $(n)_k$ Kombinationen.

Sind die Studenten *ununterscheidbar*, so handelt es sich um das Ziehen ohne Zurücklegen und *ohne* Reihenfolge. Mit anderen Worten, es wird eine k -elementige Teilmenge von $\{1, \dots, n\}$ ausgewählt. Das sind dann die Plätze, die besetzt werden. Welcher Platz von wem besetzt wird, spielt keine Rolle, denn die Studenten sind ununterscheidbar. In diesem Fall gibt es $\binom{n}{k}$ Kombinationen.

Modell 4: Ziehen ohne Reihenfolge und mit Zurücklegen. Da wir mit Zurücklegen ziehen, kann ein Ball mehrmals gezogen werden. Da wir ohne Reihenfolge ziehen, achten wir nur darauf, wie oft jeder Ball gezogen wurde, nicht aber in welcher Reihenfolge das geschah. Für $n = 4$ Bälle und $k = 2$ Ziehungen ergeben sich folgende Möglichkeiten:

(1,1)	(1,2)	(1,3)	(1,4)
	(2,2)	(2,3)	(2,4)
		(3,3)	(3,4)
			(4,4)

Beachte: Elemente $(1, 1), \dots, (4, 4)$ sind präsent, denn es wird mit Zurücklegen gezogen. Elemente $(1, 2)$ und $(2, 1)$ gelten als identisch, denn die Reihenfolge wird nicht berücksichtigt. Deshalb haben wir nur $(1, 2)$ in der Tabelle aufgeführt. Die Grundmenge ist gegeben durch:

$$\Omega = \{(a_1, \dots, a_k) : a_1 \leq a_2 \leq \dots \leq a_k, a_i \in \{1, \dots, n\}\}.$$

Elemente von Ω können als ungeordnete k -elementige Teilmengen von $\{1, \dots, n\}$ angesehen werden, wobei Wiederholungen der Elemente in der Teilmenge erlaubt sind. Im nächsten Beispiel zeigen wir, dass

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

BEISPIEL 2.2.10. k Vögel setzen sich auf n Bäume. Mehrfachbesetzungen sind möglich. Die Vögel sind ununterscheidbar. Wie viele Besetzungen gibt es?

LÖSUNG. Wir werden Vögel als Kreuze darstellen. Vögel, die auf verschiedenen Bäumen sitzen, trennen wir durch eine Trennwand. Gibt es zum Beispiel $n = 4$ Bäume und sitzen auf diesen Bäumen 2, 3, 0, 1 Vögel, so stellen wir das wie folgt dar:

$$\times \times \mid \times \times \times \parallel \times .$$

Im Allgemeinen haben wir $n - 1$ Trennwände (da n Bäume), und k Kreuze (=Vögel). Insgesamt haben wir $n + k - 1$ Elemente. Aus diesen $n + k - 1$ Elementen müssen diejenigen k

Elemente ausgewählt werden, die Kreuze sind. Die Anzahl der Konfigurationen ist somit

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

BEISPIEL 2.2.11. Wieviele Möglichkeiten gibt es, eine Zahl k als Summe von n Summanden zu schreiben? Reihenfolge der Summanden wird berücksichtigt, Nullen sind erlaubt. Beispielsweise kann man $k = 4$ wie folgt als Summe von $n = 2$ Summanden darstellen:

$$4 = 4 + 0 = 3 + 1 = 2 + 2 = 1 + 3 = 0 + 4.$$

LÖSUNG. Jede Darstellung von k als Summe von n Summanden entspricht genau einer Besetzung von n Bäumen durch k ununterscheidbare Vögel. Dabei entspricht der i -te Summand der Anzahl der Vögel auf dem i -ten Baum. Somit gibt es $\binom{n+k-1}{k}$ Möglichkeiten.

2.3. Hypergeometrische Verteilung

BEISPIEL 2.3.1 (Hypergeometrische Verteilung). Betrachte einen Teich, in dem n Fische schwimmen. Von den Fischen seien n_1 rot und n_2 gelb, mit $n_1 + n_2 = n$. Ein Fischer fängt k verschiedene Fische (ohne Zurücklegen). Betrachte das Ereignis

$A =$ “es wurden genau k_1 rote Fische gefangen (und somit $k_2 := k - k_1$ gelbe)”.

Wie groß ist die Wahrscheinlichkeit von A ?

LÖSUNG. Die Grundmenge Ω ist die Menge aller k -elementigen Teilmengen von $\{1, \dots, n\}$. Somit gilt

$$\#\Omega = \binom{n}{k}.$$

Nun bestimmen wir die Anzahl der Elemente in A . Damit A eintritt, muss der Fischer k_1 rote und k_2 gelbe Fische fangen. Er kann sich aus der Menge der roten Fische k_1 Fische aussuchen, dafür gibt es $\binom{n_1}{k_1}$ Möglichkeiten. Dann kann er sich aus der Menge der gelben Fische k_2 Fische aussuchen, dafür gibt es $\binom{n_2}{k_2}$ Möglichkeiten. Da man jede Auswahl der roten Fische mit jeder Auswahl der gelben Fische beliebig kombinieren kann, ergibt sich für die Anzahl der Elemente in A :

$$\#A = \binom{n_1}{k_1} \cdot \binom{n_2}{k_2}.$$

Für die Wahrscheinlichkeit von A erhalten wir dann

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega} = \frac{\binom{n_1}{k_1} \cdot \binom{n_2}{k_2}}{\binom{n}{k}}.$$

Diese Formel nennt man die hypergeometrische Verteilung. Genauer: man sagt, dass die Anzahl der roten Fische, die der Fischer gefangen hat, eine hypergeometrische Verteilung hat.

BEISPIEL 2.3.2 (Lotto). Es werden 6 Kugeln aus einem Topf mit 49 nummerierten Kugeln ohne Zurücklegen gezogen. Man darf auf 6 verschiedene Nummern tippen. Bestimme die Wahrscheinlichkeit von

$A =$ “man hat genau 3 Nummern richtig getippt”.

Auf die Reihenfolge, in der die Kugeln gezogen werden, wird beim Lotto nicht getippt.

LÖSUNG. Kugeln = Fische. Ohne Einschränkung der Allgemeinheit tippen wir auf die Kugeln $1, 2, \dots, 6$. Das sind die roten Fische. Alle anderen Kugeln, nämlich $7, \dots, 49$, sind die gelben Fische. Die Kugeln liegen in der Urne (= die Fische schwimmen im Teich). Es werden 6 Kugeln zufällig ohne Zurücklegen gezogen (= 6 Fische gefangen). Wir interessieren uns für die Wahrscheinlichkeit, dass unter diesen 6 Fischen genau 3 rot sind. Es ergibt sich

$$\mathbb{P}[A] = \frac{\binom{6}{3} \cdot \binom{43}{3}}{\binom{49}{6}} \approx 0.01765.$$

Wir können das Beispiel mit den Fischen verallgemeinern.

BEISPIEL 2.3.3 (Eine allgemeinere Form der hypergeometrischen Verteilung). Wir betrachten einen Teich mit n Fischen. Jeder Fisch habe eine der $r \geq 2$ Farben. Es gebe im Teich n_1 Fische von Farbe 1, n_2 Fische von Farbe 2, \dots , n_r von Farbe r , wobei $n_1 + \dots + n_r = n$. Ein Fischer fängt ohne Zurücklegen k Fische. Man bestimme die Wahrscheinlichkeit des Ereignisses

$$A = \text{“es wurden } k_1 \text{ Fische von Farbe 1,} \\ k_2 \text{ Fische von Farbe 2,} \\ \dots \\ k_r \text{ Fische von Farbe } r \text{ gefangen”}.$$

Dabei seien k_1, \dots, k_r gegeben und es gelte $k_1 + \dots + k_r = k$.

LÖSUNG. Es gilt

$$\#\Omega = \binom{n}{k}, \quad \#A = \binom{n_1}{k_1} \cdot \dots \cdot \binom{n_r}{k_r}.$$

Somit erhalten wir

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega} = \frac{\binom{n_1}{k_1} \cdot \dots \cdot \binom{n_r}{k_r}}{\binom{n}{k}}.$$

BEISPIEL 2.3.4. Ein Kartenspiel aus 52 Karten wird auf 2 Spieler verteilt, jeder erhält 26 Karten. Bestimme die Wahrscheinlichkeit von

$$A = \text{“Erster Spieler erhält genau 3 Asse, genau 2 Könige und genau 1 Dame”}.$$

LÖSUNG. Die Grundmenge besteht aus allen 26-elementigen Teilmengen einer 52-elementigen Menge. Diese Teilmenge ist die Menge der Karten, die der erste Spieler bekommt, der zweite bekommt dann automatisch den Rest. Es gilt somit $\#\Omega = \binom{52}{26}$. Damit A eintritt, muss der erste Spieler 3 der 4 Asse, 2 der 4 Könige, 1 der 4 Damen, und 20 der 40 restlichen Karten bekommen. Somit erhalten wir

$$\#A = \binom{4}{3} \cdot \binom{4}{2} \cdot \binom{4}{1} \cdot \binom{40}{20}.$$

Die Wahrscheinlichkeit von A ist

$$\mathbb{P}[A] = \frac{\binom{4}{3} \cdot \binom{4}{2} \cdot \binom{4}{1} \cdot \binom{40}{20}}{\binom{52}{26}}.$$

2.4. Binomialverteilung und Multinomialverteilung

BEISPIEL 2.4.1. In einem Teich schwimmen n Fische, davon seien n_1 rot und n_2 gelb, mit $n_1 + n_2 = n$. Es werden k Fische *mit Zurücklegen* gefangen. Bestimme die Wahrscheinlichkeit des Ereignisses

$A =$ “Es wurde genau k_1 -mal ein roter Fisch aus dem Teich gezogen”.

BEMERKUNG 2.4.2. Definiert man $k_2 = k - k_1$, so kann man das Ereignis A auch so beschreiben:

$A =$ “Es wurde genau k_2 -mal ein gelber Fisch aus dem Teich gezogen”.

LÖSUNG. Es handelt sich um eine k -fache Wiederholung (unter gleichen Bedingungen) des Experiments “ein Fisch wird gezogen, Farbe notiert, Fisch freigelassen”. Somit gilt

$$\Omega = \{1, \dots, n\}^k.$$

Die Anzahl der Ausgänge ist $\#\Omega = n^k$. Um die Anzahl der Elemente in A zu bestimmen, schauen wir uns zuerst ein anderes Ereignis an:

$B =$ “Bei den ersten k_1 Versuchen wurden rote Fische gefangen
und bei den restlichen k_2 Versuchen wurden gelbe Fische gefangen”.

Der Unterschied zwischen den Ereignissen A und B besteht darin, dass bei B die Nummern der Versuche, bei denen rote (bzw. gelbe) Fische gefangen werden sollen, explizit angegeben sind. Bei A hingegen dürfen diese Nummern beliebig sein. Es gilt

$$\#B = n_1 \cdot n_1 \cdot \dots \cdot n_1 \cdot n_2 \cdot \dots \cdot n_2 = n_1^{k_1} \cdot n_2^{k_2}.$$

Somit errechnet sich die Wahrscheinlichkeit von B zu

$$\mathbb{P}[B] = \frac{n_1^{k_1} \cdot n_2^{k_2}}{n^k}.$$

Bei A kann man sich zusätzlich die Versuche, bei denen ein roter Fisch gefangen werden soll, frei aussuchen. Es gibt dafür $\binom{k}{k_1} = \binom{k}{k_2}$ Möglichkeiten. Somit besteht A aus $\binom{k}{k_1}$ disjunkten “Kopien” von B und wir erhalten

$$\#A = \binom{k}{k_1} \cdot \#B = \binom{k}{k_1} \cdot n_1^{k_1} \cdot n_2^{k_2}.$$

Für die Wahrscheinlichkeit von A erhalten wir

$$\mathbb{P}[A] = \binom{k}{k_1} \frac{n_1^{k_1} \cdot n_2^{k_2}}{n^k} = \binom{k}{k_1} \left(\frac{n_1}{n}\right)^{k_1} \left(\frac{n_2}{n}\right)^{k_2}.$$

Man sagt, dass die Anzahl der Versuche, bei denen ein roter Fisch gefangen wurde, binomialverteilt ist.

Nun werden wir das obige Beispiel erweitern, indem wir eine beliebige Anzahl an Farben zulassen. Dafür brauchen wir eine Verallgemeinerung der Binomialkoeffizienten.

BEISPIEL 2.4.3 (Multinomialkoeffizienten). Es seien k unterscheidbare (z.B. nummerierte) Gegenstände gegeben. Diese will man auf r unterscheidbare (z.B. nummerierte) Schubladen verteilen. In eine Schublade können mehrere Gegenstände gelegt werden. Leere Schubladen

sind zugelassen. Zwei Verteilungen, die sich nur durch die Reihenfolge der Gegenstände innerhalb der Schubladen unterscheiden, gelten als identisch. Wie viele Möglichkeiten gibt es, die Gegenstände auf die Schubladen zu verteilen, so dass die erste Schublade k_1 Gegenstände, die zweite k_2, \dots , die r -te Schublade k_r Gegenstände enthält? Dabei seien k_1, \dots, k_r vorgegeben mit $k_1 + \dots + k_r = k$.

LÖSUNG. Die gesuchte Anzahl N ist gegeben durch

$$N = \binom{k}{k_1} \binom{k - k_1}{k_2} \cdots \binom{k - k_1 - \dots - k_{r-1}}{k_r}.$$

Der Faktor $\binom{k}{k_1}$ ist die Anzahl der Möglichkeiten, die k_1 Gegenstände auszusuchen, die in die erste Schublade gelegt werden sollen. Danach stehen uns nur noch $k - k_1$ Gegenstände zu Verfügung. Der Faktor $\binom{k - k_1}{k_2}$ ist die Anzahl der Möglichkeiten, die k_2 Gegenstände auszuwählen, die in die zweite Schublade gelegt werden sollen. Und so weiter. Übrigens ist der letzte Faktor, nämlich $\binom{k - k_1 - \dots - k_{r-1}}{k_r}$, gleich 1, da wir bei der letzten Schublade keine Wahl mehr haben. Dies kann man schreiben als

$$N = \frac{k!}{k_1!(k - k_1)!} \cdot \frac{(k - k_1)!}{k_2!(k - k_1 - k_2)!} \cdots \frac{(k - k_1 - \dots - k_{r-1})!}{k_r!0!},$$

oder, nachdem Terme gekürzt wurden,

$$N = \frac{k!}{k_1! \cdot \dots \cdot k_r!}.$$

DEFINITION 2.4.4. Die *Multinomialkoeffizienten* sind definiert durch

$$\binom{k}{k_1, \dots, k_r} = \frac{k!}{k_1! \cdot \dots \cdot k_r!}.$$

Dabei wird $k_1 + \dots + k_r = k$ vorausgesetzt.

BEMERKUNG 2.4.5. Im Spezialfall $r = 2$ haben wir $\binom{k}{k_1, k_2} = \binom{k}{k_1} = \binom{k}{k_2}$, mit $k_1 + k_2 = k$.

SATZ 2.4.6 (Eigenschaften der Multinomialkoeffizienten). *Es gelten folgende Formeln:*

- (1) $\binom{k}{k_1, \dots, k_r} = \binom{k-1}{k_1-1, k_2, \dots, k_r} + \binom{k-1}{k_1, k_2-1, \dots, k_r} + \dots + \binom{k-1}{k_1, \dots, k_r-1}$.
- (2) $\binom{k}{k_1, \dots, k_r}$ ändert sich nicht, wenn man die Zahlen k_1, \dots, k_r permutiert.
- (3) $(x_1 + \dots + x_r)^k = \sum_{k_1 + \dots + k_r = k} \binom{k}{k_1, \dots, k_r} x_1^{k_1} \dots x_r^{k_r}$.

KOROLLAR 2.4.7. *Es gilt* $\sum_{k_1 + \dots + k_r = k} \binom{k}{k_1, \dots, k_r} = r^k$.

BEISPIEL 2.4.8. 33 Schüler sollen auf 3 Fußballmannschaften (jeweils 11 Schüler) verteilt werden. Wieviele Möglichkeiten gibt es?

LÖSUNG. Das Problem kann auf zwei verschiedene Weisen verstanden werden. Wenn die 3 Mannschaften (= Schubladen) *unterscheidbar* sind, gibt es

$$\binom{33}{11, 11, 11} = \frac{33!}{(11!)^3} = 136526995463040$$

Möglichkeiten. Unterscheidbar könnte z.B. heißen, dass die erste Mannschaft in der ersten Liga Spielen soll, die zweite in der zweiten, und die dritte in der dritten. In diesem Fall müssen zwei mögliche Verteilungen der Schüler auch dann als verschieden angesehen werden, wenn sie sich nur durch das Permutieren der Mannschaften unterscheiden.

Wenn die 3 Mannschaften *ununterscheidbar* sind, so gibt es weniger Möglichkeiten. Es müssen nämlich jeweils $3! = 6$ Möglichkeiten, die sich nur durch das Permutieren der Mannschaften unterscheiden, als gleich angesehen werden. Die Anzahl der Möglichkeiten ist dann gegeben durch

$$\frac{1}{3!} \binom{33}{11, 11, 11} = 22754499243840.$$

BEISPIEL 2.4.9 (Multinomialverteilung). In einem Teich schwimmen n Fische. Jeder dieser Fische hat eine der r möglichen Farben. Die Anzahl der Fische von Farbe i sei mit n_i bezeichnet, wobei $i = 1, \dots, r$. Dabei gelte $n_1 + \dots + n_r = n$. Ein Fischer fängt k Fische *mit Zurücklegen*. Bestimme die Wahrscheinlichkeit von

$$A = \begin{array}{l} \text{“es wurden } k_1 \text{ Fische von Farbe 1,} \\ \quad k_2 \text{ Fische von Farbe 2,} \\ \quad \dots \\ \quad k_r \text{ Fische von Farbe } r \text{ gefangen”}. \end{array}$$

LÖSUNG. Es handelt sich um ein Produktexperiment und somit gilt $\#\Omega = n^k$. Für die Anzahl der Elemente in A gilt

$$\#A = \binom{k}{k_1, \dots, k_r} \cdot n_1^{k_1} \cdot \dots \cdot n_r^{k_r}.$$

Somit erhalten wir

$$\mathbb{P}[A] = \binom{k}{k_1, \dots, k_r} \cdot \left(\frac{n_1}{n}\right)^{k_1} \cdot \dots \cdot \left(\frac{n_r}{n}\right)^{k_r}.$$

BEISPIEL 2.4.10. Es werde 12-mal mit einem fairen Würfel gewürfelt. Bestimme die Wahrscheinlichkeit des Ereignisses

$$A = \text{“Jede der 6 möglichen Augenzahlen wurde genau 2-mal gewürfelt”}.$$

LÖSUNG. Mit $n = 6$, $r = 6$, $k_1 = \dots = k_6 = 2$, $k = 12$ erhalten wir

$$\mathbb{P}[A] = \binom{12}{2, 2, 2, 2, 2, 2} \frac{1}{6^{12}}.$$

KAPITEL 3

Zufallsvariablen

Ein Zufallsexperiment ergibt oft eine zufällige Zahl. Diese Zahl bezeichnen wir als Zufallsvariable.

BEISPIEL 3.0.11. Beim zweimaligen Würfeln kann man folgende Zufallsvariablen betrachten: Augensumme, größere Augenzahl, kleinere Augenzahl, Differenz der Augenzahlen, ...

DEFINITION 3.0.12. Sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum. Dann heißt jede Funktion

$$X : \Omega \rightarrow \mathbb{R}$$

eine *Zufallsvariable*. Für $\omega \in \Omega$ heißt $X(\omega)$ der Wert von X zum Ausgang ω .

BEISPIEL 3.0.13. Wir betrachten das zweimalige Würfeln. Die Grundmenge ist dann

$$\Omega = \{(a_1, a_2) : a_1, a_2 \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^2.$$

Dann kann man z.B. folgende Zufallsvariablen $X_1 : \Omega \rightarrow \mathbb{R}$ und $X_2 : \Omega \rightarrow \mathbb{R}$ definieren:

$$X_1(a_1, a_2) = a_1 \quad (\text{“erste Augenzahl”})$$

$$X_2(a_1, a_2) = a_2 \quad (\text{“zweite Augenzahl”}).$$

Wird z.B. der Ausgang $(2, 6)$ gewürfelt, so nimmt X_1 den Wert 2 und X_2 den Wert 6 an. Definieren wir nun $Y = X_1 + X_2$, dann gilt

$$Y(a_1, a_2) = a_1 + a_2 \quad (\text{“die Augensumme”}).$$

Sei $Z = \max(X_1, X_2)$, dann gilt

$$Z(a_1, a_2) = \max(a_1, a_2) \quad (\text{“größere Augenzahl”}).$$

Hier sind X_1, X_2, Y, Z Beispiele von Zufallsvariablen.

DEFINITION 3.0.14. Sei $A \subset \Omega$ ein Ereignis. Die Indikatorfunktion von A ist die Zufallsvariable $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$, die wie folgt definiert wird:

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

Die Indikatorfunktion nimmt den Wert 1 genau dann an, wenn das Ereignis A eintritt. Ansonsten nimmt sie den Wert 0 an.

BEMERKUNG 3.0.15. Für beliebige Ereignisse $A, B \subset \Omega$ gelten folgende Eigenschaften:

- (1) $\mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$.
- (2) $\mathbb{1}_{A \cup B} = \max\{\mathbb{1}_A, \mathbb{1}_B\}$.
- (3) $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B$ falls $A \cap B = \emptyset$.
- (4) $\mathbb{1}_{A \Delta B} = \mathbb{1}_A + \mathbb{1}_B \pmod{2}$.

$$(5) \mathbb{1}_A + \mathbb{1}_{A^c} = 1 \text{ und } \mathbb{1}_A \cdot \mathbb{1}_{A^c} = 0.$$

DEFINITION 3.0.16. Sei $Z : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Die *Zähldichte* oder die *Verteilung* von Z ist die Funktion $p_Z : \mathbb{R} \rightarrow [0, 1]$ mit

$$p_Z(y) = \mathbb{P}[Z = y] = \mathbb{P}[\{\omega \in \Omega : Z(\omega) = y\}].$$

Mit anderen Worten, $p_Z(y)$ ist die Wahrscheinlichkeit, dass die Zufallsvariable Z den Wert y annimmt.

DEFINITION 3.0.17. Sei Z eine Zufallsvariable. Das Bild von Z ist die Menge

$$\text{Im}(Z) = \{y \in \mathbb{R} : \exists \omega \in \Omega \text{ mit } Z(\omega) = y\}.$$

Mit anderen Worten, $\text{Im } Z$ ist die Menge aller Werte von Z . Für $y \notin \text{Im}(Z)$ gilt $p_Z(y) = 0$.

BEMERKUNG 3.0.18. Für die Zähldichte gelten folgende zwei Eigenschaften:

- (1) $p_Z(y) \in [0, 1]$ für alle $y \in \mathbb{R}$.
- (2) $\sum_{y \in \text{Im}(Z)} p_Z(y) = 1$.

BEISPIEL 3.0.19. Es sei Z die Augensumme beim Würfeln mit 2 fairen Würfeln. Bestimme die Zähldichte von Z .

LÖSUNG. Die Grundmenge dieses Experiments ist $\Omega = \{1, \dots, 6\}^2$ mit $\#\Omega = 36$. Die Menge der Werte, die Z annehmen kann, ist $\text{Im}(Z) = \{2, \dots, 12\}$. Nun bestimmen wir die Zähldichte von Z :

$$\begin{aligned} p_Z(2) &= \mathbb{P}[Z = 2] = \frac{1}{36}, \\ p_Z(3) &= \mathbb{P}[Z = 3] = \frac{2}{36}, \\ &\vdots \\ p_Z(7) &= \mathbb{P}[Z = 7] = \frac{6}{36}, \\ &\vdots \\ p_Z(11) &= \mathbb{P}[Z = 11] = \frac{2}{36}, \\ p_Z(12) &= \mathbb{P}[Z = 12] = \frac{1}{36}. \end{aligned}$$

Zusammenfassend gilt

$$p_Z(y) = \begin{cases} \frac{y-1}{36}, & y \in \{2, \dots, 7\}, \\ \frac{12-(y-1)}{36}, & y \in \{7, \dots, 12\}. \end{cases}$$

BEISPIEL 3.0.20. Sei $A \subset \Omega$ ein Ereignis und $Z = \mathbb{1}_A$ die Indikatorfunktion von A . Für die Zähldichte von Z gilt:

$$p_Z(y) = \begin{cases} 1 - \mathbb{P}[A], & y = 0, \\ \mathbb{P}[A], & y = 1, \\ 0, & \text{sonst.} \end{cases}$$

Unabhängigkeit

4.1. Unabhängigkeit von Ereignissen

Wir stellen uns vor, dass zwei Personen jeweils eine Münze werfen. In vielen Fällen kann man annehmen, dass die eine Münze die andere nicht beeinflusst (d.h., es gibt keine Interaktion zwischen den Münzen). Das bedeutet, dass jedes Ereignis, das mit der ersten Münze zusammenhängt, von jedem Ereignis der anderen Münze unabhängig ist.

Welche Ereignisse sind dann abhängig? Man stelle sich vor, die beiden Münzen sind durch einen Stab miteinander verbunden, sodass sie immer das gleiche Symbol zeigen. In diesem Fall hängt das Symbol, das die erste Münze zeigt, vom Symbol der zweiten Münze ab.

Nun stellen wir uns vor, dass die erste Münze in 60% aller Fälle Kopf zeigt, während die zweite Münze in 50% aller Fälle Kopf zeigt. Sind die Münzen unabhängig, so müsste das Ereignis “beide Münzen zeigen Kopf” in 50% derjenigen Fälle eintreten, wo die erste Münze Kopf zeigt. Das heißt, die Wahrscheinlichkeit, dass beide Münzen Kopf zeigen, sollte sich als das Produkt $0.6 \cdot 0.5$ errechnen. Wir werden diese Überlegung als eine Definition benutzen.

DEFINITION 4.1.1. Seien (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum und $A, B \subset \Omega$ zwei Ereignisse. Dann heißen A und B *unabhängig*, wenn

$$(4.1.1) \quad \mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

Wenn dies nicht gilt, heißen die Ereignisse A und B abhängig.

Es zeigt sich, dass manche Ereignisse, die im ersten Moment abhängig erscheinen, doch unabhängig sind.

BEISPIEL 4.1.2. Wir betrachten das einmalige Werfen eines fairen Würfels und legen folgende Ereignisse fest:

$$A = \text{“Augenzahl ist } \geq 5\text{”} = \{5, 6\}, \quad B = \text{“Augenzahl ist gerade”} = \{2, 4, 6\}.$$

Es gilt dann

$$\mathbb{P}[A] = \frac{2}{6} = \frac{1}{3}, \quad \mathbb{P}[B] = \frac{3}{6} = \frac{1}{2}.$$

Sind diese beiden Ereignisse nun unabhängig oder nicht? Es gilt $A \cap B = \{6\}$ und somit

$$\mathbb{P}[A \cap B] = \frac{1}{6} = \frac{1}{3} \cdot \frac{1}{2} = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

Es folgt, dass A und B per Definition unabhängig sind (obwohl es intuitiv nicht so scheint). Betrachte nun zusätzlich das Ereignis

$$C = \text{“Augenzahl ist } \geq 4\text{”} = \{4, 5, 6\}.$$

Man sieht, dass $\mathbb{P}[B] = \mathbb{P}[C] = \frac{1}{2}$, während $\mathbb{P}[B \cap C] = \frac{2}{6} \neq \mathbb{P}[B] \cdot \mathbb{P}[C]$. Ereignisse B und C sind somit abhängig.

BEISPIEL 4.1.3. Wir betrachten das Würfeln mit 3 fairen Würfeln. Die Würfel muss man sich immer als unterscheidbar vorstellen. Beispielsweise, kann man sie mit verschiedenen Farben färben. Dieses Experiment ergibt 3 Augenzahlen. Die Grundmenge dieses Experiments ist $\Omega = \{1, \dots, 6\}^3$ mit $\#\Omega = 6^3$. Nun können wir 3 Zufallsvariablen X_1, X_2, X_3 definieren:

$$X_i : \Omega \rightarrow \mathbb{R}, \quad X_i(a_1, a_2, a_3) = a_i = \text{“Augenzahl, die Würfel } i \text{ zeigt”},$$

wobei $i = 1, 2, 3$ und $a_1, a_2, a_3 \in \{1, \dots, 6\}$. Nun definieren wir zwei Ereignisse A und B :

$$A = \{X_1 = X_2\} = \text{“die ersten beiden Würfel zeigen die gleiche Augenzahl”},$$

$$B = \{X_2 = X_3\} = \text{“der zweite und dritte Würfel zeigen die gleiche Augenzahl”}.$$

Sind die Ereignisse A und B nun unabhängig oder abhängig? Es scheint, dass beide Ereignisse von der zweiten Augenzahl X_2 abhängen. Dennoch sind sie unabhängig, was wir nun nachweisen. Es gilt

$$A = \{(a, a, b) : a, b \in \{1, \dots, 6\}\},$$

$$B = \{(a, b, b) : a, b \in \{1, \dots, 6\}\}.$$

Somit erhalten wir, dass $\#A = \#B = 6^2$ und

$$\mathbb{P}[A] = \mathbb{P}[B] = \frac{6^2}{6^3} = \frac{1}{6}.$$

Für das Ereignis $A \cap B$ gilt

$$A \cap B = \{X_1 = X_2 = X_3\} = \{(a, a, a) : a \in \{1, \dots, 6\}\}.$$

Somit erhalten wir $\#(A \cap B) = 6$ und

$$\mathbb{P}[A \cap B] = \frac{6}{6^3} = \frac{1}{6^2} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

Daher sind die Ereignisse A und B per Definition unabhängig.

BEMERKUNG 4.1.4. Disjunktheit und Unabhängigkeit sind völlig verschiedene Begriffe. Disjunkte Ereignisse sind nämlich niemals unabhängig (außer eines der Ereignisse hat die Wahrscheinlichkeit 0). Wir beweisen das. Seien A und B disjunkt (d.h. $A \cap B = \emptyset$) mit $\mathbb{P}[A] \neq 0$ und $\mathbb{P}[B] \neq 0$. Aus unseren Annahmen folgt, dass

$$\mathbb{P}[A \cap B] = \mathbb{P}[\emptyset] = 0 \neq \mathbb{P}[A] \cdot \mathbb{P}[B].$$

Somit sind A und B abhängig.

BEMERKUNG 4.1.5. Ereignis Ω ist von jedem Ereignis A unabhängig, denn

$$\mathbb{P}[\Omega \cap A] = \mathbb{P}[A] = \mathbb{P}[A] \cdot \mathbb{P}[\Omega], \text{ da } \mathbb{P}[\Omega] = 1.$$

Ereignis \emptyset ist ebenfalls von jedem Ereignis A unabhängig, denn

$$\mathbb{P}[\emptyset \cap A] = \mathbb{P}[\emptyset] = \mathbb{P}[A] \cdot \mathbb{P}[\emptyset], \text{ da } \mathbb{P}[\emptyset] = 0.$$

Die beiden Aussagen sind ziemlich natürlich. Zum Beispiel tritt das sichere Ereignis Ω immer ein, unabhängig davon, ob irgend ein anderes Ereignis A eintritt oder nicht.

Wir haben definiert, wann *zwei* Ereignisse unabhängig sind. Nun wollen wir definieren, wann *viele* Ereignisse unabhängig sind.

DEFINITION 4.1.6. Eine Familie $A_1, A_2, \dots \subset \Omega$ von Ereignissen heißt *unabhängig*, wenn für alle $k \in \mathbb{N}$ und alle Indizes $i_1 < \dots < i_k$ gilt, dass

$$(4.1.2) \quad \mathbb{P}[A_{i_1} \cap \dots \cap A_{i_k}] = \mathbb{P}[A_{i_1}] \cdot \dots \cdot \mathbb{P}[A_{i_k}].$$

DEFINITION 4.1.7. Ereignisse $A_1, A_2, \dots \subset \Omega$ heißen *paarweise unabhängig*, wenn für alle $i_1, i_2 \in \mathbb{N}$ mit $i_1 \neq i_2$ gilt:

$$(4.1.3) \quad \mathbb{P}[A_{i_1} \cap A_{i_2}] = \mathbb{P}[A_{i_1}] \cdot \mathbb{P}[A_{i_2}].$$

Unabhängige Ereignisse sind paarweise unabhängig. Das folgt aus den beiden Definitionen. Wir wollen nun an einem Beispiel zeigen, dass die Umkehrung nicht gilt.

BEISPIEL 4.1.8. Wir betrachten das Würfeln mit 3 fairen Würfeln. Wir bezeichnen mit X_i die Augenzahl, die der i -te Würfel zeigt, wobei $i = 1, 2, 3$. Wir legen die Ereignisse A, B und C wie folgt fest:

$$A = \{X_1 = X_2\}, \quad B = \{X_2 = X_3\}, \quad C = \{X_3 = X_1\}.$$

Wir haben bereits in Beispiel 4.1.3 gezeigt, dass A und B unabhängig sind. Analog sind auch A und C unabhängig, und das Gleiche gilt für B und C . Somit sind Ereignisse A, B, C paarweise unabhängig.

Wir zeigen nun, dass die Familie A, B, C nicht unabhängig ist. Der Schnitt der 3 Mengen ist gegeben durch:

$$A \cap B \cap C = \{X_1 = X_2 = X_3\} = \{(a, a, a) : a \in \{1, \dots, 6\}\}.$$

Somit gilt $\#(A \cap B \cap C) = 6$. Die Wahrscheinlichkeit der Schnittmenge $A \cap B \cap C$ ist somit

$$\mathbb{P}[A \cap B \cap C] = \frac{6}{6^3} = \frac{1}{6^2} \neq \mathbb{P}[A] \cdot \mathbb{P}[B] \cdot \mathbb{P}[C] = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{6^3}$$

Die Familie A, B, C ist paarweise unabhängig, aber nicht unabhängig.

BEMERKUNG 4.1.9. Von nun an wird nur noch der Begriff der Unabhängigkeit benutzt. Die paarweise Unabhängigkeit spielt keine Rolle in der Zukunft.

SATZ 4.1.10. *Es seien $A, B \subset \Omega$ zwei unabhängige Ereignisse. Dann gilt:*

- (1) A und B^c sind unabhängig.
- (2) A^c und B sind unabhängig.
- (3) A^c und B^c sind unabhängig.

BEWEIS. Wir zeigen, dass A und B^c unabhängig sind. Wir setzen $\mathbb{P}[A] = p$, $\mathbb{P}[B] = q$. Da A und B unabhängig sind, folgt $\mathbb{P}[A \cap B] = p \cdot q$. Somit gilt

$$\mathbb{P}[A \cap B^c] = \mathbb{P}[A \setminus B] = \mathbb{P}[A] - \mathbb{P}[A \cap B] = p - p \cdot q = p \cdot (1 - q) = \mathbb{P}[A] \cdot \mathbb{P}[B^c].$$

Es folgt, dass A und B^c unabhängig sind. Die beiden anderen Aussagen lassen sich analog beweisen. \square

SATZ 4.1.11. *Es seien $A, B, C \subset \Omega$ unabhängige Ereignisse. Dann gilt*

- (1) $A, B \cup C$ sind unabhängig.
- (2) $A, B \cap C$ sind unabhängig.

BEMERKUNG 4.1.12. Diese Aussage kann auch verallgemeinert werden. Betrachte eine unabhängige Familie von Ereignissen $A_1, \dots, A_n, B_1, \dots, B_m$. Es sei A irgendein Ereignis, das aus Ereignissen A_1, \dots, A_n durch die Anwendung von mengentheoretischen Operationen $\cup, \cap, ^c$ in irgendeiner Reihenfolge entsteht. Sei B ein Ereignis, das aus B_1, \dots, B_m durch Anwendung von $\cup, \cap, ^c$ entsteht. Dann sind A und B unabhängig.

4.2. Produkträume

Wir betrachten n Experimente $(\Omega_1, \mathbb{P}_1), \dots, (\Omega_n, \mathbb{P}_n)$ mit Wahrscheinlichkeitsfunktionen

$$p_i(a) = \mathbb{P}_i(\{a\}), \quad a \in \Omega_i, \quad i = 1, \dots, n.$$

Wir stellen uns nun vor, dass diese Experimente unabhängig voneinander ausgeführt werden. Die Unabhängigkeit kann man beispielsweise erreichen, indem man die Experimente räumlich voneinander trennt.

Werden nun alle Experimente ausgeführt, so ist die Grundmenge gegeben durch

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{(a_1, \dots, a_n) : a_i \in \Omega_i\}.$$

Wegen der Unabhängigkeit liegt es nun nahe, die Wahrscheinlichkeit eines Ausgangs $(a_1, \dots, a_n) \in \Omega$ wie folgt zu definieren:

$$p(a_1, \dots, a_n) \stackrel{def}{=} p_1(a_1) \cdot p_2(a_2) \cdot \dots \cdot p_n(a_n),$$

wobei $p_i(a_i)$ die Wahrscheinlichkeit des Ausgangs $a_i \in \Omega_i$ im i -ten Experiment ist. Die Wahrscheinlichkeit eines beliebigen Ereignisses $A \subset \Omega$ definieren wir dann wie folgt:

$$\mathbb{P}[A] \stackrel{def}{=} \sum_{a \in A} p(a).$$

Der Wahrscheinlichkeitsraum (Ω, \mathbb{P}) heißt der *Produktraum* von $(\Omega_1, \mathbb{P}_1), \dots, (\Omega_n, \mathbb{P}_n)$ und wird auch mit $(\Omega_1 \times \dots \times \Omega_n, \mathbb{P}_1 \times \dots \times \mathbb{P}_n)$ bezeichnet.

BEISPIEL 4.2.1. Wir betrachten Ereignisse $A_1 \subset \Omega_1, \dots, A_n \subset \Omega_n$. Das Ereignis A_i ist somit mit Experiment i verbunden. Nun betrachten wir das folgende Ereignis: "Im ersten Experiment tritt A_1 ein, im zweiten Experiment tritt A_2 ein, usw.". Dieses Ereignis kann man auch wie folgt darstellen:

$$A_1 \times \dots \times A_n \stackrel{def}{=} \{(a_1, \dots, a_n) : a_1 \in A_1, \dots, a_n \in A_n\} \subset \Omega.$$

Dann gilt für die Wahrscheinlichkeit dieses Ereignisses

$$\begin{aligned} \mathbb{P}[A_1 \times \dots \times A_n] &= \sum_{(a_1, \dots, a_n) \in A_1 \times \dots \times A_n} p(a_1, \dots, a_n) \\ &= \sum_{a_1 \in A_1, \dots, a_n \in A_n} p_1(a_1) \cdot \dots \cdot p_n(a_n) \\ &= \left(\sum_{a_1 \in A_1} p_1(a_1) \right) \cdot \dots \cdot \left(\sum_{a_n \in A_n} p_n(a_n) \right) \\ &= \mathbb{P}_1[A_1] \cdot \dots \cdot \mathbb{P}_n[A_n]. \end{aligned}$$

4.3. Bedingte Wahrscheinlichkeiten

BEISPIEL 4.3.1. Stellen wir uns vor, dass jemand mit 2 fairen Würfeln würfelt. Die Grundmenge ist $\Omega = \{1, \dots, 6\}^2$. Wir betrachten zwei Ereignisse:

$$\begin{aligned} A &= \text{“erster Würfel zeigt 6”} = \{(6, 1), (6, 2), \dots, (6, 6)\}, \\ B &= \text{“Augensumme = 10”} = \{(6, 4), (5, 5), (4, 6)\}. \end{aligned}$$

Stellen wir uns vor, dass das Experiment durchgeführt wurde und dass uns mitgeteilt wurde, dass das Ereignis B eingetreten ist. Ob das Ereignis A eingetreten ist, wissen wir aber nicht. Wir wollen nun die Wahrscheinlichkeit von A bestimmen, gegeben, dass B eingetreten ist. So etwas nennt man “bedingte Wahrscheinlichkeit von A gegeben B ” und bezeichnet mit $\mathbb{P}[A|B]$. Da B eingetreten ist, kommen nur Ausgänge $\{(6, 4), (5, 5), (4, 6)\}$ in Frage. Alle anderen Ausgänge sind durch die Information, dass B eingetreten ist, ausgeschlossen. Die Grundmenge hat sich also auf das Ereignis B verkleinert. Von den drei gleichwahrscheinlichen Ausgängen $\{(6, 4), (5, 5), (4, 6)\}$ führt aber nur der Ausgang $(6, 4)$ zum Eintreten von A . Die bedingte Wahrscheinlichkeit von A gegeben B ist also

$$\mathbb{P}[A|B] = \frac{1}{3}.$$

Zum Vergleich: die Wahrscheinlichkeit von A ohne Bedingungen ist $\mathbb{P}[A] = \frac{1}{6}$.

DEFINITION 4.3.2. Seien $A, B \subset \Omega$ zwei Ereignisse mit $\mathbb{P}[B] \neq 0$. Die *bedingte Wahrscheinlichkeit von A gegeben B* ist definiert durch

$$(4.3.1) \quad \mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

BEMERKUNG 4.3.3. Beachte: $A|B$ ist kein Ereignis, sondern lediglich eine Notation für eine neue Art von Wahrscheinlichkeit.

SATZ 4.3.4. Sei $B \subset \Omega$ ein Ereignis mit $\mathbb{P}[B] \neq 0$.

- (1) Für jedes Ereignis $A \subset \Omega$ gilt $\mathbb{P}[A|B] \in [0, 1]$.
- (2) Es gilt $\mathbb{P}[\Omega|B] = \mathbb{P}[B|B] = 1$ und $\mathbb{P}[\emptyset|B] = 0$.
- (3) Für paarweise disjunkte Ereignisse $A_1, A_2, \dots \subset \Omega$ gilt:

$$\mathbb{P}[\cup_{i=1}^{\infty} A_i|B] = \sum_{i=1}^{\infty} \mathbb{P}[A_i|B].$$

- (4) Für jedes Ereignis $A \subset \Omega$ gilt $\mathbb{P}[A^c|B] = 1 - \mathbb{P}[A|B]$.
- (5) Sind Ereignisse A und B unabhängig, so gilt $\mathbb{P}[A|B] = \mathbb{P}[A]$.

BEWEIS. Zu (1): Folgt aus $0 \leq \mathbb{P}[A \cap B] \leq \mathbb{P}[B]$ und (4.3.1). Zu (3):

$$\mathbb{P}[\cup_{i=1}^{\infty} A_i|B] = \frac{\mathbb{P}[(\cup_{i=1}^{\infty} A_i) \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[\cup_{i=1}^{\infty} (A_i \cap B)]}{\mathbb{P}[B]} = \sum_{i=1}^{\infty} \frac{\mathbb{P}[A_i \cap B]}{\mathbb{P}[B]} = \sum_{i=1}^{\infty} \mathbb{P}[A_i|B].$$

Zu (5): Sind A und B unabhängig, so heißt es, dass $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$. Es folgt, dass

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A] \cdot \mathbb{P}[B]}{\mathbb{P}[B]} = \mathbb{P}[A].$$

□

BEMERKUNG 4.3.5. Aus (4.3.1) sieht man, dass $\mathbb{P}[A|B]$ und $\mathbb{P}[B|A]$ im Allgemeinen nicht gleich sein müssen.

SATZ 4.3.6 (Satz von der totalen Wahrscheinlichkeit). Die Grundmenge sei als $\Omega = B_1 \cup \dots \cup B_n$ dargestellt, wobei B_1, \dots, B_n paarweise disjunkte Ereignisse sind und $\mathbb{P}[B_i] \neq 0$ für alle $i = 1, \dots, n$. Sei $A \subset \Omega$ ein weiteres Ereignis. Dann gilt für die Wahrscheinlichkeit von A :

$$\mathbb{P}[A] = \sum_{i=1}^n \mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i].$$

BEWEIS. Das Ereignis A ist eine disjunkte Vereinigung der Ereignisse $A \cap B_1, \dots, A \cap B_n$. Somit gilt

$$\mathbb{P}[A] = \mathbb{P}[\cup_{i=1}^n (A \cap B_i)] = \sum_{i=1}^n \mathbb{P}[A \cap B_i] = \sum_{i=1}^n \mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i].$$

□

BEISPIEL 4.3.7 (Gripptest). Bei einer kranken Person schlägt ein Grippteschnelltest mit Wahrscheinlichkeit 0.9 an. Bei einer gesunden Person kann der Test allerdings ebenfalls anschlagen, und zwar mit einer Wahrscheinlichkeit von 0.2. Wenn nun 1% aller Personen in einer Population tatsächlich krank sind, wie groß ist dann die Wahrscheinlichkeit, dass der Test bei einer zufällig gewählten Person anschlägt?

LÖSUNG. Wir legen zunächst passende Ereignisse fest:

$$\begin{aligned} A &= \text{“Person wird positiv getestet”}, \\ B_1 &= \text{“Person hat Grippe”}, \\ B_2 &= \text{“Person hat keine Grippe”}. \end{aligned}$$

Also sind die Ereignisse B_1 und B_2 disjunkt, d.h. $B_1 \cap B_2 = \emptyset$, und es gilt $\Omega = B_1 \cup B_2$. Laut Aufgabenstellung gilt

$$\begin{aligned} \mathbb{P}[A|B_1] &= 0.9, \\ \mathbb{P}[A|B_2] &= 0.2. \end{aligned}$$

Da zusätzlich noch bekannt ist, dass 1% aller Personen krank sind, gilt außerdem:

$$\begin{aligned} \mathbb{P}[B_1] &= 0.01, \\ \mathbb{P}[B_2] &= 1 - 0.01 = 0.99. \end{aligned}$$

Mit der Formel der totalen Wahrscheinlichkeit folgt:

$$\mathbb{P}[A] = \mathbb{P}[A|B_1] \cdot \mathbb{P}[B_1] + \mathbb{P}[A|B_2] \cdot \mathbb{P}[B_2] = 0.9 \cdot 0.01 + 0.2 \cdot 0.99 = 0.207.$$

Eine Person wird also mit Wahrscheinlichkeit 0.207 positiv getestet.

SATZ 4.3.8 (Bayes-Formel). Die Grundmenge sei als $\Omega = B_1 \cup \dots \cup B_n$ dargestellt, wobei B_1, \dots, B_n paarweise disjunkte Ereignisse sind und $\mathbb{P}[B_i] \neq 0$ für alle $i = 1, \dots, n$. Sei $A \subset \Omega$ ein weiteres Ereignis mit Wahrscheinlichkeit $\mathbb{P}[A] \neq 0$. Dann gilt für alle $i = 1, \dots, n$:

$$(4.3.2) \quad \mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]}{\sum_{k=1}^n \mathbb{P}[A|B_k] \cdot \mathbb{P}[B_k]}.$$

BEWEIS. Wir wenden die Definition der bedingten Wahrscheinlichkeit (4.3.1) zweimal an:

$$\mathbb{P}[B_i|A] = \frac{\mathbb{P}[B_i \cap A]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A \cap B_i]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]}{\mathbb{P}[A]}$$

Das beweist die erste Hälfte von (4.3.2). Die zweite Hälfte folgt aus dem Satz von der totalen Wahrscheinlichkeit. \square

BEISPIEL 4.3.9 (Fortsetzung von Beispiel 4.3.7). Eine Person, über die nicht bekannt ist, ob sie gesund oder krank ist, wurde positiv getestet. Mit welcher Wahrscheinlichkeit ist sie tatsächlich krank?

LÖSUNG. Gegeben ist, dass die Person positiv getestet wurde. Das Ereignis A ist also eingetreten. Gegeben diese Information wollen wir wissen, mit welcher Wahrscheinlichkeit das Ereignis B_1 eintritt. Gefragt wird also nach der bedingten Wahrscheinlichkeit $\mathbb{P}[B_1|A]$. Die Bayes-Formel (4.3.2) ergibt

$$\mathbb{P}[B_1|A] = \frac{\mathbb{P}[A|B_1] \cdot \mathbb{P}[B_1]}{\mathbb{P}[A]} = \frac{0.9 \cdot 0.01}{0.207} \approx 0.043.$$

Wir erkennen also, dass dieser Schnelltest ziemlich schlecht ist. Man kann auch die Wahrscheinlichkeit berechnen, dass eine Person gesund ist, gegeben, dass sie positiv getestet wurde:

$$\mathbb{P}[B_2|A] = 1 - \mathbb{P}[B_1|A] \approx 1 - 0.043 \approx 0.957.$$

4.4. Unabhängigkeit von Zufallsvariablen

Wir haben Unabhängigkeit von Ereignissen definiert. Man kann aber auch Unabhängigkeit von Zufallsvariablen definieren. Sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum.

DEFINITION 4.4.1. Die Zufallsvariablen $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ heißen *unabhängig*, wenn für alle $y_1, \dots, y_n \in \mathbb{R}$ gilt:

$$(4.4.1) \quad \mathbb{P}[X_1 = y_1, \dots, X_n = y_n] = \mathbb{P}[X_1 = y_1] \cdot \dots \cdot \mathbb{P}[X_n = y_n].$$

Diese Definition lässt sich in folgender äquivalenter Form darstellen.

SATZ 4.4.2. Seien $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Folgende Aussagen sind äquivalent.

- (1) Die Zufallsvariablen X_1, \dots, X_n sind unabhängig.
- (2) Für beliebige Mengen $B_1, \dots, B_n \subset \mathbb{R}$ gilt:

$$(4.4.2) \quad \mathbb{P}[X_1 \in B_1, \dots, X_n \in B_n] = \mathbb{P}[X_1 \in B_1] \cdot \dots \cdot \mathbb{P}[X_n \in B_n].$$

- (3) Die Ereignisse $\{X_1 = y_1\}, \dots, \{X_n = y_n\}$ sind unabhängig für alle $y_1, \dots, y_n \in \mathbb{R}$.
- (4) Die Ereignisse $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ sind unabhängig für alle $B_1, \dots, B_n \subset \mathbb{R}$.

Wenn man jedoch eine unendliche Familie von Zufallsvariablen betrachtet, dann heißen diese Zufallsvariablen unabhängig, wenn jede endliche Teilfamilie unabhängig ist.

DEFINITION 4.4.3. Die Zufallsvariablen X_1, X_2, \dots heißen unabhängig, wenn für alle $n \in \mathbb{N}$ gilt, dass X_1, \dots, X_n unabhängig sind.

BEISPIEL 4.4.4. Wir würfeln n -mal mit einem fairen Würfel. Die Grundmenge lautet $\Omega = \{1, \dots, 6\}^n$ und wir gehen von der Laplace-Annahme aus, dass alle Ausgänge in Ω die gleiche Wahrscheinlichkeit $1/6^n$ haben. Wir bezeichnen mit X_i die Augenzahl beim i -ten Wurf:

$$X_i(a_1, \dots, a_n) = a_i, \quad (a_1, \dots, a_n) \in \Omega.$$

Wir zeigen nun, dass die Zufallsvariablen X_1, \dots, X_n unabhängig sind. (Das ist im Einklang mit dem gesunden Verstand). Es gilt

$$\begin{aligned} \mathbb{P}[X_1 = y_1, \dots, X_n = y_n] &= \begin{cases} \frac{1}{6^n}, & \text{falls } y_1, \dots, y_n \in \{1, \dots, 6\}, \\ 0, & \text{sonst} \end{cases} \\ &= \mathbb{P}[X_1 = y_1] \cdot \dots \cdot \mathbb{P}[X_n = y_n]. \end{aligned}$$

BEISPIEL 4.4.5. In diesem Beispiel betrachten wir die Augensumme $S = X_1 + \dots + X_n$. Wir zeigen, dass die Zufallsvariablen S und X_1 abhängig sind. Zu diesem Zweck müssen wir einen Fall finden, für den die Produktformel nicht gilt. Wir betrachten die Ereignisse $\{X_1 = 1\}$ und $\{S = 6n\}$. Die Wahrscheinlichkeiten der beiden Ereignisse sind strikt positiv:

$$\mathbb{P}[X_1 = 1] = \frac{1}{6}, \quad \mathbb{P}[S = 6n] = \frac{1}{6^n}.$$

Auf der anderen Seite, können beide Ereignisse gleichzeitig nicht eintreten, somit

$$\mathbb{P}[X_1 = 1, S = 6n] = 0.$$

Daraus folgt dann, dass S und X_1 abhängig sind.

BEMERKUNG 4.4.6. Es sei $X_1, \dots, X_n, Y_1, \dots, Y_m$ eine unabhängige Familie von Zufallsvariablen. Man kann zeigen, dass für beliebige Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$ und $g : \mathbb{R}^m \rightarrow \mathbb{R}$ die Zufallsvariablen $f(X_1, \dots, X_n)$ und $g(Y_1, \dots, Y_m)$ unabhängig sind.

KAPITEL 5

Erwartungswert

Wir betrachten einen diskreten Wahrscheinlichkeitsraum (Ω, \mathbb{P}) und eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ auf diesem Wahrscheinlichkeitsraum. Die Grundmenge Ω hat also nur endlich oder abzählbar viele Elemente. Für die Zufallsvariable X bedeutet es, dass sie nur endlich oder abzählbar viele Werte annehmen kann. Diese Werte kann man aufzählen:

$$\begin{aligned} \text{Werte von } X : & \quad y_1 \quad y_2 \quad y_3 \quad \dots \\ \text{Wahrscheinlichkeiten :} & \quad p_1 \quad p_2 \quad p_3 \quad \dots \end{aligned}$$

Dabei bezeichnen wir mit $p_i = \mathbb{P}[X = y_i]$ die Wahrscheinlichkeit des Ereignisses $\{X = y_i\}$. Es gilt dann

- (1) $p_i \in [0, 1]$ für alle i .
- (2) $\sum_i p_i = 1$.

DEFINITION 5.0.7. Die Zufallsvariable X heißt *integrierbar*, wenn $\sum_i p_i |y_i| < \infty$. Ist X integrierbar, so definieren wir den *Erwartungswert* von X wie folgt:

$$(5.0.3) \quad \mathbb{E}X \stackrel{\text{def}}{=} \sum_i p_i y_i.$$

BEMERKUNG 5.0.8. Nimmt X nur endlich viele Werte an, so ist die Bedingung $\sum_i p_i |y_i| < \infty$ erfüllt und X ist integrierbar.

BEMERKUNG 5.0.9. Eine Reihe $\sum_i a_i$ heißt *absolut konvergent*, falls $\sum_i |a_i| < \infty$. Es ist bekannt, dass eine absolut konvergente Reihe konvergiert, und dass die Summe einer absolut konvergenten Reihe von der Reihenfolge der Summanden unabhängig ist. In der Definition des Erwartungswerts fordern wir die absolute Konvergenz der Reihe $\sum p_i y_i$. Diese Forderung stellt sicher, dass die Summe dieser Reihe nicht von der Reihenfolge der Terme in abhängt.

Bei Reihen, die nicht absolut konvergieren, kann sich die Summe ändern, wenn man die Reihenfolge der Summanden ändert.

BEISPIEL 5.0.10. Betrachte die alternierende harmonische Reihe:

$$(5.0.4) \quad 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \dots = \ln 2.$$

Beweis: Setze $x = 1$ in der Formel $\ln(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^n}{n}$. Diese Reihe ist konvergent (gegen $\ln 2$), aber nicht absolut konvergent, denn $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$. Nun betrachte die Reihe:

$$(5.0.5) \quad 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \dots = \frac{3}{2} \log 2.$$

Beweis: Übung. Die Summen der Reihen (5.0.4) und (5.0.5) sind unterschiedlich. Dabei haben beide Reihen die gleichen Summanden, lediglich die Reihenfolge der Summanden ist unterschiedlich.

BEMERKUNG 5.0.11. Die Definition des Erwartungswerts kann man auch so schreiben:

$$\mathbb{E}X = \sum_{y \in \text{Im}(X)} y \cdot \mathbb{P}[X = y],$$

falls die Reihe absolut konvergiert.

BEISPIEL 5.0.12. Wir würfeln mit einem fairen Würfel. Sei X die Augenzahl. Der Erwartungswert von X ist

$$\mathbb{E}X = \frac{1}{6}(1 + \dots + 6) = 3,5.$$

BEISPIEL 5.0.13. Sei $A \subset \Omega$ ein Ereignis. Es sei $X = \mathbb{1}_A$ die Indikatorfunktion von A . Die Indikatorfunktion nimmt den Wert 1 genau dann an, wenn das Ereignis eingetreten ist. Ansonsten nimmt sie den Wert 0 an. Demnach gilt:

$$\mathbb{E}\mathbb{1}_A = 0 \cdot \mathbb{P}[\mathbb{1}_A = 0] + 1 \cdot \mathbb{P}[\mathbb{1}_A = 1] = 0 \cdot (1 - \mathbb{P}[A]) + 1 \cdot \mathbb{P}[A] = \mathbb{P}[A].$$

SATZ 5.0.14. Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann ist X integrierbar genau dann, wenn

$$(5.0.6) \quad \sum_{\omega \in \Omega} p(\omega) |X(\omega)| < \infty.$$

Ist (5.0.6) erfüllt, so gilt

$$(5.0.7) \quad \mathbb{E}X = \sum_{\omega \in \Omega} p(\omega) X(\omega).$$

BEWEIS. Für jedes $y \in \text{Im}(X)$ definieren wir das Ereignis

$$A_y \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = y\}.$$

Es gilt dann:

- (1) $\Omega = \cup_{y \in \text{Im}(X)} A_y$.
- (2) Die Ereignisse A_y sind paarweise disjunkt.

Laut Definition ist X integrierbar genau dann, wenn

$$(5.0.8) \quad \sum_{y \in \text{Im}(X)} |y| \cdot \mathbb{P}[X = y] < \infty.$$

Bei Reihen mit nicht-negativen Termen kann man die Summanden vertauschen, ohne dass sich die Summe ändert. Es folgt, dass

$$\begin{aligned} \sum_{y \in \text{Im}(X)} |y| \cdot \mathbb{P}[X = y] &= \sum_{y \in \text{Im}(X)} |y| \cdot \sum_{\omega \in A_y} p(\omega) \\ &= \sum_{y \in \text{Im}(X)} \sum_{\omega \in A_y} p(\omega) |X(\omega)| \\ &= \sum_{\omega \in \Omega} p(\omega) |X(\omega)|. \end{aligned}$$

Somit sind Bedingungen (5.0.6) und (5.0.8) äquivalent. Es folgt, dass X genau dann integrierbar ist, wenn (5.0.6) gilt.

Nun nehmen wir an, dass (5.0.6) gilt. Es folgt, dass

$$\begin{aligned}
 \mathbb{E}X &= \sum_{y \in \text{Im}(X)} y \cdot \mathbb{P}[A_y] \\
 &= \sum_{y \in \text{Im}(X)} y \cdot \sum_{\omega \in A_y} p(\omega) \\
 &= \sum_{y \in \text{Im}(X)} \sum_{\omega \in A_y} p(\omega) X(\omega) \\
 &= \sum_{\omega \in \Omega} p(\omega) X(\omega).
 \end{aligned}$$

Die absolute Konvergenz wurde hier dadurch benutzt, dass die Summanden vertauscht wurden. \square

SATZ 5.0.15 (Linearität des Erwartungswerts). (1) *Seien $X, Y : \Omega \rightarrow \mathbb{R}$ integrierbare Zufallsvariablen. Dann ist $X + Y$ integrierbar und es gilt*

$$(5.0.9) \quad \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

(2) *Sei $X : \Omega \rightarrow \mathbb{R}$ eine integrierbare Zufallsvariable. Sei $a \in \mathbb{R}$. Dann ist aX integrierbar und es gilt*

$$(5.0.10) \quad \mathbb{E}(aX) = a\mathbb{E}X.$$

BEWEIS. Teil 1: Wir zeigen, dass $X + Y$ integrierbar ist:

$$\begin{aligned}
 \sum_{\omega \in \Omega} p(\omega) |X(\omega) + Y(\omega)| &\leq \sum_{\omega \in \Omega} (p(\omega) |X(\omega)| + p(\omega) |Y(\omega)|) \\
 &= \sum_{\omega \in \Omega} p(\omega) |X(\omega)| + \sum_{\omega \in \Omega} p(\omega) |Y(\omega)|.
 \end{aligned}$$

Die rechte Seite ist endlich, da X und Y integrierbar sind. Somit ist auch die linke Seite endlich. Es folgt, dass $X + Y$ integrierbar ist.

Wir berechnen nun $\mathbb{E}(X + Y)$. Mit Satz 5.0.14 gilt

$$\begin{aligned}
 \mathbb{E}(X + Y) &= \sum_{\omega \in \Omega} p(\omega) (X(\omega) + Y(\omega)) \\
 &= \sum_{\omega \in \Omega} (p(\omega) X(\omega) + p(\omega) Y(\omega)) \\
 &= \sum_{\omega \in \Omega} p(\omega) X(\omega) + \sum_{\omega \in \Omega} p(\omega) Y(\omega) \\
 &= \mathbb{E}X + \mathbb{E}Y.
 \end{aligned}$$

Auch hier haben wir die absolute Konvergenz benutzt, indem wir die Summanden vertauscht haben.

Beweis von Teil 2 ist analog. \square

BEMERKUNG 5.0.16. Der eben bewiesene Satz gilt auch für n Summanden: sind $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ integrierbare Zufallsvariablen, so ist auch die Summe $X_1 + \dots + X_n$ integrierbar und es gilt

$$\mathbb{E}[X_1 + \dots + X_n] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n].$$

BEISPIEL 5.0.17. Wir würfeln n -mal mit einem fairen Würfel. Sei X_i die Augenzahl bei Wurf i , wobei $i = 1, \dots, n$. Es sei $S = X_1 + \dots + X_n$ die Augensumme. Dann gilt für den Erwartungswert von S :

$$\mathbb{E}S = \mathbb{E}X_1 + \dots + \mathbb{E}X_n = n \cdot \mathbb{E}X_1 = 3,5 \cdot n.$$

BEISPIEL 5.0.18 (Lotto). In einer Urne liegen 49 nummerierte Kugeln, es werden 6 Kugeln ohne Zurücklegen gezogen. Wir tippen auf 6 verschiedene Kugeln. Es sei S die Anzahl der Richtigen. Bestimme den Erwartungswert von S .

LÖSUNG. Wir tippen oBdA auf die Kombination $\{1, \dots, 6\}$. Dann definieren wir die Zufallsvariablen

$$X_i = \begin{cases} 1, & \text{falls Kugel } i \text{ gezogen wurde,} \\ 0, & \text{sonst,} \end{cases} \quad i = 1, \dots, 6.$$

Für den Erwartungswert von X_i gilt:

$$\mathbb{E}X_i = \mathbb{P}[X_i = 1] = \frac{\binom{48}{5}}{\binom{49}{6}} = \frac{\frac{48 \cdot \dots \cdot 44}{5 \cdot \dots \cdot 1}}{\frac{49 \cdot \dots \cdot 44}{6 \cdot \dots \cdot 1}} = \frac{6}{49}.$$

Die Anzahl der Richtigen ist dann $S = X_1 + \dots + X_6$. Mit Satz 5.0.15 gilt für den Erwartungswert von S :

$$\mathbb{E}S = \mathbb{E}X_1 + \dots + \mathbb{E}X_6 = 6 \cdot \frac{6}{49} = \frac{36}{49}.$$

SATZ 5.0.19. Seien $X, Y : \Omega \rightarrow \mathbb{R}$ unabhängige integrierbare Zufallsvariablen. Dann ist auch das Produkt $X \cdot Y$ integrierbar und es gilt

$$(5.0.11) \quad \mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

BEWEIS. Seien a_1, a_2, \dots alle Werte von X mit dazugehörigen Wahrscheinlichkeiten p_1, p_2, \dots . Analog seien b_1, b_2, \dots alle Werte von Y mit Wahrscheinlichkeiten q_1, q_2, \dots .

$$\begin{array}{c|ccc} X & a_1 & a_2 & \dots \\ \hline \mathbb{P}[X] & p_1 & p_2 & \dots \end{array} \quad \begin{array}{c|ccc} Y & b_1 & b_2 & \dots \\ \hline \mathbb{P}[Y] & q_1 & q_2 & \dots \end{array}$$

Nun definieren wir das Ereignis A_{ij} , welches eintritt, wenn X den Wert a_i annimmt und gleichzeitig Y den Wert b_j :

$$A_{ij} = \{X = a_i, Y = b_j\}.$$

Dann bilden die Ereignisse A_{ij} eine disjunkte Zerlegung von Ω . Die Wahrscheinlichkeit von A_{ij} ist, wegen der Unabhängigkeit von X und Y ,

$$\mathbb{P}[A_{ij}] = \mathbb{P}[X = a_i] \cdot \mathbb{P}[Y = b_j] = p_i \cdot q_j.$$

Wir zeigen, dass XY integrierbar ist:

$$\begin{aligned}
\sum_{\omega \in \Omega} p(\omega) \cdot |X(\omega) \cdot Y(\omega)| &= \sum_{i,j} \sum_{\omega \in A_{i,j}} p(\omega) \cdot |X(\omega) \cdot Y(\omega)| \\
&= \sum_{i,j} \sum_{\omega \in A_{i,j}} p(\omega) |a_i| |b_j| \\
&= \sum_{i,j} |a_i| |b_j| \cdot \sum_{\omega \in A_{i,j}} p(\omega) \\
&= \sum_{i,j} |a_i| |b_j| \cdot \mathbb{P}[A_{i,j}] \\
&= \sum_{i,j} |a_i| |b_j| \cdot p_i q_j \\
&= \sum_{i,j} |a_i| p_i \cdot |b_j| q_j \\
&= \left(\sum_i |a_i| p_i \right) \cdot \left(\sum_j |b_j| q_j \right).
\end{aligned}$$

Die rechte Seite ist endlich, da X und Y integrierbar sind. Somit ist auch die linke Seite endlich. Mit Satz 5.0.14 folgt, dass XY integrierbar ist.

Für den Erwartungswert von $X \cdot Y$ gilt nun mit Satz 5.0.14:

$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \sum_{\omega \in \Omega} p(\omega) \cdot X(\omega) \cdot Y(\omega) \\
&= \sum_{i,j} \sum_{\omega \in A_{i,j}} p(\omega) \cdot X(\omega) \cdot Y(\omega) \\
&= \sum_{i,j} \sum_{\omega \in A_{i,j}} p(\omega) \cdot a_i \cdot b_j \\
&= \sum_{i,j} a_i b_j \sum_{\omega \in A_{i,j}} p(\omega) \\
&= \sum_{i,j} a_i b_j \cdot \mathbb{P}[A_{i,j}] \\
&= \sum_{i,j} a_i b_j \cdot p_i q_j \\
&= \sum_{i,j} a_i p_i \cdot b_j q_j \\
&= \left(\sum_i a_i p_i \right) \cdot \left(\sum_j b_j q_j \right) \\
&= \mathbb{E}X \cdot \mathbb{E}Y.
\end{aligned}$$

Die absolute Konvergenz haben wir dabei mehrmals benutzt, indem wir die Summanden vertauscht haben. \square

BEMERKUNG 5.0.20. Der eben bewiesene Satz gilt auch für n Faktoren: sind $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ unabhängige integrierbare Zufallsvariablen, so ist auch das Produkt $X_1 \dots X_n$ integrierbar und es gilt

$$\mathbb{E}[X_1 \cdot \dots \cdot X_n] = \mathbb{E}[X_1] \cdot \dots \cdot \mathbb{E}[X_n].$$

Diskrete Verteilungen

Nun werden wir verschiedene Beispiele von diskreten Zufallsvariablen betrachten.

6.1. Gleichverteilung

DEFINITION 6.1.1. Eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt *gleichverteilt* (oder *Laplace-verteilt*) auf einer endlichen Menge $\{y_1, \dots, y_n\} \subset \mathbb{R}$, wenn

$$\mathbb{P}[X = y_i] = \frac{1}{n} \text{ für alle } i = 1, \dots, n.$$

BEMERKUNG 6.1.2. Eine Zufallsvariable ist also gleichverteilt, wenn sie n Werte annehmen kann und die Wahrscheinlichkeiten dieser n Werte gleich sind.

BEMERKUNG 6.1.3. Für den Erwartungswert dieser Zufallsvariable gilt

$$\mathbb{E}X = \frac{y_1 + \dots + y_n}{n}.$$

Dies ist das arithmetische Mittel von y_1, \dots, y_n .

BEMERKUNG 6.1.4. Definition 6.1.1 funktioniert nur für endliches n . Eine Zufallsvariable kann nicht unendlich viele Werte mit gleicher Wahrscheinlichkeit annehmen. Hätte jeder Wert die gleiche, strikt positive Wahrscheinlichkeit $p > 0$, so wäre die Summe aller Wahrscheinlichkeiten unendlich. Hätte jeder Wert Wahrscheinlichkeit 0, so wäre die Summe aller Wahrscheinlichkeiten 0. Die Summe sollte aber 1 sein. In beiden Fällen ergibt sich ein Widerspruch. Eine Gleichverteilung (im obigen Sinne) auf einer unendlichen Menge gibt es also nicht.

6.2. Bernoulli-Experimente und die Binomialverteilung

DEFINITION 6.2.1. Ein *Bernoulli-Experiment* ist ein Zufallsexperiment mit zwei Ausgängen:

$$0 \text{ (‘‘Misserfolg’’)} \quad \text{und} \quad 1 \text{ (‘‘Erfolg’’)}.$$

Die Wahrscheinlichkeit von ‘‘Erfolg’’ bezeichnen wir mit $p \in [0, 1]$. Die Wahrscheinlichkeit von ‘‘Misserfolg’’ ist dann $q := 1 - p$.

DEFINITION 6.2.2. Eine Zufallsvariable X heißt *Bernoulli-verteilt* mit Parameter $p \in [0, 1]$, falls

$$\mathbb{P}[X = 1] = p, \quad \mathbb{P}[X = 0] = 1 - p.$$

BEZEICHNUNG: $X \sim \text{Bern}(p)$.

DEFINITION 6.2.3. Ein *n -faches Bernoulli-Experiment* ist ein Bernoulli-Experiment, das n -mal unabhängig voneinander ausgeführt wurde.

BEISPIEL 6.2.4. Wir können eine (faire oder unfaire) Münze n -mal werfen und zum Beispiel “Kopf” als “Erfolg” auffassen.

BEISPIEL 6.2.5. Wir können einen Würfel n -mal werfen. Fassen wir eine 6 als einen “Erfolg” auf, so erhalten wir ein n -faches Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p = \frac{1}{6}$.

BEMERKUNG 6.2.6. Die Grundmenge eines n -fachen Bernoulli-Experiments ist $\Omega = \{0, 1\}^n$. Wegen der Unabhängigkeit der einzelnen Experimente ist die Wahrscheinlichkeit eines Ausgangs $(a_1, \dots, a_n) \in \Omega$ gegeben durch

$$p(a_1, \dots, a_n) = p^k(1-p)^{n-k},$$

wobei $k = \sum_{i=1}^n a_i$ die Anzahl der Einsen unter a_1, \dots, a_n ist. Für $p \neq 1/2$ sind die Ausgänge nicht gleichwahrscheinlich.

SATZ 6.2.7. Sei X die Anzahl der “Erfolge” in einem n -fachen Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p . Dann gilt:

$$(6.2.1) \quad \mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \text{ für alle } k = 0, 1, \dots, n.$$

BEWEIS. Sei $k \in \{0, \dots, n\}$. Wir betrachten das Ereignis $\{X = k\}$. Es besteht aus allen Ausgängen $(a_1, \dots, a_n) \in \{0, 1\}^n$ mit genau k Einsen. Es gibt genau $\binom{n}{k}$ solche Ausgänge. Jeder dieser Ausgänge hat Wahrscheinlichkeit von jeweils $p^k(1-p)^{n-k}$. Für die Wahrscheinlichkeit des Ereignisses $\{X = k\}$ ergibt sich somit Formel (6.2.1). \square

BEMERKUNG 6.2.8. Eine Zufallsvariable X , die (6.2.1) erfüllt, heißt *binomialverteilt* mit Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$.

BEZEICHNUNG: $X \sim \text{Bin}(n, p)$.

BEMERKUNG 6.2.9. Die Summe der Wahrscheinlichkeiten aller Werte einer diskreten Zufallsvariable sollte 1 ergeben. Dies ist bei der Binomialverteilung der Fall, denn

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Dabei haben wir die binomische Formel benutzt, daher die Bezeichnung “Binomialverteilung”.

SATZ 6.2.10. Für $X \sim \text{Bin}(n, p)$ gilt $\mathbb{E}X = np$. In Worten: Die erwartete Anzahl von “Erfolgen” in einem n -fachen Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p ist gleich np .

BEWEIS. Wir definieren die Zufallsvariablen $X_1, \dots, X_n : \{0, 1\}^n \rightarrow \mathbb{R}$ wie folgt:

$$X_i(a_1, \dots, a_n) = a_i, \text{ wobei } (a_1, \dots, a_n) \in \{0, 1\}^n.$$

In Worten:

$$X_i = \begin{cases} 1, & \text{falls Experiment } i \text{ ein “Erfolg” ist,} \\ 0, & \text{falls Experiment } i \text{ ein “Misserfolg” ist.} \end{cases}$$

Da die Erfolgswahrscheinlichkeit in jedem Experiment gleich p ist, gilt

$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = 0] = 1 - p \text{ für alle } i = 1, \dots, n.$$

Für den Erwartungswert von X_i gilt somit:

$$\mathbb{E}X_i = p \cdot 1 + (1 - p) \cdot 0 = p \text{ für alle } i = 1, \dots, n.$$

Die Anzahl der ‘Erfolge’ im n -fachen Bernoulli-Experiment ist gegeben durch

$$X = X_1 + \dots + X_n.$$

Aus der Additivität des Erwartungswerts folgt, dass $\mathbb{E}X = \mathbb{E}X_1 + \dots + \mathbb{E}X_n = np$. □

6.3. Poisson-Verteilung

DEFINITION 6.3.1. Eine Zufallsvariable X hat *Poisson-Verteilung* mit Parameter (auch Intensität genannt) $\lambda > 0$, wenn

$$(6.3.1) \quad \mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!} \text{ für alle } k = 0, 1, 2, \dots$$

BEZEICHNUNG: $X \sim \text{Poi}(\lambda)$.

BEMERKUNG 6.3.2. Die Wahrscheinlichkeiten in (6.3.1) summieren sich zu 1, denn

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

SATZ 6.3.3. Für $X \sim \text{Poi}(\lambda)$ gilt $\mathbb{E}X = \lambda$.

BEWEIS. Wir verwenden die Definition des Erwartungswertes:

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

Dabei haben wir $m = k - 1$ gesetzt. □

Die Poisson-Verteilung entsteht als Grenzwert der Binomialverteilung. Das wird im folgenden Satz beschrieben.

SATZ 6.3.4 (Poisson-Grenzwertsatz). Sei $p_n \in (0, 1)$ eine Folge mit

$$(6.3.2) \quad \lim_{n \rightarrow \infty} np_n = \lambda \in (0, \infty).$$

Sei S_n eine Zufallsvariable mit $S_n \sim \text{Bin}(n, p_n)$. Für jedes $k = 0, 1, 2, \dots$ gilt dann

$$(6.3.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}[S_n = k] = e^{-\lambda} \frac{\lambda^k}{k!}.$$

BEISPIEL 6.3.5. Man stelle sich S_n vor, als die Anzahl der ‘Erfolge’ in einem n -fachen Bernoulli-Experiment mit einem sehr großen n und einer sehr kleinen Erfolgswahrscheinlichkeit

$$p_n \approx \frac{\lambda}{n}.$$

Der Poisson-Grenzwertsatz besagt, dass die Anzahl der “Erfolge” in einem solchen Experiment approximativ Poisson-verteilt ist. Beispiele von Zufallsvariablen, die approximativ Poisson-verteilt sind:

- (1) Anzahl der Schäden, die einer Versicherung gemeldet werden (viele Versicherungsverträge, jeder Vertrag erzeugt mit einer sehr kleinen Wahrscheinlichkeit einen Schaden).
- (2) Anzahl der Druckfehler in einem Buch (viele Buchstaben, jeder Buchstabe kann mit einer sehr kleinen Wahrscheinlichkeit ein Druckfehler sein).
- (3) Anzahl der Zugriffe auf einen Webserver (viele User, jeder User greift mit einer sehr kleinen Wahrscheinlichkeit zu).

BEWEIS VON SATZ 6.3.4. Sei $k \in \mathbb{N}_0$ fest. Da S_n binomialverteilt ist, gilt:

$$\begin{aligned} \mathbb{P}[S_n = k] &= \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} \\ &= \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} \cdot p_n^k \cdot (1 - p_n)^{n-k} \\ &= \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{n^k} \cdot \frac{(np_n)^k}{k!} \cdot \left(1 - \frac{np_n}{n}\right)^n \cdot (1 - p_n)^{-k} \\ &\xrightarrow{n \rightarrow \infty} 1 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1. \end{aligned}$$

Dabei haben wir benutzt, dass $\lim_{n \rightarrow \infty} p_n = 0$. Dies folgt aus der Annahme (6.3.2). Außerdem haben wir die folgende Formel benutzt:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda_n}{n}\right)^n = e^{-\lambda},$$

für jede Folge λ_n mit $\lim_{n \rightarrow \infty} \lambda_n = \lambda \in (0, \infty)$. In unserem Fall war $\lambda_n = np_n$. □

BEISPIEL 6.3.6. Im Hörsaal befinden sich $n = 100$ Personen. Betrachte das Ereignis

$A =$ “mindestens eine Person im Hörsaal hat heute Geburtstag”.

Bestimme die Wahrscheinlichkeit von A .

Wir werden zwei Lösungen präsentieren. Die erste Lösung ist exakt, die zweite approximativ.

LÖSUNG 1 (EXAKT). Wir nummerieren die Personen mit $1, \dots, n$. Wir betrachten das Ereignis “Person i hat heute Geburtstag” als “Erfolg” im i -ten Bernoulli-Experiment. Die Wahrscheinlichkeit von “Erfolg” ist für jede Person i gegeben durch

$$p := \mathbb{P}[\text{Person } i \text{ hat heute Geburtstag}] = \frac{1}{365}.$$

Dabei können wir die Geburtstage der verschiedenen Personen als unabhängig betrachten. Es handelt sich also um ein $n = 100$ -faches Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p = \frac{1}{365}$. Die Anzahl der Personen im Hörsaal, die heute Geburtstag haben, ist $\text{Bin}(100, \frac{1}{365})$ -verteilt.

Für die Wahrscheinlichkeit von A^c erhalten wir:

$$\mathbb{P}[A^c] = \mathbb{P}[\text{keine Person im Hörsaal hat heute Geburtstag}] = (1 - p)^n.$$

Es folgt, dass

$$\mathbb{P}[A] = 1 - (1 - p)^n = 1 - \left(1 - \frac{1}{365}\right)^{100} \approx 0.239933.$$

LÖSUNG 2 (APPROXIMATIV). Die Anzahl der Personen im Hörsaal, die heute Geburtstag haben, ist binomialverteilt mit $n = 100$ und $p = \frac{1}{365}$. Die Wahrscheinlichkeit p ist sehr klein, die Anzahl der Personen n ist sehr groß. Deshalb benutzen wir die Poisson-Approximation:

$$\text{Bin}\left(100, \frac{1}{365}\right) \approx \text{Poi}\left(\frac{100}{365}\right).$$

Somit ist die Anzahl der Personen, die heute Geburtstag haben, approximativ Poissonverteilt mit Parameter $\lambda = np = \frac{100}{365}$. Für die Wahrscheinlichkeit von A^c erhalten wir aus der Formel (6.3.1) mit $k = 0$:

$$\mathbb{P}[A^c] = \mathbb{P}[\text{keine Person im Hörsaal hat heute Geburtstag}] \approx e^{-\lambda} \cdot \frac{\lambda^0}{0!} = e^{-\lambda}.$$

Es folgt, dass

$$\mathbb{P}[A] \approx 1 - e^{-\lambda} = 1 - e^{-\frac{100}{365}} \approx 0.239647.$$

6.4. Geometrische Verteilung

Wir betrachten ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p \in (0, 1]$, das unendlich oft und unabhängig wiederholt wird. Die Grundmenge ist dann die Menge aller unendlichen Folgen aus Nullen und Einsen:

$$\Omega = \{0, 1\}^\infty \stackrel{\text{def}}{=} \{(a_1, a_2, \dots) : a_i \in \{0, 1\} \text{ für alle } i \in \mathbb{N}\}.$$

Diese Menge ist überabzählbar. Solche Experimente werden wir später genauer betrachten. Nun legen wir eine Zufallsvariable $T : \Omega \rightarrow \mathbb{R}$ fest:

$$T(a_1, a_2, \dots) \stackrel{\text{def}}{=} \min\{n \in \mathbb{N} : a_n = 1\}.$$

Die Zufallsvariable T ist somit die Wartezeit auf den ersten “Erfolg”.

BEISPIEL 6.4.1. Gehen die Experimente wie folgt aus:

$$0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, \dots,$$

so erhalten wir $T = 4$.

Was für eine Verteilung hat nun T ?

SATZ 6.4.2. Die Wartezeit auf T auf den ersten “Erfolg” in einem unendlich oft wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p ist wie folgt verteilt:

$$(6.4.1) \quad \mathbb{P}[T = k] = (1 - p)^{k-1} p \text{ für alle } k = 1, 2, \dots$$

BEMERKUNG 6.4.3. Eine Zufallsvariable T , die (6.4.1) mit einem $p \in (0, 1]$ erfüllt, heißt *geometrisch verteilt* mit Parameter p .

BEZEICHNUNG: $T \sim \text{Geo}(p)$.

BEMERKUNG 6.4.4. Die Wahrscheinlichkeiten in (6.4.1) summieren sich zu 1, denn

$$\sum_{k=1}^{\infty} (1-p)^{k-1} p = p \cdot \sum_{k=1}^{\infty} (1-p)^{k-1} = p \cdot \frac{1}{1-(1-p)} = 1.$$

Dabei haben wir eine geometrische Reihe summiert, daher die Bezeichnung “geometrische Verteilung”.

BEWEIS VON SATZ 6.4.2. Wir benutzen die Notation 0 = “Misserfolg” und 1 = “Erfolg”. Sei $k \in \mathbb{N}$ fest. Damit das Ereignis $\{T = k\}$ eintritt, müssen die ersten k Experimente so ausgehen:

$$0, 0, \dots, 0, 1.$$

Dabei ist es egal, wie alle anderen Experimente ausgehen. Wegen der Unabhängigkeit der einzelnen Bernoulli-Experimente gilt für die Wahrscheinlichkeit davon:

$$\mathbb{P}[T = k] = (1-p) \cdot \dots \cdot (1-p) \cdot p = (1-p)^{k-1} p.$$

□

SATZ 6.4.5. Für $T \sim \text{Geo}(p)$ gilt $\mathbb{E}T = \frac{1}{p}$. In Worten: Die durchschnittliche Wartezeit auf den ersten Erfolg in einem unendlich oft wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p ist gleich $\frac{1}{p}$.

BEWEIS. Wiederum wird die Definition des Erwartungswerts benutzt:

$$\mathbb{E}T = \sum_{k=1}^{\infty} k \cdot \mathbb{P}[T = k] = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1},$$

wobei $q = 1 - p$. Die Summe auf der rechten Seite können wir wie folgt berechnen:

$$\sum_{k=1}^{\infty} kq^{k-1} = \left(\sum_{k=1}^{\infty} q^k \right)' = \left(\sum_{k=0}^{\infty} q^k \right)' = \left(\frac{1}{1-q} \right)' = \frac{1}{(1-q)^2}.$$

Es folgt, dass

$$\mathbb{E}T = p \cdot \frac{1}{(1-q)^2} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

□

BEISPIEL 6.4.6. Wir würfeln mit einem fairen Würfel so lange, bis eine 1 kommt. Wie lange müssen wir im Durchschnitt warten?

LÖSUNG. Die Wartezeit T ist $\text{Geo}(\frac{1}{6})$ -verteilt. Der Erwartungswert von T ist:

$$\mathbb{E}T = \frac{1}{1/6} = 6.$$

Dieses Ergebnis steht im Einklang mit der Intuition: Im Durchschnitt ist jeder sechste Wurf eine 1, deshalb brauchen wir im Durchschnitt 6 Würfe, um eine 1 zu würfeln.

BEMERKUNG 6.4.7 (zum folgenden Satz). Angenommen, wir haben 100 Mal gewürfelt ohne auch ein einziges Mal eine 1 zu erzielen. Dies ist zwar sehr unwahrscheinlich, jedoch nicht unmöglich. Die Frage ist nun, wie lange müssen wir jetzt noch warten, bis eine 1 kommt? Man könnte meinen, dass aufgrund dessen, dass die 1 schon sehr lange überfällig ist, diese nun sehr bald kommen muss. Das ist jedoch nicht der Fall, da der Würfel kein Gedächtnis hat und von der Geschichte der bereits ausgeführten Würfe nichts weiß. Die Anzahl der Würfe, die nun noch benötigt werden, bis eine 1 kommt, ist nach wie vor geometrisch verteilt mit Parameter $\frac{1}{6}$. Diese Eigenschaft wird nun im folgenden Satz beschrieben.

SATZ 6.4.8 (Gedächtnislosigkeit der geometrischen Verteilung). *Sei $T \sim \text{Geo}(p)$, dann gilt:*

$$\mathbb{P}[T - n > k | T > n] = \mathbb{P}[T > k] \text{ für alle } n, k \in \mathbb{N}.$$

BEWEIS. Sei $m \in \mathbb{N}$. Zuerst berechnen wir die Wahrscheinlichkeit, dass $T > m$. Dieses Ereignis tritt genau dann ein, wenn die ersten m Experimente “Misserfolge” sind. Die Ausgänge der anderen Experimente sind dabei egal. Wegen der Unabhängigkeit der einzelnen Experimente hat dieses Ereignis Wahrscheinlichkeit $(1 - p)^m$. Man kann auch direkt vorgehen:

$$\mathbb{P}[T > m] = \sum_{i=m+1}^{\infty} \mathbb{P}[T = i] = \sum_{i=m+1}^{\infty} p(1-p)^{i-1} = (1-p)^m \sum_{i=1}^{\infty} p(1-p)^{i-1} = (1-p)^m.$$

Nun erhalten wir mit der Definition der bedingten Wahrscheinlichkeit, dass

$$\mathbb{P}[T - n > k | T > n] = \frac{\mathbb{P}[T > n+k, T > n]}{\mathbb{P}[T > n]} = \frac{\mathbb{P}[T > n+k]}{\mathbb{P}[T > n]} = \frac{(1-p)^{n+k}}{(1-p)^n} = (1-p)^k.$$

Dies stimmt mit $\mathbb{P}[T > k]$ überein. □

6.5. Negative Binomialverteilung

Wir betrachten wieder ein unendlich oft wiederholtes Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p \in (0, 1]$. Für $r \in \mathbb{N}$ bezeichnen wir mit T_r die Wartezeit auf den r -ten “Erfolg”.

BEISPIEL 6.5.1. Gehen die Experimente wie folgt aus:

$$0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, \dots,$$

so erhalten wir $T_1 = 4$, $T_2 = 7$, $T_3 = 9$, $T_4 = 10$. Dabei steht 0 für “Misserfolg” und 1 für “Erfolg”.

Wir haben bereits gezeigt, dass $T_1 \sim \text{Geo}(p)$. Wie ist nun T_r für ein allgemeines $r \in \mathbb{N}$ verteilt?

SATZ 6.5.2. *Für jedes $r \in \mathbb{N}$ ist die Wartezeit T_r auf den r -ten “Erfolg” in einem unendlich oft wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p wie folgt verteilt:*

$$(6.5.1) \quad \mathbb{P}[T_r = k] = \binom{k-1}{r-1} p^r (1-p)^{k-r} \text{ für alle } k = r, r+1, \dots$$

BEMERKUNG 6.5.3. Die Mindestanzahl an Experimenten, die man benötigt, um r “Erfolge” zu erzielen, ist r . Daher ist der kleinste mögliche Wert von T_r gleich r .

DEFINITION 6.5.4. Eine Zufallsvariable T_r , die (6.5.1) mit einem $r \in \mathbb{N}$ und einem $p \in (0, 1]$ erfüllt, heißt *negativ binomialverteilt* mit Parametern r und p .

BEZEICHNUNG: $T_r \sim \text{NB}(r, p)$.

BEMERKUNG 6.5.5. Die geometrische Verteilung ist ein Spezialfall der negativen Binomialverteilung: $\text{Geo}(p) = \text{NB}(1, p)$.

BEMERKUNG 6.5.6. Negative Binomialverteilung wird auch Pascal- oder Polya-Verteilung genannt.

BEWEIS VON SATZ 6.5.2. Seien $r \in \mathbb{N}$ und $k \geq r$ fest. Das Ereignis $\{T_r = k\}$ tritt genau dann ein, wenn die beiden folgenden Ereignisse eintreten:

$A =$ "Das k -te Experiment ist ein "Erfolg"",

$B =$ "In den Experimenten $1, \dots, k - 1$ werden genau $r - 1$ "Erfolge" erzielt".

Die Geschichte der Bernoulli-Experimente muss also wie folgt aussehen:

$$\underbrace{\begin{matrix} ??? & \dots & \dots & \dots & ? & 1, \\ 1 & 2 & 3 & & k-1 & k \end{matrix}}_{r-1 \text{ "Erfolge"}}$$

wobei das Fragezeichen für 0 oder 1 steht und genau $r - 1$ Fragezeichen Einsen sein sollen. Die Wahrscheinlichkeit von A ist p . Die Wahrscheinlichkeit von B berechnen wir mit Hilfe der Binomialverteilung:

$$\mathbb{P}[B] = \binom{k-1}{r-1} p^{r-1} (1-p)^{(k-1)-(r-1)} = \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r}.$$

Dabei sind A und B unabhängig. Es folgt:

$$\mathbb{P}[T_r = k] = \mathbb{P}[A] \cdot \mathbb{P}[B] = p \cdot \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

□

BEMERKUNG 6.5.7. Wir zeigen noch, dass die Summe aller Wahrscheinlichkeiten in (6.5.1) gleich 1 ist:

$$\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = 1.$$

Eigentlich folgt das aus Satz 6.5.2. Wir geben aber einen direkten Beweis. Wir nehmen uns die Binomische Formel zur Hilfe:

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k, \quad |x| < 1.$$

Diese Formel gilt für alle $\alpha \in \mathbb{R}$. Dabei muss α nicht unbedingt ganz und nicht unbedingt positiv sein. Der Binomialkoeffizient $\binom{\alpha}{k}$ ist definiert durch

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!}, \quad \alpha \in \mathbb{R}, k \in \mathbb{N}_0.$$

Wir setzen $\alpha = -r$ und $-x$ anstatt von x in die Formel ein:

$$\begin{aligned}
 (1-x)^{-r} &= \sum_{k=0}^{\infty} \binom{-r}{k} (-x)^k \\
 &= \sum_{k=0}^{\infty} \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!} \cdot (-1)^k x^k \\
 &= \sum_{k=0}^{\infty} \frac{r(r+1)\cdots(r+k-1)}{k!} \cdot x^k \\
 &= \sum_{k=0}^{\infty} \binom{r+k-1}{k} x^k.
 \end{aligned}$$

Mit $m = k + r$ erhalten wir dann

$$(1-x)^{-r} = \sum_{m=r}^{\infty} \binom{m-1}{m-r} x^{m-r} = \sum_{m=r}^{\infty} \binom{m-1}{r-1} x^{m-r}.$$

Somit erhalten wir schließlich mit $x = 1 - p$:

$$\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = p^r \sum_{m=r}^{\infty} \binom{m-1}{r-1} x^{m-r} = p^r (1-x)^{-r} = 1.$$

Die Verteilung heißt “negative Binomialverteilung”, da wir in der Binomischen Formel für α einen negativen Wert eingesetzt haben.

SATZ 6.5.8. Für $T_r \sim NB(r, p)$ gilt $\mathbb{E}T_r = \frac{r}{p}$.

BEWEISIDEE. Auf einen “Erfolg” muss man im Durchschnitt $\frac{1}{p}$ Experimente warten. Auf r “Erfolge” wartet man dementsprechend im Durchschnitt $\frac{r}{p}$ Experimente. \square

Wahrscheinlichkeitstheorie und Maßtheorie

7.1. Vorüberlegungen

Die folgenden drei Beispiele sind Spezialfälle des Oberbegriffs *Maß*.

BEISPIEL 7.1.1 (Verteilung der Ladung oder der Masse). Man stelle sich eine positive elektrische Ladung, die sich über eine Menge Ω verteilt hat. Dabei kann Ω zum Beispiel ein Gebiet im drei- oder zweidimensionalen Raum sein. Ist A eine Teilmenge von Ω , so kann man mit $\mu(A)$ die Gesamtladung bezeichnen, die sich in der Menge A befindet. Da wir nur positive Ladungen betrachten, kann $\mu(A)$ nur Werte im Bereich $[0, +\infty]$ annehmen. Dabei ist der Wert $+\infty$ zugelassen. Außerdem kann man davon ausgehen, dass die folgende Eigenschaft, genannt σ -Additivität, gilt: Für beliebige paarweise disjunkte Teilmengen $A_1, A_2, \dots \subset \Omega$ gilt

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

Das bedeutet, dass sich die Gesamtladung einer disjunkten Vereinigung $\cup_{i=1}^{\infty} A_i$ als die Summe der Ladungen der einzelnen Mengen A_i ausrechnen lässt. Analog kann man sich anstatt einer Verteilung der Ladung auch eine Verteilung der Masse im Gebiet Ω vorstellen. In diesem Fall ist $\mu(A)$ die Masse der Menge A .

BEISPIEL 7.1.2 (Wahrscheinlichkeit). Man stelle sich ein Zufallsexperiment mit Grundmenge Ω vor. In diesem Fall kann man mit $\mathbb{P}[A]$ die Wahrscheinlichkeit des Ereignisses $A \subset \Omega$ bezeichnen. Wir werden als Axiom annehmen, dass die σ -Additivität gilt: Für beliebige paarweise disjunkte Teilmengen $A_1, A_2, \dots \subset \Omega$ gilt

$$\mathbb{P}[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

Außerdem gilt noch eine Eigenschaft, die Normiertheit genannt wird: $\mathbb{P}[\Omega] = 1$.

BEISPIEL 7.1.3 (Volumen, Flächeninhalt, Länge). Für eine Menge A im dreidimensionalen Raum kann man mit $\lambda(A)$ das Volumen von A bezeichnen. Das Volumen $\lambda(A)$ nimmt Werte in $[0, +\infty]$ an. Man kann hoffen, dass die σ -Additivität gilt: Für beliebige paarweise disjunkte Teilmengen $A_1, A_2, \dots \subset \mathbb{R}^3$ gilt

$$\lambda(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \lambda(A_i).$$

Analog kann man im zweidimensionalen Raum den Flächeninhalt, im eindimensionalen Raum die Länge, und allgemeiner im d -dimensionalen Raum das d -dimensionale Volumen betrachten.

Die oben genannten Begriffe werden in der Maßtheorie als Spezialfälle des Begriffs Maß exakt definiert. Erstaunlicherweise stellt es sich heraus, dass sich der Begriff "Volumen" nicht für alle Teilmengen von \mathbb{R}^d vernünftig erklären lässt, sondern nur für sogenannte Borel-Mengen. Diese werden im Folgenden definiert. Analog kann man in einigen Situationen die Wahrscheinlichkeit nicht für alle Ereignisse erklären, sondern nur für sogenannte messbare Ereignisse. Bislang haben wir nur Experimente mit einer endlichen oder abzählbaren Grundmenge betrachtet. Die Frage der Messbarkeit spielt für solche Experimente keine Rolle. Es gibt aber auch Experimente mit einer überabzählbaren Grundmenge, Beispiele werden im Folgenden gegeben.

7.2. Geometrische Wahrscheinlichkeiten

Wir betrachten hier einige Beispiele von Experimenten mit einer überabzählbaren Grundmenge.

BEISPIEL 7.2.1. Sei Ω ein Quadrat in der Ebene. Stellen wir uns vor, dass jemand zufällig einen Punkt S im Quadrat Ω auswählt. Wie kann man die Wahrscheinlichkeit ausrechnen, dass S in einer Teilmenge $A \subset \Omega$ (zum Beispiel, einem kleineren Quadrat) landet? Früher haben wir den Laplace-Ansatz benutzt:

$$(7.2.1) \quad \mathbb{P}[S \in A] = \frac{\#A}{\#\Omega},$$

wobei $\#A$ für die Anzahl der Elemente in A steht. In diesem Beispiel funktioniert der Ansatz allerdings nicht, denn $\#A = \#\Omega = \infty$. Anstatt (7.2.1) ist es natürlich, die folgende Formel zu verwenden:

$$(7.2.2) \quad \mathbb{P}[S \in A] = \frac{\lambda(A)}{\lambda(\Omega)},$$

wobei $\lambda(A)$ der Flächeninhalt von A ist. Gilt (7.2.2), so sagen wir, dass der Punkt S *gleichverteilt* in Ω ist.

Beachte, dass der Flächeninhalt eines einzelnen Punktes gleich Null ist und demnach gilt für jeden einzelnen Punkt $\omega \in \Omega$:

$$\mathbb{P}[S = \omega] = 0.$$

BEISPIEL 7.2.2. Ein Zufallsgenerator erzeugt zwei zufällige reelle Zahlen x, y zwischen 0 und 1 unabhängig voneinander. Bestimme die Wahrscheinlichkeit des Ereignisses A , dass

$$x + y < 1.$$

LÖSUNG. Der Wahrscheinlichkeitsraum ist das Einheitsquadrat:

$$\Omega = [0, 1]^2 = \{(x, y) \in \mathbb{R}^2 : x \in [0, 1], y \in [0, 1]\}.$$

Das Ereignis A können wir wie folgt darstellen:

$$A = \{(x, y) \in [0, 1]^2 : x + y < 1\}.$$

Von einem idealen Zufallsgenerator erwartet man, dass der Punkt $S = (x, y)$ "gleichverteilt" in $[0, 1]^2$ ist. Gehen wir von dieser Annahme aus, so erhalten wir

$$\mathbb{P}[S \in A] = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{1/2}{1} = \frac{1}{2}.$$

BEISPIEL 7.2.3. Zwei Freunde wollen sich an einem bestimmten Ort treffen. Jeder der beiden Freunde kommt zu einem zufälligen Zeitpunkt zwischen 10:00 und 11:00 Uhr an und wartet 20 Minuten lang auf die Ankunft des anderen Freundes. Wenn der andere Freund innerhalb dieser 20 Minuten nicht erscheint, findet das Treffen nicht statt. Bestimme die Wahrscheinlichkeit des Ereignisses

$$A = \text{“Freunde treffen sich”}.$$

LÖSUNG. Die Ankunftszeit des ersten Freundes bezeichnen wir mit $10 + x$ (in Stunden), wobei $x \in [0, 1]$. Analog sei die Ankunftszeit des zweiten Freundes $10 + y$, mit $y \in [0, 1]$. Als Wahrscheinlichkeitsraum können wir dann das Einheitsquadrat betrachten:

$$\Omega = [0, 1]^2 = \{(x, y) \in \mathbb{R}^2 : x \in [0, 1], y \in [0, 1]\}.$$

Die Freunde treffen sich genau dann, wenn der Abstand zwischen x und y nicht größer als $\frac{1}{3}$ ist, wobei $\frac{1}{3}$ Stunde = 20 Minuten ist. Das Ereignis A ist also die Menge

$$A = \left\{ (x, y) \in [0, 1]^2 : |x - y| \leq \frac{1}{3} \right\}.$$

Wir werden nun davon ausgehen, dass der Punkt (x, y) “gleichverteilt” auf $[0, 1]^2$ ist. Dann können wir die Wahrscheinlichkeit von A wie folgt ausrechnen:

$$\mathbb{P}[A] = \frac{\lambda(A)}{\lambda(\Omega)} = \lambda(A) = 1 - \frac{4}{9} = \frac{5}{9}.$$

Dabei bezeichnet $\lambda(A)$ den Flächeninhalt von A .

7.3. Algebren

Sei Ω eine beliebige nichtleere Menge.

DEFINITION 7.3.1. Mit 2^Ω bezeichnen wir die Menge aller Teilmengen von Ω .

DEFINITION 7.3.2. Eine Teilmenge von 2^Ω (d.h. eine Menge, deren Elemente Teilmengen von Ω sind) werden wir als eine *Mengenfamilie* bezeichnen.

DEFINITION 7.3.3. Eine Mengenfamilie $\mathcal{F} \subset 2^\Omega$ heißt *Algebra* (oder *Boolsche Algebra*), wenn folgende drei Bedingungen erfüllt sind

- (1) $\Omega \in \mathcal{F}$.
- (2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$. (Komplementstabilität).
- (3) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$. (Vereinigungstabilität).

BEISPIEL 7.3.4. Folgende Mengenfamilien sind Algebren:

- (1) $\mathcal{F} = \{\emptyset, \Omega\}$.
- (2) $\mathcal{F} = 2^\Omega$.

BEISPIEL 7.3.5. Sei $\Omega = \{1, 2, 3\}$. Folgende Mengenfamilie ist eine Algebra:

$$\mathcal{F} = \{\emptyset, \{1, 2, 3\}, \{1\}, \{2, 3\}\}.$$

BEISPIEL 7.3.6. Sei $\Omega = [0, 1)$. Betrachte die Familie aller halbabgeschlossenen Teilintervalle von $[0, 1)$:

$$\mathcal{F} = \{[a, b) : 0 \leq a \leq b \leq 1\}.$$

Diese Mengenfamilie ist keine Algebra, denn \mathcal{F} ist weder komplementstabil noch vereinigungsstabil.

Betrachte nun die Familie aller endlichen Vereinigungen von halbabgeschlossenen Teilintervallen von $[0, 1)$:

$$\mathcal{G} = \{\cup_{k=1}^n [a_k, b_k) : n \in \mathbb{N}, 0 \leq a_1 \leq b_1 \leq 1, \dots, 0 \leq a_n \leq b_n \leq 1\}.$$

Diese Mengenfamilie ist eine Algebra.

BEISPIEL 7.3.7. Sei $\Omega = [0, 1)^d$. Ein (halbabgeschlossenes) Quader ist eine Menge der Form

$$Q = [a_1, b_1) \times \dots \times [a_d, b_d) \subset \mathbb{R}^d,$$

wobei $a_1 \leq b_1, \dots, a_d \leq b_d$. Die folgende Familie ist keine Algebra:

$$\mathcal{F} = \{Q : Q \subset [0, 1]^d \text{ und } Q \text{ ist Quader}\}.$$

Allerdings ist die folgende Familie aller endlichen Vereinigungen von Quadern eine Algebra:

$$\mathcal{G} = \{Q = Q_1 \cup \dots \cup Q_n : n \in \mathbb{N} \text{ und } Q_1, \dots, Q_n \subset [0, 1)^d \text{ sind Quader}\}.$$

Man kann diese Familie auch so beschreiben:

$$\mathcal{G} = \{Q = Q_1 \cup \dots \cup Q_n : n \in \mathbb{N}, \text{ und } Q_1, \dots, Q_n \subset [0, 1)^d \text{ sind disjunkte Quader}\}.$$

BEMERKUNG 7.3.8. Nimmt man eine endliche Anzahl von Elementen einer Algebra und wendet man auf diese Elemente beliebige mengentheoretische Operationen (wie z. B. $\cup, \cap, \Delta, \setminus, ^c$) in irgendeiner Reihenfolge an, so erhält man wieder ein Element aus der Algebra. Einige Spezialfälle dieser Aussage werden im folgenden Satz bewiesen.

SATZ 7.3.9. Sei $\mathcal{F} \subset 2^\Omega$ eine Algebra. Dann gilt:

- (1) $\emptyset \in \mathcal{F}$.
- (2) $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.
- (3) $A, B \in \mathcal{F} \Rightarrow A \setminus B \in \mathcal{F}$ und $A \Delta B \in \mathcal{F}$.
- (4) $A_1, \dots, A_n \in \mathcal{F} \Rightarrow A_1 \cap \dots \cap A_n \in \mathcal{F}$.
- (5) $A_1, \dots, A_n \in \mathcal{F} \Rightarrow A_1 \cup \dots \cup A_n \in \mathcal{F}$.

BEWEIS.

- (1) \mathcal{F} ist komplementstabil und $\Omega \in \mathcal{F}$ nach Definition. Es folgt, dass $\emptyset = \Omega^c \in \mathcal{F}$.
- (2) $A, B \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}, B^c \in \mathcal{F} \Rightarrow A^c \cup B^c \in \mathcal{F} \Rightarrow (A^c \cup B^c)^c \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$. Im letzten Schritt haben wir die Regel von de Morgan benutzt.
- (3) $A, B \in \mathcal{F} \Rightarrow A \in \mathcal{F}, B^c \in \mathcal{F} \Rightarrow A \setminus B = A \cap B^c \in \mathcal{F}$. Analog gilt auch $B \setminus A \in \mathcal{F}$. Es folgt, dass $A \Delta B = (A \setminus B) \cup (B \setminus A) \in \mathcal{F}$.
- (4) Induktion.
- (5) Induktion.

□

7.4. σ -Algebren

DEFINITION 7.4.1. Eine Mengenfamilie $\mathcal{F} \subset 2^\Omega$ heißt σ -Algebra, wenn folgende drei Bedingungen erfüllt sind:

- (1) $\Omega \in \mathcal{F}$.
- (2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$. (Komplementstabilität).
- (3) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow A_1 \cup A_2 \cup \dots \in \mathcal{F}$. (σ -Vereinigungsstabilität).

BEISPIEL 7.4.2. Folgende Mengenfamilien sind σ -Algebren:

- (1) $\mathcal{F} = \{\emptyset, \Omega\}$.
- (2) $\mathcal{F} = 2^\Omega$.

SATZ 7.4.3. Ist \mathcal{F} eine σ -Algebra, so ist \mathcal{F} auch eine Algebra.

BEWEIS. Sei \mathcal{F} eine σ -Algebra. Da \mathcal{F} nach Definition komplementstabil ist und $\Omega \in \mathcal{F}$, bleibt es nur noch zu zeigen, dass \mathcal{F} vereinigungsstabil ist. Seien dazu $A, B \in \mathcal{F}$. Wir zeigen, dass $A \cup B \in \mathcal{F}$. Zunächst gilt $\Omega \in \mathcal{F}$. Wegen Komplementstabilität von σ -Algebren gilt auch $\emptyset = \Omega^c \in \mathcal{F}$. Aus Eigenschaft (3) der σ -Algebren folgt nun, dass

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \dots \in \mathcal{F}.$$

Dies beweist die Vereinigungsstabilität von \mathcal{F} . □

BEISPIEL 7.4.4. Die Umkehrung von Satz 7.4.3 gilt nicht. Sei $\Omega = \mathbb{N}$. Die Mengenfamilie

$$\mathcal{F} = \{A \subset \mathbb{N} : A \text{ endlich oder } A^c \text{ endlich}\}$$

ist eine Algebra, aber keine σ -Algebra.

BEISPIEL 7.4.5. Die Mengenfamilie \mathcal{G} aus Beispiel 7.3.7 ist eine Algebra, aber keine σ -Algebra.

BEISPIEL 7.4.6. Sei $\Omega = \mathbb{R}$. Die folgende Mengenfamilie ist eine σ -Algebra (und somit auch eine Algebra):

$$\mathcal{F} = \{A \subset \mathbb{R} : A \text{ abzählbar oder } A^c \text{ abzählbar}\}.$$

SATZ 7.4.7. Sei \mathcal{F} eine σ -Algebra. Für beliebige $A_1, A_2, \dots \in \mathcal{F}$ gilt auch $A_1 \cap A_2 \cap \dots \in \mathcal{F}$. In Worten: Eine σ -Algebra ist σ -schnittstabil.

BEWEIS. Folgt aus der Regel von de Morgan: $A_1 \cap A_2 \cap \dots = (A_1^c \cup A_2^c \cup \dots)^c \in \mathcal{F}$. □

Der nächste Satz zeigt, wie man eine σ -Algebra auf eine Teilmenge einschränken kann.

SATZ 7.4.8. Sei $\mathcal{F} \subset 2^\Omega$ eine σ -Algebra. Sei $A \subset \Omega$ nichtleer. Definiere die folgende Mengenfamilie:

$$\mathcal{F}_A := \{A \cap B : B \in \mathcal{F}\} \subset 2^A.$$

Dann ist \mathcal{F}_A eine σ -Algebra.

BEWEIS. Wir beweisen nur die σ -Vereinigungsstabilität der Mengenfamilie \mathcal{F}_A . (Andere Eigenschaften sind Übung). Seien dazu $C_1, C_2, \dots \in \mathcal{F}_A$. Aus der Definition von \mathcal{F}_A folgt: es existieren $B_1, B_2, \dots \in \mathcal{F}$ mit $C_n = A \cap B_n$ für alle $n \in \mathbb{N}$. Wir haben die Darstellung

$$\bigcup_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} (A \cap B_n) = A \cap \left(\bigcup_{n=1}^{\infty} B_n\right).$$

Da $B_1, B_2, \dots \in \mathcal{F}$ und \mathcal{F} eine σ -Algebra ist, erhalten wir, dass $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$. Es folgt aus der Definition von \mathcal{F}_A , dass $A \cap (\bigcup_{n=1}^{\infty} B_n) \in \mathcal{F}_A$. Somit gilt $\bigcup_{n=1}^{\infty} C_n \in \mathcal{F}_A$. \square

7.5. Limes superior und Limes inferior für Folgen von Mengen

Es wird in diesem Abschnitt gezeigt, dass σ -Algebren bezüglich der Limesbildung von Folgen von Mengen abgeschlossen sind.

DEFINITION 7.5.1. Seien $A_1, A_2, \dots \subset \Omega$. Dann ist $\limsup_{n \rightarrow \infty} A_n$ eine Teilmenge von Ω , die wie folgt definiert wird:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} A_i = \{\omega \in \Omega : \text{für jedes } k \in \mathbb{N} \text{ existiert ein } i \geq k \text{ mit } \omega \in A_i\}.$$

In Worten kann man das Ereignis $\limsup_{n \rightarrow \infty} A_n$ wie folgt beschreiben:

$$\limsup_{n \rightarrow \infty} A_n = \text{“Das Ereignis } A_i \text{ tritt für unendliche viele } i \in \mathbb{N} \text{ ein”}.$$

DEFINITION 7.5.2. Seien $A_1, A_2, \dots \subset \Omega$. Dann ist $\liminf_{n \rightarrow \infty} A_n$ eine Teilmenge von Ω , die wie folgt definiert wird:

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} A_i = \{\omega \in \Omega : \text{es gibt ein } k \in \mathbb{N} \text{ so dass } \omega \in A_i \text{ für alle } i \geq k\}.$$

In Worten lässt sich $\liminf_{n \rightarrow \infty} A_n$ wie folgt beschreiben:

$$\liminf_{n \rightarrow \infty} A_n = \text{“Das Ereignis } A_i \text{ tritt für alle bis auf endlich viele Werte von } i \in \mathbb{N} \text{ ein”}.$$

BEISPIEL 7.5.3. Eine Münze werde unendlich oft geworfen. Die Grundmenge ist

$$\Omega = \{K, Z\}^{\infty} = \{(a_1, a_2, \dots) : a_n \in \{K, Z\} \text{ für alle } n \in \mathbb{N}\}.$$

Definiere Ereignisse A_1, A_2, \dots wie folgt:

$$A_n = \text{“Münze zeigt Kopf bei Wurf Nummer } n\text{”} = \{(a_1, a_2, \dots) \in \{K, Z\}^{\infty} : a_n = K\}.$$

Dann gilt:

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &= \text{“Münze zeigt unendlich oft Kopf”}, \\ \liminf_{n \rightarrow \infty} A_n &= \text{“Ab irgendwann zeigt Münze nur noch Kopf”} \\ &= \text{“Münze zeigt nur endlich oft Zahl”}. \end{aligned}$$

BEMERKUNG 7.5.4. Es gilt:

- (1) $\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n$.
- (2) $(\liminf_{n \rightarrow \infty} A_n)^c = \limsup_{n \rightarrow \infty} (A_n^c)$ und $(\limsup_{n \rightarrow \infty} A_n)^c = \liminf_{n \rightarrow \infty} (A_n^c)$.

BEMERKUNG 7.5.5. Sind $A_1, A_2, \dots \in \mathcal{F}$ und ist \mathcal{F} eine σ -Algebra, so gilt auch $\limsup_{n \rightarrow \infty} A_n \in \mathcal{F}$ und $\liminf_{n \rightarrow \infty} A_n \in \mathcal{F}$.

BEMERKUNG 7.5.6. Die Indikatorfunktion von $\limsup_{n \rightarrow \infty} A_n$ ist $\limsup_{n \rightarrow \infty}$ von Indikatorfunktionen:

$$\mathbb{1}_{\limsup_{n \rightarrow \infty} A_n} = \limsup_{n \rightarrow \infty} \mathbb{1}_{A_n}.$$

Analog für \liminf :

$$\mathbb{1}_{\liminf_{n \rightarrow \infty} A_n} = \liminf_{n \rightarrow \infty} \mathbb{1}_{A_n}.$$

7.6. Borel- σ -Algebra

Der nächste Satz besagt, dass der Schnitt von σ -Algebren wieder eine σ -Algebra ist.

SATZ 7.6.1. *Sei I eine beliebige nichtleere Menge. Für jedes $i \in I$ sei $\mathcal{F}_i \subset 2^\Omega$ eine σ -Algebra. Dann ist auch die Mengenfamilie*

$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i = \{A \subset \Omega : A \in \mathcal{F}_i \text{ für alle } i \in I\}$$

eine σ -Algebra.

BEWEIS. Wir zeigen nur, dass \mathcal{F} σ -vereinigungsstabil ist. Andere Bedingungen werden analog gezeigt. Seien $A_1, A_2, \dots \in \mathcal{F}$. Dann gilt $A_1, A_2, \dots \in \mathcal{F}_i$ für jedes $i \in I$. Da \mathcal{F}_i eine σ -Algebra ist, erhalten wir, dass $A_1 \cup A_2 \cup \dots \in \mathcal{F}_i$ für jedes $i \in I$. Nach Definition von \mathcal{F} heißt es aber, dass $A_1 \cup A_2 \cup \dots \in \mathcal{F}$. \square

DEFINITION 7.6.2. Sei $\mathcal{E} \subset 2^\Omega$ eine beliebige nichtleere Mengenfamilie. Die Mengenfamilie

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{F}: \mathcal{E} \subset \mathcal{F} \subset 2^\Omega \\ \mathcal{F} \text{ ist } \sigma\text{-Algebra}}} \mathcal{F}$$

heißt die *von \mathcal{E} erzeugte σ -Algebra*. Der Schnitt wird über alle σ -Algebren $\mathcal{F} \subset 2^\Omega$, die \mathcal{E} enthalten, genommen.

BEMERKUNG 7.6.3. Aus Satz 7.6.1 folgt, dass $\sigma(\mathcal{E})$ tatsächlich eine σ -Algebra ist.

BEMERKUNG 7.6.4. Eine äquivalente Beschreibung: $\sigma(\mathcal{E})$ ist die kleinste σ -Algebra, die alle Mengen aus \mathcal{E} enthält. Dabei heißt “die kleinste” folgendes: ist $\mathcal{F} \subset 2^\Omega$ irgendeine σ -Algebra, die \mathcal{E} enthält, so gilt $\sigma(\mathcal{E}) \subset \mathcal{F}$.

DEFINITION 7.6.5. Sei $\Omega = \mathbb{R}$ und \mathcal{E} die Familie aller Intervalle der Form $(-\infty, a]$ mit $a \in \mathbb{R}$. Die von \mathcal{E} erzeugte σ -Algebra $\mathcal{B} := \sigma(\mathcal{E}) \subset 2^\mathbb{R}$ heißt die *Borel- σ -Algebra* auf \mathbb{R} . Elemente von \mathcal{B} heißen *Borel-Mengen*.

BEMERKUNG 7.6.6. Man kann zeigen, dass die Borel- σ -Algebra \mathcal{B} von jeder der folgenden Mengenfamilien erzeugt wird:

- (1) Halbabgeschlossene Intervalle $(a, b]$.
- (2) Halbabgeschlossene Intervalle $[a, b)$.
- (3) Offene Intervalle (a, b) .
- (4) Abgeschlossene Intervalle $[a, b]$.
- (5) Offene Teilmengen von \mathbb{R} .
- (6) Abgeschlossene Teilmengen von \mathbb{R} .

Die obige Definition kann man auf höhere Dimensionen verallgemeinern.

DEFINITION 7.6.7. Sei $\Omega = \mathbb{R}^d$ und \mathcal{E} die Familie aller “Oktanten” der Form

$$(-\infty, a_1] \times \dots \times (-\infty, a_d]$$

mit $a_1, \dots, a_d \in \mathbb{R}$. Die von \mathcal{E} erzeugte σ -Algebra $\mathcal{B}^d := \sigma(\mathcal{E}) \subset 2^{\mathbb{R}^d}$ heißt die *Borel- σ -Algebra* auf \mathbb{R}^d . Elemente von \mathcal{B}^d heißen *Borel-Mengen*.

BEMERKUNG 7.6.8. Man kann zeigen, dass die Borel- σ -Algebra \mathcal{B}^d von jeder der folgenden Mengenfamilien erzeugt wird:

- (1) Halbabgeschlossene Quader $(a_1, b_1] \times \dots \times (a_d, b_d]$.
- (2) Halbabgeschlossene Quader $[a_1, b_1) \times \dots \times [a_d, b_d)$.
- (3) Offene Quader $(a_1, b_1) \times \dots \times (a_d, b_d)$.
- (4) Abgeschlossene Quader $[a_1, b_1] \times \dots \times [a_d, b_d]$.
- (5) Offene Teilmengen von \mathbb{R}^d .
- (6) Abgeschlossene Teilmengen von \mathbb{R}^d .

BEMERKUNG 7.6.9. Somit ist jede offene Teilmenge und jede abgeschlossene Teilmenge von \mathbb{R}^d eine Borel-Menge. Man kann sich dann fragen, ob es überhaupt nicht-Borel Mengen gibt. Man kann zeigen, dass

- (1) Die Familie der Borel-Mengen \mathcal{B}^d ist gleichmächtig mit \mathbb{R} .
- (2) Die Familie $2^{\mathbb{R}^d}$ aller Teilmengen von \mathbb{R}^d hat eine strikt größere Mächtigkeit, als \mathbb{R} .

Somit gibt es Teilmengen von \mathbb{R}^d , die keine Borel-Mengen sind. Es ist allerdings nicht einfach, solche Mengen zu konstruieren. Im Weiteren werden wir es nur mit Borel-Mengen zu tun haben.

7.7. Maße

DEFINITION 7.7.1. Sei Ω eine nichtleere Menge und $\mathcal{F} \subset 2^\Omega$ eine σ -Algebra. Dann heißt das Paar (Ω, \mathcal{F}) ein *Messraum*. Mengen (oder Ereignisse) $A \subset \Omega$ mit $A \in \mathcal{F}$ heißen *messbar*.

DEFINITION 7.7.2. Sei (Ω, \mathcal{F}) ein Messraum. Ein *Maß* μ auf (Ω, \mathcal{F}) ist eine Funktion $\mu : \mathcal{F} \rightarrow [0, +\infty]$ mit der folgenden Eigenschaft, die *σ -Additivität* genannt wird: Für alle paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{F}$ gilt

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Das Tripel $(\Omega, \mathcal{F}, \mu)$ heißt ein *Maßraum*.

BEISPIEL 7.7.3. Sei $\omega_1, \omega_2, \dots \in \Omega$ eine beliebige Folge und $m_1, m_2, \dots \geq 0$ beliebige Zahlen. Man kann dann das folgende Maß definieren:

$$\mu(A) = \sum_{i \in \mathbb{N}: \omega_i \in A} m_i, \quad A \in \mathcal{F}.$$

Man kann sich vorstellen, dass μ eine Verteilung der positiven elektrischen Ladung auf Ω beschreibt, bei der sich die Ladung nur in den Punkten $\omega_1, \omega_2, \dots$ sammelt, wobei die Ladung des Punktes ω_i gleich m_i ist.

Wir geben nun eine Definition des Volumens einer Menge $A \subset \mathbb{R}^d$.

DEFINITION 7.7.4. Sei $A \subset \mathbb{R}^d$ beliebig. Definiere $\lambda(A)$, das *Lebesgue-Maß* von A , wie folgt:

- (1) Ist $B = [a_1, b_1] \times \dots \times [a_d, b_d]$ ein Quader, so sei

$$\lambda(B) = (b_1 - a_1) \cdot \dots \cdot (b_d - a_d).$$

(2) Ist $A \subset \mathbb{R}^d$ beliebig, so sei

$$\lambda(A) = \inf \left\{ \sum_{i=1}^{\infty} \lambda(B_i) : B_1, B_2, \dots \text{ sind Quader mit } A \subset \bigcup_{i=1}^{\infty} B_i \right\}.$$

Allerdings ist die so konstruierte Funktion λ kein Maß auf $(\mathbb{R}^d, 2^{\mathbb{R}^d})$: sie ist nicht σ -additiv und sogar nicht additiv.

SATZ 7.7.5 (Vitali). *Es existieren zwei Mengen $A_1, A_2 \in \mathbb{R}^d$ mit $A_1 \cap A_2 = \emptyset$, so dass*

$$\lambda(A_1 \cup A_2) \neq \lambda(A_1) + \lambda(A_2).$$

Wenn wir aber die Mengenfunktion λ nur auf die σ -Algebra der Borel-Mengen einschränken, wird sie σ -additiv.

SATZ 7.7.6 (Satz von Lebesgue). *λ ist ein Maß auf $(\mathbb{R}^d, \mathcal{B}^d)$.*

7.8. Wahrscheinlichkeitsmaße

Die axiomatische Begründung der Wahrscheinlichkeitstheorie auf der Grundlage der Maßtheorie wurde von A. N. Kolmogorov im Jahr 1929 gegeben.

DEFINITION 7.8.1. Sei (Ω, \mathcal{F}) ein Messraum. Ein *Wahrscheinlichkeitsmaß* \mathbb{P} auf (Ω, \mathcal{F}) ist eine Funktion $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ mit folgenden zwei Eigenschaften:

- (1) Normiertheit: $\mathbb{P}[\Omega] = 1$.
- (2) σ -Additivität: Für alle paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{F}$ gilt

$$\mathbb{P}[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

Das Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ heißt ein *Wahrscheinlichkeitsraum*.

BEISPIEL 7.8.2 (diskrete Wahrscheinlichkeitsräume). Sei Ω eine endliche oder abzählbare Menge. Sei $p : \Omega \rightarrow [0, 1]$ eine Funktion mit $\sum_{\omega \in \Omega} p(\omega) = 1$. Definiere $\mathcal{F} = 2^{\Omega}$ und $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$:

$$\mathbb{P}[A] = \sum_{\omega \in A} p(\omega), \quad A \in 2^{\Omega}.$$

Dann ist \mathbb{P} ein Wahrscheinlichkeitsmaß auf $(\Omega, 2^{\Omega})$. Wahrscheinlichkeitsräume $(\Omega, 2^{\Omega}, \mathbb{P})$, die auf diese Weise mit einem endlichen oder abzählbaren Ω konstruiert wurden, heißen *diskrete Wahrscheinlichkeitsräume*.

BEISPIEL 7.8.3 (geometrische Wahrscheinlichkeiten, siehe Kapitel 7.2). Sei $\Omega \subset \mathbb{R}^d$ eine Borel-Menge mit $\lambda(\Omega) \neq 0$ und $\lambda(\Omega) \neq \infty$. Stellen wir uns vor, dass ein zufälliger, "gleichverteilter" Punkt S in der Menge Ω ausgewählt wird. Als Grundmenge dieses Experiments können wir dann Ω betrachten. Als σ -Algebra auf Ω wählen wir die Einschränkung

$$\mathcal{B}_{\Omega}^d = \{B \cap \Omega : B \in \mathcal{B}^d\}$$

der Borel- σ -Algebra \mathcal{B}^d auf Ω ; siehe Satz 7.4.8. Dann definieren wir folgendes Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{B}_{\Omega}^d)$:

$$\mathbb{P}[A] = \frac{\lambda(A)}{\lambda(\Omega)}, \quad A \in \mathcal{B}_{\Omega}^d.$$

BEMERKUNG 7.8.4 (idealer Zufallsgenerator). Im obigen Beispiel sei $\Omega = [0, 1]$. Für jeden einzelnen Punkt $\omega \in [0, 1]$ gilt $\lambda(\{\omega\}) = 0$ und somit

$$\mathbb{P}[\{\omega\}] = 0.$$

Sei nun $A = \{\omega_1, \omega_2, \dots\} \subset [0, 1]$ abzählbar, z. B. $A = \mathbb{Q} \cap [0, 1]$, wobei \mathbb{Q} die Mengen der rationalen Zahlen ist. Aus der σ -Additivität folgt dann, dass

$$\mathbb{P}[A] = \sum_{i \in \mathbb{N}} \mathbb{P}[\{\omega_i\}] = 0.$$

Somit ist die Wahrscheinlichkeit, dass ein idealer Zufallsgenerator eine rationale Zahl erzeugt, gleich 0.

Alle Eigenschaften der Wahrscheinlichkeit, die wir in Kapitel 1 bewiesen haben, wie z. B. die Monotonie, die Additivität, die Siebformel, gelten nach wie vor. Man muss nur annehmen, dass alle betrachteten Ereignisse messbar sind. Wir beweisen nun einige weitere Eigenschaften der Wahrscheinlichkeit.

SATZ 7.8.5 (σ -Subadditivität). Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien $A_1, A_2, \dots \in \mathcal{F}$ beliebig und nicht notwendigerweise disjunkt. Dann gilt

$$\mathbb{P}[\cup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

BEWEIS. Definiere:

$$B_1 = A_1, \quad B_2 = A_2 \setminus B_1, \quad B_3 = A_3 \setminus (A_1 \cup A_2), \quad \dots, \quad B_n = A_n \setminus (A_1 \cup \dots \cup A_{n-1}), \quad \dots$$

Da $A_1, A_2, \dots \in \mathcal{F}$ und \mathcal{F} eine σ -Algebra ist, sind die Mengen B_1, B_2, \dots messbar. Die Mengen B_1, B_2, \dots sind disjunkt und es gilt $\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i$. Aus der σ -Additivität von \mathbb{P} folgt, dass

$$\mathbb{P}[\cup_{i=1}^{\infty} A_i] = \mathbb{P}[\cup_{i=1}^{\infty} B_i] = \sum_{i=1}^{\infty} \mathbb{P}[B_i] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

Dabei haben wir im letzten Schritt benutzt, dass $B_i \subset A_i$ und somit $\mathbb{P}[B_i] \leq \mathbb{P}[A_i]$. \square

DEFINITION 7.8.6. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Ein Ereignis $A \in \mathcal{F}$ mit $\mathbb{P}[A] = 0$ heißt ein *Nullereignis*. Ein Ereignis $A \in \mathcal{F}$ mit $\mathbb{P}[A] = 1$ heißt ein *fast sicheres Ereignis*.

SATZ 7.8.7. Die Vereinigung von abzählbar vielen Nullereignissen ist wieder ein Nullereignis. Der Schnitt von abzählbar vielen fast sicheren Ereignissen ist wieder ein fast sicheres Ereignis.

BEWEIS. Übung. Die erste Aussage folgt aus der σ -Subadditivität. Die zweite Aussage: Regel von de Morgan. \square

SATZ 7.8.8 (Stetigkeit von \mathbb{P}). Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien $A_1, A_2, \dots \in \mathcal{F}$ messbare Mengen. Dann gelten folgende zwei Aussagen.

$$(1) \text{ Aus } A_1 \subset A_2 \subset \dots \text{ folgt } \mathbb{P}[\cup_{i=1}^{\infty} A_i] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n].$$

(2) Aus $A_1 \supset A_2 \supset \dots$ folgt $\mathbb{P}[\cap_{i=1}^{\infty} A_i] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n]$.

BEWEIS VON TEIL 1. Es gelte $A_1 \subset A_2 \subset \dots$. Sei $A_0 = \emptyset$. Definiere:

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad \dots, \quad B_n = A_n \setminus A_{n-1}, \dots$$

Es gilt: die Mengen B_1, B_2, \dots sind messbar, disjunkt und $\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i$. Aus der Additivität von \mathbb{P} folgt, dass

$$\mathbb{P}[\cup_{i=1}^{\infty} A_i] = \mathbb{P}[\cup_{i=1}^{\infty} B_i] = \sum_{i=1}^{\infty} \mathbb{P}[B_i] = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}[B_i].$$

Da nun $\mathbb{P}[B_i] = \mathbb{P}[A_i] - \mathbb{P}[A_{i-1}]$ für jedes $i \in \mathbb{N}$ gilt, erhalten wir

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}[B_i] = \lim_{n \rightarrow \infty} (\mathbb{P}[A_1] + \mathbb{P}[A_2] - \mathbb{P}[A_1] + \dots + \mathbb{P}[A_n] - \mathbb{P}[A_{n-1}]) = \lim_{n \rightarrow \infty} \mathbb{P}[A_n].$$

BEWEIS VON TEIL 2. Es gelte $A_1 \supset A_2 \supset \dots$. Mit den Regeln von de Morgan erhalten wir

$$\mathbb{P}[\cap_{i=1}^{\infty} A_i] = 1 - \mathbb{P}[(\cap_{i=1}^{\infty} A_i)^c] = 1 - \mathbb{P}[\cup_{i=1}^{\infty} (A_i^c)].$$

Allerdings gilt $A_1^c \subset A_2^c \subset \dots$ und somit können wir die bereits bewiesene Aussage von Teil 1 auf die Folge A_1^c, A_2^c, \dots anwenden:

$$1 - \mathbb{P}[\cup_{i=1}^{\infty} (A_i^c)] = 1 - \lim_{n \rightarrow \infty} \mathbb{P}[A_n^c] = 1 - \lim_{n \rightarrow \infty} (1 - \mathbb{P}[A_n]) = \lim_{n \rightarrow \infty} \mathbb{P}[A_n].$$

□

7.9. Das Lemma von Borel–Cantelli

LEMMA 7.9.1 (Borel–Cantelli). Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien $A_1, A_2, \dots \in \mathcal{F}$ Ereignisse.

(1) Angenommen, dass $\sum_{i=1}^{\infty} \mathbb{P}[A_i] < \infty$. Dann gilt $\mathbb{P}[\limsup A_n] = 0$. Mit anderen Worten:

$$\mathbb{P}[\text{“Es treten unendlich viele } A_n \text{ ein”}] = 0.$$

(2) Angenommen, dass $\sum_{i=1}^{\infty} \mathbb{P}[A_i] = \infty$ und dass zusätzlich die Ereignisse A_1, A_2, \dots unabhängig sind. Dann gilt $\mathbb{P}[\limsup A_n] = 1$. Mit anderen Worten:

$$\mathbb{P}[\text{“Es treten unendlich viele } A_n \text{ ein”}] = 1.$$

BEISPIEL 7.9.2. Wir betrachten ein unendlich oft wiederholtes Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p \in (0, 1)$. Betrachte die Ereignisse

$$A_n = \text{“Erfolg bei Experiment } n\text{”}.$$

Bestimme $\mathbb{P}[\limsup A_n]$ und $\mathbb{P}[\liminf A_n]$.

LÖSUNG. Es ist $\mathbb{P}[A_i] = p > 0$. Somit gilt

$$\sum_{i=1}^{\infty} \mathbb{P}[A_i] = \sum_{i=1}^{\infty} p = \infty.$$

Da die Ereignisse A_1, A_2, \dots unabhängig sind, können wir den zweiten Teil des Borel–Cantelli–Lemmas anwenden:

$$\mathbb{P}[\limsup A_n] = 1.$$

In Worten: Die Wahrscheinlichkeit, dass man in einem unendlich oft wiederholten Bernoulli-Experiment unendlich viele Erfolge erzielt, ist 1. Dies ist im Einklang mit der Intuition: werfen wir eine Münze unendlich oft, so werden wir mit Wahrscheinlichkeit 1 unendlich oft “Kopf” sehen. Analog zeigt man, dass

$$\mathbb{P}[\limsup(A_n^c)] = 1.$$

In Worten: Die Wahrscheinlichkeit, dass man in einem unendlich oft wiederholten Bernoulli-Experiment unendlich viele Misserfolge erzielt, ist 1.

Für die Wahrscheinlichkeit von $\liminf A_n$ erhalten wir dann

$$\begin{aligned} \mathbb{P}[\liminf A_n] &= \mathbb{P}[\text{“Ab irgendwann nur noch Erfolge”}] \\ &= \mathbb{P}[\text{“Nur endlich viele Misserfolge”}] \\ &= 1 - \mathbb{P}[\text{“Unendlich viele Misserfolge”}] \\ &= 1 - \mathbb{P}[\limsup(A_n^c)] \\ &= 0. \end{aligned}$$

Dies steht im Einklang mit der Intuition: Die Wahrscheinlichkeit, dass die Münze beim unendlichen Münzwurf ab irgendwann nur noch “Kopf” zeigt, ist 0. Analog gilt

$$\mathbb{P}[\liminf(A_n^c)] = \mathbb{P}[\text{“Ab irgendwann nur noch Misserfolge”}] = 0.$$

BEISPIEL 7.9.3. Es seien A_1, A_2, \dots unabhängige Ereignisse mit

$$\mathbb{P}[A_n] = \frac{1}{n^\alpha},$$

wobei $\alpha > 0$ ein Parameter ist. Bestimme $\mathbb{P}[\limsup A_n]$.

LÖSUNG. Es gilt

$$\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} \text{ ist } \begin{cases} \text{unendlich,} & \text{falls } \alpha \leq 1, \\ \text{endlich,} & \text{falls } \alpha > 1. \end{cases}$$

Mit dem Lemma von Borel–Cantelli (Teil 2 im Fall $\alpha \leq 1$ und Teil 1 im Fall $\alpha > 1$) erhalten wir

$$\mathbb{P}[\limsup A_n] = \mathbb{P}[\text{“Unendlich viele } A_n \text{ treten ein”}] = \begin{cases} 1, & \text{falls } \alpha \leq 1, \\ 0, & \text{falls } \alpha > 1. \end{cases}$$

BEWEIS DES LEMMAS VON BOREL–CANTELLI, TEIL 1. Definiere $B_k = \cup_{n \geq k} A_n$. Dann sind B_1, B_2, \dots messbar und es gilt $B_1 \supset B_2 \supset \dots$. Mit der Definition von \limsup und mit dem Stetigkeitssatz 7.8.8 erhalten wir

$$\mathbb{P}[\limsup A_n] = \mathbb{P}[\cap_{k=1}^{\infty} B_k] = \lim_{k \rightarrow \infty} \mathbb{P}[B_k].$$

Nun benutzen wir die σ -Subadditivität (Satz 7.8.5):

$$\lim_{k \rightarrow \infty} \mathbb{P}[B_k] \leq \lim_{k \rightarrow \infty} \sum_{n \geq k} \mathbb{P}[A_n] = 0.$$

Der letzte Schritt folgt aus der Konvergenz der Reihe $\sum_{n=1}^{\infty} \mathbb{P}[A_n]$. □

BEWEIS DES LEMMAS VON BOREL–CANTELLI, TEIL 2.

Seien nun A_1, A_2, \dots unabhängig mit $p_n = \mathbb{P}[A_n]$ und $\sum_{n=1}^{\infty} p_n = \infty$. Für $k, m \in \mathbb{N}$ definiere das Ereignis

$$B_{k,m} = \cup_{n=k}^{k+m} A_n.$$

Diese Ereignisse sind messbar. Sei nun $k \in \mathbb{N}$ fest. Es gilt

$$B_{k,1} \subset B_{k,2} \subset \dots \text{ und } \cup_{m=1}^{\infty} B_{k,m} = \cup_{n=k}^{\infty} A_n \stackrel{def}{=} B_k.$$

Aus dem Stetigkeitssatz (Satz 7.8.8) folgt nun, dass

$$\mathbb{P}[B_k] = \lim_{m \rightarrow \infty} \mathbb{P}[B_{k,m}] = \lim_{m \rightarrow \infty} (1 - \mathbb{P}[(\cup_{n=k}^{k+m} A_n)^c]) = \lim_{m \rightarrow \infty} (1 - \mathbb{P}[\cap_{n=k}^{k+m} (A_n^c)]).$$

Ereignisse A_1, A_2, \dots sind unabhängig nach Voraussetzung. Also sind auch Ereignisse A_1^c, A_2^c, \dots unabhängig. Es folgt, dass

$$\mathbb{P}[B_k] = \lim_{m \rightarrow \infty} \left(1 - \prod_{n=k}^{k+m} \mathbb{P}[A_n^c] \right) = \lim_{m \rightarrow \infty} \left(1 - \prod_{n=k}^{k+m} (1 - p_n) \right).$$

Nun wenden wir auf die rechte Seite die Ungleichung $1 - p \leq e^{-p}$ an:

$$\mathbb{P}[B_k] \geq \liminf_{m \rightarrow \infty} \left(1 - \prod_{n=k}^{k+m} e^{-p_n} \right) = \liminf_{m \rightarrow \infty} (1 - e^{-(p_k + p_{k+1} + \dots + p_{k+m})}) = 1,$$

wobei der letzte Schritt aus der Divergenz der Reihe $p_k + p_{k+1} + \dots$ folgt. Wir haben gezeigt, dass $\mathbb{P}[B_k] = 1$ für alle $k \in \mathbb{N}$. Da der Schnitt von abzählbar vielen fast sicheren Ereignissen ebenfalls fast sicher ist (Satz 7.8.7), erhalten wir

$$\mathbb{P}[\cap_{k=1}^{\infty} B_k] = 1.$$

Die Ereignisse $\limsup A_n$ und $\cap_{k=1}^{\infty} B_k$ sind aber nach der Definition von \limsup gleich. □

Zufallsvariablen: Die allgemeine Definition

8.1. Zufallsvariablen

Bis zu diesem Zeitpunkt haben wir ausschließlich Zufallsvariablen mit endlich oder abzählbar vielen Werten (also diskrete Zufallsvariablen) betrachtet. Jetzt werden wir allgemeine Zufallsvariablen einführen.

DEFINITION 8.1.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt *messbar*, wenn für alle $a \in \mathbb{R}$ gilt:

$$\{X \leq a\} \in \mathcal{F}.$$

Hierbei ist $\{X \leq a\}$ die Menge aller Punkte im Wahrscheinlichkeitsraum, wo die Funktion X einen Wert $\leq a$ annimmt:

$$\{X \leq a\} = \{\omega \in \Omega : X(\omega) \leq a\} \subset \Omega.$$

Eine messbare Funktion nennen wir auch eine *Zufallsvariable*.

Für eine Zufallsvariable X ist also die Wahrscheinlichkeit $\mathbb{P}[X \leq a]$ wohldefiniert. Der nächste Satz besagt, dass auch die Wahrscheinlichkeit $\mathbb{P}[X \in B]$ wohldefiniert ist, wobei $B \subset \mathbb{R}$ eine beliebige Borel-Menge ist.

SATZ 8.1.2. Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann gilt für jede Borel-Menge $B \subset \mathbb{R}$:

$$\{X \in B\} \in \mathcal{F}.$$

Hierbei ist

$$\{X \in B\} = X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}.$$

BEMERKUNG 8.1.3. Aus diesem Satz folgt, dass für eine Zufallsvariable X gilt:

- (1) Das Ereignis $\{X = a\}$ ist messbar, für alle $a \in \mathbb{R}$.
- (2) Das Ereignis $\{X \in A\}$ ist messbar, für jede höchstens abzählbare Menge $A \subset \mathbb{R}$.
- (3) Somit ist auch das Ereignis $\{X \notin A\} = \{X \in A\}^c$ ebenfalls messbar, für jede höchstens abzählbare Menge $A \subset \mathbb{R}$.
- (4) Insbesondere ist das Ereignis $\{X \in \mathbb{Q}\}$ messbar.
- (5) Ereignisse $\{a < X < b\}$, $\{a \leq X \leq b\}$, $\{a < X \leq b\}$, $\{a \leq X < b\}$ sind messbar.

Für den Beweis von Satz 8.1.2 benötigen wir eine Hilfsaussage.

PROPOSITION 8.1.4. Seien (Ω, \mathcal{F}) ein Messraum und E eine Menge. Außerdem seien $X : \Omega \rightarrow E$ eine Abbildung und $\mathcal{E} \subset 2^E$ eine Mengenfamilie mit der Eigenschaft, dass $X^{-1}(B) \in \mathcal{F}$ für jede Menge $B \in \mathcal{E}$. Dann gilt auch $X^{-1}(B) \in \mathcal{F}$ für jede Menge $B \in \sigma(\mathcal{E})$.

BEMERKUNG 8.1.5. Mit anderen Worten: Um zu zeigen, dass die Urbilder aller Mengen aus einer σ -Algebra messbar sind, reicht es zu zeigen, dass die Urbilder aller Mengen aus einem Erzeuger dieser σ -Algebra messbar sind.

BEWEIS VON PROPOSITION 8.1.4. Wir wollen zeigen, dass das Urbild jeder Menge aus $\sigma(\mathcal{E})$ ein Element von \mathcal{F} ist. Deshalb betrachten wir die Familie

$$\mathcal{A} = \{B \subset E : X^{-1}(B) \in \mathcal{F}\} \subset 2^E.$$

Wir werden im Weiteren zeigen, dass die Familie \mathcal{A} eine σ -Algebra ist. Außerdem gilt $\mathcal{E} \subset \mathcal{A}$ laut Voraussetzung. Die Familie \mathcal{A} ist also eine σ -Algebra, die \mathcal{E} enthält. Die von \mathcal{E} erzeugte σ -Algebra $\sigma(\mathcal{E})$ ist (laut Definition der erzeugten σ -Algebra) die *kleinste* σ -Algebra, die \mathcal{E} enthält. Somit muss $\sigma(\mathcal{E}) \subset \mathcal{A}$ gelten. Für jede Menge $B \in \sigma(\mathcal{E})$ gilt dann $B \in \mathcal{A}$. Laut Definition von \mathcal{A} bedeutet das, dass $X^{-1}(B) \in \mathcal{F}$ für jede Menge $B \in \sigma(\mathcal{E})$. Das beweist die Behauptung der Proposition.

Wir werden nun zeigen, dass für die Familie \mathcal{A} alle drei Bedingungen aus der Definition einer σ -Algebra gelten.

Bedingung 1. Es gilt $E \in \mathcal{A}$, denn $X^{-1}(E) = \Omega$ und $\Omega \in \mathcal{F}$.

Bedingung 2. Wir zeigen, dass \mathcal{A} komplementstabil ist. Sei also $A \in \mathcal{A}$. Wir zeigen, dass $A^c \in \mathcal{A}$. Es gilt

$$X^{-1}(A^c) = \{\omega \in \Omega : X(\omega) \in A^c\} = \{\omega \in \Omega : X(\omega) \notin A\} = \{\omega \in \Omega : X(\omega) \in A\}^c = (X^{-1}(A))^c.$$

Aus $A \in \mathcal{A}$ folgt, dass $X^{-1}(A) \in \mathcal{F}$. Außerdem ist die Familie \mathcal{F} eine σ -Algebra und somit komplementstabil. Es folgt, dass $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathcal{F}$. Das bedeutet aber, dass $A^c \in \mathcal{A}$.

Bedingung 3. Schließlich zeigen wir, dass die Familie \mathcal{A} σ -vereinigungsstabil ist. Seien also $A_1, A_2, \dots \in \mathcal{A}$. Wir zeigen, dass $\cup_{n \in \mathbb{N}} A_n \in \mathcal{A}$. Es gilt

$$X^{-1}(\cup_{n \in \mathbb{N}} A_n) = \{\omega \in \Omega : X(\omega) \in \cup_{n \in \mathbb{N}} A_n\} = \cup_{n \in \mathbb{N}} \{\omega \in \Omega : X(\omega) \in A_n\} = \cup_{n \in \mathbb{N}} X^{-1}(A_n).$$

Aus $A_n \in \mathcal{A}$ folgt, dass $X^{-1}(A_n) \in \mathcal{F}$ für alle $n \in \mathbb{N}$. Außerdem ist die Familie \mathcal{F} eine σ -Algebra und somit σ -vereinigungsstabil. Es folgt, dass $X^{-1}(\cup_{n \in \mathbb{N}} A_n) = \cup_{n \in \mathbb{N}} X^{-1}(A_n) \in \mathcal{F}$. Das bedeutet, dass $\cup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Somit haben wir gezeigt, dass die Mengenfamilie \mathcal{A} eine σ -Algebra ist. □

BEWEIS VON SATZ 8.1.2. Betrachte die Mengenfamilie

$$\mathcal{E} = \{(-\infty, a], a \in \mathbb{R}\} \subset 2^{\mathbb{R}}.$$

Da $X : \Omega \rightarrow \mathbb{R}$ messbar ist, gilt $X^{-1}(B) = \{X \leq a\} \in \mathcal{F}$ für jedes $B = (-\infty, a] \in \mathcal{E}$. Proposition 8.1.4 besagt, dass $X^{-1}(B) \in \mathcal{F}$ für alle $B \in \sigma(\mathcal{E})$. Dabei ist aber $\sigma(\mathcal{E})$ nichts anderes als die Borel- σ -Algebra. □

BEISPIEL 8.1.6. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{F}$ ein messbares Ereignis. Wir zeigen, dass die Indikatorfunktion von A

$$X(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c \end{cases}$$

eine Zufallsvariable ist.

LÖSUNG. Sei $a \in \mathbb{R}$ beliebig. Wir betrachten das Ereignis

$$\{X \leq a\} = \begin{cases} \Omega, & a \geq 1, \\ \emptyset, & a < 0, \\ A^c, & a \in [0, 1). \end{cases}$$

Es gilt $\Omega, \emptyset, A^c \in \mathcal{F}$, denn $A \in \mathcal{F}$ und \mathcal{F} ist eine σ -Algebra. Somit ist X messbar.

8.2. Zufallsvektoren

DEFINITION 8.2.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien $X_1, \dots, X_d : \Omega \rightarrow \mathbb{R}$ Funktionen, wobei $d \in \mathbb{N}$. Betrachte nun die Funktion $X : \Omega \rightarrow \mathbb{R}^d$ mit

$$X(\omega) = (X_1(\omega), \dots, X_d(\omega)) \in \mathbb{R}^d.$$

Die Funktion X heißt ein d -dimensionaler *Zufallsvektor* (oder *messbar*), wenn X_1, \dots, X_d messbar sind.

SATZ 8.2.2. Sei $X : \Omega \rightarrow \mathbb{R}^d$ eine Funktion. Folgende Bedingungen sind äquivalent:

- (1) X ist ein Zufallsvektor.
- (2) Für jede Borel-Menge $B \subset \mathbb{R}^d$ gilt $\{X \in B\} \in \mathcal{F}$.

BEWEIS VON (1) \Rightarrow (2). Seien X_1, \dots, X_d messbar. Sei $A = (-\infty, a_1] \times \dots \times (-\infty, a_d]$ ein "Oktant". Dann gilt:

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} = \{\omega \in \Omega : X_1(\omega) \leq a_1, \dots, X_d(\omega) \leq a_d\} = \bigcap_{k=1}^d \{X_k \leq a_k\}.$$

Wegen der Messbarkeit von X_k gilt $\{X_k \leq a_k\} \in \mathcal{F}$ für alle $k = 1, \dots, d$. Da \mathcal{F} eine σ -Algebra ist, folgt, dass $X^{-1}(A) \in \mathcal{F}$. Wir haben gezeigt, dass das Urbild jedes Oktanten messbar ist. Die Familie der Oktanten erzeugt die Borel- σ -Algebra \mathcal{B}^d . Mit Proposition 8.1.4 folgt daraus, dass das Urbild jeder Borel-Menge messbar ist. \square

BEWEIS VON (2) \Rightarrow (1). Wir nehmen an, dass für jede Borel-Menge $B \subset \mathbb{R}^d$ gilt, dass $\{X \in B\} \in \mathcal{F}$. Sei $k \in \{1, \dots, d\}$ fest. Sei $B = \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_k \leq a\}$. Diese Menge ist Borel, da abgeschlossen. Es folgt, dass $X^{-1}(B) = \{X_k \leq a\} \in \mathcal{F}$. Somit ist die Funktion X_k messbar. Das gilt für jedes $k \in \{1, \dots, d\}$. Somit ist X messbar. \square

Die Familie der Borel-Mengen in \mathbb{R}^d wird mit \mathcal{B}^d bezeichnet.

DEFINITION 8.2.3. Eine Funktion $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ heißt *Borel-messbar* (oder *Borel-Funktion*), wenn gilt:

$$f^{-1}(A) \in \mathcal{B}^{d_1} \text{ für alle } A \in \mathcal{B}^{d_2}.$$

BEMERKUNG 8.2.4. Eine Funktion ist also Borel-messbar, wenn das Urbild jeder Borel-Menge wieder eine Borel-Menge ist. Zum Vergleich: Eine Funktion ist stetig, wenn das Urbild jeder offenen Menge offen ist.

PROPOSITION 8.2.5. *Jede stetige Funktion $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ ist Borel-messbar.*

BEWEIS. Die Funktion f sei stetig. Es folgt, dass für jede offene Menge $A \subset \mathbb{R}^{d_2}$ das Urbild $f^{-1}(A)$ offen ist. Das Urbild jeder offenen Menge ist also eine Borel-Menge. Die Familie der offenen Mengen erzeugt die Borel- σ -Algebra. Mit Proposition 8.1.4 folgt, dass auch das Urbild jeder Borel-Menge eine Borel-Menge ist. Somit ist f Borel-messbar. \square

SATZ 8.2.6. *Sei $X : \Omega \rightarrow \mathbb{R}^{d_1}$ ein Zufallsvektor und $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$ eine Borel-Funktion. Dann ist auch die Verknüpfung*

$$f \circ X : \Omega \rightarrow \mathbb{R}^{d_2}$$

ein Zufallsvektor.

BEWEIS. Sei $A \in \mathcal{B}^{d_2}$. Dann gilt $f^{-1}(A) \in \mathcal{B}^{d_1}$, denn f ist eine Borel-Funktion. Es gilt

$$(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A)) \in \mathcal{F},$$

da X messbar ist. Nach Satz 8.2.2 ist $f \circ X$ ein Zufallsvektor. \square

KOROLLAR 8.2.7. *Sind X, Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, so sind auch $X + Y$, $X \cdot Y$ und $a \cdot X$, wobei $a \in \mathbb{R}$, Zufallsvariablen.*

BEWEIS. Die Funktionen $(x, y) \mapsto x + y$, $(x, y) \mapsto xy$, $x \mapsto ax$ sind Borel-Funktionen, da sie stetig sind. Die Behauptung folgt aus Satz 8.2.6. \square

8.3. Verteilungsfunktion einer Zufallsvariable

Die Familie der Borel-Teilmengen von \mathbb{R} wird mit \mathcal{B} bezeichnet.

DEFINITION 8.3.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable.

(1) Die *Verteilung* von X ist die Funktion

$$P_X : \mathcal{B} \rightarrow [0, 1] \text{ mit } P_X(A) = \mathbb{P}[X \in A], \quad A \in \mathcal{B}.$$

P_X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$.

(2) Die *Verteilungsfunktion* von X ist die Funktion

$$F_X : \mathbb{R} \rightarrow [0, 1] \text{ mit } F_X(t) = \mathbb{P}[X \leq t], \quad t \in \mathbb{R}.$$

BEISPIEL 8.3.2. Im Einheitskreis werde ein Punkt (X, Y) zufällig und gleichverteilt gewählt. Es sei R der Abstand von (X, Y) zum Mittelpunkt des Kreises. Bestimme die Verteilungsfunktion von R .

LÖSUNG. Als Grundmenge wählen wir den Einheitskreis $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Sei $\mathcal{F} = \mathcal{B}_\Omega^2$ die σ -Algebra der Borel-Teilmengen von Ω . Als Wahrscheinlichkeitsmaß wählen wir

$$\mathbb{P}[A] = \frac{\lambda(A)}{\pi}, \quad A \in \mathcal{F},$$

wobei λ das Lebesgue-Maß ist. Das entspricht der Annahme, dass der Punkt gleichverteilt ist. Der Abstand zum Ursprung ist dann die Zufallsvariable $R : \Omega \rightarrow \mathbb{R}$ mit

$$R(x, y) = \sqrt{x^2 + y^2}, \quad (x, y) \in \Omega.$$

Beachte, dass R stetig und somit messbar ist. Um die Verteilungsfunktion von R zu bestimmen, schauen wir uns das Ereignis $\{R \leq t\}$ an:

$$\{R \leq t\} = \{(x, y) \in \Omega : \sqrt{x^2 + y^2} \leq t\} = \begin{cases} \Omega, & t \geq 1, \\ \emptyset, & t < 0, \\ \text{Kreis vom Radius } t, & t \in [0, 1]. \end{cases}$$

Es folgt, dass

$$F_R(t) = \mathbb{P}[R \leq t] = \begin{cases} 1, & t \geq 1, \\ 0, & t < 0, \\ \frac{\pi t^2}{\pi}, & t \in [0, 1] \end{cases} = \begin{cases} 1, & t \geq 1, \\ 0, & t < 0, \\ t^2, & t \in [0, 1]. \end{cases}$$

Dies ist die Verteilungsfunktion von R .

BEISPIEL 8.3.3. Ein Zufallsgenerator erzeugt zwei unabhängige und in $[0, 1]$ gleichverteilte Zufallszahlen X, Y . Bestimme die Verteilungsfunktion von $Z := X + Y$.

LÖSUNG. Als Grundmenge wählen wir $\Omega = [0, 1]^2$. Sei $\mathcal{F} = \mathcal{B}_\Omega^2$ die Einschränkung der Borel- σ -Algebra \mathcal{B}^2 auf Ω . Die Bedingung der Gleichverteilung und Unabhängigkeit von X und Y wird so interpretiert: die Wahrscheinlichkeit eines Ereignisses $A \in \mathcal{F}$ ist

$$\mathbb{P}[A] = \lambda(A).$$

Wir können die Zufallsvariablen $X, Y, Z : \Omega \rightarrow \mathbb{R}$ definieren:

$$X(x, y) = x, \quad Y(x, y) = y, \quad Z(x, y) = x + y, \quad (x, y) \in [0, 1]^2.$$

Das Ereignis, das uns hier interessiert, ist

$$\{Z \leq t\} = \{(x, y) \in [0, 1]^2 : x + y \leq t\}.$$

Es gilt

$$\{Z \leq t\} = \begin{cases} \emptyset, & t < 0, \\ \text{gleichschenkliges rechtwinkliges Dreieck mit Kathetenlänge } t, & t \in [0, 1], \\ \text{das Komplement eines solchen Dreiecks mit Kathetenlänge } 2 - t, & t \in [1, 2], \\ \Omega, & t \geq 2. \end{cases}$$

Somit erhalten wir

$$F_Z(t) = \mathbb{P}[Z \leq t] = \begin{cases} 0, & t < 0, \\ \frac{t^2}{2}, & t \in [0, 1], \\ 1 - \frac{(2-t)^2}{2}, & t \in [1, 2], \\ 1, & t \geq 2. \end{cases}$$

Dies ist die Verteilungsfunktion von Z . Sie ist stetig.

BEISPIEL 8.3.4. Betrachte eine konstante Zufallsvariable, also $X = c$. Wie sieht dann die Verteilungsfunktion F_X aus?

LÖSUNG. Es gilt

$$F_X(t) = \mathbb{P}[c \leq t] = \begin{cases} 0, & t < c, \\ 1, & t \geq c. \end{cases}$$

Dieses Beispiel zeigt, dass eine Verteilungsfunktion Unstetigkeitsstellen haben kann.

SATZ 8.3.5. Sei X eine Zufallsvariable mit Verteilungsfunktion F_X . Dann hat F_X die folgenden drei Eigenschaften:

- (1) Grenzwerte: $\lim_{t \rightarrow -\infty} F_X(t) = 0$ und $\lim_{t \rightarrow +\infty} F_X(t) = 1$.
- (2) Monotonie: Für alle $t_1 \leq t_2$ gilt $F_X(t_1) \leq F_X(t_2)$.
- (3) Rechtsstetigkeit: Für alle $t_0 \in \mathbb{R}$ gilt $\lim_{t \downarrow t_0} F_X(t) = F_X(t_0)$.

BEMERKUNG 8.3.6. Der linksseitige Grenzwert $\lim_{t \uparrow t_0} F_X(t)$ existiert ebenfalls, da die Funktion F_X monoton ist. Allerdings muss der linksseitige Grenzwert nicht mit $F_X(t_0)$ übereinstimmen, siehe Beispiel 8.3.4 mit $t_0 = c$.

BEWEIS VON (2). Seien $t_1 \leq t_2$ beliebig. Betrachte die Ereignisse

$$\{X \leq t_1\} = \{\omega \in \Omega : X(\omega) \leq t_1\} \subset \{\omega \in \Omega : X(\omega) \leq t_2\} = \{X \leq t_2\}$$

Deshalb gilt für die Wahrscheinlichkeiten dieser Ereignisse $\mathbb{P}[X \leq t_1] \leq \mathbb{P}[X \leq t_2]$, was gleichbedeutend ist mit $F_X(t_1) \leq F_X(t_2)$. \square

BEWEIS VON (1). Wir betrachten den Fall $t \rightarrow -\infty$. Führe die Ereignisse $A_n = \{X \leq -n\}$ mit $n \in \mathbb{N}$ ein. Dann gilt $A_1 \supset A_2 \supset \dots$ und $\bigcap_{n \in \mathbb{N}} A_n = \emptyset$. Aufgrund der Stetigkeit der Wahrscheinlichkeit folgt daraus, dass

$$\lim_{n \rightarrow \infty} F_X(-n) = \lim_{n \rightarrow \infty} \mathbb{P}[A_n] = \mathbb{P}[\emptyset] = 0.$$

Dabei darf allerdings n nur natürliche Werte annehmen. Für ein beliebiges $t < 0$ kann man immer ein $n \in \mathbb{N}$ mit $-n \leq t < -n + 1$ finden. Wegen der Monotonie von F und des Sandwichprinzips gilt dann

$$0 = \lim_{n \rightarrow \infty} F_X(-n) \leq \lim_{t \rightarrow -\infty} F_X(t) \leq \lim_{n \rightarrow \infty} F_X(-n + 1) = 0.$$

Somit ist $\lim_{t \rightarrow -\infty} F_X(t) = 0$, wie behauptet. \square

BEWEIS VON (3). Sei $t_0 \in \mathbb{R}$ beliebig. Wir definieren die Ereignisse $A_n = \{X \leq t_0 + 1/n\}$ mit $n \in \mathbb{N}$. Es gilt dann $A_1 \supset A_2 \supset \dots$ und $\bigcap_{n \in \mathbb{N}} A_n = \{X \leq t_0\}$. Aufgrund der Stetigkeit der Wahrscheinlichkeit gilt für den Grenzwert:

$$\lim_{n \rightarrow \infty} F_X\left(t_0 + \frac{1}{n}\right) = \lim_{n \rightarrow \infty} \mathbb{P}[A_n] = \mathbb{P}[X \leq t_0] = F_X(t_0).$$

Das gilt wieder nur für $n \in \mathbb{N}$. Für ein beliebiges $t > t_0$ können wir $n \in \mathbb{N}$ mit $t_0 + \frac{1}{n} \leq t < t_0 + \frac{1}{n-1}$ finden. Aufgrund der Monotonie von F_X und des Sandwichprinzips erhalten wir

$$F_X(t_0) = \lim_{n \rightarrow \infty} F_X\left(t_0 + \frac{1}{n}\right) \leq \lim_{t \downarrow t_0} F_X(t) \leq \lim_{n \rightarrow \infty} F_X\left(t_0 + \frac{1}{n-1}\right) = F_X(t_0).$$

Somit ist $\lim_{t \downarrow t_0} F_X(t) = F_X(t_0)$, wie behauptet. \square

Der nächste Satz besagt, dass die Verteilung einer Zufallsvariable durch ihre Verteilungsfunktion eindeutig festgelegt wird.

SATZ 8.3.7. Seien X_1 und X_2 Zufallsvariablen mit

$$F_{X_1}(t) = F_{X_2}(t) \text{ für alle } t \in \mathbb{R}.$$

Dann gilt für alle Borel-Mengen $B \subset \mathbb{R}$:

$$\mathbb{P}[X_1 \in B] = \mathbb{P}[X_2 \in B].$$

Für den Beweis benötigen wir einen Satz aus der Maßtheorie.

SATZ 8.3.8 (Eindeutigkeit der Maß-Fortsetzung). Sei Ω eine Menge und $\mathcal{E} \subset 2^\Omega$ eine schnittstabile Mengenfamilie. Die Schnittstabilität bedeutet: für $A, B \in \mathcal{E}$ gilt auch $A \cap B \in \mathcal{E}$. Es sei $\mathcal{A} = \sigma(\mathcal{E})$ die von \mathcal{E} erzeugte σ -Algebra. Seien \mathbb{P}_1 und \mathbb{P}_2 zwei Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) mit der Eigenschaft, dass

$$\mathbb{P}_1[A] = \mathbb{P}_2[A] \text{ für alle } A \in \mathcal{E}.$$

Dann gilt sogar

$$\mathbb{P}_1[A] = \mathbb{P}_2[A] \text{ für alle } A \in \mathcal{A}.$$

BEMERKUNG 8.3.9. Mit anderen Worten: stimmen zwei Wahrscheinlichkeitsmaße auf dem Erzeuger einer σ -Algebra überein, so stimmen sie auch auf der ganzen σ -Algebra überein, wenn der Erzeuger schnittstabil ist.

BEWEIS VON SATZ 8.3.7. Die Mengenfamilie $\mathcal{E} = \{(-\infty, t], t \in \mathbb{R}\}$ ist schnittstabil, denn

$$(-\infty, t] \cap (-\infty, s] = (-\infty, \min(t, s)] \in \mathcal{E}.$$

Für die Verteilungen von X_1 und X_2 gilt nun:

$$P_{X_1}((-\infty, t]) = \mathbb{P}[X_1 \leq t] = F_{X_1}(t) = F_{X_2}(t) = \mathbb{P}[X_2 \leq t] = P_{X_2}((-\infty, t]).$$

Die Wahrscheinlichkeitsmaße P_{X_1} und P_{X_2} stimmen also auf \mathcal{E} überein. Nach der Eindeutigkeit der Maß-Fortsetzung stimmen sie auch auf der Borel- σ -Algebra $\mathcal{B} = \sigma(\mathcal{E})$ überein.

Somit gilt $P_{X_1}(B) = P_{X_2}(B)$ für alle $B \in \mathcal{B}$. Mit anderen Worten, $\mathbb{P}[X_1 \in B] = \mathbb{P}[X_2 \in B]$ für alle $B \in \mathcal{B}$. \square

BEISPIEL 8.3.10. Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(t) = \mathbb{P}[X \leq t]$. Bestimme $\mathbb{P}[X < t]$.

LÖSUNG. Betrachte Ereignisse $A_n = \{X \leq t - \frac{1}{n}\}$ mit $n \in \mathbb{N}$. Es gilt $A_1 \subset A_2 \subset \dots$ und $\cup_{n \in \mathbb{N}} A_n = \{X < t\}$. Mit der Stetigkeit der Wahrscheinlichkeit folgt, dass

$$\mathbb{P}[X < t] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n] = \lim_{n \rightarrow \infty} \mathbb{P}\left[X \leq t - \frac{1}{n}\right] = \lim_{s \uparrow t} F_X(s).$$

Beachte: dieser Grenzwert muss im Allgemeinen nicht mit $F(t)$ übereinstimmen.

BEISPIEL 8.3.11. Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(t) = \mathbb{P}[X \leq t]$. Bestimme $\mathbb{P}[X = t]$.

LÖSUNG. Es gilt

$$\mathbb{P}[X = t] = \mathbb{P}[X \leq t] - \mathbb{P}[X < t] = F_X(t) - \lim_{s \uparrow t} F_X(s).$$

Die Wahrscheinlichkeit, dass $X = t$ ist, ist also gleich dem Sprung, den die Verteilungsfunktion F_X an der Stelle t macht. Die Funktion F_X ist stetig an der Stelle t genau dann wenn $\mathbb{P}[X = t] = 0$.

DEFINITION 8.3.12. Sei X eine Zufallsvariable. Ein Wert $t \in \mathbb{R}$ mit $\mathbb{P}[X = t] > 0$ heißt ein *Atom* von X . Atome sind also Unstetigkeitsstellen der Verteilungsfunktion F_X .

PROPOSITION 8.3.13. *Jede Zufallsvariable hat höchstens abzählbar viele Atome. Mit anderen Worten, jede Verteilungsfunktion hat höchstens abzählbar viele Unstetigkeitsstellen (Sprünge).*

BEWEIS. Es gilt: F_X ist monoton, $\lim_{t \rightarrow -\infty} F_X(t) = 0$ und $\lim_{t \rightarrow +\infty} F_X(t) = 1$. Für jedes $n \in \mathbb{N}$ kann es also höchstens n Sprünge geben, die eine Höhe von $\geq 1/n$ haben. Sonst wäre die Summe der Sprunghöhen > 1 , was ein Widerspruch ist. Die Menge der Sprünge ist eine abzählbare Vereinigung endlicher Mengen und somit selbst abzählbar. \square

BEISPIEL 8.3.14. Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(t) = \mathbb{P}[X \leq t]$. Für $a < b$ bestimme $\mathbb{P}[a \leq X \leq b]$, $\mathbb{P}[a \leq X < b]$, $\mathbb{P}[a < X \leq b]$, $\mathbb{P}[a < X < b]$.

LÖSUNG. Es gilt

$$\begin{aligned} \mathbb{P}[a \leq X \leq b] &= \mathbb{P}[X \leq b] - \mathbb{P}[X < a] = F_X(b) - \lim_{s \uparrow a} F_X(s), \\ \mathbb{P}[a \leq X < b] &= \mathbb{P}[X < b] - \mathbb{P}[X < a] = \lim_{s \uparrow b} F_X(s) - \lim_{s \uparrow a} F_X(s), \\ \mathbb{P}[a < X \leq b] &= \mathbb{P}[X \leq b] - \mathbb{P}[X \leq a] = F_X(b) - F_X(a), \\ \mathbb{P}[a < X < b] &= \mathbb{P}[X < b] - \mathbb{P}[X \leq a] = \lim_{s \uparrow b} F_X(s) - F_X(a). \end{aligned}$$

8.4. Definition und Eigenschaften des Erwartungswerts

Wir haben den Erwartungswert nur für Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum definiert. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Eine der Definitionen des Erwartungswerts war:

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}[\{\omega\}].$$

Nun definieren wir den Erwartungswert für beliebige Zufallsvariablen. Sei dazu $(\Omega, \mathcal{F}, \mathbb{P})$ ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable. Da der Wahrscheinlichkeitsraum Ω im allgemeinen nicht abzählbar ist, können wir die Summe $\sum_{\omega \in \Omega}$ nicht bilden. Wir werden die Summe durch das Integral (das sogenannte Lebesgue-Integral) ersetzen:

$$(8.4.1) \quad \mathbb{E}X = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

Nun definieren wir den Erwartungswert (bzw. das Lebesgue-Integral). Das soll in mehreren Schritten geschehen.

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable (d.h. eine messbare Funktion) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$.

SCHRITT 1. (Elementarfunktionen).

DEFINITION 8.4.1. Seien $A_1, \dots, A_n \in \mathcal{F}$ paarweise disjunkte Mengen mit $\Omega = A_1 \cup \dots \cup A_n$. Außerdem seien $y_1, \dots, y_n \in \mathbb{R}$. Eine Elementarfunktion ist eine Funktion der Form

$$X = \sum_{k=1}^n y_k \mathbb{1}_{A_k}.$$

Ist nun $X : \Omega \rightarrow \mathbb{R}$ eine Elementarfunktion, so definieren wir den Erwartungswert von X wie folgt:

$$\mathbb{E}X \stackrel{def}{=} \sum_{k=1}^n y_k \mathbb{P}[A_k].$$

SCHRITT 2. (Nicht-negative, messbare Funktionen)

Sei $X : \Omega \rightarrow [0, \infty)$ eine messbare, nichtnegative Funktion. Nun definieren wir

$$\mathbb{E}X \stackrel{def}{=} \sup\{\mathbb{E}Y \mid Y : \Omega \rightarrow [0, \infty), \text{ Elementarfunktion mit } 0 \leq Y(\omega) \leq X(\omega) \forall \omega \in \Omega\}.$$

Die Idee ist also, dass wir X von unten mit Elementarfunktionen approximieren. Der so definierte Erwartungswert $\mathbb{E}X$ nimmt nichtnegative Werte oder den Wert $+\infty$ an.

SCHRITT 3. (Beliebige, messbare Funktionen)

Sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige messbare Funktion. Wir werden X als eine Differenz von zwei nichtnegativen Funktionen X^+ und X^- darstellen. Definiere

$$X^+(\omega) \stackrel{def}{=} \begin{cases} X(\omega), & \text{falls } X(\omega) \geq 0, \\ 0, & \text{falls } X(\omega) < 0, \end{cases} \quad X^-(\omega) \stackrel{def}{=} \begin{cases} 0, & \text{falls } X(\omega) \geq 0, \\ |X(\omega)|, & \text{falls } X(\omega) < 0. \end{cases}$$

Dann sind X^+ und X^- messbar (Übung) und es gilt

$$X^+, X^- \geq 0, \quad X = X^+ - X^-, \quad |X| = X^+ + X^-.$$

Für die nichtnegativen Zufallsvariablen X^+ und X^- haben wir den Erwartungswert bereits in Schritt 2 definiert. Mit dieser Definition können wir nun folgende Fälle betrachten:

FALL 1: Gilt $\mathbb{E}X^+ < \infty$ und $\mathbb{E}X^- < \infty$, so heißt die Zufallsvariable X *integrierbar*. Bezeichnung: $X \in L^1$. Für eine integrierbare Zufallsvariable X definieren wir

$$\mathbb{E}X \stackrel{def}{=} \mathbb{E}X^+ - \mathbb{E}X^-.$$

In allen anderen Fällen heißt die Zufallsvariable X nicht integrierbar.

FALL 2: Gilt $\mathbb{E}X^+ = +\infty$ und $\mathbb{E}X^- < +\infty$, so definieren wir

$$\mathbb{E}X \stackrel{def}{=} +\infty.$$

FALL 3: Gilt $\mathbb{E}X^+ < +\infty$ und $\mathbb{E}X^- = +\infty$, so definieren wir

$$\mathbb{E}X \stackrel{def}{=} -\infty.$$

FALL 4: Gilt: $\mathbb{E}X^+ = \mathbb{E}X^- = +\infty$, so kann man den Erwartungswert nicht definieren.

Wir werden einige Eigenschaften des Erwartungswerts (bzw. des Lebesgue-Integrals) ohne Beweis auflisten.

SATZ 8.4.2. *Für eine beliebige Zufallsvariable X gilt $\mathbb{E}|X| = \mathbb{E}X^+ + \mathbb{E}X^-$. Insbesondere ist X genau dann integrierbar, wenn $\mathbb{E}|X| < \infty$. Für eine integrierbare Zufallsvariable X gilt die Ungleichung*

$$|\mathbb{E}X| \leq \mathbb{E}|X|.$$

SATZ 8.4.3. *Der Erwartungswert ist linear:*

- (1) *Sind X und Y integrierbare Zufallsvariablen, so ist auch $X + Y$ integrierbar und es gilt*

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

- (2) *Ist X integrierbar und ist $a \in \mathbb{R}$, so ist auch aX integrierbar und es gilt*

$$\mathbb{E}[aX] = a\mathbb{E}[X].$$

In dem Fall, wenn die Zufallsvariable höchstens abzählbar viele Werte annimmt, stimmt die neue Definition des Erwartungswerts mit der alten Definition überein.

SATZ 8.4.4. *Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, die höchstens abzählbar viele Werte y_1, y_2, \dots annimmt. Dann ist X integrierbar genau dann, wenn*

$$\mathbb{E}|X| = \sum_n |y_n| \cdot \mathbb{P}[X = y_n] < \infty.$$

Ist X integrierbar, so gilt

$$\mathbb{E}X = \sum_n y_n \cdot \mathbb{P}[X = y_n].$$

Der Erwartungswert ist monoton:

SATZ 8.4.5. Sei $Y : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}Y < \infty$ und X eine weitere Zufallsvariable mit $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$. Dann gilt $\mathbb{E}X \leq \mathbb{E}Y$.

Der Erwartungswert einer Zufallsvariable ändert sich nicht, wenn man die Werte der Zufallsvariable auf einer Nullmenge verändert. Dies wird im nächsten Satz beschrieben.

DEFINITION 8.4.6. Zwei Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{R}$ heißen *fast überall gleich*, wenn

$$\mathbb{P}[\{\omega \in \Omega : X(\omega) \neq Y(\omega)\}] = 0.$$

SATZ 8.4.7. Sind X und Y fast überall gleich und eine der Zufallsvariablen integrierbar, so ist auch die andere Zufallsvariable integrierbar und es gilt $\mathbb{E}X = \mathbb{E}Y$.

8.5. Diskrete und absolut stetige Verteilungen

DEFINITION 8.5.1. Eine Zufallsvariable X heißt *diskret*, wenn X nur endlich oder abzählbar viele Werte annimmt. Die *Zähldichte* von X ist die Funktion

$$p_X(y) = \mathbb{P}[X = y].$$

BEMERKUNG 8.5.2. Für die Verteilungsfunktion einer diskreten Zufallsvariable gilt

$$F_X(t) = \sum_{y \leq t} p_X(y) \text{ für alle } t \in \mathbb{R}.$$

Die Verteilungsfunktion einer diskreten Zufallsvariable ist eine ‘‘Sprungfunktion’’.

DEFINITION 8.5.3. Eine Zufallsvariable X mit Verteilungsfunktion F_X heißt *absolut stetig*, wenn es eine Borel-Funktion $f_X : \mathbb{R} \rightarrow [0, \infty)$ gibt, so dass

$$F_X(t) = \int_{-\infty}^t f_X(y) dy \text{ für alle } t \in \mathbb{R}.$$

Die Funktion f_X heißt die *Dichte von X* .

BEMERKUNG 8.5.4. Für die Dichte gelten die folgenden zwei Eigenschaften:

- (1) $f_X(y) \geq 0$ für alle $y \in \mathbb{R}$.
- (2) $\int_{\mathbb{R}} f_X(y) dy = 1$.

SATZ 8.5.5. Sei X eine absolut stetige Zufallsvariable. Dann gilt für jede Borel-Menge $B \subset \mathbb{R}$

$$\mathbb{P}[X \in B] = \int_B f_X(y) dy.$$

BEMERKUNG 8.5.6. Zum Vergleich: Im diskreten Fall gilt

$$\mathbb{P}[X \in B] = \sum_{y \in B} p_X(y).$$

BEWEIS. Definiere zwei Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$:

$$\mathbb{P}_1(B) = \mathbb{P}[X \in B], \quad \mathbb{P}_2(B) = \int_B f_X(y) dy, \quad B \in \mathcal{B}.$$

Diese Wahrscheinlichkeitsmaße stimmen auf allen Mengen der Form $B = (-\infty, t]$ überein, denn

$$\mathbb{P}_2(B) = \int_{-\infty}^t f_X(y) dy = F_X(t) = \mathbb{P}[X \leq t] = \mathbb{P}_1(B).$$

Die Mengen der Form $(-\infty, t]$ bilden einen schnittstabilen Erzeuger der Borel- σ -Algebra. Durch die Eindeutigkeit der Maßfortsetzung folgt, dass

$$\mathbb{P}_1(B) = \mathbb{P}_2(B) \text{ für alle } B \in \mathcal{B}.$$

Somit gilt $\mathbb{P}[X \in B] = \int_B f_X(y)dy$ für alle $B \in \mathcal{B}$. □

Die Verteilungsfunktion ist das Integral der Dichte. Umgekehrt, ist die Dichte die Ableitung der Verteilungsfunktion. Das gilt allerdings nicht an allen, sondern an fast allen Stellen.

SATZ 8.5.7. *Sei X eine absolut stetige Zufallsvariable mit Dichte f_X und Verteilungsfunktion F_X . Dann gibt es eine Borel-Menge $N \subset \mathbb{R}$ mit Lebesgue-Maß 0, so dass die Funktion F_X differenzierbar an allen Stellen $t \in \mathbb{R} \setminus N$ ist und*

$$F'_X(t) = f_X(t) \text{ für alle } t \in \mathbb{R} \setminus N.$$

OHNE BEWEIS.

Die wichtigsten Eigenschaften der diskreten und der absolut stetigen Verteilungen werden in der folgenden Tabelle zusammengefasst. Einige dieser Formeln werden wir später beweisen.

Diskrete Zufallsvariablen	Absolut stetige Zufallsvariablen
Zähldichte p_X	Dichte f_X
Verteilungsfunktion: $F_X(t) = \sum_{y \leq t} p_X(y)$	Verteilungsfunktion: $F_X(t) = \int_{-\infty}^t f_X(y)dy$
$p_X(y) \in [0, 1]$ für alle $y \in \mathbb{R}$	$f_X(y) \geq 0$ für alle $y \in \mathbb{R}$
$\sum_{y \in \mathbb{R}} p_X(y) = 1$	$\int_{-\infty}^{\infty} f_X(y)dy = 1$
$p_X(y) = \mathbb{P}[X = y], y \in \mathbb{R}$	$f_X(y) = F'_X(y) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}[y \leq X \leq y + \varepsilon]}{\varepsilon}$ f. ü.
$\mathbb{P}[X \in B] = \sum_{y \in B} p_X(y), B \subset \mathbb{R}$ Borel	$\mathbb{P}[X \in B] = \int_B f_X(y)dy, B \subset \mathbb{R}$ Borel
X integrierbar, wenn $\sum_{y \in \mathbb{R}} y \cdot p_X(y) < \infty$	X integrierbar, wenn $\int_{\mathbb{R}} y \cdot f_X(y)dy < \infty$
Erwartungswert: $\mathbb{E}X = \sum_{y \in \mathbb{R}} y \cdot p_X(y)$	Erwartungswert: $\mathbb{E}X = \int_{\mathbb{R}} y \cdot f_X(y)dy$

8.6. Beispiele von absolut stetigen Verteilungen

Gleichverteilung auf einem Intervall

DEFINITION 8.6.1. Eine Zufallsvariable X heißt *gleichverteilt* auf einem Intervall $[a, b]$, wobei $a, b \in \mathbb{R}$ und $a < b$, wenn X absolut stetig ist und die Dichte von X durch die folgende Formel gegeben ist:

$$f_X(y) = \begin{cases} \frac{1}{b-a}, & y \in [a, b], \\ 0, & y \notin [a, b]. \end{cases}$$

BEZEICHNUNG: $X \sim U[a, b]$. Dabei steht “U” für “uniform”.

BEMERKUNG 8.6.2. Die Dichte ist also konstant auf dem Intervall $[a, b]$. Die Werte außerhalb von $[a, b]$ werden nicht angenommen, da die Dichte außerhalb von $[a, b]$ verschwindet. Die Konstante $\frac{1}{b-a}$ wurde so gewählt, dass die Bedingung $\int_{\mathbb{R}} f_X(y)dy = 1$ erfüllt ist.

BEMERKUNG 8.6.3. Die Verteilungsfunktion einer gleichverteilten Zufallsvariable $X \sim U[a, b]$ ist

$$F_X(t) = \int_{-\infty}^t f_X(y) dy = \begin{cases} 0, & t \leq a, \\ \frac{t-a}{b-a}, & t \in [a, b], \\ 1, & t \geq b. \end{cases}$$

Die Verteilungsfunktion ist also linear auf dem Intervall $[a, b]$ und konstant sonst. Die Verteilungsfunktion ist stetig. Leitet man die Verteilungsfunktion ab, so erhält man die Dichte. Es gibt allerdings zwei Ausnahmestellen 0 und 1, an denen die Verteilungsfunktion nicht differenzierbar ist.

SATZ 8.6.4. Sei X eine Zufallsvariable, die auf $[a, b]$ gleichverteilt ist. Dann gilt

$$\mathbb{E}X = \frac{a+b}{2}.$$

BEWEIS.

$$\mathbb{E}X = \int_{\mathbb{R}} y f_X(y) dy = \int_a^b \frac{y}{b-a} dy = \frac{1}{b-a} \int_a^b y dy = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.$$

□

Exponentialverteilung

DEFINITION 8.6.5. Eine Zufallsvariable X heißt *exponentialverteilt* mit Parameter $\lambda > 0$, wenn X absolut stetig ist und die Dichte von X durch die folgende Formel gegeben ist:

$$f_X(y) = \begin{cases} \lambda e^{-\lambda y}, & y > 0, \\ 0, & y < 0. \end{cases}$$

BEZEICHNUNG: $X \sim \text{Exp}(\lambda)$.

BEMERKUNG 8.6.6. Das Integral der Dichte ist gleich 1, denn $\int_0^{\infty} \lambda e^{-\lambda y} dy = 1$. Eine Exponentialverteilte Zufallsvariable kann nur positive Werte annehmen: $\mathbb{P}[X > 0] = 1$.

BEMERKUNG 8.6.7. Die Verteilungsfunktion einer Zufallsvariable $X \sim \text{Exp}(\lambda)$ ist gegeben durch

$$F_X(t) = \int_{-\infty}^t f_X(y) dy = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-\lambda t}, & t \geq 0. \end{cases}$$

BEWEIS. Für $t < 0$ gilt $F_X(t) = \int_{-\infty}^t f_X(y) dy = \int_{-\infty}^t 0 = 0$, denn die Dichte f_X verschwindet auf der negativen Halbachse. Für $t \geq 0$ gilt

$$F_X(t) = \int_{-\infty}^t f_X(y) dy = \int_0^t \lambda e^{-\lambda y} dy = -e^{-\lambda y} \Big|_0^t = 1 - e^{-\lambda t}.$$

BEMERKUNG 8.6.8. Alternativ kann man die Exponentialverteilung durch die folgende Formel definieren:

$$\mathbb{P}[X > t] = e^{-\lambda t}, \quad t \geq 0.$$

SATZ 8.6.9. Für eine $\text{Exp}(\lambda)$ -verteilte Zufallsvariable X gilt

$$\mathbb{E}X = \frac{1}{\lambda}.$$

BEWEIS.

$$\mathbb{E}X = \int_0^\infty y f_X(y) dy = \int_0^\infty \lambda y e^{-\lambda y} dy = \int_0^\infty y (-e^{-\lambda y})' dy.$$

Nun benutzen wir die partielle Integration:

$$\mathbb{E}X = \int_0^\infty y (-e^{-\lambda y})' dy = -ye^{-\lambda y} \Big|_0^\infty + \int_0^\infty e^{-\lambda y} dy = \frac{1}{\lambda}.$$

□

SATZ 8.6.10 (Vergessenseigenschaft der Exponentialverteilung). Sei X eine Zufallsvariable. Die folgenden zwei Eigenschaften sind äquivalent:

- (1) X ist exponentialverteilt mit einem Parameter $\lambda > 0$.
- (2) Für alle $s, t > 0$ gilt:

$$\mathbb{P}[X > t + s | X > t] = \mathbb{P}[X > s].$$

BEWEIS VON (1) \Rightarrow (2). Sei $X \sim \text{Exp}(\lambda)$. Dann gilt für alle $t, s > 0$:

$$\mathbb{P}[X > t + s | X > t] = \frac{\mathbb{P}[X > t + s | X > t]}{\mathbb{P}[X > t]} = \frac{\mathbb{P}[X > t + s]}{\mathbb{P}[X > t]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = \mathbb{P}[X > s].$$

□

Für den Beweis der Rückrichtung benötigen wir ein Lemma.

LEMMA 8.6.11 (Cauchy-Funktionalgleichung). Sei $g : [0, \infty) \rightarrow \mathbb{R}$ eine monotone Funktion mit $g(t + s) = g(t) + g(s)$ für alle $t, s \geq 0$. Dann ist g linear, d.h. es gibt ein $\lambda \in \mathbb{R}$ mit $g(t) = \lambda t$ für alle $t \geq 0$.

BEWEIS. Sei $a \in \mathbb{R}$. Dann gilt

$$g(2a) = g(a) + g(a) = 2g(a).$$

Analog ergibt sich

$$g(3a) = g(2a) + g(a) = 2g(a) + g(a) = 3g(a).$$

Induktiv erhalten wir für jedes $n \in \mathbb{N}$

$$g(na) = ng(a).$$

Nun sei $a = 1/n$. Es ergibt sich $g(1) = ng(\frac{1}{n})$ und somit

$$g\left(\frac{1}{n}\right) = \frac{g(1)}{n} \stackrel{\text{def}}{=} \frac{\lambda}{n},$$

wobei wir $\lambda := g(1)$ gesetzt haben. Sei $m \in \mathbb{N}$, dann gilt

$$g\left(\frac{m}{n}\right) = g\left(m \cdot \frac{1}{n}\right) = m \cdot g\left(\frac{1}{n}\right) = \frac{m}{n} \cdot \lambda.$$

Wir haben somit gezeigt, dass $g(r) = \lambda r$ für alle $r \in \mathbb{Q}$, $r \geq 0$ gilt. Nun müssen wir zeigen, dass das auch für irrationale Werte gilt. Sei $t \in \mathbb{R} \setminus \mathbb{Q}$, $t > 0$ beliebig. Es existieren Folgen $\{r_i\} \subset \mathbb{Q}$ und $\{s_i\} \subset \mathbb{Q}$ mit den folgenden Eigenschaften:

- (1) r_i ist monoton fallend mit $r_i \downarrow t$ für $i \rightarrow \infty$.
- (2) s_i ist monoton steigend mit $s_i \uparrow t$ für $i \rightarrow \infty$.

Die Funktion g ist nach Voraussetzung monoton. Sei g zum Beispiel monoton fallend. (Für g monoton steigend ist der Beweis analog). Dann gilt

- (1) $g(r_i)$ ist monoton steigend mit $g(t) \geq g(r_i)$ für alle $i \in \mathbb{N}$.
- (2) $g(s_i)$ ist monoton fallend mit $g(t) \leq g(s_i)$ für alle $i \in \mathbb{N}$.

Da die Zahlen r_i und s_i rational sind, gilt $g(r_i) = \lambda r_i$ und $g(s_i) = \lambda s_i$. Somit erhalten wir $\lambda r_i \leq t \leq \lambda s_i$. Nun lassen wir $i \rightarrow \infty$ und benutzen den Sandwich-Satz: $g(t) = \lambda t$. \square

BEWEIS VON (2) \Rightarrow (1) IN SATZ 8.6.10. Sei F_X die Verteilungsfunktion von X . Betrachte die Funktion

$$g(t) \stackrel{\text{def}}{=} \log(1 - F_X(t)), \quad t \geq 0.$$

Diese Funktion ist monoton fallend, da die Verteilungsfunktion F_X monoton steigend ist. Es gilt

$$\mathbb{P}[X > t] = 1 - F_X(t) = e^{g(t)}, \quad t \geq 0.$$

Mit der Vergessenseigenschaft erhalten wir

$$\frac{e^{g(t+s)}}{e^{g(t)}} = \mathbb{P}[X > t + s | X > t] = \mathbb{P}[X > s] = e^{g(s)}.$$

Somit erfüllt g die Cauchy-Funktionalgleichung $g(t+s) = g(t) + g(s)$ und ist monoton fallend. Es gibt nach Lemma 8.6.11 ein λ mit $g(t) = -\lambda t$ für alle $t \geq 0$. Dabei muss λ positiv sein, da g monoton fallend ist. Es folgt, dass

$$F_X(t) = 1 - e^{-\lambda t}, \quad t \geq 0$$

Damit ist gezeigt, dass $X \sim \text{Exp}(\lambda)$. \square

BEISPIEL 8.6.12 (Radioaktiver Zerfall). Die Lebensdauer eines Atoms bis zum radioaktiven Zerfall sei eine Zufallsvariable $X \sim \text{Exp}(\lambda)$. Die Halbwertszeit m lässt sich bestimmen durch

$$\mathbb{P}[X > m] = \frac{1}{2}.$$

Löst man diese Gleichung auf, so erhält man

$$e^{-\lambda m} = \frac{1}{2} \Rightarrow \lambda m = \log(2) \Rightarrow m = \frac{\log(2)}{\lambda}.$$

Warum benutzt man Exponentialverteilung als die Verteilung der Lebensdauer eines Atoms? Weil man davon ausgeht, dass für die Lebensdauer die Vergessenseigenschaft gilt. Und daraus folgt, dass die Lebensdauer exponentialverteilt sein muss.

BEMERKUNG 8.6.13. Sei X eine beliebige Zufallsvariable mit Dichte f_X . Dann ist die Zerfallrate (oder die Ausfallrate) definiert durch

$$r_X(t) = \frac{f_X(t)}{1 - F_X(t)}.$$

Bei einem Atom geht man davon aus, dass die Zerfallrate konstant (d.h. unabhängig von t) ist. Die Exponentialverteilung erfüllt diese Eigenschaft: Für $X \sim \text{Exp}(\lambda)$ gilt

$$r_X(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Normalverteilung (oder Gauß-Verteilung)

DEFINITION 8.6.14. Eine Zufallsvariable X heißt *standardnormalverteilt*, wenn X absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad y \in \mathbb{R}.$$

BEZEICHNUNG: $X \sim N(0, 1)$.

Die Funktion f_X ist trivialerweise nicht negativ. Aber warum ist das Integral dieser Funktion gleich 1? Das beantwortet der folgende Satz.

SATZ 8.6.15. Es gilt $I := \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$.

BEWEIS. Die Stammfunktion von $e^{-y^2/2}$ lässt sich nicht mit Hilfe der 4 Grundrechenarten und Potenzen als eine endliche Kombination von y , e^y , $\log y$, $\sin y$, $\cos y$, usw. darstellen. Man muss also einen Trick anwenden. Um das Integral I zu berechnen, betrachten wir I^2 :

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \int_{\mathbb{R}^2} e^{-\left(\frac{x^2}{2} + \frac{y^2}{2}\right)} dx dy.$$

Und nun gehen wir zu Polarkoordinaten über:

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\varphi = 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr = 2\pi \left(-e^{-\frac{r^2}{2}} \Big|_0^{\infty} \right) = 2\pi.$$

Da $I \geq 0$, können wir die Wurzel ziehen: $I = \sqrt{2\pi}$. □

BEMERKUNG 8.6.16. Die Verteilungsfunktion der Standardnormalverteilung ist

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{y^2}{2}} dy.$$

Ihre Werte können numerisch berechnet werden.

BEMERKUNG 8.6.17. Aus $X \sim N(0, 1)$ folgt für den Erwartungswert $\mathbb{E}X = 0$.

BEWEIS. Erstens, ist X integrierbar, denn $\int_{\mathbb{R}} |y| e^{-\frac{y^2}{2}} dy < \infty$. Für den Erwartungswert erhalten wir dann

$$\mathbb{E}X = \int_{\mathbb{R}} y e^{-\frac{y^2}{2}} dy = 0,$$

denn es wird eine ungerade Funktion integriert. □

DEFINITION 8.6.18. Eine Zufallsvariable X heißt *normalverteilt* mit Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, wenn X absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad y \in \mathbb{R}.$$

BEZEICHNUNG: $X \sim N(\mu, \sigma^2)$.

BEMERKUNG 8.6.19. Ist $Y \sim N(0, 1)$ standardnormalverteilt, so ist $\mu + \sigma Y$ normalverteilt mit Parametern μ und σ^2 (Übung). Daraus folgt, dass der Erwartungswert einer $N(\mu, \sigma^2)$ -verteilten Zufallsvariable gleich μ ist. Später werden wir zeigen, dass der zweite Parameter σ^2 mit der Varianz der Zufallsvariable übereinstimmt.

Cauchy-Verteilung

DEFINITION 8.6.20. Eine Zufallsvariable X heißt *Cauchy-verteilt*, wenn X absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\pi} \frac{1}{1+y^2}, \quad y \in \mathbb{R}.$$

BEZEICHNUNG: $X \sim \text{Cauchy}$.

BEMERKUNG 8.6.21. Die Verteilungsfunktion der Cauchy-Verteilung ist gegeben durch

$$F_X(t) = \int_{-\infty}^t f_X(y) dy = \frac{1}{\pi} \int_{-\infty}^t \frac{1}{1+y^2} dy = \frac{1}{\pi} \arctan y \Big|_{-\infty}^t = \frac{1}{2} + \frac{1}{\pi} \arctan t,$$

da $\arctan(-\infty) = -\frac{\pi}{2}$. Insbesondere gilt

$$\int_{-\infty}^{\infty} f_X(y) dy = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{1+y^2} dy = \frac{1}{2} + \frac{1}{\pi} \arctan(+\infty) = 1,$$

so dass f_X tatsächlich eine Dichte ist.

BEISPIEL 8.6.22. Sei $\varphi \sim U[-\pi/2, \pi/2]$ gleichverteilt auf dem Intervall $[-\pi/2, \pi/2]$ und sei $X = \tan(\varphi)$. Nun behaupten wir, dass $X \sim \text{Cauchy}$.

BEWEIS. Wir berechnen die Verteilungsfunktion von X :

$$\mathbb{P}[X \leq t] = \mathbb{P}[\varphi \leq \arctan(t)] = \frac{\arctan t - (-\frac{\pi}{2})}{\pi} = \frac{1}{2} + \frac{1}{\pi} \arctan t.$$

BEMERKUNG 8.6.23. Sei $X \sim \text{Cauchy}$. Die Dichte von X ist eine gerade Funktion und man könnte meinen, dass der Erwartungswert von X aus diesem Grund gleich 0 ist. Dies ist jedoch nicht der Fall. Wir behaupten, dass X nicht integrierbar ist. Wir berechnen $\mathbb{E}X^+$:

$$\mathbb{E}X^+ = \int_0^{\infty} y f_X(y) dy = \frac{2}{\pi} \int_0^{\infty} \frac{y}{1+y^2} dy = \frac{1}{\pi} \log(1+y^2) \Big|_0^{\infty} = +\infty.$$

Analog ist $\mathbb{E}X^- = +\infty$. Wir haben eine Unbestimmtheit $\mathbb{E}X = (+\infty) - (+\infty)$ und die Zufallsvariable X ist somit nicht integrierbar. Dabei ist der Erwartungswert von X weder $+\infty$ noch $-\infty$, sondern er ist überhaupt nicht definiert.

8.7. Singuläre Verteilungen

Die Verteilung einer diskreten Zufallsvariable besteht ausschließlich aus Atomen. Sei nun X eine Zufallsvariable, die keine Atome hat, d.h. es gelte $\mathbb{P}[X = y] = 0$ für alle $y \in \mathbb{R}$. Eine solche Zufallsvariable ist nicht diskret. Ist sie dann absolut stetig? Es stellt sich heraus, dass die Antwort im Allgemeinen “nein” ist. Es gibt nämlich eine dritte Klasse von Verteilungen, die sogenannten singulären Verteilungen. Außerdem kann man Mischungen aus Verteilungen von diesen drei Typen (diskret, absolut stetig, singulär) bilden. Nach dem Zerlegungssatz von Lebesgue kann jede Verteilung als eine solche Mischung dargestellt werden.

DEFINITION 8.7.1. Eine Zufallsvariable X heißt *singulär*, wenn die folgenden zwei Bedingungen gelten:

- (1) X hat keine Atome, d.h. es gilt $\mathbb{P}[X = y] = 0$ für alle $y \in \mathbb{R}$.
- (2) Es gibt eine Borel-Menge $N \subset \mathbb{R}$ mit Lebesgue-Maß $\lambda(N) = 0$, so dass

$$\mathbb{P}[X \in N] = 1.$$

Eine singuläre Zufallsvariable nimmt also ausschließlich Werte in einer Nullmenge N an.

BEISPIEL 8.7.2 (Cantor-Menge). Jede Zahl $x \in [0, 1]$ kann man in einer Darstellung zur Basis 3 (d.h. im Ternärsystem) schreiben:

$$x = [0.\varepsilon_1\varepsilon_2\dots]_3 = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{3^k}.$$

Dabei sind $\varepsilon_k \in \{0, 1, 2\}$ die “Ziffern” von x . Die Cantor-Menge C ist die Menge aller Punkte $x \in [0, 1]$, dessen Ternärdarstellung ausschließlich aus den Ziffern 0 und 2 besteht (so dass die Ziffer 1 also nicht vorkommt):

$$C = \left\{ x \in [0, 1] : x = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{3^k}, \varepsilon_k \in \{0, 2\} \right\}.$$

Bei manchen Zahlen (nämlich, bei den Zahlen der Form $\frac{k}{3^m}$) ist die Ternärdarstellung nicht eindeutig, z.B.

$$1 = [1.000\dots]_3 = [0.222\dots]_3 \quad \text{und} \quad \frac{1}{3} = [0.1000\dots]_3 = [0.0222\dots]_3.$$

Solche Zahlen werden wir dann in die Cantor-Menge aufnehmen, wenn mindestens eine Darstellung nur aus den Ziffern 0 und 2 besteht. Man kann zeigen, dass die Cantor-Menge abgeschlossen ist. Wir werden nun zeigen, dass das Lebesgue-Maß von C gleich 0 ist. Betrachte die Menge C_n , die aus allen Zahlen in $[0, 1]$ besteht, die keine einzige 1 unter den ersten n Ziffern in ihrer Ternärdarstellung haben. Für die ersten n Ziffern gibt es also 2^n Möglichkeiten. Die Menge aller Zahlen in $[0, 1]$, bei denen die ersten n Ziffern festgelegt sind, ist ein Intervall der Länge $1/3^n$. Somit ist das Lebesgue-Maß von C_n gleich

$$\lambda(C_n) = \left(\frac{2}{3}\right)^n.$$

Es ist aber klar, dass $C \subset C_n$, und zwar für jedes n . Somit gilt

$$\lambda(C) \leq \left(\frac{2}{3}\right)^n \quad \text{für alle } n \in \mathbb{N}.$$

Das kann aber nur dann gelten, wenn $\lambda(C) = 0$. Das Lebesgue-Maß der Cantor-Menge ist also 0. Dabei ist die Cantor-Menge überabzählbar!

BEISPIEL 8.7.3 (Cantor-Verteilung). Nun konstruieren wir eine singuläre Verteilung. Es handelt sich dabei um eine Art Gleichverteilung auf der Cantor-Menge. Seien dazu $\varepsilon_1, \varepsilon_2, \dots$ unabhängige Zufallsvariablen mit

$$\mathbb{P}[\varepsilon_k = 0] = \mathbb{P}[\varepsilon_k = 2] = 1/2.$$

Diese interpretieren wir als Ziffern in einer Ternärdarstellung einer zufälligen Zahl. D.h., wir betrachten die Zufallsvariable

$$X = [0.\varepsilon_1\varepsilon_2\dots]_3 = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{3^k}.$$

Nach Definition gilt $\mathbb{P}[X \in C] = 1$. Dabei ist die Cantor-Menge C eine Nullmenge. Wir zeigen noch, dass X keine Atome hat. Sei $y = [0.\eta_1\eta_2\dots]_3 \in [0, 1]$. Dann gilt für jedes $n \in \mathbb{N}$

$$\mathbb{P}[X = y] \leq \mathbb{P}[\varepsilon_1 = \eta_1, \dots, \varepsilon_n = \eta_n] \leq \frac{1}{3^n}.$$

Daraus folgt, dass $\mathbb{P}[X = y] = 0$. Somit ist X singulär.

BEMERKUNG 8.7.4. Die Verteilungsfunktion der Cantor-Verteilung hat viele interessante Eigenschaften. Sie ist stetig, da die Cantor-Verteilung keine Atome hat. Außerdem ist sie konstant auf jedem Intervall, das komplett außerhalb der Cantor-Menge liegt. Somit ist die Ableitung der Verteilungsfunktion gleich 0 außerhalb der Cantor-Menge, also fast überall. Dennoch ist die Verteilungsfunktion nicht konstant, denn sie ist gleich 0 für $t = 0$ und gleich 1 für $t = 1$. Für diese Verteilungsfunktion gilt der Satz von Newton–Leibniz nicht, denn

$$1 = F_X(1) - F_X(0) \neq \int_0^1 F'_X(t) dt = 0.$$

Dabei gilt die letzte Gleichheit, weil F'_X fast überall gleich 0 ist.

Nun zeigen wir, dass die drei Klassen von Verteilungen (diskret, absolut stetig und singulär) sich nicht überschneiden. Diskrete und absolut stetige Verteilungen überschneiden sich nicht, denn die einen haben eine unstetige und die anderen eine stetige Verteilungsfunktion. Sei nun X eine singuläre Zufallsvariable. Da X keine Atome hat, kann X nicht diskret sein. Es bleibt zu zeigen, dass X nicht absolut stetig sein kann. Das wird im folgenden Satz gemacht.

SATZ 8.7.5. *Sei X eine Zufallsvariable und $N \subset \mathbb{R}$ eine Borel-Menge mit $\lambda(N) = 0$ und $\mathbb{P}[X \in N] = 1$. Dann ist X nicht absolut stetig.*

BEWEIS. Durch Widerspruch. Sei X absolut stetig und f_X die Dichte von X . Dann gilt

$$0 = \mathbb{P}[X \notin N] = \int_{\mathbb{R} \setminus N} f_X(y) dy = \int_{\mathbb{R}} f_X(y) \cdot \mathbb{1}_{\mathbb{R} \setminus N}(y) dy.$$

Die Dichte f_X ist nichtnegativ. Das Lebesgue-Integral einer nichtnegativen Funktion kann nur dann 0 sein, wenn die Funktion fast überall 0 ist. Das heißt, es muss eine Nullmenge M geben mit

$$f_X(y) \mathbb{1}_{\mathbb{R} \setminus N}(y) = 0 \text{ für alle } y \notin M.$$

Dann gilt aber $f_X(y) = 0$ für alle $y \notin N \cup M$. Dabei ist $N \cup M$ eine Nullmenge, also ist $f_X(y) = 0$ fast überall. Daraus folgt aber, dass $\int_{\mathbb{R}} f_X(y) dy = 0$. Somit ist f_X keine Dichte, denn für eine Dichte müsste dieses Integral 1 sein. Widerspruch. \square

8.8. Zerlegungssatz von Lebesgue

Der Zerlegungssatz von Lebesgue besagt, dass man jede Verteilung als eine Mischung aus einer diskreten, einer absolut stetigen und einer singulären Verteilung darstellen kann.

SATZ 8.8.1 (Lebesgue). *Sei F eine Verteilungsfunktion. Dann existieren drei Verteilungsfunktionen F_1, F_2, F_3 und drei Zahlen $p_1, p_2, p_3 \geq 0$ mit $p_1 + p_2 + p_3 = 1$, sodass*

$$F(t) = p_1 F_1(t) + p_2 F_2(t) + p_3 F_3(t) \text{ für alle } t \in \mathbb{R}$$

und dabei F_1 diskret ist, F_2 absolut stetig und F_3 singulär.

OHNE BEWEIS.

BEISPIEL 8.8.2 (Gemischte Verteilung). Man betrachte eine Bahnschranke, die für 10 Minuten offen steht, dann für 10 Minuten geschlossen, dann wieder für 10 Minuten offen, usw. Ein Fußgänger kommt zu einem zufälligen Zeitpunkt an dieser Bahnschranke an. Es sei X die Zeit, die er warten muss, bis die Bahnschranke offen ist. Wie sieht dann die Verteilungsfunktion F_X aus?

LÖSUNG. Eigentlich kann der Fußgänger zu einem beliebigen Zeitpunkt ankommen, also könnte man versuchen, die ganze Gerade \mathbb{R} mit dem Lebesgue-Maß als Wahrscheinlichkeitsraum zu nehmen. Allerdings ist das Lebesgue-Maß kein Wahrscheinlichkeitsmaß: $\lambda(\mathbb{R}) = +\infty$. Deshalb werden wir benutzen, dass die Bahnschranke periodisch funktioniert, und nur eine Periode als Grundmenge betrachten. Die Grundmenge sei also $\Omega = (0, 20)$, versehen mit der σ -Algebra der Borel-Teilmengen. Dabei sei die Bahnschranke im Zeitintervall $(0, 10)$ offen und im Zeitintervall $(10, 20)$ geschlossen. Die Ankunft des Fußgängers kann man sich nun so vorstellen: es wird im Intervall $(0, 20)$ ein zufälliger, gleichverteilter Punkt ausgewählt. Die Wahrscheinlichkeit einer Borel-Menge $A \subset (0, 20)$ ist also

$$\mathbb{P}[A] = \frac{\lambda(A)}{20},$$

wobei λ das Lebesgue-Maß sei. Kommt nun der Fußgänger zu einem Zeitpunkt $\omega \in (0, 20)$ an, so sieht seine Wartezeit wie folgt aus:

$$X(\omega) = \begin{cases} 0, & \omega \in (0, 10), \text{ denn da ist die Bahnschranke offen,} \\ 20 - \omega, & \omega \in (10, 20), \text{ denn die Bahnschranke öffnet zum Zeitpunkt 20.} \end{cases}$$

Den Wert an der Stelle 10 (der entweder als 0 oder als 10 festgelegt werden kann) kann man ignorieren, denn ein einzelner Punkt ist lediglich eine Nullmenge. Es gilt also

$$\mathbb{P}[X = 0] = \frac{10}{20} = \frac{1}{2}.$$

Nun bestimmen wir die Verteilungsfunktion von X :

$$F_X(t) = \mathbb{P}[X \leq t] = \begin{cases} 0, & t < 0, \\ \frac{1}{2}, & t = 0, \\ \frac{10+t}{20}, & t \in [0, 10], \\ 1, & t \geq 10. \end{cases}$$

Die Verteilungsfunktion ist nicht absolut stetig und nicht singulär (denn es gibt ein Atom an der Stelle 0). Sie ist auch nicht diskret, denn sie ist keine reine Sprungfunktion. Hier haben wir es mit einer gemischten Verteilung zu tun. Die Verteilung von X ist eine Mischung aus einer diskreten Verteilung (die nur den Wert 0 annimmt, ein Atom mit der Wahrscheinlichkeit 1 an der Stelle 0) und einer absolut stetigen Verteilung (Gleichverteilung auf $(0, 10)$). Bezeichnet man mit F_1 und F_2 die entsprechenden Verteilungsfunktionen, so hat man die Darstellung

$$F_X = \frac{1}{2}F_1 + \frac{1}{2}F_2.$$

8.9. Verteilungsfunktion eines Zufallsvektors

Die Familie der Borel-Teilmengen von \mathbb{R}^d wird mit \mathcal{B}^d bezeichnet.

DEFINITION 8.9.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ ein Zufallsvektor.

- (1) Die *Verteilung* von X ist die Funktion

$$P_X : \mathcal{B}^d \rightarrow [0, 1] \text{ mit } P_X(A) = \mathbb{P}[X \in A], \quad A \in \mathcal{B}^d.$$

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

- (2) Die *Verteilungsfunktion* von X ist die Funktion

$$F_X : \mathbb{R}^d \rightarrow [0, 1] \text{ mit } F_X(t) = F_X(t_1, \dots, t_d) = \mathbb{P}[X_1 \leq t_1, \dots, X_d \leq t_d],$$

wobei $t = (t_1, \dots, t_d) \in \mathbb{R}^d$.

SATZ 8.9.2. Sei $X : \Omega \rightarrow \mathbb{R}^d$ ein Zufallsvektor. Dann hat die Verteilungsfunktion von X die folgenden drei Eigenschaften.

- (1) *Grenzwerte:*

- (a) Die Verteilungsfunktion konvergiert gegen 0, wenn *mindestens eine* der Koordinaten gegen $-\infty$ konvergiert. Das heißt, für jedes $i = 1, \dots, d$ gilt

$$\lim_{t_i \rightarrow -\infty} F_X(t_1, \dots, t_d) = 0.$$

- (b) Die Verteilungsfunktion konvergiert gegen 1, wenn *alle* Koordinaten *gleichzeitig* gegen $+\infty$ konvergieren. Das heißt,

$$\lim_{t_1 \rightarrow +\infty, \dots, t_d \rightarrow +\infty} F_X(t_1, \dots, t_d) = 1.$$

- (2) *Monotonie:* Für alle $t_1 \leq t'_1, \dots, t_d \leq t'_d$ gilt

$$F_X(t_1, \dots, t_d) \leq F_X(t'_1, \dots, t'_d).$$

- (3) *Rechtsstetigkeit:* Für alle $(t_1, \dots, t_d) \in \mathbb{R}^d$ gilt

$$F_X(t_1, \dots, t_d) = \lim_{s_1 \downarrow t_1, \dots, s_d \downarrow t_d} F_X(s_1, \dots, s_d).$$

BEWEIS. Der Beweis geht analog zum eindimensionalen Fall. □

8.10. Diskrete und absolut stetige Zufallsvektoren

Wie im Fall von Zufallsvariablen lassen sich diskrete, absolut stetige und singuläre Vektoren definieren.

DEFINITION 8.10.1. Ein Zufallsvektor $X : \Omega \rightarrow \mathbb{R}^d$ heißt *diskret*, wenn X höchstens abzählbar viele Werte annimmt. Die *Zähldichte* von X ist die Funktion

$$p_X(y) = \mathbb{P}[X = y], \quad y \in \mathbb{R}^d.$$

DEFINITION 8.10.2. Ein Zufallsvektor $X : \Omega \rightarrow \mathbb{R}^d$ heißt *absolut stetig*, wenn eine Borel-Funktion $f_X : \mathbb{R}^d \rightarrow [0, \infty)$ existiert mit

$$F_X(t_1, \dots, t_d) = \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_d} f_X(y_1, \dots, y_d) dy_1 \dots dy_d$$

für alle $t_1, \dots, t_d \in \mathbb{R}$. Die Funktion f_X heißt die *Dichte* von X .

BEMERKUNG 8.10.3. Genauso wie im eindimensionalen Fall zeigt man, dass

$$\mathbb{P}[X \in B] = \int_B f_X(y_1, \dots, y_d) dy_1 \dots dy_d \text{ für alle } B \in \mathcal{B}^d.$$

BEISPIEL 8.10.4. Ein Zufallsvektor X heißt *gleichverteilt* auf einer Borel-Menge $A \subset \mathbb{R}^d$ mit $A \in \mathcal{B}^d$, $0 < \lambda(A) < \infty$, wenn X absolut stetig ist mit Dichte

$$f_X(y) = \frac{\mathbb{1}_A(y)}{\lambda(A)} = \begin{cases} \frac{1}{\lambda(A)}, & y \in A, \\ 0, & y \notin A. \end{cases}$$

8.11. Randverteilungen eines Zufallsvektors

Sei $X = (X_1, \dots, X_d)$ ein Zufallsvektor. Seine Komponenten X_1, \dots, X_d sind dann Zufallsvariablen. Die Verteilungen von X_1, \dots, X_d bezeichnet man als *Randverteilungen* von X .

BEISPIEL 8.11.1. Sei $X = (X_1, \dots, X_d)$ ein d -dimensionaler diskreter Zufallsvektor mit Zähldichte

$$p_X(t) = p_{X_1, \dots, X_d}(t_1, \dots, t_d) = \mathbb{P}[X_1 = t_1, \dots, X_d = t_d], \quad t = (t_1, \dots, t_d) \in \mathbb{R}^d.$$

Dann kann man die Zähldichten der einzelnen Komponenten X_1, \dots, X_d wie folgt berechnen:

$$p_{X_1}(t) = \mathbb{P}[X_1 = t_1] = \sum_{t_2, t_3, \dots, t_d \in \mathbb{R}} p_{X_1, \dots, X_d}(t_1, \dots, t_d),$$
$$p_{X_2}(t) = \mathbb{P}[X_2 = t_2] = \sum_{t_1, t_3, \dots, t_d \in \mathbb{R}} p_{X_1, \dots, X_d}(t_1, \dots, t_d),$$

und so weiter.

BEISPIEL 8.11.2. Wir würfeln 2-mal mit einem fairen Würfel. Dabei seien Z_1 und Z_2 die Augenzahlen. Weiter definiere

$$X_1 = \max\{Z_1, Z_2\}, \quad X_2 = \min\{Z_1, Z_2\}.$$

Dann ist $X = (X_1, X_2)$ ein diskreter Zufallsvektor. Wir bestimmen die Randverteilung (also in diesem Fall die Zähldichte) von X_1 .

LÖSUNG. Die Zähldichte von (X_1, X_2) stellt sich wie folgt dar:

$$p_X(i, i) = \mathbb{P}[X = (i, i)] = \mathbb{P}[\{(i, i)\}] = \frac{1}{36}, \quad i = 1, \dots, 6,$$

$$p_X(i, j) = \mathbb{P}[X = (i, j)] = \mathbb{P}[\{(i, j), (j, i)\}] = \frac{2}{36} = \frac{1}{18}, \quad 1 \leq j < i \leq 6.$$

Somit kann man die Zähldichte von X_1 bestimmen:

$$p_{X_1}(2) = p_X(2, 1) + p_X(2, 2) = \frac{1}{18} + \frac{1}{36} = \frac{1}{12},$$

$$p_{X_1}(3) = p_X(3, 1) + p_X(3, 2) + p_X(3, 3) = \frac{1}{18} + \frac{1}{18} + \frac{1}{36} = \frac{5}{36},$$

und so weiter.

Nun zeigen wir, wie man die Randverteilungen eines absolut stetigen Zufallsvektors berechnet.

SATZ 8.11.3. Sei $X = (X_1, \dots, X_d)$ ein absolut stetiger Zufallsvektor mit Dichte f_X . Dann gilt: Für alle $i = 1, \dots, d$ ist X_i absolut stetig mit Dichte

$$(8.11.1) \quad f_{X_i}(t_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(t_1, \dots, t_d) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_d$$

für fast alle $t_i \in \mathbb{R}$.

BEWEIS. Wir beweisen den Satz für $i = 1$. Der allgemeine Fall ist analog. Betrachte die Menge $B = \{(X_1, \dots, X_d) : X_1 \leq s\}$. Die Verteilungsfunktion von X_1 ist

$$\begin{aligned} F_{X_1}(s) &= \mathbb{P}[X_1 \leq s] \\ &= \int_B f_X(t_1, \dots, t_d) dy_1 \dots dy_d \\ &= \int_{-\infty}^s \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(y_1, \dots, y_d) dy_1 \dots dy_d \\ &= \int_{-\infty}^s g(y_1) dy_1, \end{aligned}$$

wobei g die Funktion auf der rechten Seite von (8.11.1) ist:

$$g(y_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(y_1, \dots, y_d) dy_2 \dots dy_d.$$

Also ist g die Dichte von X_1 . □

BEISPIEL 8.11.4. Die Zufallsvariable X sei gleichverteilt auf dem Einheitskreis $D = \{(y_1, y_2) : y_1^2 + y_2^2 \leq 1\}$. Die Dichte von X ist

$$f_{X_1, X_2}(y_1, y_2) = \begin{cases} \frac{1}{\pi}, & (y_1, y_2) \in D, \\ 0, & (y_1, y_2) \notin D. \end{cases}$$

Wie berechnet man die Randdichten f_{X_1}, f_{X_2} ?

LÖSUNG. Die Randdichte von X_1 ist

$$f_{X_1}(t_1) = \int_{-\infty}^{\infty} f_{X_1, X_2}(t_1, t_2) dt_2 = \begin{cases} 0, & |t_1| > 1, \\ \int_{-\sqrt{1-t_1^2}}^{+\sqrt{1-t_1^2}} \frac{1}{\pi} dt_2 = \frac{2}{\pi} \sqrt{1-t_1^2}, & \text{sonst.} \end{cases}$$

Die Randdichte von X_2 bestimmt man genauso.

8.12. Unabhängigkeit und Produktformeln

Für diskrete Zufallsvariablen haben wir die Unabhängigkeit wie folgt definiert: Zufallsvariablen X_1, \dots, X_d heißen unabhängig, wenn für alle $y_1, \dots, y_d \in \mathbb{R}$ gilt

$$\mathbb{P}[X_1 = y_1, \dots, X_d = y_d] = \mathbb{P}[X_1 = y_1] \cdot \dots \cdot \mathbb{P}[X_d = y_d].$$

Diese Definition macht im Allgemeinen Fall allerdings keinen Sinn. Sind z.B. X_1, \dots, X_d absolut stetig, so sind beide Seiten der Gleichung gleich 0. Somit wären beliebige absolut stetige Zufallsvariablen unabhängig, was dem intuitiven Verständnis der Unabhängigkeit nicht entspricht. Für allgemeine Zufallsvariablen benutzt man deshalb eine andere Definition.

DEFINITION 8.12.1. Seien X_1, \dots, X_d Zufallsvariablen. Diese Zufallsvariablen heißen *unabhängig*, wenn für alle Borel-Mengen $B_1, \dots, B_d \subset \mathbb{R}$ gilt:

$$\mathbb{P}[X_1 \in B_1, \dots, X_d \in B_d] = \mathbb{P}[X_1 \in B_1] \cdot \dots \cdot \mathbb{P}[X_d \in B_d].$$

Eine unendliche Familie X_1, X_2, \dots von Zufallsvariablen heißt unabhängig, wenn für jedes $d \in \mathbb{N}$ die Zufallsvariablen X_1, \dots, X_d unabhängig sind.

BEMERKUNG 8.12.2. Wenn die Zufallsvariablen X_1, \dots, X_d unabhängig sind, so folgt daraus, dass

$$\begin{aligned} F_{X_1, \dots, X_d}(t_1, \dots, t_d) &= \mathbb{P}[X_1 \leq t_1, \dots, X_d \leq t_d] \\ &= \mathbb{P}[X_1 \leq t_1] \cdot \dots \cdot \mathbb{P}[X_d \leq t_d] \\ &= F_{X_1}(t_1) \cdot \dots \cdot F_{X_d}(t_d). \end{aligned}$$

Die gemeinsame Verteilungsfunktion eines Zufallsvektors mit unabhängigen Komponenten ist also das Produkt der Verteilungsfunktionen der einzelnen Komponenten.

Eine ähnliche Produktformel gilt auch für Dichten, wie der nächste Satz zeigt.

SATZ 8.12.3. Seien X_1, \dots, X_d unabhängige absolut stetige Zufallsvariablen mit Dichten f_{X_1}, \dots, f_{X_d} . Dann ist der Zufallsvektor $X = (X_1, \dots, X_d)$ absolut stetig mit

$$f_{X_1, \dots, X_d}(y_1, \dots, y_d) = f_{X_1}(y_1) \cdot \dots \cdot f_{X_d}(y_d) \text{ für fast alle } (y_1, \dots, y_d) \in \mathbb{R}^d.$$

Das heißt: die gemeinsame Dichte eines Vektors mit unabhängigen Komponenten ist das Produkt der Randdichten.

BEWEIS. Wir können die Verteilungsfunktion des Vektors (X_1, \dots, X_d) wie folgt darstellen:

$$\begin{aligned} F_{X_1, \dots, X_d}(t_1, \dots, t_d) &= F_{X_1}(t_1) \cdot \dots \cdot F_{X_d}(t_d) \\ &= \int_{-\infty}^{t_1} f_{X_1}(y_1) dy_1 \cdot \dots \cdot \int_{-\infty}^{t_d} f_{X_d}(y_d) dy_d \\ &= \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_d} f_{X_1}(y_1) \cdot \dots \cdot f_{X_d}(y_d) dy_1 \dots dy_d. \end{aligned}$$

Es folgt aus der Definition der absoluten Stetigkeit, dass $f_{X_1}(y_1) \cdot \dots \cdot f_{X_d}(y_d)$ die Dichte von (X_1, \dots, X_d) ist. \square

BEISPIEL 8.12.4. Seien die Zufallsvariablen X_1, X_2 unabhängig und gleichverteilt auf dem Intervall $[0, 1]$. Dann können wir die Dichte des Vektors (X_1, X_2) wie folgt bestimmen:

$$f_{X_1, X_2}(y_1, y_2) = f_{X_1}(y_1) \cdot f_{X_2}(y_2) = \mathbb{1}_{y_1 \in [0, 1]} \cdot \mathbb{1}_{y_2 \in [0, 1]} = \mathbb{1}_{(y_1, y_2) \in [0, 1]^2}.$$

Somit ist der Vektor (X_1, X_2) gleichverteilt auf dem Quadrat $[0, 1]^2$.

8.13. Transformationsformel für die Dichte

Sei X eine absolut stetige Zufallsvariable und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion. Wie bestimmt man die Dichte von $\varphi(X)$?

SATZ 8.13.1. *Seien*

- (1) X eine absolut stetige Zufallsvariable mit Dichte f_X .
- (2) $I \subset \mathbb{R}$ ein offenes Intervall mit $\mathbb{P}[X \in I] = 1$, wobei I auch unendlich sein darf.
- (3) $\varphi : I \rightarrow \mathbb{R}$ eine stetig differenzierbare Funktion mit $\varphi'(x) \neq 0$ für alle $x \in I$.

Dann ist $\varphi(X)$ eine absolut stetige Zufallsvariable mit Dichte

$$f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y)) \cdot |(\varphi^{-1})'(y)| \text{ für fast alle } y \in \varphi(I).$$

Dabei ist φ^{-1} die inverse Funktion von φ .

BEWEIS. Wegen $\varphi'(x) \neq 0$ für alle $x \in I$ ist φ entweder monoton steigend oder monoton fallend. Sei o.B.d.A. φ monoton steigend. Die Verteilungsfunktion von $\varphi(X)$, an der Stelle $t \in \varphi(I)$, ist

$$F_{\varphi(X)}(t) = \mathbb{P}[\varphi(X) \leq t] = \mathbb{P}[X \leq \varphi^{-1}(t)] = F_X(\varphi^{-1}(t)).$$

Die Dichte von $\varphi(X)$ erhalten wir, indem wir die Verteilungsfunktion ableiten:

$$f_{\varphi(X)}(t) = \frac{d}{dt} F_{\varphi(X)}(t) = \frac{d}{dt} F_X(\varphi^{-1}(t)) = F_X'(\varphi^{-1}(t)) (\varphi^{-1})'(t) = f_X(\varphi^{-1}(t)) (\varphi^{-1})'(t).$$

\square

BEMERKUNG 8.13.2. Für monoton fallendes φ geht dies analog, jedoch mit $|(\varphi^{-1})'(t)|$ anstelle von $(\varphi^{-1})'(t)$.

BEISPIEL 8.13.3. Sei X eine absolut stetige Zufallsvariable mit Dichte f_X und $a \neq 0$ und $b \in \mathbb{R}$ zwei Konstanten. Bestimme die Dichte von $aX + b$.

LÖSUNG. Die Funktion $\varphi(X) = aX + b$ ist stetig differenzierbar auf $I = \mathbb{R}$. Die Ableitung ist $\varphi'(X) = a \neq 0$. Die Umkehrfunktion und ihre Ableitung sind

$$\varphi^{-1}(x) = \frac{x - b}{a}, \quad (\varphi^{-1})'(x) = \frac{1}{a}.$$

Nun setzt man das in die Transformationformel ein und erhält für die Dichte von $aX + b$:

$$f_{aX+b}(t) = \frac{1}{|a|} \cdot f_X\left(\frac{t - b}{a}\right).$$

BEISPIEL 8.13.4. Sei $X \sim N(0, 1)$ standardnormalverteilt und $\mu \in \mathbb{R}, \sigma > 0$ zwei Konstanten. Die Dichte von $Y := \mu + \sigma X$ ist, mit der Formel aus dem obigen Beispiel,

$$f_Y(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

Also ist $Y \sim N(\mu, \sigma^2)$. Umgekehrt folgt aus $Y \sim N(\mu, \sigma^2)$, dass die Zufallsvariable

$$X = \frac{Y - \mu}{\sigma}$$

standardnormalverteilt ist. Mit einer linearen Transformation kann man also eine beliebige Normalverteilung auf die Standardnormalverteilung zurückführen und umgekehrt.

8.14. Faltungsformeln

Seien X_1, X_2 unabhängige Zufallsvariablen. Wie bestimmt man dann die Verteilung der Zufallsvariablen

$$X_1 + X_2, X_1 - X_2, X_1 \cdot X_2, \frac{X_1}{X_2}?$$

Zuerst bestimmen wir die Verteilung von $X_1 + X_2$.

SATZ 8.14.1 (Faltungsformel für diskrete Zufallsvariablen). *Seien X_1, X_2 unabhängige Zufallsvariablen. Es seien beide Zufallsvariablen diskret. Dann gilt für die Zähldichte von $X_1 + X_2$ die Formel*

$$(8.14.1) \quad p_{X_1+X_2}(z) = \sum_{y \in \text{Im}(X_1)} p_{X_1}(y) \cdot p_{X_2}(z - y) \text{ für alle } z \in \mathbb{R}.$$

DEFINITION 8.14.2. Die Zähldichte $p_{X_1+X_2} = p_{X_1} * p_{X_2}$ heißt die *Faltung* der Zähldichten p_{X_1} und p_{X_2} .

BEMERKUNG 8.14.3. Zur Erinnerung: $p_{X_1}(y) = \mathbb{P}[X_1 = y]$ ist die Zähldichte von X_1 und $\text{Im}(X_1) = \{y \in \mathbb{R} : \mathbb{P}[X_1 = y] \neq 0\}$ ist die Menge der Werte, die X_1 annehmen kann.

BEWEIS VON SATZ 8.14.1. Sei $z \in \mathbb{R}$. Es gilt dann

$$\begin{aligned}
 p_{X_1+X_2}(z) &= \mathbb{P}[X_1 + X_2 = z] \\
 &= \sum_{y \in \text{Im}(X_1)} \mathbb{P}[X_1 = y, X_2 = z - y] \\
 &= \sum_{y \in \text{Im}(X_1)} \mathbb{P}[X_1 = y] \cdot \mathbb{P}[X_2 = z - y] \\
 &= \sum_{y \in \text{Im}(X_1)} p_{X_1}(y) \cdot p_{X_2}(z - y).
 \end{aligned}$$

□

SATZ 8.14.4. Es seien $X_1 \sim \text{Poi}(\lambda_1)$ und $X_2 \sim \text{Poi}(\lambda_2)$ unabhängige, Poisson-verteilte Zufallsvariablen. Dabei sind $\lambda_1, \lambda_2 > 0$ Parameter. Dann ist auch die Summe $X_1 + X_2$ Poisson-verteilt:

$$X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2).$$

Das heißt: $\text{Poi}(\lambda_1) * \text{Poi}(\lambda_2) = \text{Poi}(\lambda_1 + \lambda_2)$.

BEWEIS. Für alle $n \in \mathbb{N}_0$ gilt

$$\begin{aligned}
 \mathbb{P}[X_1 + X_2 = n] &= \sum_{k=0}^n \mathbb{P}[X_1 = k] \cdot \mathbb{P}[X_2 = n - k] \\
 &= \sum_{k=0}^n e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_1^k \cdot \lambda_2^{n-k}}{k!(n-k)!} \frac{n!}{n!} \\
 &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda_1^k \lambda_2^{n-k} \\
 &= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}.
 \end{aligned}$$

Somit ist $X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$.

□

ÜBUNG 8.14.5. Seien $X_1 \sim \text{Bin}(n_1, p)$ und $X_2 \sim \text{Bin}(n_2, p)$ unabhängig. Dann gilt

$$X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p).$$

Das heißt: $\text{Bin}(n_1, p) * \text{Bin}(n_2, p) = \text{Bin}(n_1 + n_2, p)$.

ÜBUNG 8.14.6. Seien $X_1 \sim \text{NB}(r_1, p)$ und $X_2 \sim \text{NB}(r_2, p)$ unabhängig. Dann gilt

$$X_1 + X_2 \sim \text{NB}(r_1 + r_2, p).$$

Das heißt: $\text{NB}(r_1, p) * \text{NB}(r_2, p) = \text{NB}(r_1 + r_2, p)$.

Als Spezialfall dieser Aussage ergibt sich

ÜBUNG 8.14.7. Seien $X_1, \dots, X_n \sim \text{Geo}(p)$ unabhängig. Dann gilt

$$X_1 + \dots + X_n \sim \text{NB}(n, p).$$

Das heißt: $\text{Geo}(p) * \dots * \text{Geo}(p) = \text{NB}(n, p)$.

SATZ 8.14.8 (Faltungsformel für absolut stetige Zufallsvariablen). Seien X_1, X_2 unabhängige und absolut stetige Zufallsvariablen mit Dichten f_{X_1} und f_{X_2} . Dann ist auch die Summe $X_1 + X_2$ absolut stetig mit Dichte

$$(8.14.2) \quad f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_{X_1}(y)f_{X_2}(z-y)dy \text{ für fast alle } z \in \mathbb{R}.$$

DEFINITION 8.14.9. Die Dichte $f_{X_1+X_2} = f_{X_1} * f_{X_2}$ heißt die *Faltung* der Dichten f_{X_1} und f_{X_2} .

BEMERKUNG 8.14.10. Eine intuitive Erklärung dieser Formel: Damit $X_1 + X_2 = z$ ist, muss X_1 irgendeinen Wert y annehmen (diesem Ereignis entspricht die Dichte $f_{X_1}(y)$) und dann muss $X_2 = z - y$ gelten (diesem Ereignis entspricht die Dichte $f_{X_2}(z - y)$). Da X_1 und X_2 unabhängig sind, multipliziert man beide Dichten. Da außerdem y beliebige Werte annehmen kann, integriert man zum Schluss über y .

BEWEIS VON SATZ 8.14.8. Die rechte Seite der Faltungsformel bezeichnen wir mit $g(z)$, d.h.

$$g(z) = \int_{\mathbb{R}} f_{X_1}(y)f_{X_2}(z-y)dy.$$

Für alle $t \in \mathbb{R}$ gilt mit dem Satz von Fubini (Maßtheorie)

$$\begin{aligned} \int_{-\infty}^t g(z)dz &= \int_{-\infty}^t \int_{\mathbb{R}} f_{X_1}(y)f_{X_2}(z-y)dydz \\ &= \int_{\mathbb{R}} \int_{-\infty}^t f_{X_1}(y)f_{X_2}(z-y)dzdy \\ &= \int_{\mathbb{R}} \int_{-\infty}^{t-y} f_{X_1}(y)f_{X_2}(u)dudy, \end{aligned}$$

wobei wir die neue Variable $u = z - y$ eingeführt haben. Definiere die Menge

$$B \stackrel{\text{def}}{=} \{(u, y) : u + y \leq t\}.$$

Dann kann man das Integral so darstellen:

$$\int_{-\infty}^t g(z)dz = \int_B f_{X_1}(y)f_{X_2}(u)dudy = \int_B f_{X_1, X_2}(y, u)d(u, y) = \mathbb{P}[(X_1, X_2) \in B].$$

Dabei gilt die Formel $f_{X_1}(y)f_{X_2}(u) = f_{X_1, X_2}(y, u)$ wegen der Unabhängigkeit von X_1 und X_2 . Somit haben wir gezeigt, dass

$$\int_{-\infty}^t g(z)dz = \mathbb{P}[X_1 + X_2 \leq t] = F_{X_1+X_2}(t).$$

Daraus folgt, dass g die Dichte von $X_1 + X_2$ ist. □

SATZ 8.14.11. Seien $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$ normalverteilt und unabhängig. Dann ist auch die Summe $X_1 + X_2$ normalverteilt:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

BEWEIS. SCHRITT 1. Seien zuerst $\mu_1 = \mu_2 = 0$ und somit $X_1 \sim N(0, \sigma_1^2)$ und $X_2 \sim N(0, \sigma_2^2)$. Mit der Faltungsformel ergibt sich

$$\begin{aligned} f_{X_1+X_2}(z) &= \int_{-\infty}^{\infty} f_{X_1}(y)f_{X_2}(z-y)dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y^2}{\sigma_1^2} + \frac{(z-y)^2}{\sigma_2^2}\right)} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y^2}{\sigma_1^2} + \frac{(z-y)^2}{\sigma_2^2} - \frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right)} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y(\sigma_1^2+\sigma_2^2)-z\sigma^2}{\sigma_1\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}\right)^2} dy. \end{aligned}$$

Nun führen wir eine neue Variable $w = \frac{y(\sigma_1^2+\sigma_2^2)-z\sigma^2}{\sigma_1\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}$ ein:

$$\begin{aligned} f_{X_1+X_2}(z) &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}w^2} dw \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \cdot \sqrt{2\pi} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}. \end{aligned}$$

Also gilt $X_1 + X_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$.

SCHRITT 2. Seien nun $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$ normalverteilt mit beliebigen Parametern. Betrachte die zentrierten Zufallsvariablen $X'_1 := X_1 - \mu_1 \sim N(0, \sigma_1^2)$ und $X'_2 := X_2 - \mu_2 \sim N(0, \sigma_2^2)$. Für die Zufallsvariablen X'_1 und X'_2 haben wir in Schritt 1 bereits gezeigt, dass $X'_1 + X'_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$. Somit gilt

$$X_1 + X_2 = (\mu_1 + \mu_2) + (X'_1 + X'_2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Später werden wir diesen Satz auf einem viel schönerem Weg mithilfe von charakteristischen Funktionen beweisen. \square

BEISPIEL 8.14.12. Die Dauer eines Telefongesprächs sei eine exponentialverteilte Zufallsvariable mit Parameter 1. Betrachte zwei unabhängig stattfindende Gespräche. Wie ist dann die Gesamtdauer der beiden Gespräche verteilt?

LÖSUNG. Wir betrachten also zwei unabhängige Zufallsvariablen $X_1 \sim \text{Exp}(1)$ und $X_2 \sim \text{Exp}(1)$ und bestimmen die Verteilung von $X_1 + X_2$. Die Dichten von X_1 und X_2 sind

$$f_{X_1}(y) = f_{X_2}(y) = \begin{cases} e^{-y}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

Die Zufallsvariable $X_1 + X_2$ nimmt nur positive Werte an. Sei $z > 0$. Mit der Faltungsformel ergibt sich

$$f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_{X_1}(y) f_{X_2}(z-y) dy = \int_0^z e^{-y} e^{-(z-y)} dy = e^{-z} \int_0^z 1 dy = ze^{-z}.$$

Es gilt somit

$$f_{X_1+X_2}(z) = \begin{cases} ze^{-z} & z > 0, \\ 0, & z \leq 0. \end{cases}$$

ÜBUNG 8.14.13. Seien $X_1, \dots, X_n \sim \text{Exp}(1)$ unabhängig. Dann ist die Dichte von $X_1 + \dots + X_n$ gegeben durch

$$f_{X_1+\dots+X_n}(z) = \begin{cases} \frac{1}{n!} z^n e^{-z}, & z > 0, \\ 0, & z \leq 0. \end{cases}$$

Diese Verteilung nennt man *Erlang-Verteilung*.

Seien X_1, X_2 eine unabhängige Zufallsvariablen. Wir bestimmen die Verteilung des Produkts $X_1 X_2$. Sind X_1 und X_2 diskret, so kann man die Zähldichte von $X_1 X_2$ wie folgt berechnen

$$p_{X_1 X_2}(z) = \sum_{y \in \text{Im}(X_1) \setminus \{0\}} p_{X_1}(y) p_{X_2}\left(\frac{z}{y}\right), \quad z \neq 0.$$

Sind X_1 und X_2 absolut stetig, so gilt für die Dichte von $X_1 X_2$ die Formel

$$f_{X_1 X_2}(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X_1}(y) f_{X_2}\left(\frac{z}{y}\right) dy.$$

Wir werden diese Formel nicht beweisen, sondern nur erklären, warum in ihr der Faktor $\frac{1}{|y|}$ auftaucht. Sei $\varepsilon > 0$ sehr klein. Dann gilt

$$f_{X_1 X_2}(z) \approx \frac{1}{\varepsilon} \cdot \mathbb{P}[z \leq X_1 X_2 \leq z + \varepsilon].$$

Was muss nun geschehen, damit das Ereignis $z \leq X_1 X_2 \leq z + \varepsilon$ eintritt? Zuerst muss X_1 irgendeinen Wert y annehmen. Diesem Ereignis entspricht die Dichte $f_{X_1}(y)$. Ist nun $y > 0$, so muss für X_2 gelten: $z/y \leq X_2 \leq z/y + \varepsilon/y$. Dieses Ereignis hat eine Wahrscheinlichkeit von $\approx f_{X_2}(z/y) \cdot \varepsilon/y$. Analog kommt man im Fall $y < 0$ auf das Ereignis $z/y + \varepsilon/y \leq X_2 \leq z/y$ mit einer Wahrscheinlichkeit von $\approx -f_{X_2}(z/y) \cdot \varepsilon/y$. Da nun y beliebig sein kann, integriert man über y und erhält

$$\mathbb{P}[z \leq X_1 X_2 \leq z + \varepsilon] \approx \varepsilon \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X_1}(y) f_{X_2}\left(\frac{z}{y}\right) dy.$$

Daraus ergibt sich die Formel für die Dichte von $X_1 X_2$.

Analog kann man zeigen: Sind X_1 und X_2 unabhängig und beide absolut stetig, so gilt für die Dichte von X_1/X_2 die Formel

$$f_{X_1/X_2}(z) = \int_{-\infty}^{\infty} |y| f_{X_1}(zy) f_{X_2}(y) dy.$$

8.15. Transformationsformel für den Erwartungswert

Sei X eine absolut stetige Zufallsvariable mit Dichte f_X . Sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine Borel-Funktion. Wie bestimmt man dann den Erwartungswert von $\varphi(X)$?

SATZ 8.15.1 (Transformationsformel für den Erwartungswert). *Sei X eine absolut stetige Zufallsvariable und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine Borel-Funktion. Dann gilt*

$$(8.15.1) \quad \mathbb{E}\varphi(X) = \int_{-\infty}^{\infty} \varphi(y) f_X(y) dy,$$

falls $\int_{-\infty}^{\infty} |\varphi(y)| f_X(y) dy < \infty$.

BEMERKUNG 8.15.2. Intuitiv kann man die Formel so verstehen: Die Zufallsvariable X nimmt einen Wert y mit "Dichte" $f_X(y)$ an. Ist $X = y$, so gilt $\varphi(X) = \varphi(y)$. Der entsprechende Beitrag zum Erwartungswert ist also $\varphi(y) f_X(y)$. Da nun y beliebig sein kann, integrieren wir über y .

BEWEIS VON SATZ 8.15.1. SCHRITT 1. Sei zuerst $\varphi(y) = \mathbb{1}_A(y)$ eine Indikatorfunktion, wobei $A \subset \mathbb{R}$ eine Borel-Menge ist. Dann nimmt $\varphi(X)$ nur die Werte 0 und 1 an und es gilt

$$\mathbb{E}\varphi(X) = \mathbb{E}\mathbb{1}_A(X) = 1 \cdot \mathbb{P}[X \in A] + 0 \cdot \mathbb{P}[X \notin A] = \mathbb{P}[X \in A].$$

Auf der anderen Seite, gilt

$$\int_{-\infty}^{\infty} \varphi(y) f_X(y) dy = \int_{-\infty}^{\infty} \mathbb{1}_A(y) f_X(y) dy = \int_A f_X(y) dy = \mathbb{P}[X \in A].$$

Es folgt, dass Formel (8.15.1) erfüllt ist.

SCHRITT 2. Sei $\varphi(y) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(y)$ eine Elementarfunktion. Dabei seien $A_1, \dots, A_n \subset \mathbb{R}$ disjunkte Borel-Mengen und $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. Dann gilt

$$\mathbb{E}\varphi(X) = \mathbb{E} \left[\sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(X) \right] = \sum_{i=1}^n \alpha_i \mathbb{E}[\mathbb{1}_{A_i}(X)] = \sum_{i=1}^n \alpha_i \mathbb{P}[X \in A_i],$$

wobei wir im letzten Schritt das Ergebnis von Schritt 1 benutzt haben. Auf der anderen Seite gilt

$$\int_{-\infty}^{\infty} \varphi(y) f_X(y) dy = \int_{-\infty}^{\infty} \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(y) f_X(y) dy = \sum_{i=1}^n \alpha_i \int_{-\infty}^{\infty} \mathbb{1}_{A_i}(y) f_X(y) dy = \sum_{i=1}^n \alpha_i \mathbb{P}[X \in A_i],$$

wobei wir im letzten Schritt wieder das Ergebnis von Schritt 1 benutzt haben. Also gilt die Formel (8.15.1).

SCHRITT 3. Sei nun $\varphi \geq 0$ eine beliebige, nichtnegative Borel-Funktion. Wir approximieren φ mit Elementarfunktionen. Dazu definieren wir $\varphi_n : \mathbb{R} \rightarrow \mathbb{R}$ wie folgt

$$\varphi_n(y) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbb{1}_{\frac{k-1}{2^n} \leq \varphi(y) < \frac{k}{2^n}} + n \cdot \mathbb{1}_{\varphi(y) \geq n}.$$

Die Folge φ_n hat folgende Eigenschaften:

- (1) Für alle $y \in \mathbb{R}$ gilt $\varphi_n(y) \uparrow \varphi(y)$ für $n \rightarrow \infty$.
- (2) Für alle $\omega \in \Omega$ gilt $\varphi_n(X(\omega)) \uparrow \varphi(X(\omega))$ für $n \rightarrow \infty$.

Der Satz von der monotonen Konvergenz (Maßtheorie) ergibt dann

- (1) $\int \varphi_n(y) f_X(y) dy \rightarrow \int \varphi(y) f_X(y) dy$ für $n \rightarrow \infty$.
- (2) $\mathbb{E} \varphi_n(X) \rightarrow \mathbb{E} \varphi(X)$ für $n \rightarrow \infty$.

Hier gilt $\int \varphi_n(y) f_X(y) dy = \mathbb{E} \varphi_n(X)$ nach Schritt 2 (denn φ_n ist eine Elementarfunktion), also sind die Terme $\int \varphi(y) f_X(y) dy$ und $\mathbb{E} \varphi(X)$ auch gleich und die Formel (8.15.1) gilt.

SCHRITT 4. Sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine beliebige Borel-Funktion mit $\int |\varphi(y)| f_X(y) dy \leq \infty$. Wir schreiben dann

$$\varphi = \varphi_+ - \varphi_-,$$

wobei φ_+ und φ_- der positive bzw. der negative Anteil von φ ist:

$$\varphi_+(y) = \begin{cases} \varphi(y), & \text{falls } \varphi(y) \geq 0, \\ 0, & \text{falls } \varphi(y) < 0; \end{cases} \quad \varphi_-(y) = \begin{cases} 0, & \text{falls } \varphi(y) \geq 0, \\ |\varphi(y)|, & \text{falls } \varphi(y) < 0. \end{cases}$$

Es gilt $\varphi_+ \geq 0$ und $\varphi_- \geq 0$. In Schritt 3 haben wir gezeigt, dass

$$\mathbb{E} \varphi_+(X) = \int \varphi_+(y) f_X(y) dy < \infty, \quad \mathbb{E} \varphi_-(X) = \int \varphi_-(y) f_X(y) dy < \infty.$$

Dabei sind beide Erwartungswerte endlich wegen der Annahme $\int |\varphi(y)| f_X(y) dy \leq \infty$. Wir bilden die Differenz:

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi_+(X) - \varphi_-(X)] = \int (\varphi_+(y) - \varphi_-(y)) f_X(y) dy = \int \varphi(y) f_X(y) dy.$$

Somit gilt die Formel (8.15.1). □

BEISPIEL 8.15.3. Mit $\varphi(y) = y$ erhalten wir die Formel

$$\mathbb{E}X = \int_{-\infty}^{\infty} y f_X(y) dy.$$

Etwas allgemeiner, mit $\varphi(y) = y^n$ erhalten wir

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} y^n f_X(y) dy.$$

Die Zahlen $\mathbb{E}X, \mathbb{E}[X^2], \mathbb{E}[X^3], \dots$ nennt man auch *Momente* von X .

BEISPIEL 8.15.4. Sei $X \sim \text{Exp}(1)$. Bestimme $\mathbb{E}[X^n]$.

LÖSUNG. Die Dichte von X ist $f_X(y) = e^{-y}$ für $y > 0$. Dann gilt:

$$\mathbb{E}[X^n] = \int_0^\infty y^n f_X(y) dy = \int_0^\infty y^n e^{-y} dy = n!.$$

BEMERKUNG 8.15.5. Eine ähnliche Transformationsformel gilt auch für diskrete Zufallsvariablen: ist X diskret, so haben wir

$$\mathbb{E}\varphi(X) = \sum_{y \in \text{Im}(X)} \varphi(y) p_X(y),$$

falls $\sum_{y \in \text{Im}(X)} |\varphi(y)| p_X(y) < \infty$.

8.16. Multiplikatitivität des Erwartungswerts

Sind $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ zwei (möglicherweise abhängige) integrierbare Zufallsvariablen, so ist die Summe $X + Y$ ebenfalls integrierbar und es gilt

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Dies ist eine der Eigenschaften des Lebesgue-Integrals. Nun beweisen wir, dass eine ähnliche Formel auch für das Produkt gilt, wenn man zusätzlich voraussetzt, dass die Zufallsvariablen unabhängig sind.

SATZ 8.16.1. *Die Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{R}$ seien integrierbar und unabhängig. Dann ist auch XY integrierbar und es gilt*

$$(8.16.1) \quad \mathbb{E}[XY] = (\mathbb{E}X) \cdot (\mathbb{E}Y).$$

BEWEIS. SCHRITT 1. Wenn die Zufallsvariablen nur endlich viele Werte annehmen, haben wir diesen Satz bereits im Kapitel über den Erwartungswert der diskreten Zufallsvariablen bewiesen.

SCHRITT 2. Annahme: Seien $X(\omega) \geq 0$ und $Y(\omega) \geq 0$ für alle $\omega \in \Omega$. Dann existieren zwei Folgen von Zufallsvariablen $X_n, Y_n : \Omega \rightarrow \mathbb{R}$ mit den Eigenschaften:

- (1) Für jedes $\omega \in \Omega$ gilt $X_n(\omega) \uparrow X(\omega)$ für $n \rightarrow \infty$.
- (2) Für jedes $\omega \in \Omega$ gilt $Y_n(\omega) \uparrow Y(\omega)$ für $n \rightarrow \infty$.
- (3) Für jedes $\omega \in \Omega$ gilt $X_n(\omega) \geq 0, Y_n(\omega) \geq 0$.
- (4) X_n und Y_n nehmen nur endlich viele Werte an.

Beispielsweise kann man X_n und Y_n so konstruieren: $X_n = f_n(X)$ und $Y_n = f_n(Y)$ mit

$$f_n(z) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbb{1}_{\frac{k-1}{2^n} \leq z \leq \frac{k}{2^n}} + n \cdot \mathbb{1}_{z \geq n}.$$

Damit gilt dann für alle $\omega \in \Omega$:

$$X_n(\omega) \cdot Y_n(\omega) \uparrow X(\omega) \cdot Y(\omega) \text{ für } n \rightarrow \infty.$$

Da die Zufallsvariablen X und Y unabhängig sind, sind auch $X_n = f_n(X)$ und $Y_n = f_n(Y)$ unabhängig. Aus dem Satz von der monotonen Konvergenz folgt, dass

$$\begin{aligned}\mathbb{E}[XY] &= \lim_{n \rightarrow \infty} \mathbb{E}[X_n \cdot Y_n] \\ &= \lim_{n \rightarrow \infty} \mathbb{E}[X_n] \cdot \mathbb{E}[Y_n] \\ &= \lim_{n \rightarrow \infty} \mathbb{E}[X_n] \cdot \lim_{n \rightarrow \infty} \mathbb{E}[Y_n] \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y].\end{aligned}$$

Dabei haben wir benutzt, dass $\mathbb{E}[X_n \cdot Y_n] = \mathbb{E}[X_n] \cdot \mathbb{E}[Y_n]$. Dies wurde in Schritt 1 gezeigt, denn X_n und Y_n nehmen nur endlich viele Werte an. Also gilt die Formel (8.16.1).

SCHRITT 3. Seien nun $X, Y : \Omega \rightarrow \mathbb{R}$ unabhängig und integrierbar. Dann sind $|X|$ und $|Y|$ nichtnegativ und wir können Schritt 2 anwenden:

$$\mathbb{E}|XY| = \mathbb{E}[|X| \cdot |Y|] = \mathbb{E}|X| \cdot \mathbb{E}|Y| < \infty.$$

Also ist XY integrierbar. Nun schreiben wir $X = X^+ - X^-$ und $Y = Y^+ - Y^-$ mit

$$\begin{aligned}X^+(\omega) &= \begin{cases} X(\omega), & X(\omega) \geq 0, \\ 0, & X(\omega) < 0, \end{cases} & X^-(\omega) &= \begin{cases} 0, & X(\omega) \geq 0, \\ |X(\omega)|, & X(\omega) < 0, \end{cases} \\ Y^+(\omega) &= \begin{cases} Y(\omega), & Y(\omega) \geq 0, \\ 0, & Y(\omega) < 0, \end{cases} & Y^-(\omega) &= \begin{cases} 0, & Y(\omega) \geq 0, \\ |Y(\omega)|, & Y(\omega) < 0. \end{cases}\end{aligned}$$

Da X und Y unabhängig sind, gilt auch

- (1) X^+ und Y^+ sind unabhängig.
- (2) X^+ und Y^- sind unabhängig.
- (3) X^- und Y^+ sind unabhängig.
- (4) X^- und Y^- sind unabhängig.

Da nun X^+, Y^+, X^-, Y^- alle nichtnegativ sind, können wir auf diese Zufallsvariablen Schritt 2 anwenden:

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] \\ &= \mathbb{E}[X^+Y^+ - X^-Y^+ - X^+Y^- + X^-Y^-] \\ &= \mathbb{E}[X^+Y^+] - \mathbb{E}[X^-Y^+] - \mathbb{E}[X^+Y^-] + \mathbb{E}[X^-Y^-] \\ &= \mathbb{E}[X^+] \cdot \mathbb{E}[Y^+] - \mathbb{E}[X^-] \cdot \mathbb{E}[Y^+] - \mathbb{E}[X^+] \cdot \mathbb{E}[Y^-] + \mathbb{E}[X^-] \cdot \mathbb{E}[Y^-] \\ &= \mathbb{E}[X^+ - X^-] \cdot \mathbb{E}[Y^+ - Y^-] \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y].\end{aligned}$$

Also gilt die Formel $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$. □

KAPITEL 9

Varianz und Kovarianz

9.1. Varianz

DEFINITION 9.1.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Wir benutzen die Notation

- (1) $X \in L^1$, falls $\mathbb{E}[|X|] < \infty$.
- (2) $X \in L^2$, falls $\mathbb{E}[X^2] < \infty$.

LEMMA 9.1.2. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X, Y : \Omega \rightarrow \mathbb{R}$ zwei Zufallsvariablen. Dann gilt:

- (1) $X, Y \in L^2 \Rightarrow X + Y \in L^2$.
- (2) $X \in L^2, a \in \mathbb{R} \Rightarrow aX \in L^2$.
- (3) $X \in L^2 \Rightarrow X \in L^1$.

BEWEIS VON (1). Seien $X, Y \in L^2$, also $\mathbb{E}[X^2] < \infty, \mathbb{E}[Y^2] < \infty$. Aus der Ungleichung $(X + Y)^2 \leq 2X^2 + 2Y^2$ folgt, dass

$$\mathbb{E}[(X + Y)^2] \leq 2\mathbb{E}[X^2] + 2\mathbb{E}[Y^2] < \infty.$$

Somit gilt $X + Y \in L^2$. □

BEWEIS VON (2). Sei $X \in L^2$, also $\mathbb{E}[X^2] < \infty$. Es folgt, dass $\mathbb{E}[(aX)^2] = a^2\mathbb{E}[X^2] < \infty$, somit $aX \in L^2$. □

BEWEIS VON (3). Sei $X \in L^2$, also $\mathbb{E}[X^2] < \infty$. Es gilt die Ungleichung $|X| \leq \frac{X^2+1}{2}$, also

$$\mathbb{E}[|X|] \leq \mathbb{E}\left[\frac{X^2+1}{2}\right] = \mathbb{E}\left[\frac{X^2}{2}\right] + \frac{1}{2} < \infty.$$

Somit ist $X \in L^1$. □

DEFINITION 9.1.3. Sei $X \in L^2$. Die *Varianz* von X ist:

$$\text{Var } X = \mathbb{E}[(X - \mathbb{E}[X])^2] < \infty.$$

Die *Standardabweichung* von X ist:

$$\sigma(X) = \sqrt{\text{Var } X}.$$

BEMERKUNG 9.1.4. Die Varianz beschreibt die erwartete quadratische Abweichung einer Zufallsvariable von ihrem Erwartungswert. Wird X in irgendwelchen physikalischen Einheiten, etwa in Metern, gemessen, so wird $\text{Var } X$ in Quadratmetern gemessen. Deshalb führt man die Standardabweichung von X ein. Diese wird dann wieder in Metern gemessen, hat

also die gleichen Einheiten wie X . Die Standardabweichung und die Varianz beschreiben, wie stark die Zufallsvariable um ihrem Erwartungswert streut. Für eine konstante Zufallsvariable $X = c$ gilt zum Beispiel, dass $\text{Var } X = \sigma(X) = 0$, da es in diesem Fall gar keine Streuung gibt.

SATZ 9.1.5. Für jede Zufallsvariable $X \in L^2$ gilt die Formel:

$$\text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

BEMERKUNG 9.1.6. Es gilt $\text{Var } X \geq 0$. Als Korollar erhalten wir, dass $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$.

BEWEIS. Wir benutzen die Linearität des Erwartungswerts:

$$\begin{aligned} \text{Var } X &= \mathbb{E}[(X - \mathbb{E}X)^2] \\ &= \mathbb{E}[X^2 - 2X \cdot \mathbb{E}X + (\mathbb{E}X)^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[2\mathbb{E}X \cdot X] + \mathbb{E}[(\mathbb{E}X)^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}X \cdot \mathbb{E}X + (\mathbb{E}X)^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}X)^2. \end{aligned}$$

Dabei haben wir benutzt, dass $\mathbb{E}X$ eine Konstante ist. □

BEISPIEL 9.1.7. Sei $X \sim U[0, 1]$ gleichverteilt auf dem Intervall $[0, 1]$. Wir berechnen $\text{Var } X$.

LÖSUNG. Die Dichte von X ist

$$f_X(y) = \begin{cases} 1, & y \in [0, 1], \\ 0, & y \notin [0, 1]. \end{cases}$$

Wir berechnen $\mathbb{E}X$:

$$\mathbb{E}X = \int_{\mathbb{R}} y f_X(y) dy = \int_0^1 y dy = \frac{1}{2}.$$

Wir berechnen $\mathbb{E}[X^2]$ mit der Transformationsformel für den Erwartungswert:

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} y^2 f_X(y) dy = \int_0^1 y^2 dy = \frac{1}{3}.$$

Somit erhalten wir

$$\text{Var } X = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

Für die Standardabweichung ergibt sich $\sigma(X) = \frac{1}{\sqrt{12}}$.

SATZ 9.1.8. Sei $X \in L^2$ eine Zufallsvariable und $a, b \in \mathbb{R}$ Konstanten. Dann gilt:

- (1) $\text{Var}(aX + b) = a^2 \text{Var } X$.
- (2) $\sigma(aX + b) = |a|\sigma(X)$.

BEWEIS VON (1). Wir benutzen die Linearität des Erwartungswerts:

$$\text{Var}(aX+b) = \mathbb{E}[(aX+b-\mathbb{E}(aX+b))^2] = \mathbb{E}[(aX+b-a\mathbb{E}X-b)^2] = \mathbb{E}[a^2(X-\mathbb{E}X)^2] = a^2 \text{Var } X. \quad \square$$

BEWEIS VON (2). $\sigma(aX+b) = \sqrt{\text{Var}(aX+b)} = \sqrt{a^2 \text{Var } X} = |a|\sigma(X). \quad \square$

BEISPIEL 9.1.9. Sei $X \sim U[a, b]$ gleichverteilt auf einem Intervall $[a, b]$. Es gilt $\mathbb{E}X = \frac{a+b}{2}$. Wir berechnen $\text{Var } X$.

LÖSUNG. Wir haben die Darstellung $X = (b-a)Y + a$, wobei $Y \sim U[0, 1]$.

$$\text{Var } X = \text{Var}((b-a)Y + a) = (b-a)^2 \text{Var } Y = \frac{(b-a)^2}{12}.$$

Für die Standardabweichung ergibt sich $\sigma(X) = \frac{b-a}{\sqrt{12}}$.

BEMERKUNG 9.1.10. Die Varianz einer Zufallsvariable ist immer ≥ 0 . Für eine konstante Zufallsvariable $X = c$ gilt $\text{Var } X = 0$. Können wir die Umkehrung dieser Aussage beweisen? Sei X eine Zufallsvariable mit $\text{Var } X = 0$. Dann kann man behaupten, dass $\mathbb{P}[X = \mathbb{E}X] = 1$. Das heißt: die Zufallsvariable X ist *fast sicher* konstant. Wir werden das später mit Hilfe der Tschebyscheff-Ungleichung beweisen.

SATZ 9.1.11. Sei $X \sim N(\mu, \sigma^2)$ normalverteilt mit Parametern (μ, σ^2) . Dann gilt $\mathbb{E}X = \mu$ und $\text{Var } X = \sigma^2$.

BEWEIS. Zuerst betrachten wir den Fall $\mu = 0, \sigma^2 = 1$. Die Dichte von X ist dann

$$f_X(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad y \in \mathbb{R}.$$

Für den Erwartungswert erhalten wir

$$\mathbb{E}X = \int_{\mathbb{R}} y f_X(y) dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} y e^{-y^2/2} dy = 0,$$

da die Funktion $y e^{-y^2/2}$ ungerade ist. Für die Varianz erhalten wir

$$\text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \int_{\mathbb{R}} y^2 f_X(y) dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} y^2 e^{-y^2/2} dy.$$

Um dieses Integral zu berechnen, benutzen wir partielle Integration:

$$\text{Var } X = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} y (-e^{-y^2/2})' dy = \frac{1}{\sqrt{2\pi}} \left(-y e^{-y^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} dy \right) = 1.$$

Sei nun $X \sim N(\mu, \sigma^2)$ normalverteilt mit beliebigen Parametern (μ, σ^2) . Dann haben wir die Darstellung $X = \sigma Y + \mu$, wobei $Y \sim N(0, 1)$. Es gilt dann

$$\mathbb{E}X = \mathbb{E}[\sigma Y + \mu] = \mu$$

und

$$\text{Var } X = \text{Var}(\sigma Y + \mu) = \sigma^2 \text{Var } Y = \sigma^2.$$

Für die Standardabweichung ergibt sich übrigens $\sigma(X) = \sigma$ (wobei wir immer annehmen, dass $\sigma > 0$). \square

SATZ 9.1.12. Sei $X \sim \text{Poi}(\lambda)$ Poisson-verteilt mit Parameter $\lambda > 0$. Dann gilt $\mathbb{E}X = \text{Var } X = \lambda$.

BEWEIS. Wir haben bereits gezeigt, dass $\mathbb{E}X = \lambda$. Wir berechnen $\text{Var } X$. Betrachte zunächst

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

Für $\mathbb{E}[X^2]$ gilt dann

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

Für die Varianz erhalten wir

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

\square

9.2. Kovarianz und Korrelationskoeffizient

DEFINITION 9.2.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X, Y : \Omega \rightarrow \mathbb{R}$ zwei Zufallsvariablen mit $X, Y \in L^2$. Dann ist die *Kovarianz* von X und Y wie folgt definiert:

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

BEMERKUNG 9.2.2. $\text{Cov}(X, X) = \text{Var } X$.

BEMERKUNG 9.2.3. Wir zeigen, dass die Kovarianz wohldefiniert ist. Seien $X, Y \in L^2$. Nach Lemma 9.1.2 existieren $\mathbb{E}X$ und $\mathbb{E}Y$. Außerdem gilt dann

$$\tilde{X} := X - \mathbb{E}X \in L^2 \text{ und } \tilde{Y} := Y - \mathbb{E}Y \in L^2.$$

Wir zeigen, dass $\mathbb{E}[\tilde{X}\tilde{Y}]$ existiert. Es gilt

$$|\tilde{X}\tilde{Y}| \leq \frac{\tilde{X}^2 + \tilde{Y}^2}{2}.$$

Daraus folgt, dass

$$\mathbb{E}[|\tilde{X}\tilde{Y}|] \leq \frac{1}{2} (\mathbb{E}[\tilde{X}^2] + \mathbb{E}[\tilde{Y}^2]) < \infty.$$

Somit ist die Zufallsvariable $\tilde{X}\tilde{Y} = (X - \mathbb{E}X)(Y - \mathbb{E}Y)$ integrierbar und die Kovarianz existiert. \square

SATZ 9.2.4. Seien $X, Y \in L^2$. Dann gilt:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

BEWEIS. Wir benutzen die Linearität des Erwartungswerts:

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\ &= \mathbb{E}[XY - X \cdot \mathbb{E}Y - Y \cdot \mathbb{E}X + (\mathbb{E}X)(\mathbb{E}Y)] \\ &= \mathbb{E}[XY] - \mathbb{E}[\mathbb{E}Y \cdot X] - \mathbb{E}[\mathbb{E}X \cdot Y] + \mathbb{E}[(\mathbb{E}X)(\mathbb{E}Y)] \\ &= \mathbb{E}[XY] - \mathbb{E}Y \cdot \mathbb{E}X - \mathbb{E}X \cdot \mathbb{E}Y + (\mathbb{E}X)(\mathbb{E}Y) \\ &= \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]. \end{aligned}$$

Dabei haben wir $\mathbb{E}X$ und $\mathbb{E}Y$ wie Konstanten behandelt. \square

BEMERKUNG 9.2.5. Für die Berechnung der Kovarianz kann man folgende Formeln benutzen:

(1) Für X, Y diskret:

$$\mathbb{E}[XY] = \sum_{\substack{s \in \text{Im } X \\ t \in \text{Im } Y}} st \cdot \mathbb{P}[X = s, Y = t].$$

(2) Für X, Y absolut stetig mit gemeinsamer Dichte $f_{X,Y}$:

$$\mathbb{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} st \cdot f_{X,Y}(s, t) ds dt.$$

SATZ 9.2.6. Seien $X, Y \in L^2$ Zufallsvariablen und $a, b, c, d \in \mathbb{R}$ Konstanten. Dann gilt:

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y).$$

BEWEIS. Übungsaufgabe. \square

Für die Kovarianz gelten ähnliche Rechenregeln wie für das Skalarprodukt.

SATZ 9.2.7. Seien $X_1, \dots, X_n, Y_1, \dots, Y_m \in L^2$ Zufallsvariablen und $a_1, \dots, a_n, b_1, \dots, b_m \in \mathbb{R}$ Konstanten. Dann gilt:

$$\text{Cov}(a_1X_1 + \dots + a_nX_n, b_1Y_1 + \dots + b_mY_m) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j).$$

BEWEIS. Übungsaufgabe. \square

SATZ 9.2.8 (Cauchy-Schwarz-Ungleichung). Seien $X, Y \in L^2$ Zufallsvariablen. Dann gilt

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

BEMERKUNG 9.2.9. Zum Vergleich: Für zwei Vektoren $x, y \in \mathbb{R}^d$ gilt die Ungleichung

$$|\langle x, y \rangle| = |x| \cdot |y| \cdot |\cos(\phi)| \leq |x||y| = \sqrt{\langle x, x \rangle} \sqrt{\langle y, y \rangle}.$$

BEWEIS VON SATZ 9.2.8. Für beliebige Konstante $a \in \mathbb{R}$ gilt

$$0 \leq \mathbb{E}[(aX + Y)^2] = \mathbb{E}[a^2X^2 + 2aXY + Y^2] = a^2\mathbb{E}[X^2] + 2a\mathbb{E}[XY] + \mathbb{E}[Y^2].$$

Die rechte Seite ist eine quadratische Funktion von a . Diese Funktion ist für alle a nichtnegativ. Die Diskriminante D muss somit ≤ 0 sein:

$$D = 4(\mathbb{E}[XY])^2 - 4\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2] \leq 0.$$

Das beweist die Behauptung. □

KOROLLAR 9.2.10. Für beliebige Zufallsvariablen $X, Y \in L^2$ gilt

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var } X} \sqrt{\text{Var } Y}.$$

BEWEIS. Betrachte die zentrierten Zufallsvariablen $\tilde{X} := X - \mathbb{E}X \in L^2$ und $\tilde{Y} := Y - \mathbb{E}Y \in L^2$. Nun wenden wir die Cauchy-Schwarz-Ungleichung auf die Zufallsvariablen \tilde{X} und \tilde{Y} an:

$$\text{Cov}(X, Y) = \mathbb{E}[\tilde{X}\tilde{Y}] \leq \sqrt{\mathbb{E}[\tilde{X}^2]} \sqrt{\mathbb{E}[\tilde{Y}^2]} = \sqrt{\text{Var } X} \sqrt{\text{Var } Y}.$$

□

DEFINITION 9.2.11. Seien $X, Y \in L^2$ zwei Zufallsvariablen. Der *Korrelationskoeffizient* von X und Y ist

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X} \sqrt{\text{Var } Y}}.$$

Dabei nehmen wir an, dass X und Y nicht fast sicher konstant sind, so dass $\text{Var } X \neq 0$ und $\text{Var } Y \neq 0$.

BEMERKUNG 9.2.12. Aus Korollar 9.2.10 folgt, dass $|\rho(X, Y)| \leq 1$. Der Korrelationskoeffizient ist ein Maß für den linearen Zusammenhang zwischen X und Y . Später werden wir zeigen, dass der Korrelationskoeffizient genau dann ± 1 ist, wenn zwischen X und Y ein linearer Zusammenhang besteht, d.h., wenn $Y = aX + b$.

BEMERKUNG 9.2.13. Seien $a, b, c, d \in \mathbb{R}$ und $a, c > 0$. Dann gilt:

$$\rho(aX + b, cY + d) = \rho(X, Y).$$

BEMERKUNG 9.2.14. $\rho(X, X) = 1$, $\rho(X, -X) = -1$.

DEFINITION 9.2.15. Zwei Zufallsvariablen $X, Y \in L^2$ heißen unkorreliert, wenn $\text{Cov}(X, Y) = 0$ (und somit auch $\rho(X, Y) = 0$).

SATZ 9.2.16. Seien $X, Y \in L^2$ unabhängig. Dann sind X und Y unkorreliert, d.h. $\text{Cov}(X, Y) = 0$.

BEWEIS. Für unabhängige Zufallsvariablen gilt $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$. Wir erhalten

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0.$$

Somit sind X und Y unkorreliert. □

BEISPIEL 9.2.17. Unabhängige Zufallsvariablen sind unkorreliert. Die Umkehrung dieser Aussage gilt jedoch nicht. Sei (X, Y) ein Vektor, der gleichverteilt auf dem Einheitskreis

$$A = \{(t, s) : t^2 + s^2 < 1\}$$

ist. Wir behaupten, dass $\text{Cov}(X, Y) = 0$ und dass dennoch X und Y abhängig sind.

LÖSUNG. Die gemeinsame Dichte von (X, Y) ist

$$f_{X,Y}(t, s) = \frac{1}{\pi} \mathbb{1}_A(t, s).$$

Die Randdichten (d.h. die Dichten von X und Y) sind

$$f_X(t) = f_Y(t) = \frac{2}{\pi} \sqrt{1-t^2} \mathbb{1}_{|t|<1}.$$

Wegen Symmetrie sind die Erwartungswerte von X und Y gleich 0:

$$\mathbb{E}X = \mathbb{E}Y = \frac{2}{\pi} \int_{-1}^1 t \sqrt{1-t^2} dt = 0.$$

Nun berechnen wir $\mathbb{E}[XY]$:

$$\mathbb{E}[XY] = \int_A ts f_{X,Y}(t, s) dt ds = \frac{1}{\pi} \int_{\mathbb{R}^2} ts \mathbb{1}_A(t, s) dt ds = 0.$$

Dabei ist das Integral gleich 0, da sich die Funktion $ts \mathbb{1}_A(t, s)$ bei der Transformation $(t, s) \mapsto (-t, s)$ in $-ts \mathbb{1}_A(t, s)$ verwandelt. Somit sind X und Y unkorreliert: $\text{Cov}(X, Y) = 0$.

Und doch sind X, Y abhängig, denn wären X und Y unabhängig, dann müsste die gemeinsame Dichte $f_{X,Y}(t, s)$ mit dem Produkt der Randdichten $f_X(t)f_Y(s)$ für alle (t, s) außerhalb einer Menge vom Lebesgue-Maß 0 übereinstimmen. Dies ist jedoch nicht der Fall, denn

$$f_{X,Y}(t, s) = \frac{1}{\pi} \mathbb{1}_A(t, s) \neq \left(\frac{2}{\pi}\right)^2 \sqrt{1-t^2} \sqrt{1-s^2} \mathbb{1}_{|t|<1, |s|<1} = f_X(t) \cdot f_Y(s)$$

zum Beispiel für alle (t, s) mit $|t| < 1, |s| < 1$ und $t^2 + s^2 > 1$. Somit sind X und Y unkorreliert, aber abhängig.

SATZ 9.2.18. Für beliebige Zufallsvariablen $X, Y \in L^2$ gilt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y).$$

BEWEIS. Mit $\tilde{X} := X - \mathbb{E}X$ und $\tilde{Y} := Y - \mathbb{E}Y$ gilt

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X - \mathbb{E}X) + (Y - \mathbb{E}Y)]^2 \\ &= \mathbb{E}[(\tilde{X} + \tilde{Y})^2] \\ &= \mathbb{E}[\tilde{X}^2] + 2\mathbb{E}[\tilde{X}\tilde{Y}] + \mathbb{E}[\tilde{Y}^2] \\ &= \text{Var}(X) + 2 \text{Cov}(X, Y) + \text{Var}(Y). \end{aligned}$$

□

KOROLLAR 9.2.19. Für unabhängige Zufallsvariablen $X, Y \in L^2$ gilt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

BEMERKUNG 9.2.20. Zum Vergleich: Für beliebige Zufallsvariablen $X, Y \in L^1$ gilt

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

BEMERKUNG 9.2.21. Für beliebige $X_1, \dots, X_n \in L^2$ gilt

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

Für unabhängige $X_1, \dots, X_n \in L^2$ gilt sogar

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

BEISPIEL 9.2.22. Sei $X \sim \text{Bin}(n, p)$ binomialverteilt mit Parametern n, p . Berechne $\text{Var} X$.

LÖSUNG. Wir können X als die Anzahl der Erfolge in einem n -fachen Bernoulli-Experiment auffassen. Wir haben also die Darstellung $X = X_1 + \dots + X_n$, wobei

$$X_i = \begin{cases} 1, & \text{Erfolg beim } i\text{-ten Experiment,} \\ 0, & \text{sonst,} \end{cases} \quad i = 1, \dots, n.$$

Dabei sind X_1, \dots, X_n unabhängig mit $\mathbb{P}[X_i = 1] = p$ und $\mathbb{P}[X_i = 0] = 1 - p$. Für jedes X_i gilt

$$\mathbb{E}[X_i] = 1 \cdot p + 0 \cdot (1 - p) = p$$

und

$$\text{Var} X_i = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \mathbb{E}[X_i] - (\mathbb{E}[X_i])^2 = p(1 - p).$$

Für die Varianz von X erhalten wir (wegen der Unabhängigkeit von X_1, \dots, X_n)

$$\text{Var} X = \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = np(1 - p).$$

BEMERKUNG 9.2.23. Die maximale Varianz wird angenommen, wenn $p = \frac{1}{2}$. Die minimale Varianz tritt ein, wenn $p = 0$ oder $p = 1$. (In diesem Fall ist X konstant und $\text{Var} X = 0$).

Im nächsten Satz zeigen wir, dass der Korrelationskoeffizient genau dann gleich ± 1 ist, wenn es einen linearen Zusammenhang zwischen X und Y gibt. Dabei entspricht der Wert $+1$ einem positiven Zusammenhang und -1 einem negativen Zusammenhang.

SATZ 9.2.24. Seien $X, Y : \Omega \rightarrow \mathbb{R}$ zwei Zufallsvariablen mit $X, Y \in L^2$ und $\text{Var} X \neq 0$, $\text{Var} Y \neq 0$. Es gilt:

- (1) $\rho(X, Y) = +1 \Rightarrow$ es existieren $a > 0$ und $b \in \mathbb{R}$ mit $\mathbb{P}[Y = aX + b] = 1$.
- (2) $\rho(X, Y) = -1 \Rightarrow$ es existieren $a < 0$ und $b \in \mathbb{R}$ mit $\mathbb{P}[Y = aX + b] = 1$.

BEWEIS VON (1). Sei $\rho(X, Y) = 1$. Definiere

$$a = \frac{\sqrt{\text{Var} Y}}{\sqrt{\text{Var} X}} > 0 \text{ und } b = \mathbb{E}Y - a\mathbb{E}X \in \mathbb{R}.$$

Definiere die Zufallsvariable $\Delta = Y - (aX + b)$. Wir zeigen, dass $\mathbb{P}[\Delta = 0] = 1$. Es gilt

$$\Delta = (Y - \mathbb{E}Y) - a(X - \mathbb{E}X) = \tilde{Y} - a\tilde{X}.$$

Somit ist $\mathbb{E}[\Delta] = \mathbb{E}\tilde{Y} - a\mathbb{E}\tilde{X} = 0$. Weiterhin gilt

$$\begin{aligned}\text{Var } \Delta &= \mathbb{E}[\Delta^2] \\ &= \mathbb{E}[\tilde{Y}^2 - 2a\tilde{X}\tilde{Y} + a^2\tilde{X}^2] \\ &= \text{Var } Y - 2a \text{Cov}(X, Y) + a^2 \text{Var } X \\ &= \text{Var } Y - 2 \frac{\sqrt{\text{Var } Y}}{\sqrt{\text{Var } X}} \cdot \sqrt{\text{Var } X} \cdot \sqrt{\text{Var } Y} + \frac{\text{Var } Y}{\text{Var } X} \cdot \text{Var } X \\ &= 0.\end{aligned}$$

Somit gilt $\mathbb{P}[\Delta = \mathbb{E}\Delta] = 1$, also $\mathbb{P}[\Delta = 0] = 1$. □

KAPITEL 10

Gesetz der großen Zahlen

10.1. Zwei Beispiele

BEISPIEL 10.1.1. Wir betrachten ein Bernoulli-Experiment, das unendlich oft wiederholt wird. Die Wahrscheinlichkeit für einen Erfolg sei p . Die Zufallsvariable, die den Ausgang des i -ten Experiments beschreibt, ist:

$$X_i = \begin{cases} 1, & \text{falls Experiment } i \text{ Erfolg,} \\ 0, & \text{sonst.} \end{cases}$$

Die Anzahl der Erfolge in den ersten n Experimenten ist $S_n = X_1 + \dots + X_n$. Dann kann man erwarten, dass

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{} p.$$

Dies ist ein Spezialfall des sogenannten Gesetzes der großen Zahlen.

BEISPIEL 10.1.2. Ein fairer Würfel wird unendlich oft geworfen. Die Zufallsvariable X_i sei die Augenzahl bei Wurf i . Mit $S_n = X_1 + \dots + X_n$ bezeichnen wir die Summe der Augenzahlen der ersten n Würfe. Der Erwartungswert von X_i ist

$$\mathbb{E}[X_i] = \frac{1}{6} \cdot (1 + \dots + 6) = 3.5.$$

Man kann dann erwarten, dass

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{} 3.5.$$

Dies ist ein anderer Spezialfall des Gesetzes der großen Zahlen.

In beiden Aussagen geht es um Konvergenz von Folgen von Zufallsvariablen. Damit wir das Gesetz der großen Zahlen formulieren können, müssen wir definieren, was wir unter einer solchen Konvergenz verstehen. Es stellt sich heraus, dass es viele verschiedene Konvergenzbegriffe für Folgen von Zufallsvariablen gibt. Vier dieser Begriffe (*Konvergenz in Wahrscheinlichkeit*, *Konvergenz in L^p* , *fast sichere Konvergenz*, *Konvergenz in Verteilung*) werden im Folgenden behandelt.

10.2. Konvergenz in Wahrscheinlichkeit und L^2 -Konvergenz

DEFINITION 10.2.1. Eine Folge von Zufallsvariablen $Z_1, Z_2, \dots : \Omega \rightarrow \mathbb{R}$ konvergiert *in Wahrscheinlichkeit* (oder *stochastisch*) gegen eine Zufallsvariable $Z : \Omega \rightarrow \mathbb{R}$, wenn für alle $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}[\{\omega \in \Omega : |Z_n(\omega) - Z(\omega)| > \varepsilon\}] = 0.$$

In einer vereinfachten Schreibweise:

$$\lim_{n \rightarrow \infty} \mathbb{P}[|Z_n - Z| > \varepsilon] = 0.$$

BEZEICHNUNG: $Z_n \xrightarrow[n \rightarrow \infty]{P} Z$.

DEFINITION 10.2.2. Eine Folge von Zufallsvariablen $Z_1, Z_2, \dots : \Omega \rightarrow \mathbb{R}$ konvergiert in L^2 (oder *im quadratischen Mittel*) gegen eine Zufallsvariable $Z : \Omega \rightarrow \mathbb{R}$, wenn $Z_n \in L^2$, $Z \in L^2$ und

$$\lim_{n \rightarrow \infty} \mathbb{E}[(Z_n - Z)^2] = 0.$$

BEZEICHNUNG: $Z_n \xrightarrow[n \rightarrow \infty]{L^2} Z$.

Im Folgenden werden wir zeigen, dass aus der L^2 -Konvergenz die stochastische Konvergenz folgt, jedoch nicht umgekehrt.

10.3. Ungleichungen von Markow und Tschebyschew

LEMMA 10.3.1. Sei $Z \geq 0$ eine Zufallsvariable. Dann gilt für alle $a > 0$:

$$\mathbb{P}[Z \geq a] \leq \frac{\mathbb{E}Z}{a}.$$

BEWEIS. Für die Zufallsvariable Z gilt:

$$Z(\omega) \geq \begin{cases} a, & \text{falls } Z(\omega) \geq a, \\ 0, & \text{falls } Z(\omega) < a. \end{cases}$$

Mit anderen Worten, $Z \geq a \cdot \mathbb{1}_{\{Z \geq a\}}$. Für den Erwartungswert von Z gilt somit:

$$\mathbb{E}Z \geq \mathbb{E}[a \cdot \mathbb{1}_{\{Z \geq a\}}] = a \cdot \mathbb{E}[\mathbb{1}_{\{Z \geq a\}}] = a \cdot \mathbb{P}[Z \geq a].$$

Durch Umformen erhält man die Ungleichung im Lemma. □

SATZ 10.3.2 (Tschebyschew-Ungleichung). Sei $X \in L^2$ eine Zufallsvariable. Dann gilt für alle $a > 0$:

$$\mathbb{P}[|X - \mathbb{E}X| \geq a] \leq \frac{\text{Var } X}{a^2}.$$

BEWEIS. Da $Z := (X - \mathbb{E}X)^2 \geq 0$, folgt mit Lemma 10.3.1:

$$\mathbb{P}[(X - \mathbb{E}X)^2 \geq a^2] \leq \frac{\mathbb{E}[(X - \mathbb{E}X)^2]}{a^2} = \frac{\text{Var } X}{a^2}.$$

Die Ungleichung $(X - \mathbb{E}X)^2 \geq a^2$ gilt genau dann, wenn $|X - \mathbb{E}X| \geq a$. Damit erhält man die Tschebyschew-Ungleichung. □

SATZ 10.3.3 (Markow-Ungleichung). Sei $Z \geq 0$ eine Zufallsvariable und $f : [0, \infty) \rightarrow [0, \infty)$ eine monoton steigende Funktion. Dann gilt für alle $a > 0$:

$$\mathbb{P}[Z \geq a] \leq \frac{\mathbb{E}[f(Z)]}{f(a)}.$$

BEWEIS. Da f monoton steigend ist, gilt für die Wahrscheinlichkeit:

$$\mathbb{P}[Z \geq a] = \mathbb{P}[f(Z) \geq f(a)] \leq \frac{\mathbb{E}[f(Z)]}{f(a)}.$$

Die letzte Ungleichung erhält man aus Lemma 10.3.1, indem man dort $f(Z)$ anstelle von Z einsetzt. \square

BEISPIEL 10.3.4. Mit $f(a) = a^p$, wobei $p > 0$, erhält man die Ungleichung

$$\mathbb{P}[Z \geq a] \leq \frac{\mathbb{E}[Z^p]}{a^p}.$$

SATZ 10.3.5. Seien Z_1, Z_2, \dots und Z Zufallsvariablen mit $Z_n \xrightarrow{L^2} Z$. Dann gilt: $Z_n \xrightarrow{P} Z$.

BEWEIS. Aus $Z_n \xrightarrow{L^2} Z$ folgt mit der Definition der L^2 -Konvergenz, dass

$$\lim_{n \rightarrow \infty} \mathbb{E}[(Z_n - Z)^2] \xrightarrow{n \rightarrow \infty} 0.$$

Sei $\varepsilon > 0$. Mit dem obigen Beispiel (mit $p = 2$) erhalten wir

$$\mathbb{P}[|Z_n - Z| \geq \varepsilon] \leq \frac{\mathbb{E}[(Z_n - Z)^2]}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Wir haben gezeigt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}[|Z_n - Z| \geq \varepsilon] = 0$$

und somit $Z_n \xrightarrow{P} Z$. \square

10.4. Schwaches Gesetz der großen Zahlen

SATZ 10.4.1 (Schwaches Gesetz der großen Zahlen). Seien $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ unabhängige Zufallsvariablen mit $\mathbb{E}X_i = \mu$ und $\text{Var } X_i = \sigma^2$ für alle $i \in \mathbb{N}$. Dann gilt:

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{L^2, P} \mu$$

Das heißt, das arithmetische Mittel der Zufallsvariablen X_1, \dots, X_n konvergiert in L^2 bzw. in Wahrscheinlichkeit gegen den Erwartungswert.

BEWEIS. Sei $S_n = X_1 + \dots + X_n$ die Summe der ersten n Zufallsvariablen. Dann gilt für den Erwartungswert und die Varianz der Summe

$$\mathbb{E}S_n = n\mu \text{ und } \text{Var } S_n = n\sigma^2,$$

wobei im Fall der Varianz benutzt wird, dass die Zufallsvariablen unabhängig sind. Wir zeigen zuerst die L^2 -Konvergenz:

$$\mathbb{E} \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] = \mathbb{E} \left[\left(\frac{S_n - \mathbb{E}S_n}{n} \right)^2 \right] = \frac{1}{n^2} \cdot \text{Var } S_n = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Daraus folgt $\frac{S_n}{n} \xrightarrow{L^2} \mu$. Und mit Satz 10.3.5 folgt nun auch die stochastische Konvergenz:
 $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{P} \mu$. □

BEISPIEL 10.4.2. Ein fairer Würfel wird unendlich oft geworfen. Es sei S_n die Augensumme nach n Würfeln, also $S_n = X_1 + \dots + X_n$, wobei die Zufallsvariable X_i die Augenzahl im i -ten Wurf beschreibt. Der Erwartungswert eines Wurfes ist $\mu = 3.5$. Mit Satz 10.4.1 folgt nun

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{P} 3.5.$$

Man kann entsprechend der Definition der stochastischen Konvergenz z.B. mit $\varepsilon = 0.01$ das folgende Ergebnis erhalten:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[3,49 < \frac{S_n}{n} < 3,51 \right] = 1.$$

Die Wahrscheinlichkeit, dass die mittlere Augenzahl außerhalb des Intervalls $(3,49, 3,51)$ liegt, geht gegen 0, wenn die Anzahl der Würfe gegen ∞ geht.

BEMERKUNG 10.4.3. Das Wort “schwach” in der Bezeichnung “schwaches Gesetz der großen Zahlen” bezieht sich auf die Art der Konvergenz (in Wahrscheinlichkeit oder L^2). Im folgenden werden wir stärkere Versionen des Gesetzes der großen Zahlen beweisen.

10.5. Fast sichere Konvergenz

DEFINITION 10.5.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Ein Ereignis $A \in \mathcal{F}$ heißt ein *fast sicheres* Ereignis, wenn $\mathbb{P}[A] = 1$.

DEFINITION 10.5.2. Ein Ereignis $B \in \mathcal{F}$ heißt ein *fast unmögliches* Ereignis (oder ein *Nullereignis*), wenn $\mathbb{P}[B] = 0$.

BEISPIEL 10.5.3. Das Ereignis Ω ist fast sicher (und in Wirklichkeit, sogar sicher). Das Ereignis \emptyset ist fast unmöglich (und in Wirklichkeit, sogar unmöglich). Das Komplement einer fast sicheren Ereignisses ist fast unmöglich und umgekehrt, das Komplement eines fast unmöglichen Ereignisses ist fast sicher.

BEMERKUNG 10.5.4. Man geht davon aus, dass ein Nullereignis in der Praxis nicht beobachtet werden kann. Es kann zwar sein, dass es im Wahrscheinlichkeitsraum gewisse Ausgänge gibt, die zum Eintreten dieses Ereignisses führen, diese bilden allerdings eine so “kleine” Menge, dass man diese Ausgänge nicht beobachten kann, egal, wie oft man das Experiment wiederholt. Analog geht man davon aus, dass ein fast sicheres Ereignis bei jeder Ausführung des Experiments beobachtet wird.

BEMERKUNG 10.5.5. Eine Vereinigung von abzählbar vielen Nullereignissen ist ein Nullereignis (auch dann, wenn die Ereignisse nicht disjunkt sind). Ein Schnitt von höchstens abzählbar vielen fast sicheren Ereignissen ist wieder ein fast sicheres Ereignis. Diese Aussagen gelten allerdings nicht für überabzählbare Schnitte bzw. Vereinigungen.

BEISPIEL 10.5.6. Man betrachte einen idealen Zufallsgenerator, der eine im Intervall $[0, 1]$ gleichverteilte Zahl erzeugt. Die Grundmenge dieses Experiments ist $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}_{[0,1]}$ ist die Borel- σ -Algebra auf $[0, 1]$ und $\mathbb{P} = \lambda$ ist das Lebesgue-Maß.

Es sei $B = \{q_1, q_2, \dots\} \subset [0, 1]$ ein Ereignis, das aus abzählbar vielen Ausgängen besteht. Zum Beispiel, kann man das Ereignis

$$B = \mathbb{Q} \cap [0, 1] = \text{“es wurde eine rationale Zahl erzeugt”}$$

betrachten. Für jedes q_i ist $\mathbb{P}[q_i] = 0$ und somit, mit der σ -Additivität,

$$\mathbb{P}[B] = \sum_{i=1}^{\infty} \mathbb{P}[q_i] = 0.$$

Daraus folgt, dass B fast unmöglich ist. Ein idealer Zufallsgenerator erzeugt also nie eine rationale Zahl. Natürlich erfüllt kein realer Zufallsgenerator (z.B. in einem Rechner) diese Anforderung. Im Gegenteil: Ein realer Zufallsgenerator erzeugt nur Zahlen mit einer endlichen Dezimaldarstellung, also nur rationale Zahlen.

Es sei nun $A = [0, 1] \setminus B$ ein Ereignis, das ein Komplement einer abzählbaren Menge ist. Zum Beispiel kann man das Ereignis

$$A = [0, 1] \setminus \mathbb{Q} = \text{“es wurde eine irrationale Zahl erzeugt”}$$

betrachten. Dann gilt für die Wahrscheinlichkeit von A : $\mathbb{P}[A] = 1 - \mathbb{P}[B] = 1 - 0 = 1$. Daraus folgt, dass A fast sicher ist. Ein idealer Zufallsgenerator erzeugt also nur irrationale Zahlen.

DEFINITION 10.5.7. Eine Folge von Zufallsvariablen $Z_1, Z_2, \dots : \Omega \rightarrow \mathbb{R}$ konvergiert *fast sicher* gegen eine Zufallsvariable $Z : \Omega \rightarrow \mathbb{R}$, wenn:

$$\mathbb{P}[\{\omega \in \Omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)\}] = 1.$$

In einer vereinfachten Schreibweise:

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} Z_n = Z \right] = 1.$$

BEZEICHNUNG: $Z_n \xrightarrow{f.s.} Z$ oder $\lim_{n \rightarrow \infty} Z_n = Z$ f.s.

BEMERKUNG 10.5.8. Damit diese Definition Sinn macht, muss man zeigen, dass das Ereignis

$$A := \{\omega \in \Omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)\}$$

messbar ist. Wir zeigen also, dass $A \in \mathcal{F}$. Die Gleichheit $\lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)$ gilt genau dann, wenn für alle $\varepsilon > 0$ ein $k \in \mathbb{N}$ existiert, so dass für alle $n \geq k$:

$$|Z_n(\omega) - Z(\omega)| \leq \varepsilon.$$

Man kann sich auch nur auf Werte $\varepsilon = \frac{1}{m}$ mit $m \in \mathbb{N}$ einschränken. Dann kann man A folgendermaßen beschreiben:

$$\begin{aligned} A &= \left\{ \omega \in \Omega : \forall m \in \mathbb{N} \exists k \in \mathbb{N} \forall n \geq k : |Z_n(\omega) - Z(\omega)| \leq \frac{1}{m} \right\} \\ &= \bigcap_{m \in \mathbb{N}} \bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} \left\{ \omega \in \Omega : |Z_n(\omega) - Z(\omega)| \leq \frac{1}{m} \right\}. \end{aligned}$$

Dabei haben wir \forall durch \bigcap und \exists durch \bigcup ersetzt. Da $|Z_n - Z|$ eine messbare Funktion ist, sind alle Mengen der Form

$$\left\{ \omega \in \Omega : |Z_n(\omega) - Z(\omega)| \leq \frac{1}{m} \right\}$$

messbare Ereignisse. Da die messbaren Ereignisse eine σ -Algebra bilden, sind auch beliebige abzählbare Schnitte und Vereinigungen dieser Mengen messbar, also auch A . Es folgt, dass $A \in \mathcal{F}$.

SATZ 10.5.9. Seien Z_1, Z_2, \dots und Z Zufallsvariablen mit $Z_n \xrightarrow[n \rightarrow \infty]{f.s.} Z$. Dann gilt: $Z_n \xrightarrow[n \rightarrow \infty]{P} Z$.

BEWEIS. Es gelte, dass $Z_n \xrightarrow[n \rightarrow \infty]{f.s.} Z$. Dann hat das Ereignisses

$$A = \{\omega \in \Omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)\}$$

Wahrscheinlichkeit $\mathbb{P}[A] = 1$. Sei $\varepsilon > 0$. Wir definieren Ereignisse

$$A_k = \{\omega \in \Omega : \forall n \geq k : |Z_n(\omega) - Z(\omega)| \leq \varepsilon\}, \quad k \in \mathbb{N}.$$

Es gilt $A \subset \cup_{k \in \mathbb{N}} A_k$. Aus $1 = \mathbb{P}[A] \leq \mathbb{P}[\cup_{k \in \mathbb{N}} A_k]$ folgt, dass

$$\mathbb{P}[\cup_{k \in \mathbb{N}} A_k] = 1.$$

Außerdem gilt $A_1 \subset A_2 \subset A_3 \subset \dots$. Mit der Stetigkeit des Wahrscheinlichkeitsmaßes folgt

$$\lim_{k \rightarrow \infty} \mathbb{P}[A_k] = \mathbb{P}[\cup_{k \in \mathbb{N}} A_k] = 1.$$

Es ist nun

$$\lim_{k \rightarrow \infty} \mathbb{P}[|Z_k - Z| \leq \varepsilon] = \lim_{k \rightarrow \infty} \mathbb{P}[\{\omega \in \Omega : |Z_k(\omega) - Z(\omega)| \leq \varepsilon\}] \geq \lim_{k \rightarrow \infty} \mathbb{P}[A_k] = 1.$$

Also ist $\lim_{k \rightarrow \infty} \mathbb{P}[|Z_k - Z| > \varepsilon] = 0$. Daraus folgt $Z_n \xrightarrow[n \rightarrow \infty]{P} Z$. □

BEISPIEL 10.5.10. Aus $Z_n \xrightarrow[n \rightarrow \infty]{P} Z$ folgt nicht $Z_n \xrightarrow[n \rightarrow \infty]{f.s.} Z$. Das zeigen wir an einem Beispiel. Die Zufallsvariablen Z_1, Z_2, \dots seien unabhängig mit

$$\mathbb{P}[Z_n = 1] = \frac{1}{n^\alpha}, \quad \mathbb{P}[Z_n = 0] = 1 - \frac{1}{n^\alpha}.$$

Dabei sei $\alpha \in (0, 1]$ ein Parameter.

(1) Es gilt $Z_n \xrightarrow[n \rightarrow \infty]{P} 0$, denn für alle $\varepsilon > 0$ gilt

$$\mathbb{P}[|Z_n| > \varepsilon] \leq \mathbb{P}[Z_n = 1] = \frac{1}{n^\alpha} \xrightarrow[n \rightarrow \infty]{} 0.$$

(2) Allerdings gilt nicht $Z_n \xrightarrow[n \rightarrow \infty]{f.s.} 0$. Es ist $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} = \infty$ (da $\alpha \leq 1$). Mit dem Lemma von Borel–Cantelli (Teil 2) folgt

$$\mathbb{P}[Z_n = 1 \text{ für unendlich viele } n] = 1.$$

Daraus folgt $\mathbb{P}[\lim_{n \rightarrow \infty} Z_n \neq 0] = 1$ und deshalb gilt nicht $Z_n \xrightarrow[n \rightarrow \infty]{f.s.} 0$.

10.6. Starkes Gesetz der großen Zahlen: Erste Version

SATZ 10.6.1 (Starkes Gesetz der großen Zahlen: Erste Version). *Seien X_1, X_2, \dots unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathbb{E}[X_n^2] < \infty$ für alle $n \in \mathbb{N}$. Außerdem sei*

$$\sum_{n=1}^{\infty} \frac{\text{Var } X_n}{n^2} < \infty.$$

Für die Summen $S_n = X_1 + \dots + X_n$ gilt dann

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

BEMERKUNG 10.6.2. Spezialfall: Seien X_1, X_2, \dots unabhängige Zufallsvariablen mit $\mathbb{E}X_n = \mu$ und $\text{Var } X_n = \sigma^2$. Dann gilt:

$$\sum_{n=1}^{\infty} \frac{\text{Var } X_n}{n^2} = \sigma^2 \cdot \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Satz 10.6.1 besagt nun, dass $\frac{S_n - n\mu}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$ und daraus folgt, dass

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \mu.$$

Die fast sichere Konvergenz impliziert die stochastische Konvergenz (ist also stärker), daher die Bezeichnung starkes Gesetz der großen Zahlen.

Im Folgenden werden wir Satz 10.6.1 beweisen. Dafür benötigen wir einige Hilfsmittel.

SATZ 10.6.3 (Kolmogorow-Ungleichung). *Seien X_1, X_2, \dots, X_n unabhängige Zufallsvariablen mit Erwartungswert $\mathbb{E}X_i = 0$ und $X_i \in L^2$ für $i = 1, \dots, n$. Sei $S_i = X_1 + \dots + X_i$ die Summe der ersten i Zufallsvariablen. Dann gilt für alle $a > 0$ die Ungleichung*

$$\mathbb{P}[\max\{|S_1|, |S_2|, \dots, |S_n|\} \geq a] \leq \frac{\text{Var } S_n}{a^2}.$$

BEMERKUNG 10.6.4. Die Tschebyschew-Ungleichung ist schwächer. Mit der kann man lediglich eine der Summen abschätzen, z.B.

$$\mathbb{P}[|S_n| \geq a] \leq \frac{\text{Var } S_n}{a^2}.$$

BEWEIS VON SATZ 10.6.3. SCHRITT 1. Wir definieren das Ereignis:

$$A := \left\{ \max_{k=1, \dots, n} |S_k| \geq a \right\}$$

und, für $k = 1, \dots, n$, die Ereignisse

$$A_k := \{|S_1| < a, |S_2| < a, \dots, |S_{k-1}| < a, |S_k| \geq a\}$$

= "Erster Austritt aus dem Intervall $(-a, a)$ findet zum Zeitpunkt k statt."

Es gilt $A = \cup_{k=1}^n A_k$. (Ereignis A tritt genau dann ein, wenn es ein k gibt, sodass der erste Austritt aus dem Intervall $(-a, a)$ zum Zeitpunkt k stattfindet). Außerdem sind A_1, \dots, A_n

disjunkt. (Wenn der erste Austritt zum Zeitpunkt k stattfindet, kann er nicht zu einem anderen Zeitpunkt stattfinden). Also ist

$$\mathbb{P}[A] = \sum_{k=1}^n \mathbb{P}[A_k].$$

SCHRITT 2. Die Varianz von S_n ist: $\text{Var } S_n = \mathbb{E}[S_n^2] - (\mathbb{E}S_n)^2 = \mathbb{E}[S_n^2]$, denn $\mathbb{E}S_n = \mathbb{E}[X_1 + \dots + X_n] = \mathbb{E}X_1 + \dots + \mathbb{E}X_n = 0$. Wir betrachten also $\mathbb{E}[S_n^2]$:

$$\mathbb{E}[S_n^2] \geq \mathbb{E}[S_n^2 \cdot \mathbb{1}_A] = \mathbb{E} \left[\sum_{k=1}^n S_n^2 \cdot \mathbb{1}_{A_k} \right] = \sum_{k=1}^n \mathbb{E}[S_n^2 \cdot \mathbb{1}_{A_k}].$$

Für die Summanden auf der rechten Seite gilt

$$\begin{aligned} \mathbb{E}[S_n^2 \cdot \mathbb{1}_{A_k}] &= \mathbb{E}[(S_k + S_n - S_k)^2 \cdot \mathbb{1}_{A_k}] \\ &= \mathbb{E}[S_k^2 \cdot \mathbb{1}_{A_k}] + 2 \cdot \mathbb{E}[S_k \cdot (S_n - S_k) \cdot \mathbb{1}_{A_k}] + \mathbb{E}[(S_n - S_k)^2 \cdot \mathbb{1}_{A_k}] \\ &= (1) + (2) + (3). \end{aligned}$$

SCHRITT 3. Die drei Terme auf der rechten Seite kann man folgendermaßen abschätzen. Der erste Term:

$$(1) = \mathbb{E}[S_k^2 \cdot \mathbb{1}_{A_k}] \geq \mathbb{E}[a^2 \cdot \mathbb{1}_{A_k}] = a^2 \cdot \mathbb{P}[A_k],$$

denn wenn A_k eintritt, ist $\mathbb{1}_{A_k} = 1$ und $|S_k| \geq a$.

Der zweite Term:

$$(2) = 2 \cdot \mathbb{E}[S_k \cdot \mathbb{1}_{A_k} \cdot (S_n - S_k)] = 2 \cdot \mathbb{E}[S_k \cdot \mathbb{1}_{A_k}] \cdot \mathbb{E}[S_n - S_k] = 0.$$

Die Umformung gilt, da die Zufallsvariable $S_k \cdot \mathbb{1}_{A_k}$ nur von X_1, \dots, X_k und die Zufallsvariable $(S_n - S_k)$ nur von X_{k+1}, \dots, X_n abhängt, also sind diese zwei Zufallsvariablen unabhängig. Außerdem gilt $\mathbb{E}[S_n - S_k] = 0$, da $\mathbb{E}S_n = \mathbb{E}S_k = 0$.

Schließlich ist der dritte Term trivialerweise nichtnegativ:

$$(3) \geq 0.$$

SCHRITT 4. Somit erhalten wir

$$\mathbb{E}[S_n^2 \cdot \mathbb{1}_{A_k}] = (1) + (2) + (3) \geq a^2 \cdot \mathbb{P}[A_k].$$

Betrachte wir nun wieder die Varianz von S_n :

$$\text{Var } S_n = \mathbb{E}[S_n^2] \geq \sum_{k=1}^n \mathbb{E}[S_n^2 \cdot \mathbb{1}_{A_k}] \geq \sum_{k=1}^n a^2 \cdot \mathbb{P}[A_k].$$

Damit folgt: $\mathbb{P}[A] \leq \frac{\text{Var } S_n}{a^2}$. □

BEWEIS VON SATZ 10.6.1. SCHRITT 1. O.B.d.A. können wir annehmen, dass $\mathbb{E}X_n = 0$ für alle $n \in \mathbb{N}$. Sonst betrachte man die zentrierten Zufallsvariablen

$$\widetilde{X}_n = X_n - \mathbb{E}X_n, \quad \widetilde{S}_n = \widetilde{X}_1 + \dots + \widetilde{X}_n.$$

Es gilt dann $\mathbb{E}\widetilde{X}_n = 0$ und

$$\sum_{n=1}^{\infty} \frac{\text{Var } \widetilde{X}_n}{n^2} = \sum_{n=1}^{\infty} \frac{\text{Var } X_n}{n^2} < \infty,$$

denn $\text{Var } \widetilde{X}_n = \text{Var}[X_n - \mathbb{E}X_n] = \text{Var } X_n$, da $\mathbb{E}X_n$ eine Konstante ist. Außerdem ist

$$\frac{\widetilde{S}_n - \mathbb{E}\widetilde{S}_n}{n} = \frac{\widetilde{S}_n}{n} = \frac{S_n - \mathbb{E}S_n}{n}.$$

Somit gilt Satz 10.6.1 für die Zufallsvariablen X_n genau dann, wenn er für die Zufallsvariablen \widetilde{X}_n gilt.

SCHRITT 2. Sei also $\mathbb{E}S_n = 0$ für alle $n \in \mathbb{N}$. Wir zeigen, dass

$$\frac{|S_k|}{k} \xrightarrow[k \rightarrow \infty]{f.s.} 0.$$

Dazu definieren wir die Zufallsvariablen

$$U_n := \max_{k=1, \dots, 2^n} |S_k|, \quad n \in \mathbb{N}.$$

Nun gibt es für alle $k \in \mathbb{N}$ genau ein $n \in \mathbb{N}$ mit $2^{n-1} \leq k < 2^n$. Daraus folgt, dass

$$\frac{|S_k|}{k} \leq \frac{U_n}{k} \leq \frac{U_n}{2^{n-1}} = 2 \cdot \frac{U_n}{2^n},$$

also reicht es zu zeigen, dass

$$\frac{U_n}{2^n} \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

SCHRITT 3. Sei $\varepsilon > 0$ fest. Betrachte das Ereignis $A_n(\varepsilon) = \left\{ \frac{U_n}{2^n} > \varepsilon \right\}$. Wir zeigen nun, dass dieses Ereignis mit Wahrscheinlichkeit 1 nur endlich oft eintritt, also

$$\mathbb{P}[\text{“}A_n(\varepsilon) \text{ tritt für unendlich viele } n \in \mathbb{N} \text{ ein”}] = \mathbb{P}[\limsup_{n \rightarrow \infty} A_n(\varepsilon)] = 0.$$

Dafür benutzen wir den ersten Teil des Lemmas von Borel–Cantelli. Mit der Kolmogorov-Ungleichung erhalten wir

$$\sum_{n=1}^{\infty} \mathbb{P}[A_n(\varepsilon)] = \sum_{n=1}^{\infty} \mathbb{P}[U_n > 2^n \varepsilon] = \sum_{n=1}^{\infty} \mathbb{P} \left[\max_{k=1, \dots, 2^n} |S_k| > 2^n \varepsilon \right] \leq \sum_{n=1}^{\infty} \frac{\text{Var } S_{2^n}}{2^{2n} \cdot \varepsilon^2}.$$

Wir definieren noch $\sigma_k^2 = \text{Var } X_k$. Somit gilt

$$\begin{aligned}
 \sum_{n=1}^{\infty} \mathbb{P}[A_n(\varepsilon)] &\leq \frac{1}{\varepsilon^2} \cdot \sum_{n=1}^{\infty} \frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_{2^n}^2}{2^{2n}} \quad (\text{gilt, da } X_1, X_2, \dots \text{ unabhängig}) \\
 &= \frac{1}{\varepsilon^2} \cdot \sum_{k=1}^{\infty} \sigma_k^2 \cdot \sum_{n: 2^n \geq k} \frac{1}{2^{2n}} \\
 &\leq \frac{1}{\varepsilon^2} \cdot \sum_{k=1}^{\infty} \sigma_k^2 \cdot \frac{1}{k^2} \cdot \sum_{m=1}^{\infty} \frac{1}{4^m} \\
 &= \frac{4}{3} \cdot \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \\
 &< \infty. \quad (\text{gilt nach Voraussetzung})
 \end{aligned}$$

Mit dem Lemma von Borel–Cantelli folgt, dass $\mathbb{P}[\limsup_{n \rightarrow \infty} A_n(\varepsilon)] = 0$.

SCHRITT 4. Setzt man nun $\varepsilon = \frac{1}{m}$ mit $m \in \mathbb{N}$, so gilt für alle $m \in \mathbb{N}$:

$$P \left[\limsup_{n \rightarrow \infty} \frac{U_n}{2^n} > \frac{1}{m} \right] = 0.$$

Da eine Vereinigung von abzählbar vielen Nullereignissen wieder ein Nullereignis ist, gilt somit auch

$$\mathbb{P} \left[\limsup_{n \rightarrow \infty} \frac{U_n}{2^n} > 0 \right] = 0.$$

Daraus folgt:

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \frac{U_n}{2^n} = 0 \right] = 1$$

und somit $\frac{U_n}{2^n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$. □

10.7. Starkes Gesetz der großen Zahlen: Zweite Version

In der ersten Version des starken Gesetzes der großen Zahlen haben wir vorausgesetzt, dass die Varianzen der Zufallsvariablen endlich sein sollen. In der Behauptung des Satzes, nämlich

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$$

taucht die Varianz allerdings nicht auf. Man kann sich deshalb fragen, ob die Voraussetzung der Endlichkeit der Varianz nicht überflüssig ist. Das ist in der Tat der Fall, wie der nächste Satz zeigt.

SATZ 10.7.1 (Starkes Gesetz der großen Zahlen: Zweite Version). *Seien X_1, X_2, \dots unabhängige und identisch verteilte Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Die Zufallsvariablen seien außerdem integrierbar, d.h. $\mathbb{E}|X_1| < \infty$. Dann gilt:*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E}[X_1]$$

Das heißt, das arithmetische Mittel der Zufallsvariablen konvergiert fast sicher gegen den Erwartungswert.

BEMERKUNG 10.7.2. Die Zufallsvariablen X_1, X_2, \dots heißen identisch verteilt, wenn ihre Verteilungsfunktionen gleich sind, d.h.

$$F_{X_1}(t) = F_{X_2}(t) = \dots \quad \text{für alle } t \in \mathbb{R}.$$

Identisch verteilte Zufallsvariablen haben den gleichen Erwartungswert: $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \dots$

Für den Beweis von Satz 10.7.1 benötigen wir mehrere Lemmata.

LEMMA 10.7.3. Seien $X_n, Y_n, X, Y : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen mit $X_n \xrightarrow[n \rightarrow \infty]{f.s.} X$ und $Y_n \xrightarrow[n \rightarrow \infty]{f.s.} Y$. Dann gilt:

$$\begin{aligned} X_n + Y_n &\xrightarrow[n \rightarrow \infty]{f.s.} X + Y, \\ X_n \cdot Y_n &\xrightarrow[n \rightarrow \infty]{f.s.} X \cdot Y, \\ a \cdot X_n &\xrightarrow[n \rightarrow \infty]{f.s.} a \cdot X, \quad a \in \mathbb{R}. \end{aligned}$$

BEWEIS. Wir beweisen nur die erste Aussage. Dazu definieren wir zwei Ereignisse:

$$\begin{aligned} A &:= \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}, \\ B &:= \{\omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)\}. \end{aligned}$$

Es gilt mit der Definition der fast sicheren Konvergenz:

$$X_n \xrightarrow[n \rightarrow \infty]{f.s.} X \Rightarrow \mathbb{P}[A] = 1 \quad \text{und} \quad Y_n \xrightarrow[n \rightarrow \infty]{f.s.} Y \Rightarrow \mathbb{P}[B] = 1.$$

Daraus folgt:

$$\mathbb{P}[A^c] = 0 \quad \text{und} \quad \mathbb{P}[B^c] = 0.$$

Für Ausgänge aus dem Schnitt von A und B , also $\omega \in A \cap B$, gilt:

$$\lim_{n \rightarrow \infty} (X_n(\omega) + Y_n(\omega)) = X(\omega) + Y(\omega).$$

Für die Wahrscheinlichkeit des Ereignisses $A \cap B$ gilt:

$$\mathbb{P}[A \cap B] = 1 - \mathbb{P}[(A \cap B)^c] = 1 - \mathbb{P}[A^c \cup B^c] \geq 1 - \mathbb{P}[A^c] - \mathbb{P}[B^c] = 1,$$

wobei wir im zweiten Schritt die Regeln von de Morgan und im dritten Schritt die Subadditivität benutzt haben. Das Ereignis $A \cap B$ tritt also fast sicher ein. Damit gilt:

$$X_n + Y_n \xrightarrow[n \rightarrow \infty]{f.s.} X + Y.$$

Die anderen Aussagen werden analog bewiesen. □

LEMMA 10.7.4. Sei $Z \geq 0$ eine Zufallsvariable. Dann gilt:

$$\sum_{n=1}^{\infty} \mathbb{P}[Z \geq n] \leq \mathbb{E}Z \leq \sum_{n=0}^{\infty} \mathbb{P}[Z \geq n].$$

BEWEIS. Wir beweisen nur die Ungleichung $\sum_{n=1}^{\infty} \mathbb{P}[Z \geq n] \leq \mathbb{E}Z$. Die andere Ungleichung kann man analog beweisen. Sei $p_n = \mathbb{P}[n \leq Z < n+1]$. Es gilt

$$\mathbb{E}Z = \mathbb{E} \left[\sum_{n=0}^{\infty} Z \cdot \mathbb{1}_{n \leq Z < n+1} \right] \geq \mathbb{E} \left[\sum_{n=0}^{\infty} n \cdot \mathbb{1}_{n \leq Z < n+1} \right] = \sum_{n=0}^{\infty} np_n = \sum_{n=1}^{\infty} np_n.$$

Es gilt außerdem

$$\begin{aligned} \mathbb{P}[Z \geq 1] &= p_1 + p_2 + p_3 + p_4 + \dots, \\ \mathbb{P}[Z \geq 2] &= p_2 + p_3 + p_4 + \dots, \\ \mathbb{P}[Z \geq 3] &= p_3 + p_4 + \dots, \\ \mathbb{P}[Z \geq 4] &= p_4 + \dots, \end{aligned}$$

Addiert man diese Ungleichungen, so erhält man $\sum_{n=1}^{\infty} \mathbb{P}[Z \geq n] = \sum_{n=1}^{\infty} np_n$. Kombiniert man beide Resultate, so erhält man die gewünschte Ungleichung. \square

LEMMA 10.7.5. Für alle $k \geq 2$ gilt: $\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{2}{k}$.

BEWEIS. $\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \int_{k-1}^{\infty} \frac{1}{x^2} dx = \frac{1}{k-1} \leq \frac{2}{k}$. \square

LEMMA 10.7.6 (Cesaro). Sei x_1, x_2, \dots eine Folge in reeller Zahlen mit einem endlichen Grenzwert $\lim_{n \rightarrow \infty} x_n = x$. Dann gilt:

$$\lim_{n \rightarrow \infty} \frac{x_1 + \dots + x_n}{n} = x.$$

BEWEIS. SCHRITT 0. Wir können o.B.d.A. annehmen, dass $x = 0$. Sonst kann man $\tilde{x}_n := x_n - x$ betrachten.

SCHRITT 1. Sei $\varepsilon > 0$. Aus $\lim_{n \rightarrow \infty} x_n = 0$ folgt, dass es ein $K \in \mathbb{N}$ gibt, so dass für alle $k \geq K$ gilt:

$$|x_k| < \frac{\varepsilon}{2}.$$

Nachdem K gewählt wurde, können wir ein $N \in \mathbb{N}$ finden, so dass für alle $n \geq N$ gilt:

$$\frac{|x_1 + \dots + x_K|}{n} < \frac{\varepsilon}{2}.$$

Das liegt daran, dass der Zähler konstant ist, während der Nenner beliebig groß gewählt werden kann.

SCHRITT 2. Sei nun $n \geq \max\{K, N\}$. Dann gilt:

$$\begin{aligned} \left| \frac{x_1 + \dots + x_n}{n} \right| &\leq \left| \frac{x_1 + \dots + x_K}{n} \right| + \frac{1}{n} \sum_{k=K+1}^n |x_k| \\ &\leq \frac{\varepsilon}{2} + \frac{1}{n} \sum_{k=K+1}^n \frac{\varepsilon}{2} \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

Da $\varepsilon > 0$ beliebig ist, folgt daraus, dass $\lim_{n \rightarrow \infty} \frac{x_1 + \dots + x_n}{n} = 0$. □

Schließlich können wir das starke Gesetz der großen Zahlen (zweite Version) beweisen.

BEWEIS VON SATZ 10.7.1. Seien $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}|X_1| < \infty$. Wir zeigen, dass dann gilt:

$$\frac{S_n}{n} := \frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E}[X_1].$$

SCHRITT 1. Definiere

$$X'_n := X_n \cdot \mathbb{1}_{|X_n| < n} \text{ und } X''_n := X_n \cdot \mathbb{1}_{|X_n| \geq n},$$

dann ist $X'_n + X''_n = X_n$. Außerdem sei

$$S'_n := X'_1 + \dots + X'_n \text{ und } S''_n := X''_1 + \dots + X''_n.$$

Es gilt:

- (1) X'_1, X'_2, \dots sind unabhängig,
- (2) X''_1, X''_2, \dots sind unabhängig.

SCHRITT 2. Unser Ziel ist es, zu zeigen, dass, dass

$$\frac{S_n}{n} - \mathbb{E}X_1 \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

Es gilt:

$$\frac{S_n}{n} - \mathbb{E}X_1 = \frac{S_n - \mathbb{E}S_n}{n} = \frac{S'_n - \mathbb{E}S'_n}{n} + \frac{S''_n}{n} - \frac{\mathbb{E}S''_n}{n} =: a_n + b_n - c_n$$

wobei a_n, b_n, c_n Zufallsvariablen sind. Nach Lemma 10.7.3 reicht es zu zeigen, dass

$$a_n, b_n, c_n \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

SCHRITT 3. Wir zeigen, dass $a_n = \frac{S'_n - \mathbb{E}S'_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$, indem wir Satz 10.6.1 benutzen.

$$\begin{aligned}
 \sum_{n \geq 1} \frac{\text{Var } X'_n}{n^2} &\leq \sum_{n \geq 1} \frac{\mathbb{E}[X_n'^2]}{n^2} \\
 &= \sum_{n \geq 1} \frac{1}{n^2} \cdot \mathbb{E}[X_n^2 \cdot \mathbb{1}_{|X_n| < n}] \\
 &= \sum_{n \geq 1} \frac{1}{n^2} \cdot \mathbb{E}[X_1^2 \cdot \mathbb{1}_{|X_1| < n}] \quad (\text{gilt, da } X_n \text{ identisch verteilt}) \\
 &= \sum_{n \geq 1} \frac{1}{n^2} \cdot \mathbb{E} \left[X_1^2 \cdot \sum_{k=1}^n \mathbb{1}_{A_k} \right],
 \end{aligned}$$

wobei $A_k = \{k-1 \leq X_1 < k\}$. Nun vertauschen wir die beiden Summen:

$$\begin{aligned}
 \sum_{n \geq 1} \frac{\text{Var } X'_n}{n^2} &\leq \sum_{n \geq 1} \frac{1}{n^2} \cdot \sum_{k=1}^n \mathbb{E}[X_1^2 \cdot \mathbb{1}_{A_k}] \\
 &= \sum_{k \geq 1} \mathbb{E}[X_1^2 \cdot \mathbb{1}_{A_k}] \cdot \sum_{n \geq k} \frac{1}{n^2} \\
 &\leq 2 \cdot \sum_{k \geq 1} \frac{1}{k} \cdot \mathbb{E}[X_1^2 \cdot \mathbb{1}_{A_k}],
 \end{aligned}$$

wobei wir Lemma 10.7.5 benutzt haben: $\sum_{n \geq k} \frac{1}{n^2} \leq \frac{2}{k}$. Es folgt, dass

$$\begin{aligned}
 \sum_{n \geq 1} \frac{\text{Var } X'_n}{n^2} &\leq 2 \cdot \sum_{k \geq 1} \mathbb{E} \left[\frac{|X_1|}{k} \cdot |X_1| \cdot \mathbb{1}_{A_k} \right] \quad (\text{wobei } \frac{|X_1|}{k} \leq 1, \text{ falls } A_k \text{ eintritt}) \\
 &\leq 2 \cdot \sum_{k \geq 1} \mathbb{E}[|X_1| \cdot \mathbb{1}_{A_k}] \\
 &= 2 \cdot \mathbb{E} \left[\sum_{k \geq 1} |X_1| \cdot \mathbb{1}_{A_k} \right] \quad (\text{gilt mit monotoner Konvergenz}) \\
 &= 2 \cdot \mathbb{E}|X_1| \\
 &< \infty \quad (\text{gilt nach Voraussetzung})
 \end{aligned}$$

Mit Satz 10.6.1 folgt nun $a_n \xrightarrow[n \rightarrow \infty]{f.s.} 0$.

SCHRITT 4. Wir zeigen, dass $b_n = \frac{S_n''}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$, indem wir den ersten Teil des Lemmas von Borel–Cantelli benutzen. Es gilt:

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}[X_n'' \neq 0] &= \sum_{n \geq 1} \mathbb{P}[|X_n| \geq n] \\ &= \sum_{n \geq 1} \mathbb{P}[|X_1| \geq n] \quad (\text{gilt, da } X_n \text{ identisch verteilt}) \\ &\leq \mathbb{E}|X_1| \quad (\text{gilt mit Lemma 10.7.4}) \\ &< \infty. \end{aligned}$$

Mit dem Lemma von Borel–Cantelli folgt nun:

$$\mathbb{P}\left[\limsup_{n \rightarrow \infty} \{X_n'' \neq 0\}\right] = \mathbb{P}[X_n'' \neq 0 \text{ für unendlich viele } n] = 0.$$

Daraus folgt:

$$\mathbb{P}[X_n'' \neq 0 \text{ für endlich viele } n] = \mathbb{P}[A] = 1.$$

Für $\omega \in A$ gilt dann, dass $X_n''(\omega) \neq 0$ für endlich viele Werte von n . Also ist $S_n''(\omega)$ konstant ab irgendeinem n und damit gilt:

$$\lim_{n \rightarrow \infty} \frac{S_n''(\omega)}{n} = 0.$$

Da $\mathbb{P}[A] = 1$, folgt $b_n \xrightarrow[n \rightarrow \infty]{f.s.} 0$.

SCHRITT 5. Es bleibt noch zu zeigen, dass $c_n = \frac{\mathbb{E}S_n''}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$. Es gilt:

$$\mathbb{E}X_n'' = \mathbb{E}[X_n \cdot \mathbb{1}_{|X_n| \geq n}] = \mathbb{E}[X_1 \cdot \mathbb{1}_{|X_1| \geq n}] \quad (\text{da } X_i \text{ identisch verteilt}).$$

Nun gilt für alle $\omega \in \Omega$:

$$|X_1(\omega)| \cdot \mathbb{1}_{|X_1(\omega)| \geq n} \xrightarrow[n \rightarrow \infty]{\text{monoton}} 0.$$

Mit dem Satz von der monotonen Konvergenz erhält man:

$$\mathbb{E}[|X_1| \cdot \mathbb{1}_{|X_1| \geq n}] \xrightarrow[n \rightarrow \infty]{} 0.$$

Es folgt:

$$|\mathbb{E}X_n''| = |\mathbb{E}[X_1 \cdot \mathbb{1}_{|X_1| \geq n}]| \leq \mathbb{E}[|X_1| \cdot \mathbb{1}_{|X_1| \geq n}] \xrightarrow[n \rightarrow \infty]{} 0 \quad (\text{da } |\mathbb{E}Z| \leq \mathbb{E}|Z|).$$

Also gilt:

$$\mathbb{E}X_n'' \xrightarrow[n \rightarrow \infty]{} 0.$$

Mit Lemma 10.7.6 folgt dann:

$$c_n = \frac{\mathbb{E}X_1'' + \dots + \mathbb{E}X_n''}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0 \quad (\text{und sogar sicher}).$$

SCHRITT 6. Fügt man die Ergebnisse für a_n, b_n, c_n zusammen, so erhält man:

$$\frac{S_n}{n} - \mathbb{E}X_1 = a_n + b_n + c_n \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

Das ist die gewünschte Konvergenz. □

10.8. Der Fall eines unendlichen Erwartungswerts

In allen Versionen des Gesetzes der großen Zahlen haben wir vorausgesetzt, dass der Erwartungswert existiert. Der nächste Satz zeigt, was passiert, wenn der Erwartungswert nicht existiert. In diesem Fall gilt das Gesetz der großen Zahlen nämlich nicht.

SATZ 10.8.1. *Seien X_1, X_2, \dots unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}|X_1| = +\infty$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt:*

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} \text{ existiert und ist endlich} \right] = 0.$$

BEWEIS. SCHRITT 1. Wir definieren das Ereignis

$$A := \left\{ \omega \in \Omega : L(\omega) := \lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} \text{ existiert und ist endlich} \right\},$$

wobei wieder $S_n = X_1 + \dots + X_n$. Dieses Ereignis ist messbar (Übung). Für $\omega \in A$ gilt nun:

$$\frac{X_n(\omega)}{n} = \frac{S_n(\omega)}{n} - \frac{S_{n-1}(\omega)}{n} = \frac{S_n(\omega)}{n} - \frac{S_{n-1}(\omega)}{n-1} \cdot \frac{n-1}{n} \xrightarrow[n \rightarrow \infty]{} 0,$$

denn

$$\frac{S_n(\omega)}{n} \xrightarrow[n \rightarrow \infty]{} L(\omega), \quad \frac{S_{n-1}(\omega)}{n-1} \xrightarrow[n \rightarrow \infty]{} L(\omega), \quad \frac{n-1}{n} \xrightarrow[n \rightarrow \infty]{} 1.$$

Nun betrachten wir das Ereignis

$$B := \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{X_n(\omega)}{n} = 0 \right\}.$$

Wir haben gezeigt: falls $\omega \in A$, dann ist auch $\omega \in B$. Daraus folgt $A \subset B$.

SCHRITT 2. Als nächstes definieren wir eine Folge von Ereignissen:

$$C_n := \{\omega \in \Omega : |X_n| \geq n\}.$$

Daraus erhalten wir:

$$C = (\limsup_{n \rightarrow \infty} C_n)^c = \{\omega \in \Omega : |X_n(\omega)| \geq n \text{ für endlich viele } n\}.$$

Falls $\omega \in B$, dann ist auch $\omega \in C$. Daraus folgt $A \subset B \subset C$.

SCHRITT 3. Wir wollen zeigen, dass $\mathbb{P}[A] = 0$. Dafür reicht es nun zu zeigen, dass $\mathbb{P}[C] = 0$ bzw. $\mathbb{P}[\limsup C_n] = 1$. Wir benutzen dazu den zweiten Teil des Lemmas von Borel–Cantelli:

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}[C_n] &= \sum_{n \geq 1} \mathbb{P}[|X_n| \geq n] \\ &= \sum_{n \geq 1} \mathbb{P}[|X_1| \geq n] \quad (\text{gilt, da } X_i \text{ identisch verteilt}) \\ &\geq \mathbb{E}|X_1| - 1 \quad (\text{gilt mit Lemma 10.7.4, da } |X_1| \geq 0) \\ &= \infty. \end{aligned}$$

Da die Ereignisse C_n unabhängig sind, folgt mit dem zweiten Teil des Lemma von Borel–Cantelli, dass

$$\mathbb{P}[\limsup_{n \rightarrow \infty} C_n] = 1.$$

Daraus folgt dann $\mathbb{P}[C] = 0$. Da $A \subset C$, folgt auch $\mathbb{P}[A] = 0$. □

BEISPIEL 10.8.2. Seien X_1, X_2, \dots unabhängige, Cauchy-verteilte Zufallsvariablen mit Dichte $f_{X_k}(t) = \frac{1}{\pi} \cdot \frac{1}{1+t^2}$, $t \in \mathbb{R}$. Dann gilt für den Erwartungswert: $\mathbb{E}|X_k| = \infty$. Mit Satz 10.8.1 folgt nun, dass

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \frac{S_n}{n} \text{ existiert und ist endlich} \right] = 0.$$

Das starke Gesetz der großen Zahlen gilt also für Cauchy-verteilte Zufallsvariablen nicht.

10.9. Anwendungen des Gesetzes der großen Zahlen

Die nächsten 3 Beispiele sind Spezialfälle der sogenannten Monte-Carlo-Methode.

Berechnung von π . Man kann die Zahl π mit stochastischen Mitteln berechnen. Diese Methode ist jedoch sehr zeitaufwendig bzw. ungenau. Man erzeuge n Punkte bzw. Zufallsvektoren Z_1, Z_2, \dots, Z_n , die unabhängig und gleichverteilt auf dem Quadrat $[-1, 1]^2$ sind. Nun betrachtet man alle Punkte, die im Einheitskreis liegen, deren Bertag also kleiner 1 ist. Die Anzahl dieser Punkte ist

$$S_n := \sum_{k=1}^n \mathbb{1}_{|Z_k| \leq 1}.$$

Für den Erwartungswert der Indikatorfunktion gilt

$$\mathbb{E} \mathbb{1}_{|Z_1| \leq 1} = \mathbb{P}[|Z_1| \leq 1] = \frac{\lambda(\text{Kreis})}{\lambda(\text{Quadrat})} = \frac{\pi}{4}.$$

Also dürfen wir das starke Gesetz der großen Zahlen benutzen:

$$\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{|Z_k| \leq 1} \xrightarrow[n \rightarrow \infty]{f.s.} \frac{\pi}{4}.$$

Um die Zahl π zu berechnen, zählen wir also die Punkte, die im Kreis liegen und teilen diese Anzahl durch die Anzahl aller Punkte. Bei sehr großem n sollte das Verhältnis die Zahl $\pi/4$

approximieren. Den Fehler dieser Approximation bestimmen wir später mit dem zentralen Grenzwertsatz.

Monte-Carlo-Integration. Man betrachte eine integrierbare Funktion $\varphi : [0, 1]^d \rightarrow \mathbb{R}$. Das Integral der Funktion bezeichnen wir mit

$$I := \int_{[0,1]^d} \varphi(x) dx = \int_0^1 \dots \int_0^1 \varphi(x_1, \dots, x_d) dx_1 \dots dx_d.$$

Für kleines d , etwa $d = 1$ oder 2 , kann man zur Berechnung des Integrals z.B. die folgende Formel benutzen:

$$I \approx \frac{1}{n} \cdot \sum_{k_1=0}^{n-1} \dots \sum_{k_d=0}^{n-1} \varphi\left(\frac{k_1}{n}, \dots, \frac{k_d}{n}\right) \text{ für großes } n.$$

Wenn allerdings die Dimension d groß ist, so besteht die Summe auf der rechten Seite aus n^d Summanden, was sehr viel sein kann. Man denke z.B. an den Fall $d = 100$, wo selbst für $n = 2$ die Anzahl der Summanden $2^{100} > 10^{30}$ ist. In diesem Fall wird die oben beschriebene Methode ineffizient und man kann stattdessen die sogenannte Monte-Carlo Methode benutzen.

Man erzeuge n zufällige Punkte bzw. Zufallsvektoren Z_1, Z_2, \dots, Z_n , die unabhängig und gleichverteilt auf dem "Quader" $[0, 1]^d$ sind. Da die Funktion φ integrierbar ist, gilt

$$\mathbb{E}[\varphi(Z_1)] = \int_{\mathbb{R}^d} \varphi(x) \cdot f_{Z_1}(x) dx = \int_{[0,1]^d} \varphi(x) dx = I.$$

Also können wir das starke Gesetz der großen Zahlen benutzen:

$$\frac{1}{n} \sum_{k=1}^n \varphi(Z_k) \xrightarrow[n \rightarrow \infty]{f.s.} I.$$

Man kann also das Integral I durch das arithmetische Mittel $\frac{1}{n} \sum_{k=1}^n \varphi(Z_k)$ approximieren. Später werden wir mithilfe des zentralen Grenzwertsatzes den Fehler dieser Approximation bestimmen.

Empirische Definition der Wahrscheinlichkeit. Bei einem Experiment mit Grundmenge Ω möchte man die Wahrscheinlichkeit eines Ereignisses $A \subset \Omega$ bestimmen. Dazu wiederholt man das Experiment n -mal und erhält die Ausgänge $\omega_1, \dots, \omega_n \in \Omega$. Die Anzahl der Ausgänge, die in A liegen, sei

$$N_n(A) = \sum_{k=1}^n \mathbb{1}_{\omega_k \in A}.$$

Wir können auf die Indikatorfunktionen $\mathbb{1}_{\omega_k \in A}$ das starke Gesetz der großen Zahlen anwenden:

$$\frac{N_n(A)}{n} = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\omega_k \in A} \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E} \mathbb{1}_{\omega_1 \in A} = \mathbb{P}[A].$$

In Worten: Die relative Häufigkeit eines Ereignisses konvergiert fast sicher gegen seine Wahrscheinlichkeit, wenn die Anzahl der Experimente gegen ∞ geht.

BEISPIEL 10.9.1 (Befragung). Wir betrachten eine Population von N Personen, die eine Frage mit “ja” oder “nein” beantworten. Die Zahl N sei bekannt. Es sei N_0 (bzw. N_1) die Anzahl der Personen, die die Frage mit “nein” (bzw. mit “ja”) beantworten. Dabei sind N_0 und N_1 unbekannt. Es gilt $N_0 + N_1 = N$.

Wir kann man N_0 und N_1 bestimmen, ohne die ganze Population zu befragen? Es werden zufällig n Personen aus der Population mit Zurücklegen ausgewählt und befragt, wobei n groß aber viel kleiner als N ist. Es sei A_k das Ereignis “die k -te Person sagt ja”. Nach dem starken Gesetz der großen Zahlen gilt

$$\frac{1}{n} \cdot \sum_{k=1}^n \mathbb{1}_{A_k} \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E} \mathbb{1}_{A_1} = \mathbb{P}[A_1] = \frac{N_1}{N}.$$

Damit kann N_1 geschätzt werden.

Approximationssatz von Weierstraß. Als nächstes werden wir einen Satz aus der Analysis mit Mitteln aus der Wahrscheinlichkeitstheorie beweisen.

SATZ 10.9.2 (Weierstraß). Sei $f : [0, 1] \rightarrow \mathbb{R}$ eine stetige Funktion. Dann kann man zu jedem gegebenen $\varepsilon > 0$ ein Polynom P mit

$$\max_{x \in [0, 1]} |f(x) - P(x)| \leq \varepsilon$$

konstruieren. Das heißt, jede stetige Funktion auf einem Intervall kann beliebig genau durch Polynome approximiert werden.

BEWEIS. SCHRITT 1. Seien X_1, X_2, \dots unabhängige Zufallsvariablen mit $X_i \sim \text{Bern}(p)$, d.h.

$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = 0] = 1 - p, \quad p \in [0, 1].$$

Dann ist $S_n := X_1 + \dots + X_n \sim \text{Bin}(n, p)$, also $\mathbb{P}[S_n = k] = \binom{n}{k} p^k (1 - p)^{n-k}$. Insbesondere gilt für den Erwartungswert und die Varianz

$$\mathbb{E}S_n = n\mathbb{E}X_1 = np, \quad \text{Var } S_n = np(1 - p).$$

Wir definieren eine Funktionenfolge:

$$f_n(p) := \mathbb{E} \left[f \left(\frac{S_n}{n} \right) \right] = \sum_{k=0}^n f \left(\frac{k}{n} \right) \mathbb{P}[S_n = k] = \sum_{k=0}^n f \left(\frac{k}{n} \right) \binom{n}{k} p^k (1 - p)^{n-k}.$$

Somit ist jedes f_n ein Polynom vom Grad n in p . Es wird *Bernstein-Polynom* genannt.

SCHRITT 2. Die Idee des Beweises ist nun: für $n \rightarrow \infty$ besagt das Gesetz der großen Zahlen, dass $\frac{S_n}{n} \approx p$ und damit auch $f \left(\frac{S_n}{n} \right) \approx f(p)$. Wendet man den Erwartungswert auf die beiden Funktionswerte an, so erhält man $f_n(p) \approx f(p)$. Somit approximiert das Polynom f_n die Funktion f , jedenfalls wenn n groß ist. Nun präzisieren wir diese Idee.

SCHRITT 3. Da f stetig ist, folgt:

- (1) f ist beschränkt, d.h. es existiert $M > 0$ mit $|f(x)| < M$ für alle $x \in [0, 1]$.
- (2) f ist sogar gleichmäßig stetig, d.h. für jedes $\varepsilon > 0$ existiert ein $\delta > 0$ so dass für alle $x, y \in [0, 1]$ mit $|x - y| \leq \delta$ gilt, dass $|f(x) - f(y)| < \varepsilon$.

SCHRITT 4. Sei $\varepsilon > 0$ fest vorgegeben. Dazu konstruieren wir ein $\delta > 0$ wie in Schritt 3. Es gilt

$$|f_n(p) - f(p)| = \left| \mathbb{E} \left[f \left(\frac{S_n}{n} \right) - f(p) \right] \right| \leq \mathbb{E} \left| f \left(\frac{S_n}{n} \right) - f(p) \right| \leq \mathbb{E} \left[\varepsilon + 2M \cdot \mathbb{1}_{\left| \frac{S_n}{n} - p \right| > \delta} \right],$$

wobei wir die Fallunterscheidung $\left| \frac{S_n}{n} - p \right| \leq \delta$ bzw. $\left| \frac{S_n}{n} - p \right| > \delta$ gemacht haben. Im ersten Fall haben wir die gleichmäßige Stetigkeit von f benutzt, im zweiten Fall die Beschränktheit von f . Es gilt also

$$|f_n(p) - f(p)| \leq \varepsilon + 2M \cdot \mathbb{P} \left[\left| \frac{S_n}{n} - p \right| > \delta \right] \leq \varepsilon + 2M \cdot \frac{\text{Var } \frac{S_n}{n}}{\delta^2},$$

wobei wir die Tschebyschew-Ungleichung benutzt haben. Mit der Formel $\text{Var } S_n = np(1-p)$ erhalten wir nun für hinreichend großes n

$$|f_n(p) - f(p)| \leq \varepsilon + 2M \cdot \frac{np(1-p)}{n^2\delta^2} \leq \varepsilon + 2M \cdot \frac{1}{n\delta^2} \leq 2\varepsilon.$$

Dabei folgt die letzte Ungleichung aus der Tatsache, dass $2M \cdot \frac{1}{n\delta^2}$ für $n \rightarrow \infty$ gegen 0 konvergiert, also $< \varepsilon$ für hinreichend großes n ist. Wir haben ein approximierendes Polynom mit einem Approximationsfehler von 2ε konstruiert. Da jedoch ε beliebig war, ist der Satz damit bewiesen. \square

Normale Zahlen. Sei $\omega \in (0, 1)$ eine reelle Zahl mit der Dezimaldarstellung

$$\omega = 0.x_1x_2x_3\dots$$

Dabei ist $x_k \in \{0, \dots, 9\}$ die k -te Ziffer von ω .

DEFINITION 10.9.3. Eine Zahl $\omega \in (0, 1)$ heißt *normal*, wenn für jedes $l \in \mathbb{N}$ und für jede Ziffernfolge $(a_1, \dots, a_l) \in \{0, \dots, 9\}^l$ der Länge l gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n \mathbb{1}_{X_{i+1}=a_1, \dots, X_{i+l}=a_l} = 10^{-l}.$$

In Worten: Jede Ziffernfolge der Länge l kommt in der Dezimaldarstellung von ω mit Häufigkeit 10^{-l} vor, und das gilt für jedes l .

DEFINITION 10.9.4. Eine Zahl $\omega \in (0, 1)$ heißt *einfach normal*, wenn die obige Bedingung für $l = 1$ erfüllt ist, d.h., wenn in der Dezimaldarstellung von ω jede Ziffer mit einer Häufigkeit von $1/10$ vorkommt.

BEISPIEL 10.9.5. Die Zahl $\frac{1}{3} = 0,3333\dots$ ist nicht normal. Allgemein sind rationale Zahlen nicht normal, da sie eine periodische Dezimalbruchentwicklung besitzen.

BEISPIEL 10.9.6. Unbewiesene Vermutung: Die Zahlen

$$\begin{aligned} \pi &= 3.1415926535897932384626433832795028841971693993751\dots, \\ e &= 2.7182818284590452353602874713526624977572470937000\dots, \\ \log 2 &= 0.6931471805599453094172321214581765680755001343602\dots \end{aligned}$$

sind normal.

BEISPIEL 10.9.7. Unbewiesene Vermutung: Jede algebraische Zahl ist normal, sofern sie nicht rational ist. Zum Beispiel ist die Zahl

$$\sqrt{2} = 1.4142135623730950488016887242096980785696718753769 \dots$$

normal.

BEISPIEL 10.9.8. Die Champernowne-Zahl entsteht durch die Aneinanderreihung aller natürlichen Zahlen in ihrer Dezimaldarstellung:

$$0.123 \dots 91011 \dots 192021 \dots 99100101 \dots$$

Man kann zeigen, dass diese Zahl normal ist.

SATZ 10.9.9 (Borel, 1909). Sei $A \subset [0, 1]$ die Menge der normalen Zahlen. Dann ist das Lebesgue-Maß von A gleich 1. D.h., fast jede Zahl ist normal.

BEWEIS. SCHRITT 1. Wir betrachten den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, wobei $\Omega = (0, 1)$, \mathcal{F} die σ -Algebra der Borel-Teilmengen von $(0, 1)$ und \mathbb{P} das Lebesgue-Maß ist. Für alle $k \in \mathbb{N}$ sei

$$X_k(\omega) = \omega_k$$

die k -te Ziffer von $\omega \in (0, 1)$. Dann ist X_k eine Zufallsvariable auf Ω .

Wir zeigen nun, dass die Zufallsvariablen X_1, X_2, \dots unabhängig und gleichverteilt auf $\{0, \dots, 9\}$ sind. Betrachte für ein beliebiges $m \in \mathbb{N}$ und eine Ziffernfolge $(b_1, \dots, b_m) \in \{0, \dots, 9\}^m$ die Menge

$$\{\omega \in (0, 1) : X_1(\omega) = b_1, \dots, X_m(\omega) = b_m\}.$$

Diese Menge ist ein Intervall mit Endpunkten

$$0.b_1b_2 \dots b_m000 \dots \text{ und } 0.b_1b_2 \dots b_m999 \dots$$

Die Länge des Intervalls ist somit 10^{-m} . Es gilt somit

$$\mathbb{P}[X_1 = b_1, \dots, X_m = b_m] = 10^{-m}.$$

Sei nun $i \in \{1, \dots, m\}$ fest und betrachte die Menge

$$\{\omega \in (0, 1) : X_i(\omega) = b_i\}.$$

Das ist die Menge aller ω , die in der Dezimaldarstellung an der i -ten Stelle die Ziffer b_i haben. Für die ersten $i - 1$ Stellen gibt es 10^{i-1} Möglichkeiten, die Ziffern zu wählen. Diese Menge ist somit eine Vereinigung von 10^{i-1} Intervallen der Länge 10^{-i} . Somit ergibt sich für die Wahrscheinlichkeit dieser Menge:

$$\mathbb{P}[X_i = b_i] = 10^{i-1} \cdot 10^{-i} = \frac{1}{10}.$$

Daraus folgt, dass die Zufallsvariable X_i gleichverteilt auf $\{0, \dots, 9\}$ ist. Außerdem gilt für alle $m \in \mathbb{N}$ und alle $(b_1, \dots, b_m) \in \{0, \dots, 9\}^m$

$$\mathbb{P}[X_1 = b_1, \dots, X_m = b_m] = 10^{-m} = \mathbb{P}[X_1 = b_1] \cdot \dots \cdot \mathbb{P}[X_m = b_m].$$

Daraus folgt, dass die Zufallsvariablen X_1, X_2, \dots unabhängig sind.

SCHRITT 2. Die Idee ist nun, das starke Gesetz der großen Zahlen folgendermaßen zu benutzen. Es gilt z.B. für jede Ziffer $a \in \{0, 1, \dots, 9\}$:

$$\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k=a} \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E}[\mathbb{1}_{X_1=a}] = \mathbb{P}[X_1 = a] = \frac{1}{10}.$$

Das zeigt schon mal, dass fast jede Zahl einfach normal ist. Wir werden nun diese Argumentation auf Ziffernfolgen beliebiger Länge l ausweiten.

SCHRITT 3. Wir betrachten eine Ziffernfolge $(a_1, \dots, a_l) \in \{0, \dots, 9\}^l$ der Länge l . Wir definieren die Zufallsvariablen

$$Z_i^{(0)} := \mathbb{1}_{X_{i+l+1}=a_1, \dots, X_{i+l+l}=a_l}, \quad i = 0, 1, 2, \dots,$$

und, allgemeiner,

$$Z_i^{(j)} := \mathbb{1}_{X_{i+l+j+1}=a_1, \dots, X_{i+l+j+l}=a_l}, \quad i = 0, 1, 2, \dots, \quad j = 0, \dots, l-1.$$

Diese Zufallsvariablen sind folgendermaßen zu verstehen: Man unterteile alle Ziffern einer Zahl ω in “Pakete” der Länge l , wobei das erste “Paket” an der Stelle $j+1$ startet, und betrachte das i -te “Paket”. Entspricht dieses der Ziffernfolge (a_1, \dots, a_l) , so ist $Z_i^{(j)} = 1$. Sonst ist $Z_i^{(j)} = 0$.

Für alle $j \in \{0, \dots, l-1\}$ sind die Zufallsvariablen $Z_0^{(j)}, Z_1^{(j)}, Z_2^{(j)}, \dots$ unabhängig und identisch verteilt. Mit dem starken Gesetz der großen Zahlen erhalten wir folgendes: Es gibt eine Menge $B(l, j, a_1, \dots, a_l) \subset (0, 1)$ mit Lebesgue-Maß 0, so dass für alle $\omega \notin B(l, j, a_1, \dots, a_l)$ gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n Z_i^{(j)}(\omega) = \mathbb{E}Z_0^{(j)} = 10^{-l}.$$

Die Häufigkeit der Ziffernfolge (a_1, \dots, a_l) ist also 10^{-l} , allerdings werden dabei nur “Pakete” berücksichtigt, die an den Stellen $j+1, j+1+l, j+1+2l, \dots$ anfangen. Da dies aber für jedes $j \in \{0, \dots, l-1\}$ gilt, erhalten wir, dass für alle $\omega \notin \bigcup_{j=0}^{l-1} B(l, j, a_1, \dots, a_l)$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \mathbb{1}_{X_{k+1}=a_1, \dots, X_{k+l}=a_l}(\omega) = 10^{-l}.$$

Nun definieren wir die “Ausnahmemenge”

$$B := \bigcup_{l \in \mathbb{N}} \bigcup_{j=0}^{l-1} \bigcup_{a_1, \dots, a_l \in \{0, \dots, 9\}} B(l, j, a_1, \dots, a_l) \subset (0, 1).$$

Jede Zahl ω , die nicht in B liegt, ist normal. Da B eine abzählbare Vereinigung von Nullmengen ist, ist das Lebesgue-Maß von B gleich 0. Das Komplement von B hat also Lebesgue-Maß 1, und alle Zahlen darin sind normal. Somit ist das Lebesgue-Maß der Menge der normalen Zahlen gleich 1. \square

Erneuerungssatz. Man betrachte ein Gerät (etwa eine Glühbirne), dessen Lebensdauer eine Zufallsvariable $X_1 > 0$ ist. Sobald das Gerät kaputt geht, wird es durch ein neues Gerät

ersetzt, dessen Lebensdauer eine Zufallsvariable $X_2 > 0$ ist. Sobald das zweite Gerät kaputt geht, wird es durch ein drittes Gerät ersetzt, usw. Wir nehmen an, dass X_1, X_2, \dots (die Lebensdauern der Geräte) unabhängige Zufallsvariablen sind, da ein Gerät nichts von der Lebensdauer eines anderen wissen kann. Außerdem nehmen wir an, dass X_1, X_2, \dots identisch verteilt sind, etwa weil die Geräte vom gleichen Hersteller sind bzw. die gleiche Qualität haben.

Seien also $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ unabhängige und identisch verteilte Zufallsvariablen mit positiven Werten, d.h. $\mathbb{P}[X_k > 0] = 1$. Die Zufallsvariable $S_n = X_1 + \dots + X_n$ heißt die *n-te Erneuerungszeit*. Zum Zeitpunkt S_n geht das *n-te* Gerät kaputt und wird durch das *n + 1-te* Gerät ersetzt. Es gilt $0 < S_1 < S_2 < \dots$. Mit

$$N(T) = \sum_{n=1}^{\infty} \mathbb{1}_{S_n < T}, \quad T > 0,$$

bezeichnen wir die Anzahl der Erneuerungen im Zeitintervall $(0, T)$.

SATZ 10.9.10 (Erneuerungssatz). *Es gilt:*

$$\frac{N(T)}{T} \xrightarrow[T \rightarrow \infty]{f.s.} \frac{1}{\mathbb{E}X_1}.$$

Wir werden den Satz nicht beweisen. Die Aussage ist allerdings sehr intuitiv: die Lebensdauer eines Geräts ist im Durchschnitt gleich $\mathbb{E}X_1$, also verbraucht man im Zeitintervall $(0, T)$ ungefähr $T/\mathbb{E}X_1$ Geräte, jedenfalls dann, wenn T groß ist. Es gilt also die Approximation

$$\frac{N(T)}{T} \approx \frac{1}{\mathbb{E}X_1}.$$

KAPITEL 11

Ungleichungen

11.1. Jensen-Ungleichung

DEFINITION 11.1.1. Eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ heißt *konvex*, wenn man für jedes $x_0 \in \mathbb{R}$ ein $K_0 = K_0(x_0) \in \mathbb{R}$ finden kann, so dass für alle $x \in \mathbb{R}$ gilt:

$$g(x) \geq g(x_0) + K_0(x - x_0).$$

BEMERKUNG 11.1.2. Eine Funktion $g(x)$ ist also genau dann konvex, wenn der Graph von g die folgende Eigenschaft besitzt: zu jedem x_0 können wir eine Gerade finden, die durch den Punkt $(x_0, f(x_0))$ geht und die Eigenschaft hat, dass der Graph von g komplett oberhalb dieser Geraden liegt. Die Zahl K_0 heißt das Subdifferential von g an der Stelle x_0 und ist nicht immer eindeutig bestimmt. Falls g differenzierbar ist, dann kann K_0 durch $K_0 = g'(x_0)$ festgelegt werden.

SATZ 11.1.3 (Jensen-Ungleichung). Sei g eine konvexe Funktion und X eine Zufallsvariable mit $X \in L^1$, $g(X) \in L^1$. Dann gilt:

$$(11.1.1) \quad g(\mathbb{E}X) \leq \mathbb{E}[g(X)].$$

BEISPIEL 11.1.4. Wir betrachten eine Zufallsvariable X , die n Werte x_1, x_2, \dots, x_n mit einer Wahrscheinlichkeit von jeweils $1/n$ annimmt. Der Erwartungswert ist dann das arithmetische Mittel der x_i , also

$$\mathbb{E}X = \frac{x_1 + \dots + x_n}{n}.$$

Dann besagt die Jensen-Ungleichung für diesen Spezialfall, dass für jede konvexe Funktion g gilt

$$g\left(\frac{x_1 + \dots + x_n}{n}\right) \leq \frac{g(x_1) + \dots + g(x_n)}{n}.$$

BEWEIS VON SATZ 11.1.3. Wir setzen in die Definition der Konvexität $x_0 = \mathbb{E}X$, $x = X$ ein. Dann existiert laut Definition ein $K_0 \in \mathbb{R}$ mit der Eigenschaft, dass

$$g(X) \geq g(\mathbb{E}X) + K_0(X - \mathbb{E}X).$$

Nun bilden wir den Erwartungswert:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(\mathbb{E}X)] + K_0\mathbb{E}[X - \mathbb{E}X] = g(\mathbb{E}X) + K_0(\mathbb{E}X - \mathbb{E}X) = g(\mathbb{E}X).$$

□

11.2. Ljapunow-Ungleichung

Ein Spezialfall der Jensen-Ungleichung ist die Ljapunow-Ungleichung.

SATZ 11.2.1 (Ljapunow-Ungleichung). *Seien $0 < s < t$ und sei X eine Zufallsvariable. Dann gilt*

$$(\mathbb{E}[|X|^s])^{\frac{1}{s}} \leq (\mathbb{E}[|X|^t])^{\frac{1}{t}}.$$

DEFINITION 11.2.2. Sei $p \geq 0$. Die L^p -Norm einer Zufallsvariable X ist definiert durch

$$\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}.$$

BEMERKUNG 11.2.3. Mit dieser Notation besagt die Ljapunow-Ungleichung, dass

$$\|X\|_s \leq \|X\|_t, \quad 0 < s < t.$$

BEISPIEL 11.2.4. Wir betrachten eine Zufallsvariable X , die n Werte $x_1, x_2, \dots, x_n > 0$ mit Wahrscheinlichkeit von jeweils $1/n$ annimmt. Dann besagt die Ljapunow-Ungleichung, dass

$$\left(\frac{x_1^s + \dots + x_n^s}{n} \right)^{\frac{1}{s}} \leq \left(\frac{x_1^t + \dots + x_n^t}{n} \right)^{\frac{1}{t}}, \quad 0 < s < t.$$

Dies ist die klassische Ungleichung der verallgemeinerten Mittel. Setzen wir nun $s = 1$ und $t = 2$, so erhalten wir die Ungleichung zwischen dem arithmetischem und dem quadratischem Mittel:

$$\frac{x_1 + \dots + x_n}{n} \leq \sqrt{\frac{x_1^2 + \dots + x_n^2}{n}}.$$

BEISPIEL 11.2.5. Wir betrachten den Wahrscheinlichkeitsraum $[0, 1]$ mit der Borel- σ -Algebra und dem Lebesgue-Maß. Sei $f : [0, 1] \rightarrow \mathbb{R}$ eine messbare Funktion. Dann besagt die Ungleichung von Ljapunow, dass

$$\left(\int_0^1 |f(z)|^s dz \right)^{\frac{1}{s}} \leq \left(\int_0^1 |f(z)|^t dz \right)^{\frac{1}{t}}, \quad 0 < s < t.$$

BEWEIS VON SATZ 11.2.1. Wir führen den Beweis mit Hilfe der Jensen-Ungleichung. Sei $g(y) = |y|^{\frac{t}{s}}$. Dann ist g konvex, denn $\frac{t}{s} > 1$. Setzen wir nun $Y = |X|^s$ und nutzen die Jensen-Ungleichung, so erhalten wir

$$g(\mathbb{E}Y) \leq \mathbb{E}[g(Y)].$$

Da $g(Y) = (|X|^s)^{\frac{t}{s}} = |X|^t$, ist dies äquivalent zu

$$(\mathbb{E}[|X|^s])^{\frac{t}{s}} \leq \mathbb{E}[|X|^t].$$

Daraus ergibt sich die Ljapunow-Ungleichung, wenn man die $1/t$ -te Potenz der beiden Seiten betrachtet. \square

11.3. Young-Ungleichung

SATZ 11.3.1 (Young-Ungleichung). Seien $p, q > 1$ mit der Eigenschaft, dass $1/p + 1/q = 1$. Dann gilt für alle $a, b > 0$:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

BEISPIEL 11.3.2. Seien $p = q = 2$. Dann erhalten wir die Ungleichung $ab \leq \frac{a^2+b^2}{2}$. Diese Ungleichung kann man schnell beweisen, indem man alle Terme auf die rechte Seite bringt: $0 \leq \frac{1}{2}(a-b)^2$.

BEWEIS. Wir betrachten die Funktion $y = x^{p-1}$, $x > 0$. Dabei handelt es sich, da $p > 1$ ist, um eine monoton wachsende Funktion. Die Umkehrfunktion lautet

$$x = y^{\frac{1}{p-1}} = y^{q-1}.$$

Also hat die Umkehrfunktion eine ähnliche Form wie die Ausgangsfunktion. Graphisch ergibt sich

$$ab \leq \int_0^a x^{p-1} dx + \int_0^b y^{q-1} dy.$$

Daraus ergibt sich die Young-Ungleichung. □

11.4. Hölder-Ungleichung

SATZ 11.4.1 (Hölder-Ungleichung). Seien $p, q > 1$ mit $1/p + 1/q = 1$. Seien $X \in L^p$ und $Y \in L^q$ Zufallsvariablen. Dann gilt

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} \cdot (\mathbb{E}[|Y|^q])^{\frac{1}{q}}.$$

BEMERKUNG 11.4.2. Äquivalente Schreibweise:

$$\|XY\|_1 \leq \|X\|_p \cdot \|Y\|_q.$$

Insbesondere folgt aus $X \in L^p$ und $Y \in L^q$, dass $X \cdot Y \in L^1$.

BEMERKUNG 11.4.3. Mit $p = q = 2$ erhalten wir die Cauchy-Schwarzsche Ungleichung

$$(11.4.1) \quad \|XY\|_1 \leq \|X\|_2 \cdot \|Y\|_2.$$

BEWEIS VON SATZ 11.4.1. Falls $\|X\|_p = 0$, so gilt $\mathbb{E}[|X|^p] = 0$. Dann ist $X = 0$ fast sicher und daher auch $X \cdot Y = 0$ fast sicher. In diesem Fall gilt die Hölder-Ungleichung, da $0 \leq 0$.

Falls $\|Y\|_q = 0$, so gilt die Hölder-Ungleichung, analog zu vorherigem Fall, ebenso.

Seien nun also $\|X\|_p \neq 0$ und $\|Y\|_q \neq 0$. Wir führen die folgenden Zufallsvariablen a und b ein:

$$a = \frac{|X|}{\|X\|_p}, \quad b = \frac{|Y|}{\|Y\|_q}.$$

Dann gilt:

$$\mathbb{E}[a^p] = 1, \quad \mathbb{E}[b^q] = 1.$$

Mit der Young-Ungleichung erhalten wir

$$\mathbb{E}[ab] \leq \mathbb{E}\left[\frac{a^p}{p}\right] + \mathbb{E}\left[\frac{b^q}{q}\right] = \frac{1}{p} + \frac{1}{q} = 1.$$

Durch Einsetzen von a und b folgt

$$\mathbb{E} \left[\frac{|X||Y|}{\|X\|_p \cdot \|Y\|_q} \right] \leq 1.$$

Da $\|X\|_p \cdot \|Y\|_q$ eine Konstante ist, darf man den Term aus dem Erwartungswert herausziehen, was zu unserer Behauptung führt:

$$\mathbb{E}|X \cdot Y| \leq \|X\|_p \cdot \|Y\|_q.$$

□

BEISPIEL 11.4.4. Sei $\Omega = \{1, \dots, n\}$ ein diskreter Wahrscheinlichkeitsraum mit $\mathbb{P}[\{\omega\}] = 1/n$ für alle $\omega \in \Omega$. Betrachte zwei Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{R}$ mit

$$X(k) = x_k, \quad Y(k) = y_k, \quad x_k, y_k \in \mathbb{R}.$$

Indem wir X und Y in die Hölder-Ungleichung einsetzen, erhalten wir

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}.$$

Dies ist die klassische Form der Hölder-Ungleichung für Summen.

BEISPIEL 11.4.5. Betrachte den Wahrscheinlichkeitsraum $\Omega = [0, 1]$ mit der Borel- σ -Algebra und dem Lebesgue-Maß. Seien $f, g : [0, 1] \rightarrow \mathbb{R}$ zwei messbare Funktionen. Indem wir $X = f$ und $Y = g$ in die Hölder-Ungleichung einsetzen, erhalten wir

$$\int_0^1 |f(z)g(z)| dz \leq \left(\int_0^1 |f(z)|^p dz \right)^{\frac{1}{p}} \cdot \left(\int_0^1 |g(z)|^q dz \right)^{\frac{1}{q}}.$$

Dies ist die klassische Form der Hölder-Ungleichung für Integrale.

11.5. Minkowski-Ungleichung

SATZ 11.5.1 (Minkowski-Ungleichung). Sei $p \geq 1$ und seien $X, Y \in L^p$ Zufallsvariablen. Dann gilt

$$(\mathbb{E}[|X + Y|^p])^{\frac{1}{p}} \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} + (\mathbb{E}[|Y|^p])^{\frac{1}{p}}.$$

BEMERKUNG 11.5.2. Äquivalente Schreibweise:

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Die Minkowski-Ungleichung ist also eine Dreiecksungleichung für die L^p -Norm.

BEISPIEL 11.5.3. Sei $\Omega = \{1, \dots, n\}$ ein diskreter Wahrscheinlichkeitsraum mit $\mathbb{P}[\{\omega\}] = 1/n$ für alle $\omega \in \Omega$. Betrachte zwei Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{R}$ mit

$$X(k) = x_k, \quad Y(k) = y_k, \quad x_k, y_k \in \mathbb{R}.$$

Indem wir X und Y in die Minkowski-Ungleichung einsetzen, erhalten wir

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n |y_i|^p \right)^{\frac{1}{p}}.$$

Dies ist die klassische Minkowski-Ungleichung für Summen. Für $p = 2$ ist dies genau die klassische Dreiecksungleichung, die besagt, dass die Länge einer Summe von zwei Vektoren nicht größer sein kann, als die Summe der Längen der beiden Vektoren:

$$\left(\sum_{i=1}^n (x_i + y_i)^2 \right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} + \left(\sum_{i=1}^n y_i^2 \right)^{\frac{1}{2}}.$$

BEISPIEL 11.5.4. Betrachte den Wahrscheinlichkeitsraum $\Omega = [0, 1]$ mit der Borel- σ -Algebra und dem Lebesgue-Maß. Seien $f, g : [0, 1] \rightarrow \mathbb{R}$ zwei messbare Funktionen. Indem wir $X = f$ und $Y = g$ in die Minkowski-Ungleichung einsetzen, erhalten wir

$$\left(\int_0^1 |f(z) + g(z)|^p dz \right)^{1/p} \leq \left(\int_0^1 |f(z)|^p dz \right)^{1/p} + \left(\int_0^1 |g(z)|^p dz \right)^{1/p}.$$

Dies ist die klassische Form der Minkowski-Ungleichung für Integrale.

BEWEIS VON SATZ 11.5.1. Sei zunächst $p = 1$, dann gilt mit der Ungleichung $|X + Y| \leq |X| + |Y|$:

$$\mathbb{E}|X + Y| \leq \mathbb{E}(|X| + |Y|) = \mathbb{E}|X| + \mathbb{E}|Y|.$$

Also gilt die Minkowski-Ungleichung.

Nun betrachten wir den Fall $p > 1$. Wir definieren $q = \frac{p}{p-1} > 1$. Mit dieser Wahl von q gilt die Relation $\frac{1}{p} + \frac{1}{q} = 1$. Wir wenden nun die Ungleichung $|X + Y| \leq |X| + |Y|$ und danach die Hölder-Ungleichung an:

$$\begin{aligned} \mathbb{E}[|X + Y|^p] &= \mathbb{E}[|X + Y|^{p-1} \cdot |X + Y|] \\ &\leq \mathbb{E}[|X + Y|^{p-1} \cdot |X|] + \mathbb{E}[|X + Y|^{p-1} \cdot |Y|] \\ &\leq (\mathbb{E}[|X + Y|^{(p-1)q}])^{\frac{1}{q}} \cdot \|X\|_p + (\mathbb{E}[|X + Y|^{(p-1)q}])^{\frac{1}{q}} \cdot \|Y\|_p \\ &= (\mathbb{E}[|X + Y|^p])^{\frac{1}{q}} \cdot (\|X\|_p + \|Y\|_p). \end{aligned}$$

Mit $1 - \frac{1}{q} = \frac{1}{p}$ erhalten wir

$$(\mathbb{E}[|X + Y|^p])^{\frac{1}{p}} \leq \|X\|_p + \|Y\|_p.$$

□

11.6. L^p -Räume und L^p -Konvergenz

DEFINITION 11.6.1. Sei $p > 0$. Sei X eine Zufallsvariable. Wir schreiben $X \in L^p$, wenn $\mathbb{E}[|X|^p] < \infty$. Die L^p -Norm von $X \in L^p$ ist definiert durch

$$\|X\|_p = (\mathbb{E}[|X|^p])^{\frac{1}{p}}.$$

BEMERKUNG 11.6.2. Die Ljapunow-Ungleichung besagt, dass

$$\|X\|_s \leq \|X\|_t, \text{ wenn } 0 < s < t.$$

Somit sind die L^p -Räume ineinander geschachtelt: $L^s \supset L^t$, wenn $s < t$. Insbesondere gilt

$$L^1 \supset L^2 \supset L^3 \supset \dots$$

Wir haben früher bereits gezeigt und sehr oft benutzt, dass $L^1 \supset L^2$. (Wenn eine Zufallsvariable eine Varianz besitzt, dann besitzt sie auch einen Erwartungswert).

BEMERKUNG 11.6.3. Für $p \geq 1$ ist L^p ein Vektorraum, denn

- (1) Für $X \in L^p$ und $\lambda \in \mathbb{R}$ ist auch $\lambda X \in L^p$.
- (2) Für $X \in L^p$ und $Y \in L^p$ ist auch $X + Y \in L^p$.

Die zweite Eigenschaft folgt aus der Minkowski-Ungleichung, denn

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p < \infty.$$

Für $p < 1$ ist L^p im Allgemeinen kein Vektorraum.

DEFINITION 11.6.4. Sei $p > 1$. Der L^p -Abstand zwischen zwei Zufallsvariablen $X \in L^p$ und $Y \in L^p$ ist

$$d_p(X, Y) = \|X - Y\|_p = (\mathbb{E}[|X - Y|^p])^{\frac{1}{p}}.$$

BEMERKUNG 11.6.5. Für $X, Y \in L^p$ ist der L^p -Abstand $\|X - Y\|_p$ endlich. Das folgt aus der Minkowski-Ungleichung:

$$\|X - Y\|_p \leq \|X\|_p + \| - Y\|_p = \|X\|_p + \|Y\|_p < \infty.$$

BEMERKUNG 11.6.6. Für $p \geq 1$ ist der L^p -Abstand eine Metrik, denn es gilt

- (1) $d_p(X, Y) = 0$ genau dann, wenn $X = Y$ *fast sicher*.
- (2) $d_p(X, Y) = d_p(Y, X)$.
- (3) $d_p(X, Z) \leq d_p(X, Y) + d_p(Y, Z)$.

Die letzte Eigenschaft folgt aus der Minkowski-Ungleichung, wobei hier $p \geq 1$ benutzt wird. Somit erfüllt der L^p -Raum die drei Axiome eines metrischen Raumes bis auf eine Kleinigkeit: Es kann sein, dass $d_p(X, Y) = 0$ und dennoch $X \neq Y$. Deshalb macht man eine Konvention: man betrachtet zwei Zufallsvariablen als identisch, wenn sie *fast überall* gleich sind. Dann gelten alle drei Eigenschaften eines metrischen Raumes. Ist aber $p < 1$, so gilt die Dreiecksungleichung nicht.

DEFINITION 11.6.7. Eine Folge X_1, X_2, \dots von Zufallsvariablen *konvergiert* in L^p gegen eine Zufallsvariable X , wenn $X, X_1, X_2, \dots \in L^p$ und

$$\lim_{n \rightarrow \infty} \|X - X_n\|_p = 0.$$

BEMERKUNG 11.6.8. Äquivalente Schreibweise: $\lim_{n \rightarrow \infty} \mathbb{E}[|X - X_n|^p] = 0$.

BEZEICHNUNG: $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$.

BEMERKUNG 11.6.9. Aus der Ljapunow-Ungleichung folgt, dass

$$X_n \xrightarrow[n \rightarrow \infty]{L^t} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{L^s} X, \text{ wenn } s < t.$$

Daher gelten folgende Implikationen zwischen den L^p -Konvergenzen:

$$L^1 \Leftarrow L^2 \Leftarrow L^3 \Leftarrow \dots$$

Wenn man die Konvergenz in Wahrscheinlichkeit als L^0 -Konvergenz und die gleichmäßige Konvergenz als L^∞ -Konvergenz betrachtet, dann kann man dieses Schema vervollständigen:

$$L^0 \Leftarrow L^1 \Leftarrow L^2 \Leftarrow L^3 \Leftarrow \dots \Leftarrow L^\infty.$$

Analytische Methoden

Es seien unabhängige Zufallsvariablen X_1, \dots, X_n gegeben. Die Verteilungen dieser Zufallsvariablen seien bekannt. Wie bestimmt man dann die Verteilung der Summe $X_1 + \dots + X_n$? Man kann die Faltungsformel benutzen, jedoch ist diese ziemlich kompliziert, besonders dann, wenn n groß ist. In diesem Kapitel werden wir Methoden einführen, die dieses Problem auf eine viel elegantere Weise lösen, als die Faltungsformel.

12.1. Erzeugende Funktion

DEFINITION 12.1.1. Sei X eine Zufallsvariable mit Werten in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Wir schreiben $p_n = \mathbb{P}[X = n]$ für $n \in \mathbb{N}_0$. Dann heißt

$$g_X(z) \stackrel{\text{def}}{=} \mathbb{E}[z^X] = \sum_{n=0}^{\infty} p_n z^n$$

die *erzeugende Funktion* von X .

SATZ 12.1.2. Die obige Reihe konvergiert absolut für $z \in [-1, 1]$ und es gilt $|g_X(z)| \leq 1$ für alle $z \in [-1, 1]$. Außerdem gilt $g_X(1) = 1$.

BEMERKUNG 12.1.3. Die erzeugende Funktion ist also wohldefiniert im Intervall $[-1, 1]$. Es ist bekannt, dass die Summe einer Taylor-Reihe eine unendlich oft differenzierbare Funktion im Konvergenzbereich der Reihe ist. Somit ist $g_X(z)$ unendlich oft differenzierbar. Eigentlich kann man $g_X(z)$ sogar für komplexe Werte von z betrachten, dann ist g_X sogar eine analytische Funktion im Einheitskreis $|z| \leq 1$.

BEWEIS VON SATZ 12.1.2. Sei $z \in [-1, 1]$. Um die absolute Konvergenz zu zeigen, betrachten wir die Reihe

$$\sum_{n=0}^{\infty} |p_n z^n| \leq \sum_{n=0}^{\infty} p_n = 1.$$

Somit ist die Reihe absolut konvergent. Es gilt

$$|g_X(z)| = \left| \sum_{n=0}^{\infty} p_n z^n \right| \leq \sum_{n=0}^{\infty} |p_n z^n| \leq \sum_{n=0}^{\infty} p_n = 1.$$

Außerdem gilt $g_X(1) = \sum_{n=0}^{\infty} p_n = 1$. □

SATZ 12.1.4. Für jedes $n \in \mathbb{N}_0$ gilt

$$\mathbb{P}[X = n] = \frac{g_X^{(n)}(0)}{n!}.$$

BEMERKUNG 12.1.5. Aus diesem Satz folgt, dass die Verteilung einer Zufallsvariable eindeutig durch die erzeugende Funktion definiert ist. D. h.: sind X und Y zwei Zufallsvariablen mit $g_X(z) = g_Y(z)$ für alle $z \in [-1, 1]$, so sind die Verteilungen von X und Y gleich, d.h. es gilt $\mathbb{P}[X = n] = \mathbb{P}[Y = n]$ für alle $n \in \mathbb{N}_0$. Insbesondere sind alle Charakteristika einer Zufallsvariable, wie beispielsweise der Erwartungswert oder die Varianz, in der erzeugenden Funktion versteckt.

BEWEIS VON SATZ 12.1.4. Wir können eine konvergente Taylor-Reihe termweise n -mal ableiten:

$$g_X^{(n)}(z) = \left(\sum_{k=0}^{\infty} p_k z^k \right)^{(n)} = \sum_{k=0}^{\infty} p_k (z^k)^{(n)}.$$

Dabei ist

$$(z^k)^{(n)} = \begin{cases} 0, & k < n, \\ n!, & k = n, \\ k(k-1) \dots (k-n+1) z^{k-n}, & k > n. \end{cases}$$

Indem wir nun $z = 0$ einsetzen, erhalten wir $g_X^{(n)}(0) = n! p_n$. □

Warum betrachten wir überhaupt erzeugende Funktionen? Weil sie uns erlauben, die Faltung in ein Produkt umzuwandeln. Das wird im nächsten Satz gezeigt.

SATZ 12.1.6. Seien X, Y unabhängige Zufallsvariablen mit Werten in \mathbb{N}_0 . Dann gilt:

$$g_{X+Y}(z) = g_X(z) \cdot g_Y(z), \quad z \in [-1, 1].$$

ERSTER BEWEIS. Mit der Definition der erzeugenden Funktion erhalten wir

$$g_{X+Y}(z) = \mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X \cdot z^Y] = \mathbb{E}[z^X] \cdot \mathbb{E}[z^Y] = g_X(z) \cdot g_Y(z),$$

wobei wir benutzt haben, dass X und Y (und somit auch z^X und z^Y) unabhängig sind. □

ZWEITER BEWEIS. Wir wählen die folgende Notation:

$$p_n = \mathbb{P}[X = n] \text{ und } q_n = \mathbb{P}[Y = n].$$

Mit der Faltungsformel erhalten wir dann

$$r_n := \mathbb{P}[X + Y = n] = \sum_{k=0}^n p_k q_{n-k}.$$

Nun multiplizieren wir die erzeugenden Funktionen:

$$g_X(z)g_Y(z) = \left(\sum_{k=0}^{\infty} p_k z^k \right) \left(\sum_{l=0}^{\infty} q_l z^l \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n p_k q_{n-k} \right) z^n = \sum_{n=0}^{\infty} r_n z^n = g_{X+Y}(z).$$

□

BEISPIEL 12.1.7. Sei $X \sim \text{Poi}(\lambda)$. Dann gilt: $g_X(z) = e^{\lambda(z-1)}$.

BEWEIS. $X \sim \text{Poi}(\lambda)$ heißt, dass $p_n = \mathbb{P}[X = n] = e^{-\lambda} \frac{\lambda^n}{n!}$ für $n \in \mathbb{N}_0$. Somit gilt

$$g_X(z) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} z^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda z)^n}{n!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}.$$

BEISPIEL 12.1.8. Seien $X_1 \sim \text{Poi}(\lambda_1)$ und $X_2 \sim \text{Poi}(\lambda_2)$ unabhängige Zufallsvariablen. Dann gilt:

$$X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2).$$

BEWEIS. Wir berechnen die Verteilung von $X_1 + X_2$ nicht direkt mit der Faltungsformel, sondern benutzen die erzeugende Funktion. Die erzeugende Funktion von $X_1 + X_2$ ist

$$\begin{aligned} g_{X_1+X_2}(z) &= g_{X_1}(z) \cdot g_{X_2}(z), \quad \text{da } X_1, X_2 \text{ unabhängig} \\ &= e^{\lambda_1(z-1)} \cdot e^{\lambda_2(z-1)}, \quad \text{da } X_i \sim \text{Poi}(\lambda_i) \\ &= e^{(\lambda_1+\lambda_2)(z-1)}. \end{aligned}$$

Auf der rechten Seite erkennen wir die erzeugende Funktion einer $\text{Poi}(\lambda_1 + \lambda_2)$ -Verteilung. Die erzeugende Funktion von $X_1 + X_2$ stimmt also mit der erzeugenden Funktion einer $\text{Poi}(\lambda_1 + \lambda_2)$ -Verteilung überein. Da nun die erzeugende Funktion einer Zufallsvariable ihre Verteilung eindeutig bestimmt, muss gelten: $X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$. \square

Wie zuvor schon erwähnt, steckt die gesamte Verteilung von X , insbesondere der Erwartungswert und die Varianz, in der erzeugenden Funktion. Nun zeigen wir, wo der Erwartungswert und die Varianz versteckt sind.

DEFINITION 12.1.9. Sei X eine Zufallsvariable. Dann heißen die Zahlen

$$\mathbb{E}[X], \mathbb{E}[X^2], \dots, \mathbb{E}[X^n], \dots$$

Momente von X .

SATZ 12.1.10. Für jedes $n \in \mathbb{N}_0$ gilt

$$\mathbb{E}[X(X-1)\dots(X-n+1)] = g_X^{(n)}(1).$$

BEMERKUNG 12.1.11. Dies ist an sich kein Moment, aber X^n steckt in diesem Erwartungswert drin und wir können damit letztendlich $\mathbb{E}[X^n]$ rechnerisch isolieren.

BEISPIEL 12.1.12. Zwei Spezialfälle dieser Formel sind:

$$\begin{aligned} \mathbb{E}X &= g'_X(1), \quad n = 1, \\ \mathbb{E}[X(X-1)] &= g''_X(1), \quad n = 2. \end{aligned}$$

Damit lässt sich nun die Varianz berechnen:

$$\begin{aligned} \text{Var } X &= \mathbb{E}[X^2] - (\mathbb{E}X)^2 \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}X - (\mathbb{E}X)^2 \\ &= g''_X(1) + g'_X(1) - (g'_X(1))^2. \end{aligned}$$

BEMERKUNG 12.1.13. Die erzeugende Funktion ist im Allgemeinen nur für $|z| \leq 1$ definiert. Daher müssen wir uns die Ableitung der erzeugenden Funktion an den Stellen ± 1 als eine einseitige Ableitung vorstellen.

BEWEIS VON SATZ 12.1.10. Wir leiten die Taylor-Reihe $g_X(z)$ n -mal termweise ab:

$$g_X^{(n)}(z) = \sum_{k=0}^{\infty} p_k (z^k)^{(n)} = \sum_{k=n}^{\infty} p_k \cdot k(k-1) \dots (k-n+1) z^{k-n}.$$

Nun setzen wir $z = 1$ ein:

$$g_X^{(n)}(1) = \sum_{k=0}^{\infty} p_k \cdot k(k-1) \dots (k-n+1) = \mathbb{E}[X(X-1) \dots (X-n+1)].$$

Im letzten Schritt wurde die Transformationsformel für den Erwartungswert verwendet. \square

BEISPIEL 12.1.14. Sei $X \sim \text{Poi}(\lambda)$. Wir haben bereits gezeigt, dass $g_X(z) = e^{\lambda(z-1)}$. Es gilt also für jedes $n \in \mathbb{N}$

$$\mathbb{E}[X(X-1) \dots (X-n+1)] = g_X^{(n)}(1) = (e^{\lambda(z-1)})^{(n)} \Big|_{z=1} = \lambda^n e^{\lambda(z-1)} \Big|_{z=1} = \lambda^n.$$

Damit kann man den Erwartungswert und die Varianz von X berechnen:

$$\begin{aligned} \mathbb{E}X &= g'_X(1) = \lambda, \\ \text{Var } X &= g''_X(1) + g'_X(1) - (g'_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

12.2. Summen mit einer zufälligen Anzahl von Summanden

BEISPIEL 12.2.1. Einer Versicherung werden N Schäden gemeldet. Dabei sei N eine Zufallsvariable mit Werten in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Die einzelnen Schadenhöhen seien ebenfalls zufällig und mit X_1, X_2, \dots bezeichnet. Der Gesamtschaden beläuft sich also auf

$$S = X_1 + \dots + X_N.$$

Dabei besteht die Summe S aus einer zufälligen Anzahl (nämlich N) an Summanden. Die Summanden sind ebenfalls zufällig. Wie bestimmt man die Verteilung von S ?

SATZ 12.2.2. Die folgenden Bedingungen seien erfüllt:

- (1) N, X_1, X_2, \dots sind unabhängige Zufallsvariablen mit Werten in \mathbb{N}_0 .
- (2) X_1, X_2, \dots sind identisch verteilt.

Dann lässt sich die erzeugende Funktion der Summe $S = X_1 + \dots + X_N$ wie folgt berechnen:

$$g_S(z) = g_N(g_{X_1}(z)), \quad z \in [-1, 1].$$

BEMERKUNG 12.2.3. Ist $N = n$ konstant (und nicht zufällig), so erhalten wir die Formel $g_S(z) = (g_{X_1}(z))^n$. Diese Formel folgt direkt aus Satz 12.1.6, denn

$$g_S(z) = g_{X_1 + \dots + X_n}(z) = g_{X_1}(z) \cdot \dots \cdot g_{X_n}(z) = (g_{X_1}(z))^n.$$

BEWEIS. Wir benutzen die Notation $p_n = \mathbb{P}[N = n]$ für $n \in \mathbb{N}_0$. Nun verwenden wir die Formel von der totalen Wahrscheinlichkeit:

$$\begin{aligned} g_S(z) &= \sum_{k=0}^{\infty} \mathbb{P}[S = k] \cdot z^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} \mathbb{P}[X_1 + \dots + X_N = k | N = n] \cdot p_n \right) \cdot z^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} \mathbb{P}[X_1 + \dots + X_n = k | N = n] \cdot p_n \right) \cdot z^k. \end{aligned}$$

Da nun das Ereignis $\{N = n\}$ vom Ereignis $\{X_1 + \dots + X_n = k\}$ unabhängig ist, erhalten wir

$$\begin{aligned} g_S(z) &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} \mathbb{P}[X_1 + \dots + X_n = k] \cdot p_n \right) \cdot z^k \\ &= \sum_{n=0}^{\infty} p_n \cdot \left(\sum_{k=0}^{\infty} z^k \cdot \mathbb{P}[X_1 + \dots + X_n = k] \right) \\ &= \sum_{n=0}^{\infty} p_n \cdot g_{X_1 + \dots + X_n}(z). \end{aligned}$$

Nun sind die Zufallsvariablen X_1, \dots, X_n unabhängig und identisch verteilt, somit erhalten wir

$$g_S(z) = \sum_{n=0}^{\infty} p_n (g_{X_1}(z))^n = g_N(g_{X_1}(z)).$$

□

KOROLLAR 12.2.4 (Wald-Identität). *Unter den gleichen Voraussetzungen wie in Satz 12.2.2 gilt*

$$\mathbb{E}[X_1 + \dots + X_N] = \mathbb{E}[N] \cdot \mathbb{E}[X_1].$$

BEWEIS. Übung.

□

BEMERKUNG 12.2.5. Im Beispiel mit der Versicherung kann man den erwarteten Gesamtschaden berechnen, indem man die erwartete Anzahl an gemeldeten Schäden mit der zu erwarteten jeweiligen Schadenshöhe multipliziert.

12.3. Verzweigungsprozesse

Einen *Verzweigungsprozess* (auch *Galton–Watson Prozess* genannt) kann man sich als ein Modell für eine Kettenreaktion vorstellen. Hier folgt die Beschreibung dieses Modells. In Generation 0 gibt es 1 Teilchen. Dieses Teilchen erzeugt eine zufällige Anzahl Töchterteilchen, die zu Generation 1 gehören. Jedes Teilchen in Generation 1 erzeugt eine zufällige Anzahl Töchterteilchen, die zu Generation 2 gehören, usw. Dabei wird vorausgesetzt, dass sich alle

Teilchen unabhängig voneinander verhalten. (Die Anzahl der Töchterteilchen, die ein Teilchen produziert, ist also unabhängig davon, wieviele Töchterteilchen die anderen Teilchen produzieren). Wir bezeichnen mit p_k die Wahrscheinlichkeit, dass ein Teilchen k Töchter erzeugt. Wir nehmen an, dass sich diese Wahrscheinlichkeiten nicht von Teilchen zu Teilchen ändern.

Nun geben wir eine präzisere Beschreibung des Modells. Es sei eine Zahlenfolge p_0, p_1, \dots gegeben, so dass

- (1) $p_0, p_1, \dots \geq 0$.
- (2) $p_0 + p_1 + \dots = 1$.

Es seien $X_{i,n}$, $i \in \mathbb{N}$, $n \in \mathbb{N}_0$, unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{P}[X_{i,n} = k] = p_k$. Dabei soll $X_{i,n}$ die Anzahl der Töchterteilchen des Teilchens i in Generation n bezeichnen. Definiere nun Zufallsvariablen Z_0, Z_1, \dots induktiv durch $Z_0 = 1$ und

$$(12.3.1) \quad Z_{n+1} = \sum_{i=0}^{Z_n} X_{i,n}.$$

Somit ist Z_n die Anzahl der Teilchen in Generation n . Wie bestimmt man nun die Verteilung von Z_n ?

SATZ 12.3.1. Sei $g(t) = \sum_{k=0}^{\infty} p_k z^k$. Dann gilt für die erzeugende Funktion von Z_n :

$$g_{Z_n}(t) = g(g(\dots g(t) \dots)) \quad (n\text{-mal}).$$

BEWEIS. Für $n = 0$ gilt $g_{Z_0}(t) = t$, denn $Z_0 = 1$. Die erzeugende Funktion von $X_{i,n}$ ist g . Wendet man nun Satz 12.2.2 auf die Formel (12.3.1) an, so erhält man

$$g_{Z_{n+1}}(t) = g_{Z_n}(g(t)).$$

Wendet man das induktiv an, so erhält man die Aussage des Satzes. □

Wir können den Erwartungswert von Z_n bestimmen.

SATZ 12.3.2. Sei $\mu := \mathbb{E}[X_{1,0}] = \sum_{k=0}^{\infty} p_k k$. Dann gilt

$$\mathbb{E}[Z_n] = \mu^n.$$

BEMERKUNG 12.3.3. Diese Formel ist nicht überraschend. Jedes Teilchen erzeugt im Durchschnitt μ Töchterteilchen für die nächste Generation. Jede Generation ist also im Durchschnitt μ -Mal so groß wie die vorherige. Daher gilt $\mathbb{E}[Z_n] = \mu^n$.

BEWEIS. Für $n = 0$ ergibt sich nach Definition $\mathbb{E}[Z_0] = 1 = \mu^0$. Wendet man die Wald-Identität auf die Formel (12.3.1) an, so erhält man

$$\mathbb{E}[Z_{n+1}] = \mathbb{E}[Z_n] \cdot \mathbb{E}[X_{1,n}] = \mathbb{E}[Z_n] \cdot \mu.$$

Wendet man das induktiv an, so erhält man die Aussage des Satzes. □

DEFINITION 12.3.4. Die *Aussterbewahrscheinlichkeit* q ist definiert durch

$$q = \mathbb{P}[\exists n \in \mathbb{N} : Z_n = 0].$$

Wie lässt sich diese Aussterbewahrscheinlichkeit berechnen? Zunächst einmal gibt es einen trivialen Fall: Ist $p_0 = 0$, so erzeugt jedes Teilchen mindestens ein Töchterteilchen und der Verzweigungsprozess kann nicht aussterben. Somit gilt in diesem Fall $q = 0$. Im nächsten Satz betrachten wir den Fall $p_0 > 0$.

SATZ 12.3.5. Sei $p_0 > 0$. Dann gilt:

- (1) Ist $\mu < 1$ (der subkritische Fall), so gilt $q = 1$.
- (2) Ist $\mu = 1$ (der kritische Fall), so gilt $q = 1$.
- (3) Ist $\mu > 1$ (der superkritische Fall), so ist q die einzige Lösung der Gleichung $g(q) = q$ mit $q < 1$. Es gilt also $0 < q < 1$.

BEMERKUNG 12.3.6. Im subkritischen Fall $\mu < 1$ ist jede Generation im Durchschnitt kleiner als die vorherige. Deshalb ist es nicht überraschend, dass der Verzweigungsprozess mit Wahrscheinlichkeit 1 aussirbt.

BEMERKUNG 12.3.7. Im kritischen Fall $\mu = 1$ ist jede Generation im Durchschnitt genauso groß, wie die vorherige. Es gilt also $\mathbb{E}[Z_n] = 1$ für jedes $n \in \mathbb{N}_0$. Dennoch stirbt der Prozess mit Wahrscheinlichkeit 1 aus. Das kann man sich intuitiv so vorstellen: eine Generation, die im Durchschnitt aus einem Teilchen besteht hat eine gewisse positive Wahrscheinlichkeit, kein einziges Töchterteilchen zu produzieren. Da diese Wahrscheinlichkeit nun zu jedem Zeitpunkt besteht, wird irgendwann tatsächlich kein einziges Töchterteilchen erzeugt und der Prozess stirbt aus. Im kritischen Fall gilt $\lim_{n \rightarrow \infty} Z_n = 0$ f.s. (weil Z_n mit Wahrscheinlichkeit 1 ab irgendwann gleich 0 ist). Dabei ist $\mathbb{E}[Z_n] = 1$ für jedes $n \in \mathbb{N}_0$. Hier sehen wir, dass nicht immer $\mathbb{E}[\lim_{n \rightarrow \infty} Z_n] = \lim_{n \rightarrow \infty} \mathbb{E}[Z_n]$ gelten muss.

BEMERKUNG 12.3.8. Im superkritischen Fall $\mu > 1$ ist jede Generation im Durchschnitt größer, als die vorherige. Die Aussterbewahrscheinlichkeit ist kleiner als 1. Es sei aber bemerkt, dass auch ein superkritischer Verzweigungsprozess mit positiver Wahrscheinlichkeit aussterben kann. Das geschieht zum Beispiel dann, wenn das allererste Teilchen keine Nachkommen produziert. Das passiert mit Wahrscheinlichkeit p_0 , also ist $q \geq p_0$.

BEWEIS VON SATZ 12.3.5. Ist die n -te Generation ausgestorben, so sind auch alle nachfolgenden Generationen leer. Deshalb gelten folgende Inklusionen von Ereignissen:

$$\{Z_1 = 0\} \subset \{Z_2 = 0\} \subset \{Z_3 = 0\} \subset \dots$$

Es sei q_n die Wahrscheinlichkeit, dass die n -te Generation leer ist:

$$q_n = \mathbb{P}[Z_n = 0] = g_{Z_n}(0) = g(g(\dots g(0))\dots) \quad (n\text{-mal}).$$

Dabei ist $g(z) = \sum_{k=0}^{\infty} p_k z^k$ die erzeugende Funktion der Anzahl der Nachkommen eines Teilchens. Mit dem Satz über die Stetigkeit der Wahrscheinlichkeit gilt

$$q = \mathbb{P}[\cup_{n=1}^{\infty} \{Z_n = 0\}] = \lim_{n \rightarrow \infty} \mathbb{P}[Z_n = 0] = \lim_{n \rightarrow \infty} g(g(\dots g(0))\dots) \quad (n\text{-mal}).$$

Der Rest des Beweises ergibt sich aus dem Bild. □

12.4. Momenterzeugende Funktion (Laplace-Transformierte)

Die erzeugende Funktion hat einen Nachteil: Sie kann nur für Zufallsvariablen, die ganzzahlige, nicht-negative Werte annehmen, definiert werden. Für Zufallsvariablen, die diese Bedingung nicht erfüllen, müssen wir etwas anderes definieren.

DEFINITION 12.4.1. Sei X eine Zufallsvariable mit Werten in \mathbb{R} . Dann heißt

$$m_X(t) = \mathbb{E}[e^{tX}] \in (0, \infty], \quad t \in \mathbb{R},$$

die *Laplace-Transformierte* (oder die *momenterzeugende Funktion*) von X .

BEISPIEL 12.4.2. Sei X eine absolut stetige Zufallsvariable mit Dichte f_X . Dann ist die Laplace-Transformierte gegeben durch

$$m_X(t) = \int_{\mathbb{R}} e^{ty} f_X(y) dy.$$

BEMERKUNG 12.4.3. Es gilt immer $m_X(0) = 1$.

BEMERKUNG 12.4.4. Ein Nachteil der Laplace-Transformierten ist, dass sie auch den Wert ∞ annehmen kann. Es sei X z.B. Cauchy-verteilt, d.h. absolut stetig mit Dichte

$$f_X(y) = \frac{1}{\pi} \cdot \frac{1}{1 + y^2}, \quad y \in \mathbb{R}.$$

Dann gilt für die Laplace-Transformierte:

$$m_X(t) = \int_{\mathbb{R}} e^{ty} \cdot \frac{1}{\pi} \cdot \frac{1}{1 + y^2} dy = \begin{cases} +\infty, & t \neq 0, \\ 1, & t = 0. \end{cases}$$

Die Laplace-Transformierte ist also überall gleich $+\infty$ außer an der Stelle 0. Man kann auch andere Verteilungen konstruieren, die die gleiche Laplace-Transformierte haben. Sei z.B. X' absolut stetig mit Dichte $f_{X'}(t) = \frac{1}{2}y^{-2} \mathbb{1}_{|y|>1}$. Dann gilt $m_X(t) = m_{X'}(t)$ für alle $t \in \mathbb{R}$. Wir sehen, dass die Laplace-Transformierte die Verteilung einer Zufallsvariable nicht eindeutig bestimmt.

Im weiteren werden wir voraussetzen, dass die Laplace-Transformierte endlich in einem Intervall $(-\varepsilon, \varepsilon)$ ist. Im nächsten Satz berechnen wir die Momente von X mithilfe der Laplace-Transformierten.

SATZ 12.4.5. Sei X eine Zufallsvariable mit $m_X(t) < \infty$ für alle $t \in (-\varepsilon, \varepsilon)$ mit $\varepsilon > 0$. Dann gilt:

$$\mathbb{E}[X^n] = m_X^{(n)}(0).$$

BEISPIEL 12.4.6. Für $n = 1$ und $n = 2$ ergibt sich

$$m_X'(0) = \mathbb{E}[X], \quad m_X''(0) = \mathbb{E}[X^2].$$

Daraus ergibt sich die Formel für die Varianz:

$$\text{Var } X = m_X''(0) - (m_X'(0))^2.$$

BEMERKUNG 12.4.7. Dieser Satz erklärt auch die Bezeichnung “momenterzeugende Funktion” für m_X . Die Momente können als Ableitungen der momenterzeugenden Funktion an der Stelle 0 abgelesen werden.

IDEES DES BEWEISES VON SATZ 12.4.5. Wir betrachten die n -te Ableitung von $m_X(t)$:

$$m_X^{(n)}(t) = (\mathbb{E}[e^{tX}])^{(n)} = \mathbb{E} \left[(e^{tX})^{(n)} \right] = \mathbb{E}[X^n e^{tX}].$$

Setzen wir nun $t = 0$ ein, so erhalten wir $m_X^{(n)}(0) = \mathbb{E}[X^n]$. Allerdings haben wir hier den Erwartungswert mit dem Ableitungsoperator vertauscht. Dieser Schritt bedarf einer Begründung. Daher werden wir den Beweis auf eine andere Weise führen.

BEWEIS VON SATZ 12.4.5. SCHRITT 1. Wir entwickeln die Funktion e^{tX} in eine Taylor-Reihe:

$$e^{tX} = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!} = \lim_{m \rightarrow \infty} \sum_{k=0}^m \frac{(tX)^k}{k!}, \quad \text{wobei } S_m := \sum_{k=0}^m \frac{(tX)^k}{k!}.$$

Die Summen S_m können wie folgt abgeschätzt werden:

$$|S_m| = \left| \sum_{k=0}^m \frac{(tX)^k}{k!} \right| \leq \sum_{k=0}^m \frac{|tX|^k}{k!} = e^{|tX|} \leq S,$$

wobei $S = e^{tX} + e^{-tX}$. Für $t \in (-\varepsilon, \varepsilon)$ gilt $\mathbb{E}S < \infty$ nach Voraussetzung des Satzes. Wir haben gezeigt, dass $\lim_{m \rightarrow \infty} S_m = e^{tX}$ fast sicher (sogar sicher) und $|S_m| \leq S$ mit $\mathbb{E}S < \infty$. Daher können wir den Satz von der majorisierten Konvergenz (Maßtheorie) anwenden:

$$\lim_{m \rightarrow \infty} \mathbb{E}[S_m] = \mathbb{E}[e^{tX}].$$

SCHRITT 2. Es gilt für jedes $t \in (-\varepsilon, \varepsilon)$:

$$m_X(t) = \mathbb{E}[e^{tX}] = \lim_{m \rightarrow \infty} \mathbb{E}[S_m] = \lim_{m \rightarrow \infty} \mathbb{E} \left[\sum_{k=0}^m \frac{(tX)^k}{k!} \right] = \lim_{m \rightarrow \infty} \sum_{k=0}^m \frac{t^k}{k!} \mathbb{E}[X^k] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k].$$

SCHRITT 3. Dort, wo eine Taylor-Reihe konvergiert, kann man sie beliebig oft termweise ableiten. Es ergibt sich

$$m_X^{(n)}(t) = \sum_{k=n}^{\infty} \frac{k(k-1)\dots(k-n+1)}{k!} t^{k-n} \mathbb{E}[X^k].$$

Setzt man in diese Formel $t = 0$ ein, so verschwinden alle Summanden bis auf den ersten (mit $k = n$). Daraus ergibt sich $m_X^{(n)}(0) = \mathbb{E}[X^n]$. \square

BEISPIEL 12.4.8. Sei $X \sim \text{Poi}(\lambda)$, d.h. $\mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$ für $k \in \mathbb{N}_0$. Für die Laplace-Transformierte von X erhalten wir

$$m_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} \mathbb{P}[X = k] = \sum_{k=0}^{\infty} e^{tk} \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = e^{-\lambda} \cdot e^{e^t \lambda} = e^{\lambda(e^t - 1)}.$$

Damit lässt sich der Erwartungswert berechnen:

$$\mathbb{E}X = m'_X(0) = \left(e^{\lambda(e^t - 1)} \right)' \Big|_{t=0} = \lambda e^t e^{\lambda(e^t - 1)} \Big|_{t=0} = \lambda.$$

SATZ 12.4.9. Seien X, Y unabhängige Zufallsvariablen. Dann gilt:

$$m_{X+Y}(t) = m_X(t) \cdot m_Y(t).$$

BEWEIS. Mit der Definition der Laplace-Transformierten erhalten wir:

$$m_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} \cdot e^{tY}] = \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}] = m_X(t) \cdot m_Y(t).$$

Dabei haben wir benutzt, dass X und Y (und somit auch e^{tX} und e^{tY}) unabhängig sind. \square

12.5. Charakteristische Funktion (Fourier-Transformierte)

Ein Nachteil der Laplace-Transformierten ist, dass sie auch unendlich sein kann (siehe das Beispiel mit der Cauchy-Verteilung). Außerdem ist die Verteilung einer Zufallsvariable nicht immer eindeutig durch die Laplace-Transformierte festgelegt. Wir werden deshalb eine andere Transformation betrachten, die sich von der Laplace-Transformation dadurch unterscheidet, dass man im Exponenten e^{tX} noch die Komplexe Zahl $i = \sqrt{-1}$ hinzunimmt. Damit erreicht man, dass die Transformierte immer endlich ist.

DEFINITION 12.5.1. Sei X eine Zufallsvariable. Dann heißt

$$\varphi_X(t) = \mathbb{E}[e^{itX}] \in \mathbb{C}, \quad t \in \mathbb{R},$$

die *charakteristische Funktion* (oder die *Fourier-Transformierte*) von X .

BEMERKUNG 12.5.2. Der Erwartungswert von e^{itX} ist

$$\mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX) + i \sin(tX)] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)].$$

Die charakteristische Funktion $\varphi_X(t)$ existiert also für jedes X und jedes $t \in \mathbb{R}$, denn $|\cos(tX)| \leq 1$ und $|\sin(tX)| \leq 1$.

BEMERKUNG 12.5.3. Ist X diskret, mit den Werten y_1, y_2, \dots und den dazugehörigen Wahrscheinlichkeiten p_1, p_2, \dots , so gilt für die charakteristische Funktion von X :

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \sum_{k=1}^{\infty} e^{ity_k} p_k.$$

Ist X absolut stetig mit Dichte f_X , so gilt

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{ity} f_X(y) dy.$$

BEISPIEL 12.5.4. Nimmt X nur die zwei Werte $+1$ und -1 mit einer Wahrscheinlichkeit von jeweils $1/2$ an, so ist die charakteristische Funktion von X gegeben durch

$$\varphi_X(t) = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it} = \cos(t).$$

SATZ 12.5.5. Sei X eine Zufallsvariable. Dann hat die charakteristische Funktion φ_X die folgenden Eigenschaften:

- (1) $\varphi_X(0) = 1$ und $|\varphi_X(t)| \leq 1$ für alle $t \in \mathbb{R}$.
- (2) $\varphi_X(-t) = \overline{\varphi_X(t)}$ für alle $t \in \mathbb{R}$.
- (3) Die Funktion φ_X ist gleichmäßig stetig auf \mathbb{R} .

(4) Die Funktion φ_X ist positiv semidefinit, d.h., für alle $t_1, \dots, t_n \in \mathbb{R}$ und für alle $c_1, \dots, c_n \in \mathbb{C}$ gilt:

$$\sum_{k,j=1}^n c_k \bar{c}_j \cdot \varphi_X(t_k - t_j) \geq 0.$$

Mit anderen Worten: $(\varphi_X(t_k - t_j))_{k,j=1}^n$ ist eine positiv semidefinite Matrix.

BEWEIS VON 1. Für $t = 0$ gilt $\varphi_X(0) = \mathbb{E}[e^{i \cdot 0 \cdot X}] = \mathbb{E}1 = 1$. Für ein beliebiges $t \in \mathbb{R}$ gilt:

$$|\varphi_X(t)| = |\mathbb{E}e^{itX}| \leq \mathbb{E}|e^{itX}| = \mathbb{E}1 = 1.$$

BEWEIS VON 2. Wir betrachten die charakteristische Funktion an der Stelle $-t$:

$$\varphi_X(-t) = \mathbb{E}[e^{-itX}] = \mathbb{E}[\cos(tX) - i \sin(tX)] = \mathbb{E}[\cos(tX)] - i\mathbb{E}[\sin(tX)].$$

Auf der anderen Seite gilt

$$\overline{\varphi_X(t)} = \overline{\mathbb{E}[e^{itX}]} = \overline{\mathbb{E}[\cos(tX) + i \sin(tX)]} = \overline{\mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]}.$$

Somit gilt $\varphi_X(-t) = \overline{\varphi_X(t)}$.

BEWEIS VON 3. Seien $t, h \in \mathbb{R}$. Dann gilt

$$|\varphi_X(t+h) - \varphi_X(t)| = |\mathbb{E}[e^{i(t+h)X} - e^{itX}]| \leq \mathbb{E}|e^{itX}(e^{ihX} - 1)| = \mathbb{E}|e^{ihX} - 1| \stackrel{\text{def}}{=} g(h).$$

Nun müssen wir zeigen, dass $\lim_{h \rightarrow 0} g(h) = 0$. Zunächst bemerken wir, dass für jeden Ausgang $\omega \in \Omega$: $\lim_{h \rightarrow 0} |e^{ihX(\omega)} - 1| = 0$. Somit gilt:

$$|e^{ihX} - 1| \xrightarrow[h \rightarrow 0]{f.s.} 0.$$

Außerdem haben wir die Abschätzung $|e^{ihX(\omega)} - 1| \leq 2$. Der Satz von der majorisierten Konvergenz ergibt

$$\lim_{h \rightarrow 0} g(h) = \lim_{h \rightarrow 0} \mathbb{E}|e^{ihX} - 1| = \mathbb{E}\left[\lim_{h \rightarrow 0} |e^{ihX} - 1|\right] = \mathbb{E}0 = 0.$$

Es folgt, dass

$$\limsup_{h \rightarrow 0} \sup_{t \in \mathbb{R}} |\varphi_X(t+h) - \varphi_X(t)| = 0.$$

Also ist φ_X gleichmäßig stetig.

BEWEIS VON 4. Für beliebige t_1, \dots, t_n und $c_1, \dots, c_n \in \mathbb{C}$ gilt

$$\begin{aligned} \sum_{k,j=1}^n c_k \bar{c}_j \varphi_X(t_k - t_j) &= \sum_{k,j=1}^n c_k \bar{c}_j \mathbb{E}[e^{i(t_k - t_j)X}] \\ &= \mathbb{E}\left[\sum_{k,j=1}^n c_k \bar{c}_j e^{i(t_k - t_j)X}\right] \\ &= \mathbb{E}\left[\left(\sum_{k=1}^n c_k e^{it_k X}\right) \cdot \overline{\left(\sum_{j=1}^n c_j e^{it_j X}\right)}\right] \end{aligned}$$

Mit der Bezeichnung $A := \sum_{k=1}^n c_k e^{it_k X}$ erhalten wir somit

$$\sum_{k,j=1}^n c_k \bar{c}_j \varphi_X(t_k - t_j) = \mathbb{E}[A\bar{A}] = \mathbb{E}[|A|^2] \geq 0,$$

denn $|A|^2 \geq 0$. □

Es gilt auch eine Umkehrung des soeben bewiesenen Satzes.

SATZ 12.5.6 (Satz von Bochner). *Sei $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ eine Funktion mit den 4 Eigenschaften aus obigem Satz. Dann gibt es eine Zufallsvariable X mit $\varphi_X = \varphi$.*

OHNE BEWEIS.

Im folgenden Lemma berechnen wir die charakteristische Funktion einer linearen Transformation.

LEMMA 12.5.7. *Seien $a, b \in \mathbb{R}$ und X eine Zufallsvariable. Dann gilt:*

$$\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at).$$

BEWEIS. Wir benutzen die Definition der charakteristischen Funktion:

$$\varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = \mathbb{E}[e^{itb} \cdot e^{i(at)X}] = e^{itb} \cdot \mathbb{E}[e^{i(at)X}] = e^{itb} \cdot \varphi_X(at).$$

□

SATZ 12.5.8. *Sei $X \sim N(\mu, \sigma^2)$. Dann gilt*

$$\varphi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

BEMERKUNG 12.5.9. Spezialfall: Für $X \sim N(0, 1)$ gilt $\varphi_X(t) = e^{-\frac{t^2}{2}}$. Somit stimmt in diesem Fall die charakteristische Funktion mit der Dichte bis auf einen Vorfaktor überein.

BEWEIS. SCHRITT 1. Wir betrachten den Fall, dass $X \sim N(0, 1)$. Zuerst berechnen wir die Laplace-Transformierte:

$$\begin{aligned} m_X(t) &= \int_{-\infty}^{\infty} e^{ty} f_X(y) dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} \cdot e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2} dy \\ &= e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \\ &= e^{\frac{t^2}{2}}, \end{aligned}$$

wobei wir die neue Variable $z = y - t$ eingeführt haben.

SCHRITT 2. Aus der Laplace-Transformierten berechnen wir durch Hinzufügen des Faktors i die charakteristische Funktion:

$$\varphi_X(t) = m_X(it) = e^{\frac{(it)^2}{2}} = e^{-\frac{t^2}{2}}.$$

SCHRITT 3. Sei nun $X \sim N(\mu, \sigma^2)$ mit beliebigen Parametern μ, σ^2 . Dann haben wir die Darstellung $X = \sigma Y + \mu$, wobei $Y \sim N(0, 1)$. Mit Lemma 12.5.7 gilt dann:

$$\varphi_X(t) = e^{it\mu} \cdot \varphi_Y(\sigma t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

□

SATZ 12.5.10. Seien X, Y unabhängige Zufallsvariablen. Dann gilt

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t).$$

BEWEIS. Übung.

□

Nun berechnen wir Momente mithilfe der charakteristischen Funktion.

SATZ 12.5.11. Sei X eine Zufallsvariable mit $\mathbb{E}|X|^n < \infty$ für ein $n \in \mathbb{N}$. Dann ist die charakteristische Funktion φ_X n -mal stetig differenzierbar und es gilt:

$$\mathbb{E}[X^n] = i^{-n} \varphi_X^{(n)}(0).$$

Zum Beweis benötigen wir ein Lemma.

LEMMA 12.5.12. Für $y \in \mathbb{R}$ gilt die Ungleichung $|e^{iy} - 1| \leq |y|$.

BEWEIS. Für $y \geq 0$ gilt

$$|e^{iy} - 1| = \left| \int_0^y e^{is} ds \right| \leq \int_0^y |e^{is}| ds = y.$$

Für $y \leq 0$ benutzen wir, dass $|e^{iy} - 1| = |e^{-iy} - 1|$.

□

BEWEIS VON SATZ 12.5.11. SCHRITT 1. Wir betrachten den Fall $n = 1$. Sei also $\mathbb{E}|X| < \infty$. Wir werden zeigen, dass φ_X stetig differenzierbar ist und dass

$$(12.5.1) \quad \varphi_X'(t) = \mathbb{E}[iX e^{itX}].$$

Wir stellen den Differenzenquotienten der Funktion φ_X auf:

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathbb{E} \left[e^{itX} \frac{e^{ihX} - 1}{h} \right].$$

Mit der Taylor-Reihe der Exponentialfunktion erhalten wir

$$\lim_{h \rightarrow 0} \left(e^{itX} \frac{e^{ihX} - 1}{h} \right) = iX e^{itX}.$$

Außerdem erhalten wir mit Lemma 12.5.12 die Abschätzung

$$\left| e^{itX} \frac{e^{ihX} - 1}{h} \right| = \left| \frac{e^{ihX} - 1}{h} \right| \leq |X|.$$

Wir haben vorausgesetzt, dass $\mathbb{E}|X| < \infty$. Also können wir den Satz von der majorisierten Konvergenz benutzen:

$$\lim_{h \rightarrow 0} \frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathbb{E} [iX e^{itX}].$$

Somit ist φ_X differenzierbar und die Formel (12.5.1) gilt. Außerdem ist die Funktion $\mathbb{E}[iX e^{itX}]$ stetig. Der Beweis dafür ist identisch mit dem Beweis, dass φ_X stetig ist.

Nun setzen wir $t = 0$ in (12.5.1) ein: $\varphi'_X(0) = \mathbb{E}[iX]$. Das beweist den Satz für $n = 1$.

SCHRITT 2. Sei nun $n \in \mathbb{N}$ beliebig und $\mathbb{E}|X|^n < \infty$. Wendet man die Methode von Schritt 1 induktiv an, so erhält man die Formel

$$\varphi_X^{(n)}(t) = \mathbb{E}[(iX)^n e^{itX}].$$

Setzt man $t = 0$ ein, so erhält man die Aussage des Satzes. □

BEISPIEL 12.5.13. Sei $X \sim N(0, 1)$ standardnormalverteilt. Die Taylor-Entwicklung der charakteristischen Funktion lautet: $\varphi_X(t) = e^{-t^2/2} = 1 - \frac{1}{2}t^2 + \dots$. Damit gilt

$$\begin{aligned}\varphi'_X(0) &= 0 \Rightarrow \mathbb{E}X = 0, \\ \varphi''_X(0) &= -1 \Rightarrow \mathbb{E}X^2 = i^{-2} \cdot (-1) = 1.\end{aligned}$$

Also gilt für die Varianz von X : $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = 1$.

Wenn die charakteristische Funktion einer Zufallsvariable bekannt ist, dann kann man die Momente dieser Zufallsvariable ausrechnen. Wie kann man aber die ganze Verteilung, also die Verteilungsfunktion der Zufallsvariable, wiederherstellen?

SATZ 12.5.14 (Umkehrformel). Sei Z eine Zufallsvariable mit charakteristischer Funktion φ_Z und Verteilungsfunktion F_Z . Dann gilt für alle $a, b \in \mathbb{R}$ mit $a < b$ und $\mathbb{P}[Z = a] = \mathbb{P}[Z = b] = 0$ die Formel

$$\mathbb{P}[a \leq Z \leq b] = F_Z(b) - F_Z(a) = \frac{1}{2\pi} \lim_{c \rightarrow \infty} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi_Z(t) dt.$$

BEMERKUNG 12.5.15. Die Eigenschaft $\mathbb{P}[Z = a] = 0$ bedeutet, dass a kein Atom von X ist, bzw. dass die Verteilungsfunktion F_Z an der Stelle a stetig ist.

BEWEIS. SCHRITT 1. Wir verwenden das Fresnel-Integral

$$\lim_{c \rightarrow +\infty} S(c) = \frac{\pi}{2}, \text{ wobei } S(c) := \int_0^c \frac{\sin(x)}{x} dx.$$

Außerdem benötigen wir die Vorzeichenfunktion:

$$\text{sgn}(\theta) = \begin{cases} 1, & \theta > 0, \\ 0, & \theta = 0, \\ -1, & \theta < 0. \end{cases}$$

Sei $\theta \in \mathbb{R} \setminus \{0\}$, dann gilt (wobei wir die Variable $s = |\theta|t$ einführen):

$$\int_0^c \frac{\sin(\theta t)}{t} dt = \text{sgn}(\theta) \int_0^{|\theta|c} \frac{\sin(s)}{s/|\theta|} \frac{ds}{|\theta|} = \text{sgn}(\theta) \int_0^{|\theta|c} \frac{\sin(s)}{s} ds = \text{sgn}(\theta) \cdot S(|\theta|c).$$

Die Formel gilt auch für $\theta = 0$, denn beide Seiten sind dann gleich 0.

SCHRITT 2. Nun zum Beweis der Umkehrformel. Betrachte das Integral

$$\begin{aligned} I(c) &:= \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi_z(t) dt \\ &= \int_{-c}^c \mathbb{E} \left[\frac{e^{it(Z-a)} - e^{it(Z-b)}}{it} \right] dt \\ &= \mathbb{E} \left[\int_{-c}^c \frac{e^{it(Z-a)} - e^{it(Z-b)}}{it} dt \right], \end{aligned}$$

wobei wir im letzten Schritt den Satz von Fubini benutzen durften, denn

$$\left| \frac{e^{it(Z-a)} - e^{it(Z-b)}}{it} \right| = \left| \frac{e^{it(b-a)} - 1}{it} \right| \leq b - a.$$

Da $e^{ix} = \cos x + i \sin x$, wobei \cos eine gerade und \sin eine ungerade Funktion ist, erhalten wir

$$\begin{aligned} I(c) &= 2 \cdot \mathbb{E} \left[\int_0^c \frac{\sin(t(Z-a)) - \sin(t(Z-b))}{t} dt \right] \\ &= 2 \cdot \mathbb{E} [\operatorname{sgn}(z-a)S(|Z-a|c) - \operatorname{sgn}(z-b)S(|Z-b|c)]. \end{aligned}$$

Für die Funktion unter dem Erwartungswert gilt (mit dem Fresnel-Integral aus Schritt 1)

$$\lim_{c \rightarrow +\infty} (\operatorname{sgn}(z-a)S(|z-a|c) - \operatorname{sgn}(z-b)S(|z-b|c)) = \pi \psi_{a,b}(z),$$

wobei

$$\psi_{a,b}(z) = \frac{\pi}{2} (\operatorname{sgn}(z-a) - \operatorname{sgn}(z-b)) = \begin{cases} 0, & z < a, \\ \frac{1}{2}, & z = a, \\ 1, & a < z < b, \\ \frac{1}{2}, & z = b, \\ 0, & z > b. \end{cases}$$

Aus der Existenz eines endlichen Grenzwerts $\lim_{c \rightarrow +\infty} S(c)$ folgt, dass $B := \sup_{c \geq 0} |S(c)| < \infty$. Außerdem gilt $|\operatorname{sgn} t| \leq 1$. Daher haben wir die obere Schranke

$$|\operatorname{sgn}(z-a)S(|z-a|c) - \operatorname{sgn}(z-b)S(|z-b|c)| < 2B.$$

Wir können also den Satz von der majorisierten Konvergenz benutzen und erhalten damit:

$$\begin{aligned} \frac{1}{2\pi} \lim_{c \rightarrow \infty} I(c) &= \frac{2}{2\pi} \lim_{c \rightarrow \infty} \mathbb{E} [\operatorname{sgn}(z-a) \cdot S(|z-a|c) - \operatorname{sgn}(z-b) \cdot S(|z-b|c)] \\ &= \frac{1}{\pi} \mathbb{E} \left[\lim_{c \rightarrow \infty} \operatorname{sgn}(z-a) \cdot S(|z-a|c) - \operatorname{sgn}(z-b) \cdot S(|z-b|c) \right] \\ &= \mathbb{E}[\psi_{a,b}(z)]. \end{aligned}$$

Da nun a und b keine Atome von X sind, ist die rechte Seite gleich $\mathbb{P}[a < z < b]$. □

SATZ 12.5.16 (Eindeutigkeitssatz). *Die charakteristische Funktion bestimmt die Verteilungsfunktion vollständig, das heißt: Sind X, Y Zufallsvariablen mit $\varphi_X(t) = \varphi_Y(t)$ für alle $t \in \mathbb{R}$, dann gilt:*

$$F_X(t) = F_Y(t) \quad \text{für alle } t \in \mathbb{R}.$$

BEWEIS. SCHRITT 1. Die Menge aller Punkte, wo beide Verteilungsfunktionen F_X und F_Y stetig sind, bezeichnen wir mit

$$S = \{t \in \mathbb{R} : F_X \text{ und } F_Y \text{ sind stetig an der Stelle } t\}.$$

Die Menge $\mathbb{R} \setminus S$ ist höchstens abzählbar. Nun benutzen wir die Umkehrformel, die besagt, dass für alle $a, b \in S$ mit $a < b$ gilt:

$$\begin{aligned} F_X(b) - F_X(a) &= \frac{1}{2\pi} \lim_{c \rightarrow \infty} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt \\ &= \frac{1}{2\pi} \lim_{c \rightarrow \infty} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi_Y(t) dt \\ &= F_Y(b) - F_Y(a). \end{aligned}$$

Dabei haben wir benutzt, dass $\varphi_X = \varphi_Y$ laut Voraussetzung. Für alle $a, b \in S$ mit $a < b$ gilt also $F_X(b) - F_X(a) = F_Y(b) - F_Y(a)$.

SCHRITT 2. Da das Komplement von S höchstens abzählbar ist, können wir für jedes $n \in \mathbb{N}$ ein $a_n < -n$ mit $a_n \in S$ finden. Daraus folgt, dass $\lim_{n \rightarrow \infty} a_n = -\infty$. Für alle $b \in S$ gilt nun laut Schritt 1:

$$\begin{aligned} F_X(b) &= F_X(b) - \lim_{n \rightarrow \infty} F_X(a_n) \\ &= \lim_{n \rightarrow \infty} (F_X(b) - F_X(a_n)) \\ &= \lim_{n \rightarrow \infty} (F_Y(b) - F_Y(a_n)) \\ &= F_Y(b) - \lim_{n \rightarrow \infty} F_Y(a_n) \\ &= F_Y(b). \end{aligned}$$

Für alle $b \in S$ gilt also $F_X(b) = F_Y(b)$.

SCHRITT 3. Sei nun $b \in \mathbb{R}$ beliebig. Da das Komplement von S höchstens abzählbar ist, kann man für jedes $n \in \mathbb{N}$ ein $b_n \in (b, b + 1/n) \cap S$ finden. Insbesondere gilt $\lim_{n \rightarrow \infty} b_n = b$ und dabei ist $b_n > b$. Dann gilt:

$$\begin{aligned} F_X(b) &= \lim_{n \rightarrow \infty} F_X(b_n) \quad (\text{da } F_X \text{ rechtsstetig}) \\ &= \lim_{n \rightarrow \infty} F_Y(b_n) \quad (\text{mit Schritt 2}) \\ &= F_Y(b) \quad (\text{da } F_Y \text{ rechtsstetig}). \end{aligned}$$

Für alle $b \in \mathbb{R}$ gilt also $F_X(b) = F_Y(b)$. □

Die Umkehrformel gilt für beliebige Zufallsvariablen. Sie ist aber nicht sehr schön, weswegen wir noch eine einfachere Darstellung zeigen, die aber nur für absolut stetige Zufallsvariablen gilt.

SATZ 12.5.17 (Umkehrformel für die Dichte). Sei Z eine Zufallsvariable, deren charakteristische Funktion φ_Z integrierbar ist, d.h. $\int_{-\infty}^{\infty} |\varphi_Z(t)| dt < \infty$. Dann gilt: Z ist absolut stetig mit Dichte

$$f_Z(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \varphi_Z(t) dt.$$

BEWEIS. Wir bezeichnen die rechte Seite mit

$$g(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \varphi_Z(t) dt.$$

Die Funktion g ist wohldefiniert, denn $|e^{-ity} \varphi_Z(t)| = |\varphi_Z(t)|$ und $|\varphi_Z|$ ist integrierbar (laut Voraussetzung). Außerdem ist g stetig. (Beweis identisch mit dem Beweis für Stetigkeit der charakteristischen Funktion).

Nun wollen wir zeigen, dass die Zufallsvariable Z eine Dichte besitzt und dass g diese Dichte ist. Dazu betrachten wir das Integral

$$\begin{aligned} \int_a^b g(y) dy &= \int_a^b \frac{1}{2\pi} \left(\int_{-\infty}^{\infty} e^{-ity} \varphi_Z(t) dt \right) dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_a^b e^{-ity} \varphi_Z(t) dy \right) dt \quad (\text{Fubini}) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\varphi_Z(t) \cdot \int_a^b e^{-ity} dy \right) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_Z(t) \cdot \frac{e^{-ita} - e^{-itb}}{it} dt \\ &= \mathbb{P}[a < Z < b] \quad (\text{mit Umkehrformel}). \end{aligned}$$

Wir haben gezeigt, dass für alle $a, b \in \mathbb{R}$ mit $a < b$, die Stetigkeitspunkte von F_Z sind,

$$\int_a^b g(y) dy = \mathbb{P}[a \leq Z \leq b]$$

Daraus folgt (genauso wie im Beweis von Satz 12.5.16), dass g die Dichte von Z ist. \square

BEISPIEL 12.5.18. In diesem Beispiel berechnen wir die charakteristische Funktion einer Cauchy-verteilten Zufallsvariable. Sei zuerst X eine Zufallsvariable mit der Dichte $f_X(y) = \frac{1}{2} e^{-|y|}$ (zweiseitige Exponentialverteilung). Die charakteristische Funktion von X ist dann gegeben durch:

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{\infty} \frac{1}{2} e^{-|y|} e^{ity} dy = \frac{1}{2} \left(\int_0^{\infty} e^{-|y|} e^{ity} dy + \int_{-\infty}^0 e^{-|y|} e^{ity} dy \right) \\ &= \frac{1}{2} \left[\frac{1}{1-it} + \frac{1}{1+it} \right] = \frac{1}{1+t^2}. \end{aligned}$$

Nun kann man die Umkehrformel für die Dichte anwenden:

$$\frac{1}{2} e^{-|y|} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \frac{1}{1+t^2} dt.$$

Nun ersetzen wir t durch $-t$ und multiplizieren beide Seiten mit 2:

$$e^{-|y|} = \int_{-\infty}^{\infty} e^{ity} \left(\frac{1}{\pi} \cdot \frac{1}{1+t^2} \right) dt = \int_{-\infty}^{\infty} e^{ity} f_Z(t) dt = \varphi_Z(y),$$

wobei Z eine Cauchy-verteilte Zufallsvariable ist. Wir haben somit gezeigt, dass die charakteristische Funktion einer Cauchy-verteilten Zufallsvariable gleich $\varphi_Z(y) = e^{-|y|}$ ist.

BEISPIEL 12.5.19. Seien $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$ unabhängige, normalverteilte Zufallsvariablen. Dann gilt:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

BEWEIS. Wir haben das bereits mithilfe der Faltungsformel gezeigt. Hier führen wir den Beweis mit charakteristischen Funktionen. Die charakteristischen Funktionen von X_1 und X_2 sind gegeben durch

$$\varphi_{X_k}(t) = e^{i\mu_k t - \frac{1}{2}\sigma_k^2 t^2}, \quad k = 1, 2.$$

Nun berechnen wir die charakteristische Funktion von $X_1 + X_2$:

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \cdot \varphi_{X_2}(t) = e^{i\mu_1 t - \frac{1}{2}\sigma_1^2 t^2} \cdot e^{i\mu_2 t - \frac{1}{2}\sigma_2^2 t^2} = e^{i(\mu_1+\mu_2)t - \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}.$$

Auf der rechten Seite erkennen wir die charakteristische Funktion einer $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ -Verteilung. Mit dem Eindeutigkeitssatz folgt dann, dass

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz besagt, dass eine Summe von sehr vielen unabhängigen identisch verteilten Zufallsvariablen mit endlicher Varianz approximativ normalverteilt ist. Dieser Satz begründet theoretisch die herausragende Rolle, die die Normalverteilung in der Wahrscheinlichkeitstheorie und Statistik spielt. Bevor wir den zentralen Grenzwertsatz beweisen, müssen wir eine weitere Konvergenzart, die *Konvergenz in Verteilung*, definieren.

13.1. Konvergenz in Verteilung

Für eine Zufallsvariable X bezeichnen wir mit

$$S(X) = \{t \in \mathbb{R} : F_X \text{ ist stetig an der Stelle } t\}$$

die Menge der Stetigkeitspunkte der Verteilungsfunktion von X . Die Menge $\mathbb{R} \setminus S$ ist die Menge der Atome von X . Diese Menge ist höchstens abzählbar.

DEFINITION 13.1.1. Eine Folge von Zufallsvariablen X_1, X_2, \dots konvergiert *in Verteilung* gegen eine Zufallsvariable X , wenn für alle $t \in S(X)$ gilt:

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t).$$

BEZEICHNUNG: $X_n \xrightarrow[n \rightarrow \infty]{d} X$ oder $F_{X_n} \xrightarrow[n \rightarrow \infty]{d} F_X$. Dabei steht d für “distribution” (Verteilung).

BEMERKUNG 13.1.2. In der obigen Definition sind nur die Verteilungsfunktionen von X_n und X relevant. Der zugrundeliegende Wahrscheinlichkeitsraum spielt keine Rolle. Es kann z.B. sogar sein, dass die Zufallsvariablen auf verschiedenen Wahrscheinlichkeitsräumen definiert sind.

BEISPIEL 13.1.3. In diesem Beispiel zeigen wir, warum in der Definition der Verteilungskonvergenz die Einschränkung auf die Stetigkeitspunkte von F_X sinnvoll ist. Seien $c_1 > c_2 > \dots > 0$ Konstanten mit $\lim_{n \rightarrow \infty} c_n = 0$. Als Beispiel kann man $c_n = \frac{1}{n}$ betrachten. Definiere nun Zufallsvariablen $X_n := c_n$ und $X := 0$. Es gilt:

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \\ 0, & t = 0 \end{cases} = F_X(t), \text{ für alle } t \neq 0.$$

Für $t = 0$ stimmt die Gleichung $\lim_{n \rightarrow \infty} F_{X_n}(0) = F_X(0)$ nicht, denn $F_X(0) = 1$. Da allerdings $0 \notin S(X)$, ist die Bedingung in der Definition der Verteilungskonvergenz erfüllt und wir haben $X_n \xrightarrow[n \rightarrow \infty]{d} X$.

Hätten wir aber in der Definition der Verteilungskonvergenz verlangt, dass $\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t)$ für alle $t \in \mathbb{R}$ gelten soll (und nicht nur für alle $t \in S(X)$), so würde der resultierende Konvergenzbegriff die absurde Eigenschaft haben, dass $\frac{1}{n}$ nicht gegen 0 konvergiert.

Der nächste Satz zeigt, dass die Verteilungskonvergenz die schwächste der von uns eingeführten Konvergenzarten ist.

SATZ 13.1.4. Seien X_n und X Zufallsvariablen mit $X_n \xrightarrow[n \rightarrow \infty]{P} X$. Dann gilt: $X_n \xrightarrow[n \rightarrow \infty]{d} X$.

BEMERKUNG 13.1.5. Man kann nun die Beziehungen zwischen verschiedenen Konvergenzarten folgendermaßen darstellen:

$$\begin{array}{ccccccc} & & & & f.s. & \Rightarrow & P & \Rightarrow & d \\ & & & & & & \uparrow & & \\ \dots & \Rightarrow & L^3 & \Rightarrow & L^2 & \Rightarrow & L^1 & & \end{array}$$

BEWEIS VON SATZ 13.1.4. Sei $X_n \xrightarrow[n \rightarrow \infty]{P} X$. Für alle $\varepsilon > 0$ gilt also:

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0.$$

SCHRITT 1. Für alle $t \in \mathbb{R}$ gilt:

$$\begin{aligned} F_{X_n}(t) &= \mathbb{P}[X_n \leq t] \\ &= \mathbb{P}[X_n \leq t, |X_n - X| \leq \varepsilon] + \mathbb{P}[X_n \leq t, |X_n - X| > \varepsilon] \\ &\leq \mathbb{P}[X \leq t + \varepsilon] + \mathbb{P}[|X_n - X| > \varepsilon] \\ &= F_X(t + \varepsilon) + \mathbb{P}[|X_n - X| > \varepsilon]. \end{aligned}$$

Mit der Voraussetzung $X_n \xrightarrow[n \rightarrow \infty]{P} X$ gilt nun für $n \rightarrow \infty$:

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t + \varepsilon).$$

Das gilt für jedes $\varepsilon > 0$. Wir können $\varepsilon \downarrow 0$ gehen lassen und die Rechtsstetigkeit von F_X benutzen:

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t).$$

SCHRITT 2. Sei $t \in S(X)$ beliebig. Es gilt:

$$\begin{aligned} F_X(t - \varepsilon) &= \mathbb{P}[X \leq t - \varepsilon] \\ &= \mathbb{P}[X \leq t - \varepsilon, |X_n - X| \leq \varepsilon] + \mathbb{P}[X \leq t - \varepsilon, |X_n - X| > \varepsilon] \\ &\leq F_{X_n}(t) + \mathbb{P}[|X_n - X| > \varepsilon]. \end{aligned}$$

Mit der Voraussetzung $X_n \xrightarrow[n \rightarrow \infty]{P} X$ gilt nun für $n \rightarrow \infty$:

$$F_X(t - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

Das gilt für jedes $\varepsilon > 0$. Wir können also $\varepsilon \downarrow 0$ gehen lassen und benutzen, dass F_X an der Stelle t stetig ist (denn $t \in S(X)$):

$$F_X(t) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

SCHRITT 3. Fügt man Schritt 1 und Schritt 2 zusammen, so erhält man, dass für alle $t \in S(X)$

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

Daraus folgt, dass $F_X(t) = \lim_{n \rightarrow \infty} F_{X_n}(t)$ für alle $t \in S(X)$. Somit gilt $X_n \xrightarrow[n \rightarrow \infty]{d} X$. \square

BEISPIEL 13.1.6. Wir betrachten eine Folge von normalverteilten Zufallsvariablen $X_n \sim N(0, \sigma_n^2)$ mit $\lim_{n \rightarrow \infty} \sigma_n^2 = 0$. Wir zeigen, dass $X_n \xrightarrow[n \rightarrow \infty]{d} 0$.

BEWEIS. Es gilt $X_n \xrightarrow{L^2} 0$, denn

$$\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2] = \text{Var } X_n = \sigma_n^2 \xrightarrow[n \rightarrow \infty]{} 0.$$

Nun folgt aus der L^2 -Konvergenz die stochastische Konvergenz und damit auch die Konvergenz in Verteilung nach dem Schema $L^2 \Rightarrow P \Rightarrow d$. \square

BEISPIEL 13.1.7. Dieses Beispiel soll zeigen, dass die Umkehrung von Satz 13.1.4 im Allgemeinen falsch ist. Wir betrachten eine Zufallsvariable $X \sim N(0, 1)$ und die Folge $X_n = -X$ für alle $n \in \mathbb{N}$.

SCHRITT 1. Zuerst zeigen wir, dass X_n in Verteilung gegen X konvergiert. Es gilt:

$$F_{X_n}(t) = \mathbb{P}[X_n \leq t] = \mathbb{P}[-X \leq t] = \mathbb{P}[X \leq t] = F_X(t).$$

Dabei haben wir benutzt, dass $N(0, 1)$ eine symmetrische Verteilung ist, d.h. die Verteilung von X stimmt mit der Verteilung von $-X$ überein. Da $F_{X_n}(t) = F_X(t)$ für alle $t \in \mathbb{R}$, folgt, dass X_n gegen X in Verteilung konvergiert.

SCHRITT 2. Nun zeigen wir, dass X_n jedoch nicht in Wahrscheinlichkeit gegen X konvergiert. Es gilt:

$$\mathbb{P}[|X_n - X| > 1] = \mathbb{P}[2|X| > 1] = \mathbb{P}\left[|X| > \frac{1}{2}\right] = \frac{1}{\sqrt{2\pi}} \int_{|t|>1/2} e^{-\frac{t^2}{2}} dt =: c > 0.$$

Da das Integral unabhängig von n ist, folgt $\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| > 1] = c > 0$. Somit gilt nicht, dass $X_n \xrightarrow[n \rightarrow \infty]{P} X$.

Es gibt allerdings einen Spezialfall, in dem die Umkehrung von Satz 13.1.4 richtig ist. Man muss nämlich voraussetzen, dass die Grenzwertzufallsvariable konstant ist.

SATZ 13.1.8. Seien X_1, X_2, \dots Zufallsvariablen und $c \in \mathbb{R}$ eine Konstante mit $X_n \xrightarrow[n \rightarrow \infty]{d} c$.

Dann gilt: $X_n \xrightarrow[n \rightarrow \infty]{P} c$.

BEWEIS. Sei $X_n \xrightarrow[n \rightarrow \infty]{d} c$. Dann gilt für alle $t \neq c$:

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t) = \begin{cases} 1, & t > c, \\ 0, & t < c. \end{cases}$$

Sei nun $\varepsilon > 0$. Dann gilt:

$$\begin{aligned} \mathbb{P}[|X_n - c| > \varepsilon] &= \mathbb{P}[X_n < c - \varepsilon] + \mathbb{P}[X_n > c + \varepsilon] \\ &\leq \mathbb{P}[X_n \leq c - \varepsilon] + 1 - \mathbb{P}[X_n \leq c + \varepsilon] \\ &= F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

da $\lim_{n \rightarrow \infty} F_{X_n}(c - \varepsilon) = 0$ und $\lim_{n \rightarrow \infty} F_{X_n}(c + \varepsilon) = 1$. Es gilt also $\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - c| > \varepsilon] = 0$ und somit $X_n \xrightarrow[n \rightarrow \infty]{P} X$. \square

BEISPIEL 13.1.9. Für jedes $n \in \mathbb{N}$ sei X_n eine Zufallsvariable, die gleichverteilt auf der Menge $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$ ist. Sei außerdem X gleichverteilt auf $[0, 1]$. Somit ist X_n eine diskrete Zufallsvariable, wohingegen X absolut stetig ist. Wir zeigen, dass

$$X_n \xrightarrow[n \rightarrow \infty]{d} X.$$

BEWEIS. Für $t < 0$ hat man $F_{X_n}(t) = F_X(t) = 0$. Für $t > 1$ hat man $F_{X_n}(t) = F_X(t) = 1$. Für $t \in [0, 1]$ hat man

$$F_{X_n}(t) = \frac{[tn]}{n} \xrightarrow[n \rightarrow \infty]{} t = F_X(t),$$

wobei $[\cdot]$ die Gauß-Klammer ist. Also gilt für jedes $t \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t).$$

Somit folgt $X_n \xrightarrow[n \rightarrow \infty]{d} X$. \square

BEISPIEL 13.1.10 (Poisson-Grenzwertsatz). Wir betrachten eine Folge von Zufallsvariablen mit $X_n \sim \text{Bin}(n, p_n)$, wobei für die Erfolgswahrscheinlichkeit p_n gilt: $\lim_{n \rightarrow \infty} np_n = \lambda \in (0, \infty)$. Zum Beispiel kann man $p_n = \frac{\lambda}{n}$ betrachten. Sei außerdem $X \sim \text{Poi}(\lambda)$. Wir zeigen, dass

$$X_n \xrightarrow[n \rightarrow \infty]{d} X.$$

BEWEIS. Der Poisson-Grenzwertsatz besagt, dass für alle $k \in \mathbb{N}_0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n = k] = e^{-\lambda} \frac{\lambda^k}{k!} = \mathbb{P}[X = k].$$

Die Menge der Stetigkeitspunkte der Verteilungsfunktion von X ist $S(X) = \mathbb{R} \setminus \mathbb{N}_0$. Für $t < 0$ gilt $F_{X_n}(t) = 0 = F_X(t)$. Sei deshalb $t > 0$ mit $t \notin \mathbb{N}_0$. Dann kann man ein $m \in \mathbb{N}_0$ mit $m < t < m + 1$ finden. Es gilt

$$F_{X_n}(t) = \mathbb{P}[X_n \leq t] = \sum_{k=0}^m \mathbb{P}[X_n = k] \xrightarrow[n \rightarrow \infty]{} \sum_{k=0}^m e^{-\lambda} \frac{\lambda^k}{k!} = \mathbb{P}[X \leq t] = F_X(t).$$

Daraus folgt: $X_n \xrightarrow[n \rightarrow \infty]{d} X$. □

13.2. Eine Charakterisierung der Konvergenz in Verteilung

DEFINITION 13.2.1. Die Menge der stetigen und beschränkten Funktionen bezeichnen wir mit

$$C_b := \left\{ f : \mathbb{R} \rightarrow \mathbb{R} \mid f \text{ ist stetig und } \sup_{t \in \mathbb{R}} |f(t)| \leq \infty \right\}.$$

SATZ 13.2.2. Seien X, X_1, X_2, \dots Zufallsvariablen. Folgende zwei Aussagen sind äquivalent:

- (1) $X_n \xrightarrow[n \rightarrow \infty]{d} X$.
- (2) Für alle $f \in C_b$ gilt: $\lim_{n \rightarrow \infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$.

BEWEIS VON (1) \Rightarrow (2). Sei $X_n \xrightarrow[n \rightarrow \infty]{d} X$. Sei $f \in C_b$ eine stetige und beschränkte Funktion. Wir werden zeigen, dass

$$\lim_{n \rightarrow \infty} \mathbb{E}f(X_n) = \mathbb{E}f(X).$$

SCHRITT 1. Für f beschränkt ist, gilt $b := \sup_{t \in \mathbb{R}} |f(t)| < \infty$.

SCHRITT 2. Sei $\varepsilon > 0$. Es gilt $\lim_{c \rightarrow \infty} \mathbb{P}[|X| > c] = 0$. Somit existiert ein hinreichend großes c mit

$$\mathbb{P}[|X| > c] < \frac{\varepsilon}{b}.$$

Indem man c vergrößert, kann man außerdem erreichen, dass $c \in S(X)$ und $-c \in S(X)$.

SCHRITT 3. Aus $X_n \xrightarrow[n \rightarrow \infty]{d} X$ folgt $\lim_{n \rightarrow \infty} \mathbb{P}[|X_n| > c] = \mathbb{P}[|X| > c]$, denn

$$\mathbb{P}[|X_n| > c] = F_{X_n}(-c) + (1 - F_{X_n}(c)) \xrightarrow[n \rightarrow \infty]{} F_X(-c) + (1 - F_X(c)) = \mathbb{P}[|X| > c].$$

Für hinreichend großes n gilt somit:

$$\mathbb{P}[|X_n| > c] < \frac{2\varepsilon}{b}.$$

SCHRITT 4. Da die Funktion f stetig ist, kann man eine Funktion g mit folgenden Eigenschaften konstruieren:

- (1) $\sup_{t \in \mathbb{R}} |f(t) - g(t)| \leq \varepsilon$.
- (2) g ist eine Treppenfunktion:

$$g(t) = \sum_{i=1}^k a_i \cdot \mathbb{1}_{t_{i-1} < t \leq t_i},$$

wobei $-c = t_0 < t_1 < \dots < t_k = c$ und $t_0, \dots, t_k \in S(X)$.

SCHRITT 5. Für hinreichend großes n gilt:

$$\begin{aligned}
& |\mathbb{E}f(X_n) - \mathbb{E}f(X)| \\
& \leq |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| \leq c} - \mathbb{E}f(X)\mathbb{1}_{|X| \leq c}| + |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| > c}| + |\mathbb{E}f(X)\mathbb{1}_{|X| > c}| \\
& \leq |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| \leq c} - \mathbb{E}f(X)\mathbb{1}_{|X| \leq c}| + 3\varepsilon \\
& \leq |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| \leq c} - \mathbb{E}g(X_n)| + |\mathbb{E}g(X) - \mathbb{E}f(X)\mathbb{1}_{|X| \leq c}| + |\mathbb{E}g(X_n) - \mathbb{E}g(X)| + 3\varepsilon \\
& \leq |\mathbb{E}g(X_n) - \mathbb{E}g(X)| + 5\varepsilon,
\end{aligned}$$

wobei die zweite Ungleichung gilt, da $|f(X_n)| \leq b$, $|f(X)| \leq b$, $\mathbb{P}[|X_n| > c] \leq \frac{2\varepsilon}{b}$ und $\mathbb{P}[|X| > c] \leq \frac{\varepsilon}{b}$, und die letzte Ungleichung gilt, da $|f(x) - g(x)| \leq \varepsilon$.

SCHRITT 6. Wir betrachten nun den Term $\mathbb{E}g(X_n) - \mathbb{E}g(X)$. Aus $X_n \xrightarrow[n \rightarrow \infty]{d} X$ und $t_0, \dots, t_k \in S(X)$ folgt, dass

$$\begin{aligned}
\mathbb{E}g(X_n) &= \sum_{i=1}^k a_i \cdot \mathbb{E}\mathbb{1}_{t_{i-1} < X_n \leq t_i} \\
&= \sum_{i=1}^k a_i \cdot (F_{X_n}(t_i) - F_{X_n}(t_{i-1})) \\
&\xrightarrow[n \rightarrow \infty]{} \sum_{i=1}^k a_i \cdot (F_X(t_i) - F_X(t_{i-1})) \\
&= \mathbb{E}g(X).
\end{aligned}$$

Daraus folgt: $\lim_{n \rightarrow \infty} \mathbb{E}g(x_n) = \mathbb{E}g(X)$. Für hinreichend großes n gilt somit:

$$|\mathbb{E}g(X_n) - \mathbb{E}g(X)| \leq \varepsilon.$$

SCHRITT 7. Aus Schritt 5 und Schritt 6 ergibt sich, dass für hinreichend großes n gilt:

$$|\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq 6\varepsilon.$$

Dabei war $\varepsilon > 0$ beliebig. Daraus folgt, dass $\lim_{n \rightarrow \infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$.

BEWEIS VON (2) \Rightarrow (1). Sei $\lim_{n \rightarrow \infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ für alle $f \in C_b$. Wir zeigen, dass

$$X_n \xrightarrow[n \rightarrow \infty]{d} X.$$

Beweisidee: Dürften wir $f(z) = \mathbb{1}_{z \leq t}$ einsetzen, so würde gelten:

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = \lim_{n \rightarrow \infty} \mathbb{E}f(X_n) = \mathbb{E}f(X) = F_X(t).$$

Allerdings ist die Indikatorfunktion $\mathbb{1}_{z \leq t}$ nicht stetig. Also werden wir diese Indikatorfunktion durch stetige Funktionen approximieren.

SCHRITT 1. Sei $t \in \mathbb{R}$ und $\varepsilon > 0$. Dann existiert eine Funktion $f \in C_b$ mit

$$\mathbb{1}_{y \leq t} \leq f(y) \leq \mathbb{1}_{y \leq t + \varepsilon} \text{ für alle } y \in \mathbb{R}.$$

(1) Es gilt also $\mathbb{1}_{X_n \leq t} \leq f(X_n) \leq \mathbb{1}_{X_n \leq t + \varepsilon}$. Daraus folgt:

$$F_{X_n}(t) = \mathbb{E} \mathbb{1}_{X_n \leq t} \leq \mathbb{E} f(X_n).$$

(2) Außerdem gilt $\mathbb{1}_{X \leq t} \leq f(X) \leq \mathbb{1}_{X \leq t + \varepsilon}$. Es folgt:

$$\mathbb{E} f(X) \leq \mathbb{E} \mathbb{1}_{X \leq t + \varepsilon} = F_X(t + \varepsilon).$$

Fügt man nun (1) und (2) zusammen, erhält man für $n \rightarrow \infty$:

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq \lim_{n \rightarrow \infty} \mathbb{E} f(X_n) = \mathbb{E} f(X) \leq F_X(t + \varepsilon).$$

Das gilt für jedes $\varepsilon > 0$. Da nun F_X rechtsstetig ist, erhalten wir für $\varepsilon \downarrow 0$:

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t).$$

SCHRITT 2. Sei $t \in S(X)$ und $\varepsilon > 0$. Dann existiert ein $f \in C_b$ mit $\mathbb{1}_{y \leq t - \varepsilon} \leq f(y) \leq \mathbb{1}_{y \leq t}$ für alle $y \in \mathbb{R}$. Es gilt:

$$F_X(t - \varepsilon) = \mathbb{E} \mathbb{1}_{X \leq t - \varepsilon} \leq \mathbb{E} f(X) = \lim_{n \rightarrow \infty} \mathbb{E} f(X_n) \leq \liminf_{n \rightarrow \infty} \mathbb{E} \mathbb{1}_{X_n \leq t} = \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

Für $\varepsilon \downarrow 0$ erhalten wir dann:

$$F_X(t) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

SCHRITT 3. Fügt man nun Schritte 1 und 2 zusammen, so folgt für alle $t \in S(X)$:

$$F_X(t) = \lim_{n \rightarrow \infty} F_{X_n}(t).$$

Und das bedeutet, dass $X_n \xrightarrow[n \rightarrow \infty]{d} X$. □

13.3. Satz von Helly

Für eine Verteilungsfunktion G bezeichnen wir mit $S(G)$ die Menge der Stetigkeitspunkte von G .

SATZ 13.3.1 (Helly). *Seien F_1, F_2, \dots Verteilungsfunktionen. Dann gibt es eine rechtsstetige nichtfallende Funktion G und eine Teilfolge F_{n_1}, F_{n_2}, \dots , so dass für alle $t \in S(G)$ gilt:*

$$\lim_{n \rightarrow \infty} F_{n_k}(t) = G(t).$$

BEWEIS. Die Menge der rationalen Zahlen ist abzählbar, also gibt es eine Abzählung $\mathbb{Q} = \{r_1, r_2, \dots\}$. Wir werden im Folgenden eine geschachtelte Familie von unendlichen Mengen $\mathbb{N} = K_0 \supset K_1 \supset \dots$ mit gewissen (unten näher beschriebenen) Eigenschaften konstruieren.

SCHRITT 1. Sei $K_0 = \mathbb{N}$. Es gilt: $\{F_n(r_1)\}_{n \in \mathbb{N}} \subset [0, 1]$. Mit dem Satz von Bolzano–Weierstraß folgt: Es gibt eine Teilmenge $K_1 \subset K_0$ mit $|K_1| = \infty$ und

$$\lim_{n \rightarrow \infty, n \in K_1} F_n(r_1) = g_1.$$

SCHRITT 2. Genauso gilt: $\{F_n(r_2)\}_{n \in K_1} \subset [0, 1]$. Und damit folgt wiederum mit dem Satz von Bolzano–Weierstraß, dass eine Teilmenge $K_2 \subset K_1$ existiert mit $|K_2| = \infty$ und

$$\lim_{n \rightarrow \infty, n \in K_2} F_n(r_2) = g_2.$$

SCHRITT 3. Diese Argumentation kann fortgesetzt werden. Es existieren also Mengen $\mathbb{N} = K_0 \subset K_1 \subset K_2 \subset \dots$ mit $|K_j| = \infty$ und

$$\lim_{n \rightarrow \infty, n \in K_j} F_n(r_j) = g_j \text{ für alle } j \in \mathbb{N}.$$

SCHRITT 4. Nun benutzen wir die sogenannte Cantor–Diagonalisierung. In jeder Menge K_j wählen wir das kleinste Element. Die Vereinigung dieser Elemente nennen wir K , so dass $K = \{\min K_j\}_{j \in \mathbb{N}}$. Es gilt nach Konstruktion $|K| = \infty$. Außerdem gilt für alle $j \in \mathbb{N}$: K ist bis auf endlich viele Elemente eine Teilmenge von K_j und somit

$$\lim_{n \rightarrow \infty, n \in K} F_n(r_j) = g_j \text{ für alle } j \in \mathbb{N}.$$

Wir haben gezeigt, dass für alle $j \in \mathbb{N}$ der folgende Grenzwert existiert:

$$H(r_j) := \lim_{n \rightarrow \infty, n \in K} F_n(r_j).$$

Die Funktion H ist nur auf der Menge der rationalen Zahlen definiert. Wir erweitern die Funktion H auf ganz \mathbb{R} :

$$G(t) := \inf\{H(r) : r \in \mathbb{Q}, r > t\}.$$

Für diese Funktion gilt: G ist nichtfallend und rechtsstetig (Übung).

SCHRITT 5. Nun müssen wir noch zeigen, dass für $t \in S(G)$ gilt:

$$\lim_{n \rightarrow \infty, n \in K} F_n(t) = G(t).$$

Sei also $t \in S(G)$ und sei $\varepsilon > 0$. Dann existiert $y < t$ mit $G(t) - \varepsilon \leq G(y) \leq G(t)$. Außerdem existieren $r, s \in \mathbb{Q}$ mit $y < r < t < s$ und $G(s) \leq G(t) + \varepsilon$. Nun gilt also:

$$G(t) - \varepsilon \leq G(y) \leq G(r) \leq G(s) \leq G(t) + \varepsilon.$$

Es folgt, dass

$$G(r) \xleftarrow[n \rightarrow \infty]{n \in K} F_n(r) \leq F_n(t) \leq F_n(s) \xrightarrow[n \rightarrow \infty]{n \in K} G(s).$$

Somit gilt

$$G(t) - \varepsilon \leq \liminf_{n \rightarrow \infty, n \in K} F_n(t) \leq \limsup_{n \rightarrow \infty, n \in K} F_n(t) \leq G(t) + \varepsilon.$$

Für $\varepsilon \downarrow 0$ erhalten wir dann: $G(t) = \lim_{n \rightarrow \infty, n \in K} F_n(t)$. □

BEISPIEL 13.3.2. Die Funktion G aus dem Satz von Helly ist monoton nichtfallend und rechtsstetig, sie muss aber nicht unbedingt eine Verteilungsfunktion sein. Es fehlen nämlich die Eigenschaften $\lim_{t \rightarrow -\infty} G(t) = 0$ und $\lim_{t \rightarrow +\infty} G(t) = 1$. Um zu sehen, dass die beiden Eigenschaften nicht immer erfüllt sind, betrachten wir die Folge $X_n = n$. Für die Verteilungsfunktion von X_n gilt dann

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = \lim_{n \rightarrow \infty} \mathbb{1}_{[n, \infty)}(t) = 0.$$

Also ist auch der Grenzwert jeder Teilfolge von F_{X_n} gleich 0. Die Funktion $G = 0$ ist aber keine Verteilungsfunktion.

DEFINITION 13.3.3. Eine Folge F_1, F_2, \dots von Verteilungsfunktionen (bzw. eine Folge von Zufallsvariablen, die diese Verteilungsfunktionen haben) heißt *straff*, wenn für alle $\varepsilon > 0$ ein $c > 0$ existiert, so dass für alle $n \in \mathbb{N}$ gilt:

$$F_n(-c) + (1 - F_n(c)) < \varepsilon.$$

BEMERKUNG 13.3.4. Ist F_n die Verteilungsfunktion einer Zufallsvariable X_n , so kann man die obige Bedingung auch so darstellen: für alle $\varepsilon > 0$ existiert ein $c > 0$, so dass

$$\mathbb{P}[|X_n| > c] < \varepsilon \text{ für alle } n \in \mathbb{N}.$$

BEISPIEL 13.3.5. Die Folge der Verteilungsfunktionen $F_n = \mathbb{1}_{[n, \infty)}$ ist nicht straff.

BEISPIEL 13.3.6. Seien X_1, X_2, \dots beliebige Zufallsvariablen, die nur Werte in einem Intervall $[-A, A]$ annehmen. Dann ist diese Folge von Zufallsvariablen straff.

SATZ 13.3.7 (Helly, Prochorow). *Sei F_1, F_2, \dots eine straffe Folge von Verteilungsfunktionen. Dann existieren eine Verteilungsfunktion G und eine Teilfolge F_{n_1}, F_{n_2}, \dots , so dass für alle $t \in S(G)$ gilt:*

$$\lim_{k \rightarrow \infty} F_{n_k}(t) = G(t).$$

BEMERKUNG 13.3.8. Mit anderen Worten: aus einer straffen Folge von Zufallsvariablen kann man eine in Verteilung konvergente Teilfolge extrahieren. Vergleiche das mit dem Satz von Bolzano–Weierstraß: Aus einer beschränkten Folge von Vektoren in \mathbb{R}^d kann man eine konvergente Teilfolge extrahieren.

BEWEIS. Aus Satz 13.3.1 folgt, dass es eine Teilfolge n_1, n_2, \dots und eine nichtfallende, rechtstetige Funktion G mit $\lim_{k \rightarrow \infty} F_{n_k}(t) = G(t)$ für alle $t \in S(G)$ gibt. Wir berechnen die Grenzwerte $\lim_{t \rightarrow \pm\infty} G(t)$.

Für alle $\varepsilon > 0$ existiert ein $c > 0$ mit $F_{n_k}(-c) < \varepsilon$ und $1 - F_{n_k}(c) < \varepsilon$ für alle $k \in \mathbb{N}$, da die Folge F_1, F_2, \dots straff ist. Indem wir c vergrößern, können wir zusätzlich annehmen, dass $c, -c \in S(G)$. Nun lassen wir $k \rightarrow \infty$ und erhalten, dass

$$G(-c) \leq \varepsilon \text{ und } 1 - G(c) \leq \varepsilon.$$

Da das für alle $\varepsilon > 0$ gilt, ergibt sich daraus, dass $\lim_{c \rightarrow +\infty} G(-c) = 0$ und $\lim_{c \rightarrow +\infty} G(c) = 1$. Also ist G eine Verteilungsfunktion. \square

13.4. Stetigkeitssatz von Lévy

Der nächste Satz wird im Beweis des zentralen Grenzwertsatzes eine wichtige Rolle spielen. Dieser Satz besagt, dass Konvergenz in Verteilung äquivalent zur punktweisen Konvergenz der entsprechenden charakteristischen Funktionen ist.

SATZ 13.4.1 (Stetigkeitssatz von Lévy). *Seien X, X_1, X_2, \dots Zufallsvariablen mit charakteristischen Funktionen $\varphi_X, \varphi_{X_1}, \varphi_{X_2}, \dots$. Dann sind folgende zwei Aussagen äquivalent:*

- (1) $X_n \xrightarrow[n \rightarrow \infty]{d} X$.
- (2) $\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t)$ für alle $t \in \mathbb{R}$.

Im Beweis werden wir die vereinfachte Notation $\varphi_{X_n} = \varphi_n$ und $\varphi_X = \varphi$ benutzen. Für die Verteilungsfunktionen von X_n und X benutzen wir die Notation $F_{X_n} = F_n$ und $F_X = F$.

BEWEIS VON (1) \Rightarrow (2). Sei $X_n \xrightarrow[n \rightarrow \infty]{d} X$. Dann gilt mit Satz 13.2.2 für alle $t \in \mathbb{R}$:

$$\varphi_n(t) = \mathbb{E} \cos(tX_n) + i\mathbb{E} \sin(tX_n) \xrightarrow[n \rightarrow \infty]{} \mathbb{E} \cos(tX) + i\mathbb{E} \sin(tX) = \varphi(t).$$

BEWEIS VON (2) \Rightarrow (1). Es gelte $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ für alle $t \in \mathbb{R}$. Wir wollen zeigen, dass $X_n \xrightarrow[n \rightarrow \infty]{d} X$.

SCHRITT 1. Zuerst zeigen wir, dass die Folge von Verteilungsfunktionen F_1, F_2, \dots straff ist. Wir definieren eine Funktion $B_n(u)$ folgendermaßen:

$$B_n(u) := \frac{1}{u} \int_{-u}^u (1 - \varphi_n(t)) dt = \frac{1}{u} \int_{-u}^u \mathbb{E}[1 - e^{itX_n}] dt = \mathbb{E} \left[\frac{1}{u} \int_{-u}^u (1 - e^{itX_n}) dt \right],$$

wobei wir im letzten Schritt die Integrale nach dem Satz von Fubini vertauscht haben. Das war möglich, denn $|1 - e^{itX_n}| \leq 2$. Da nun $\sin x$ eine gerade Funktion ist und der Imaginärteil somit nach der Integration verschwindet, erhalten wir, dass

$$B_n(u) = 2\mathbb{E} \left(1 - \frac{\sin(uX_n)}{uX_n} \right) \geq 2\mathbb{E} \left[1 - \frac{1}{|uX_n|} \right] \geq \mathbb{P} \left[|X_n| \geq \frac{2}{u} \right].$$

Es gilt: $\varphi(0) = 1$ und φ ist stetig, da φ eine charakteristische Funktion ist. Somit existiert für jedes $\varepsilon > 0$ ein hinreichend kleines $u > 0$, so dass

$$B(u) = \frac{1}{u} \cdot \int_{-u}^u (1 - \varphi(t)) dt < \varepsilon.$$

Da $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ für alle $t \in \mathbb{R}$ gilt, folgt mit der majorisierten Konvergenz:

$$B_n(u) = \frac{1}{u} \int_{-u}^u (1 - \varphi_n(t)) dt \xrightarrow[n \rightarrow \infty]{} \frac{1}{u} \int_{-u}^u (1 - \varphi(t)) dt = B(u).$$

Wir durften die majorisierte Konvergenz benutzen, da $|1 - \varphi_n(t)| \leq 2$. Somit existiert ein n_0 , so dass $B_n(u) < 2\varepsilon$ für alle $n \geq n_0$. Dann ist auch

$$\mathbb{P} \left[|X_n| \geq \frac{2}{u} \right] \leq B_n(u) < 2\varepsilon \text{ für alle } n \geq n_0.$$

Indem wir die Schranke $\frac{2}{u}$ vergrößern, können wir erreichen, dass die obige Ungleichung sogar für alle $n \in \mathbb{N}$ gilt. Daraus folgt, dass die Folge F_1, F_2, \dots straff ist.

SCHRITT 2. Nun zeigen wir durch Widerspruch, dass $X_n \xrightarrow[n \rightarrow \infty]{d} X$. Wir nehmen also an, dass diese Konvergenz nicht gilt. Es existiert somit ein $t \in \mathbb{R}$ mit

$$(13.4.1) \quad \lim_{n \rightarrow \infty} F_n(t) \neq F(t) \text{ und } \mathbb{P}[X = t] = 0.$$

Daraus folgt: es gibt ein $\varepsilon > 0$ und eine Teilfolge F_{n_1}, F_{n_2}, \dots mit

$$|F_{n_k}(t) - F(t)| > \varepsilon \text{ für alle } k \in \mathbb{N}.$$

Da die Folge F_1, F_2, \dots und somit auch die Teilfolge F_{n_1}, F_{n_2}, \dots straff ist, folgt mit Satz 13.3.7, dass eine Verteilungsfunktion G und eine Teilteilfolge $F_{n_{k_1}}, F_{n_{k_2}}, \dots$ existieren, für die gilt:

$$(13.4.2) \quad F_{n_{k_j}} \xrightarrow{j \rightarrow \infty} G.$$

Wir betrachten zwei Fälle:

FALL 1. G ist nicht stetig an der Stelle t . Dann folgt: $G \neq F$, denn F ist stetig an der Stelle t wegen (13.4.1).

FALL 2. G ist stetig an der Stelle t . Dann folgt aus (13.4.2), dass $\lim_{j \rightarrow \infty} F_{n_{k_j}}(t) = G(t) \neq F(t)$, also $G \neq F$.

In beiden Fällen gilt $G \neq F$. Aus (13.4.2) und wegen der bereits bewiesenen Richtung (1) \Rightarrow (2) im Satz folgt, dass

$$\lim_{j \rightarrow \infty} \varphi_{n_{k_j}}(t) = \varphi_G(t) \text{ für alle } t \in \mathbb{R}.$$

Dabei bezeichnen wir mit φ_G die charakteristische Funktion von G . Nach Voraussetzung sollte aber auch gelten, dass

$$\lim_{j \rightarrow \infty} \varphi_{n_{k_j}}(t) = \varphi(t) \text{ für alle } t \in \mathbb{R}.$$

Somit gilt $\varphi_G = \varphi$. Das heißt, die charakteristischen Funktionen von G und F sind gleich. Dabei haben wir aber gezeigt, dass $G \neq F$. Dies ist ein Widerspruch, da die Verteilungsfunktion eindeutig durch die charakteristische Funktion bestimmt wird. \square

13.5. Der zentrale Grenzwertsatz

Seien X_1, X_2, \dots unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = \mu$, $\text{Var } X_k = \sigma^2$. Wir betrachten die Summe $S_n = X_1 + \dots + X_n$. Wir werden nun die Frage beantworten, wie die Summe S_n für $n \rightarrow \infty$ verteilt ist. Damit wir eine sinnvolle Aussage über die Verteilung von S_n erhalten können, müssen wir zuerst einige Vorbereitungen treffen.

SCHRITT 1. Durch Abziehen des Erwartungswerts können wir S_n zentrieren. Das heißt, wir betrachten die Zufallsvariable

$$S_n - \mathbb{E}S_n = S_n - n\mu.$$

Es gilt dann $\mathbb{E}[S_n - n\mu] = 0$.

SCHRITT 2. Indem wir durch die Standardabweichung dividieren, können wir die Zufallsvariable auch normieren. Das heißt, wir betrachten die Zufallsvariable

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\text{Var } S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Dann gilt $\mathbb{E}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}}\right] = 0$ und $\text{Var}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}}\right] = 1$.

SATZ 13.5.1 (Der zentrale Grenzwertsatz). *Seien X_1, X_2, \dots unabhängige, identisch verteilte, quadratisch integrierbare Zufallsvariablen mit $\mathbb{E}X_k = \mu$, $\text{Var } X_k = \sigma^2 \in (0, \infty)$. Sei $S_n = X_1 + \dots + X_n$. Dann gilt:*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{n \rightarrow \infty} N,$$

wobei $N \sim N(0, 1)$ eine standardnormalverteilte Zufallsvariable ist.

BEMERKUNG 13.5.2. Mit anderen Worten: Für alle $x \in \mathbb{R}$ gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right] = \Phi(x) \text{ mit } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Als Spezialfall erhalten wir den folgenden Satz von de Moivre–Laplace.

SATZ 13.5.3 (de Moivre (1733), Laplace (1812)). Sei $S_n \sim \text{Bin}(n, p)$ binomialverteilt, wobei $p \in (0, 1)$ konstant ist. Dann gilt für alle $x \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{S_n - np}{\sqrt{np(1-p)}} \leq x \right] = \Phi(x).$$

BEWEIS. Wir definieren unabhängige, identisch verteilte Zufallsvariablen X_1, X_2, \dots mit

$$\mathbb{P}[X_k = 1] = p \text{ und } \mathbb{P}[X_k = 0] = 1 - p.$$

Dann gilt für die Summe der Zufallsvariablen

$$S_n = X_1 + \dots + X_n \sim \text{Bin}(n, p).$$

Der Erwartungswert von X_k ist $\mu = p$ und die Varianz ist $\sigma^2 = p \cdot (1 - p)$. Mit dem zentralen Grenzwertsatz folgt die Behauptung. \square

BEMERKUNG 13.5.4. Mit dem Satz von de Moivre–Laplace kann man die Binomialverteilung $\text{Bin}(n, p)$ für ein großes n und ein konstantes p durch die Normalverteilung approximieren. Sei nämlich $S_n \sim \text{Bin}(n, p)$. Wir wollen die Wahrscheinlichkeit $\mathbb{P}[a \leq S_n \leq b]$, dass S_n innerhalb eines gegebenen Intervalls $[a, b]$ liegt, berechnen. Mit dem Satz von de Moivre–Laplace erhalten wir die folgende Approximation:

$$\mathbb{P}[a \leq S_n \leq b] = \mathbb{P} \left[\frac{a - n\mu}{\sigma\sqrt{n}} \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{b - n\mu}{\sigma\sqrt{n}} \right] \approx \mathbb{P}[a^* \leq N \leq b^*] = \Phi(b^*) - \Phi(a^*),$$

wobei $\frac{a - n\mu}{\sigma\sqrt{n}} = a^*$ und $\frac{b - n\mu}{\sigma\sqrt{n}} = b^*$. Wir haben benutzt, dass $\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx N \sim N(0, 1)$ für großes n ist.

BEISPIEL 13.5.5. Wir betrachten ein $n = 100$ -faches Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p = \frac{1}{2}$. Sei $S = S_{100} \sim \text{Bin}(100, \frac{1}{2})$ die Anzahl der Erfolge. Wir approximieren nun mit dem zentralen Grenzwertsatz die Wahrscheinlichkeit $\mathbb{P}[S \leq 55]$.

LÖSUNG. Für den Erwartungswert und die Varianz von S erhalten wir

$$\mathbb{E}S = np = 100 \cdot \frac{1}{2} = 50 \text{ und } \text{Var } S = np(1-p) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25.$$

Also gilt mit dem Satz von de Moivre–Laplace:

$$\mathbb{P}[S \leq 55] = \mathbb{P} \left[\frac{S - 50}{\sqrt{25}} \leq \frac{55 - 50}{\sqrt{25}} \right] \approx \mathbb{P}[N \leq 1] = \Phi(1) = 0.8413,$$

wobei wir benutzt haben, dass $\frac{S-50}{\sqrt{25}} \approx N$.

Um eine bessere Approximation zu erhalten, kann man den sogenannten $\pm \frac{1}{2}$ Trick anwenden. Da die Zufallsvariable S_n nämlich ganzzahlig ist, sind die Wahrscheinlichkeiten $\mathbb{P}[S \leq 55]$

und $\mathbb{P}[S < 56]$ gleich. Sollen wir nun 55 oder 56 in die Approximation einsetzen? Wir werden den Mittelwert 55.5 einsetzen:

$$\mathbb{P}[S \leq 55] = \mathbb{P}[S \leq 55,5] = \mathbb{P}\left[\frac{S - 50}{\sqrt{25}} \leq \frac{55,5 - 50}{\sqrt{25}}\right] \approx \mathbb{P}[N \leq 1,1] = \Phi(1,1) = 0,86433.$$

Der exakte Wert für die Wahrscheinlichkeit ist übrigens $\mathbb{P}[S \leq 55] = 0.86437$.

BEISPIEL 13.5.6. Seien $X_1, X_2, \dots \sim N(\mu, \sigma^2)$ unabhängige normalverteilte Zufallsvariablen. Dann gilt nicht nur approximativ, sondern sogar exakt, dass

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0, 1).$$

13.6. Beweis des zentralen Grenzwertsatzes

BEWEIS VON SATZ 13.5.1. Es seien X_1, X_2, \dots unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = \mu$ und $\text{Var} X_k = \sigma^2 > 0$. Definiere $S_n = X_1 + \dots + X_n$. Wir zeigen, dass

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

Anstelle von X_1, X_2, \dots betrachten wir die normierten Zufallsvariablen

$$Y_k := \frac{X_k - \mu}{\sigma} \text{ mit } \mathbb{E}Y_k = 0 \text{ und } \text{Var} Y_k = 1.$$

Wir definieren die Zufallsvariable

$$V_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}.$$

Zu zeigen ist, dass

$$V_n \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

Wir werden zeigen, dass die charakteristische Funktion von V_n gegen die charakteristische Funktion der Standardnormalverteilung punktweise konvergiert. Dann kann man den Stetigkeitssatz 13.4.1 anwenden. Sei $t \in \mathbb{R}$ fest. Für die charakteristische Funktion von V_n gilt:

$$\varphi_{V_n}(t) = \mathbb{E}e^{itV_n} = \mathbb{E}\left[e^{it\frac{Y_1 + \dots + Y_n}{\sqrt{n}}}\right] = \mathbb{E}\left[e^{it\frac{Y_1}{\sqrt{n}}}\right] \cdot \dots \cdot \mathbb{E}\left[e^{it\frac{Y_n}{\sqrt{n}}}\right] = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n,$$

wobei $\varphi := \varphi_{Y_1}$ die charakteristische Funktion von Y_1 ist. Nun wollen wir $n \rightarrow \infty$ gehen lassen. Da $\varphi(0) = 1$ und φ stetig ist, gilt

$$\lim_{n \rightarrow \infty} \varphi\left(\frac{t}{\sqrt{n}}\right) = 1.$$

Also haben wir es mit einer Unbestimmtheit von der Form 1^∞ zu tun. Um die aufzulösen, werden wir die ersten zwei Terme der Taylor-Entwicklung von φ betrachten. Es gilt

$$\varphi(0) = 1, \quad \varphi'(0) = i\mathbb{E}Y_1 = 0, \quad \varphi''(0) = -\mathbb{E}[Y_1^2] = -1.$$

Somit erhalten wir die folgende Taylorentwicklung:

$$\varphi(s) = 1 - \frac{s^2}{2} + o(s^2), \quad s \rightarrow 0.$$

Nun setzen wir $s = \frac{t}{\sqrt{n}}$ ein:

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right), \quad n \rightarrow \infty.$$

Und dann gilt für die charakteristische Funktion von V_n :

$$\varphi_{V_n}(t) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right)\right)^n \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}}.$$

Dabei haben wir die folgende Aussage benutzt: für eine Folge a_1, a_2, \dots mit $\lim_{n \rightarrow \infty} na_n = \lambda$ gilt, dass $\lim_{n \rightarrow \infty} (1 + a_n)^n = e^\lambda$. Da $e^{-t^2/2}$ die charakteristische Funktion der Standardnormalverteilung ist, folgt schließlich mit dem Stetigkeitssatz 13.4.1, dass V_n in Verteilung gegen eine Standardnormalverteilte Zufallsvariable konvergiert. Das beweist die Behauptung. \square

13.7. Sätze von Lindeberg und Ljapunow

Der folgende Satz ist eine Verallgemeinerung des zentralen Grenzwertsatzes. Er besagt, dass eine Summe von sehr vielen sehr kleinen und unabhängigen zufälligen Fehlern unter ziemlich allgemeinen Bedingungen approximativ normalverteilt ist. In diesem Satz werden wir anstatt einer Folge ein sogenanntes *Dreiecksschema* von Zufallsvariablen betrachten:

$$\begin{array}{cccc} X_{11} & & & \\ X_{21}, & X_{22} & & \\ X_{31}, & X_{32}, & X_{33} & \\ \dots & & & \\ X_{n1}, & X_{n2}, & X_{n3}, & \dots, & X_{nn} \\ \dots & & & & \end{array}$$

Wir werden annehmen, dass jede Zeile dieses Dreiecks aus unabhängigen Zufallsvariablen besteht. Abhängigkeiten zwischen den Variablen aus verschiedenen Zeilen sind jedoch erlaubt.

SATZ 13.7.1 (Lindeberg). *Für alle $n \in \mathbb{N}$ gelte, dass $X_{n1}, X_{n2}, \dots, X_{nn}$ unabhängige Zufallsvariablen mit $\mathbb{E}X_{nk} = 0$ und $\text{Var } X_{nk} = \sigma_{nk}^2$ sind, wobei $\sum_{k=1}^n \sigma_{nk}^2 = 1$. Außerdem gelte die folgende Lindeberg-Bedingung: für jedes $\varepsilon > 0$*

$$(13.7.1) \quad L_n(\varepsilon) := \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}| > \varepsilon}] = 0.$$

Dann gilt:

$$X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

BEMERKUNG 13.7.2. Die Lindeberg-Bedingung besagt, dass alle Werte der Zufallsvariablen, die größer als ε sind, einen asymptotisch verschwindenden Beitrag zur Varianz der Summe der Zufallsvariablen leisten.

BEMERKUNG 13.7.3. Wir beweisen, dass der zentrale Grenzwertsatz ein Spezialfall von Satz 13.7.1 ist. Seien X_1, X_2, \dots unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_k] = \mu$ und $\text{Var } X_k = \sigma^2 = 0$. Wir definieren die Zufallsvariablen

$$X_{nk} = \frac{X_k - \mu}{\sigma\sqrt{n}}, \quad k = 1, \dots, n.$$

Dann sind die Zufallsvariablen X_{n1}, \dots, X_{nn} für jedes $n \in \mathbb{N}$ unabhängig, da X_1, \dots, X_n nach Voraussetzung unabhängig sind. Außerdem gilt $\mathbb{E}X_{nk} = 0$ und $\sigma_{nk}^2 = \text{Var } X_{nk} = \frac{1}{n}$. Somit gilt tatsächlich $\sum_{k=1}^n \sigma_{nk}^2 = 1$.

Um zu zeigen, dass alle Voraussetzungen aus Satz 13.7.1 erfüllt sind, müssen wir noch die Lindeberg-Bedingung (13.7.1) überprüfen. Sei $\varepsilon > 0$ fest. Es gilt, da die Zufallsvariablen identisch verteilt sind,

$$L_n(\varepsilon) = \sum_{k=1}^n \mathbb{E} \left[\left(\frac{X_k - \mu}{\sigma\sqrt{n}} \right)^2 \mathbb{1}_{\left| \frac{X_k - \mu}{\sigma\sqrt{n}} \right| > \varepsilon} \right] = \frac{1}{\sigma^2} \mathbb{E}[(X_1 - \mu)^2 \mathbb{1}_{|X_1 - \mu| > \varepsilon\sigma\sqrt{n}}].$$

Wir zeigen, dass der Erwartungswert auf der rechten Seite für $n \rightarrow \infty$ gegen 0 konvergiert. Es gilt

$$(X_1 - \mu)^2 \mathbb{1}_{|X_1 - \mu| > \varepsilon\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{f.s.} 0 \quad (\text{sogar sicher}),$$

da die Indikatorfunktion für hinreichend großes n gleich 0 wird. Außerdem gilt für alle n die Abschätzung

$$(X_1 - \mu)^2 \mathbb{1}_{|X_1 - \mu| > \varepsilon\sigma\sqrt{n}} \leq (X_1 - \mu)^2 \text{ mit } \mathbb{E}[(X_1 - \mu)^2] < \infty.$$

Damit können wir den Satz von der majorisierten Konvergenz anwenden. Somit gilt die Lindeberg-Bedingung (13.7.1).

Nun können wir Satz 13.7.1 von Lindeberg anwenden:

$$\frac{X_1 - \mu}{\sigma\sqrt{n}} + \dots + \frac{X_n - \mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

Der zentrale Grenzwertsatz ist also ein Spezialfall von Satz 13.7.1.

BEMERKUNG 13.7.4. Der Satz von Lindeberg ist etwas allgemeiner, als der zentrale Grenzwertsatz. Zum Beispiel wird im Satz von Lindeberg nicht verlangt, dass die Zufallsvariablen X_{n1}, \dots, X_{nn} identisch verteilt sein sollen.

BEISPIEL 13.7.5. Definiere $X_{n1} = X$, wobei X eine beliebige (aber nicht normalverteilte) Zufallsvariable mit $\mathbb{E}X = 0$ und $\text{Var } X = 1$ ist. Die restlichen Zufallsvariablen X_{n2}, \dots, X_{nn} seien gleich 0. Dann gilt

$$X_{n1} + \dots + X_{nn} = X \not\approx N(0, 1).$$

In diesem Fall konvergieren die Summen also nicht gegen die Normalverteilung. Die Ursache dafür ist, dass die Lindeberg-Bedingung (13.7.1) nicht erfüllt ist.

BEMERKUNG 13.7.6. Wenn die Lindeberg-Bedingung (13.7.1) erfüllt ist, so gilt, dass

$$\lim_{n \rightarrow \infty} \max_{k=1, \dots, n} \sigma_{nk}^2 = 0.$$

BEWEIS. Aus der Gleichung

$$\sigma_{nk}^2 = \mathbb{E}[X_{nk}^2] = \mathbb{E}[X_{nk}^2 \cdot \mathbb{1}_{|X_{nk}| \leq \varepsilon}] + \mathbb{E}[X_{nk}^2 \cdot \mathbb{1}_{|X_{nk}| > \varepsilon}]$$

folgt, dass für die maximale Varianz folgende Abschätzung gilt:

$$M_n := \max_{k=1, \dots, n} \sigma_{nk}^2 \leq \varepsilon^2 + \sum_{k=1}^n \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}| > \varepsilon}] = \varepsilon^2 + L_n(\varepsilon).$$

Da die Lindeberg-Summe auf der rechten Seite gegen 0 für $n \rightarrow \infty$ konvergiert, erhalten wir, dass $\limsup_{n \rightarrow \infty} M_n \leq \varepsilon^2$. Da das für jedes $\varepsilon > 0$ gilt, folgt, dass $\lim_{n \rightarrow \infty} M_n = 0$. \square

Zum Beweis von Satz 13.7.1 brauchen wir zunächst einige Hilfsaussagen.

LEMMA 13.7.7. Für alle $y_1, \dots, y_n \in \mathbb{C}$ und für alle $z_1, \dots, z_n \in \mathbb{C}$ mit $|y_k| \leq 1$ und $|z_k| \leq 1$ für alle $k = 1, \dots, n$, gilt:

$$\left| \prod_{k=1}^n y_k - \prod_{k=1}^n z_k \right| \leq \sum_{k=1}^n |y_k - z_k|.$$

BEWEIS. Wir benutzen die Induktion.

INDUKTIONSANFANG. Für $n = 1$ nimmt unsere Ungleichung die Form $|y_1 - z_1| \leq |y_1 - z_1|$ an. Sie stimmt also.

INDUKTIONSSCHRITT. Wir nehmen an, dass die Ungleichung für ein $n \in \mathbb{N}$ gilt. Nun gilt es, zu zeigen, dass sie auch für $n + 1$ gilt. Mit der Dreiecksungleichung erhalten wir, dass

$$\begin{aligned} \left| \prod_{k=1}^{n+1} y_k - \prod_{k=1}^{n+1} z_k \right| &\leq \left| \prod_{k=1}^{n+1} y_k - y_{n+1} \prod_{k=1}^n z_k \right| + \left| y_{n+1} \prod_{k=1}^n z_k - \prod_{k=1}^{n+1} z_k \right| \\ &= |y_{n+1}| \left| \prod_{k=1}^n y_k - \prod_{k=1}^n z_k \right| + \left| \prod_{k=1}^n z_k \right| \cdot |y_{n+1} - z_{n+1}| \\ &\leq \left| \prod_{k=1}^n y_k - \prod_{k=1}^n z_k \right| + |y_{n+1} - z_{n+1}|, \end{aligned}$$

da $|y_{n+1}| \leq 1$ und $|\prod_{k=1}^n z_k| \leq 1$. Nun können wir die Induktionsannahme anwenden:

$$\left| \prod_{k=1}^{n+1} y_k - \prod_{k=1}^{n+1} z_k \right| \leq \sum_{k=1}^n |y_k - z_k| + |y_{n+1} - z_{n+1}| = \sum_{k=1}^{n+1} |y_k - z_k|.$$

Somit ist die Ungleichung auch für $n + 1$ Terme gültig. \square

LEMMA 13.7.8. Sei X eine Zufallsvariable mit $\mathbb{E}[|X|^n] < \infty$ für ein $n \in \mathbb{N}$. Dann gilt die Taylor-Entwicklung

$$\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E}[X^k] + R_n(t),$$

wobei für den Restterm $R_n(t)$ die folgende Abschätzung gilt

$$|R_n(t)| \leq \mathbb{E} \left[\min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, 2 \frac{|tX|^n}{n!} \right\} \right].$$

BEMERKUNG 13.7.9. Wenn man sich erlaubt, Erwartungswerte und unendliche Summen ohne Begründung zu vertauschen, dann erhält man die Entwicklung

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E} \left[\sum_{k=0}^{\infty} \frac{(itX)^k}{k!} \right] = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mathbb{E}[X^k].$$

Das Lemma ist eine exakte Aussage darüber, inwieweit diese Entwicklung möglich ist. Hat die Zufallsvariable n Momente, so können wir in der Entwicklung Terme bis zum Grad n benutzen und den Restterm wie im Lemma abschätzen.

BEWEIS VON LEMMA 13.7.8.

SCHRITT 1. Zuerst beweisen wir eine Abschätzung für den Restterm in der Taylor-Entwicklung der Funktion e^{it} . Wir werden zeigen, dass für alle $t \in \mathbb{R}$

$$(13.7.2) \quad \left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| \leq \min \left\{ \frac{|t|^{n+1}}{(n+1)!}, 2 \frac{|t|^n}{n!} \right\}.$$

Wir führen den Beweis mit Induktion.

INDUKTIONSANFANG. Für $n = 0$ stimmt die zu beweisende Ungleichung (13.7.2), denn

$$|e^{it} - 1| \leq |t| \text{ und } |e^{it} - 1| \leq 2.$$

INDUKTIONSSCHRITT. Wir nehmen an, dass die Ungleichung (13.7.2) für ein $n \in \mathbb{N}$ gilt. Nun ist zu zeigen, dass sie auch für $n + 1$ gilt. Sei o.E.d.A. $t > 0$, denn die Funktion auf der rechten Seite von (13.7.2) ist gerade. Es gilt

$$\left| e^{it} - \sum_{k=0}^{n+1} \frac{(is)^k}{k!} \right| = \left| \int_0^t \left(e^{is} - \sum_{k=0}^n \frac{(is)^k}{k!} \right) ds \right| \leq \int_0^t \left| e^{is} - \sum_{k=0}^n \frac{(is)^k}{k!} \right| ds.$$

Nun benutzen wir die Induktionsannahme:

$$\begin{aligned} \left| e^{it} - \sum_{k=0}^{n+1} \frac{(it)^k}{k!} \right| &= \int_0^t \min \left\{ \frac{s^{n+1}}{(n+1)!}, 2 \frac{s^n}{n!} \right\} ds \\ &\leq \min \left\{ \int_0^t \frac{s^{n+1}}{(n+1)!}, 2 \int_0^t \frac{s^n}{n!} ds \right\} \\ &= \min \left\{ \frac{t^{n+2}}{(n+2)!}, 2 \frac{t^{n+1}}{(n+1)!} \right\}. \end{aligned}$$

Die zu beweisende Ungleichung gilt also auch für $n + 1$.

SCHRITT 2. Nun beweisen wir die Abschätzung für den Restterm $R_n(t)$:

$$|R_n(t)| = \left| \mathbb{E} \left[e^{itX} - \sum_{k=0}^n \frac{(itX)^k}{k!} \right] \right| \leq \mathbb{E} \left| e^{itX} - \sum_{k=0}^n \frac{(itX)^k}{k!} \right| \leq \mathbb{E} \left[\min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, 2 \frac{|tX|^n}{n!} \right\} \right].$$

Das Lemma ist somit bewiesen. □

Nun können wir den Satz von Lindeberg beweisen.

BEWEIS VON SATZ 13.7.1. Es ist zu zeigen, dass

$$X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

SCHRITT 1. Nach dem Stetigkeitssatz 13.4.1 reicht es, zu zeigen, dass für alle $t \in \mathbb{R}$ gilt:

$$\lim_{n \rightarrow \infty} \prod_{k=1}^n \varphi_{nk}(t) = e^{-\frac{t^2}{2}}.$$

Dabei sei $\varphi_{nk}(t) = \mathbb{E}[e^{itX_{nk}}]$ die charakteristische Funktion von X_{nk} . Deshalb betrachten wir die Differenz dieser beiden Terme:

$$\begin{aligned} \left| \prod_{k=1}^n \varphi_{nk}(t) - e^{-\frac{t^2}{2}} \right| &= \left| \prod_{k=1}^n \varphi_{nk}(t) - \prod_{k=1}^n e^{-\frac{1}{2}\sigma_{nk}^2 t^2} \right| \quad \left(\text{da } \sum_{k=1}^n \sigma_{nk}^2 = 1 \right) \\ &\leq \sum_{k=1}^n \left| \varphi_{nk}(t) - e^{-\frac{1}{2}\sigma_{nk}^2 t^2} \right| \\ &\leq \sum_{k=1}^n \left| \varphi_{nk}(t) - 1 + \frac{1}{2}\sigma_{nk}^2 t^2 \right| - \sum_{k=1}^n \left| e^{-\frac{1}{2}\sigma_{nk}^2 t^2} - 1 + \frac{1}{2}\sigma_{nk}^2 t^2 \right|. \end{aligned}$$

Nun schätzen wir diese beiden Summen separat ab und definieren hierfür

$$c_{nk} = \left| \varphi_{nk}(t) - 1 + \frac{1}{2}\sigma_{nk}^2 t^2 \right|, \quad d_{nk} = \left| e^{-\frac{1}{2}\sigma_{nk}^2 t^2} - 1 + \frac{1}{2}\sigma_{nk}^2 t^2 \right|.$$

SCHRITT 2. Betrachten zunächst c_{nk} . Mit Lemma 13.7.8 erhalten wir

$$c_{nk} = \left| \varphi_{nk}(t) - \left(1 - \frac{1}{2}\sigma_{nk}^2 t^2 \right) \right| \leq \mathbb{E} \left[\min \{ |tX_{nk}|^3, |tX_{nk}|^2 \} \right].$$

Und nun verwenden wir für $X_{nk} \leq \varepsilon$ die Abschätzung $|tX_{nk}|^3$ und für $X_{nk} > \varepsilon$ die Abschätzung $|tX_{nk}|^2$:

$$c_{nk} \leq \mathbb{E} \left[|tX_{nk}|^3 \mathbb{1}_{|X_{nk}| \leq \varepsilon} \right] + \mathbb{E} \left[|tX_{nk}|^2 \mathbb{1}_{|X_{nk}| > \varepsilon} \right].$$

Sei $L_n(\varepsilon) = \sum_{k=1}^n \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}| > \varepsilon}]$ die Summe aus der Lindeberg-Bedingung. Nun schätzen wir die Summe über c_{nk} ab:

$$\begin{aligned} \sum_{k=1}^n c_{nk} &\leq |t|^3 \sum_{k=1}^n \mathbb{E}[|X_{nk}|^3 \mathbb{1}_{|X_{nk}| \leq \varepsilon}] + |t|^2 L_n(\varepsilon) \\ &\leq |t|^3 \sum_{k=1}^n \varepsilon \cdot \mathbb{E}[|X_{nk}|^2 \mathbb{1}_{|X_{nk}| \leq \varepsilon}] + |t|^2 L_n(\varepsilon) \\ &\leq |t|^3 \varepsilon \sum_{k=1}^n \sigma_{nk}^2 + |t|^2 L_n(\varepsilon) \\ &= |t|^3 \varepsilon + |t|^2 L_n(\varepsilon). \end{aligned}$$

Da die Lindenberg-Summe $L_n(\varepsilon)$ gegen 0 für $n \rightarrow \infty$ konvergiert und $\varepsilon > 0$ beliebig klein gewählt werden kann, erhalten wir, dass

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n c_{nk} = 0.$$

SCHRITT 3. Wir betrachten nun d_{nk} . Wir verwenden hier die Abschätzung

$$|e^z - (1 - z)| = \left| \sum_{k=2}^{\infty} \frac{z^k}{k!} \right| \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!} \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{(k-2)!} \leq |z|^2 e^{|z|}.$$

Und nun substituieren wir $z = -\frac{1}{2}\sigma_{nk}^2 t^2$:

$$d_{nk} = \left| e^{-\frac{1}{2}\sigma_{nk}^2 t^2} - \left(1 - \frac{1}{2}\sigma_{nk}^2 t^2 \right) \right| \leq \sigma_{nk}^4 |t|^4 e^{\frac{1}{2}\sigma_{nk}^2 t^2} \leq \sigma_{nk}^4 |t|^4 e^{t^2}.$$

Somit gilt für die Summe

$$\sum_{k=1}^n d_{nk} \leq |t|^4 e^{t^2} \sum_{k=1}^n \sigma_{nk}^4 \leq |t|^4 e^{t^2} \max_{k=1, \dots, n} \sigma_{nk}^2.$$

Also gilt laut Bemerkung 13.7.6

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n d_{nk} = 0.$$

SCHRITT 4. Wir haben also gezeigt, dass

$$\lim_{n \rightarrow \infty} \left| \prod_{k=1}^n \varphi_{nk}(t) - e^{-\frac{t^2}{2}} \right| = 0.$$

Mit dem Stetigkeitssatz 13.4.1 folgt daraus, dass $X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$. □

Nun beweisen wir einen Satz, der eine vereinfachte Form des Satzes von Lindeberg ist.

SATZ 13.7.10 (Ljapunow). Für alle $n \in \mathbb{N}$ seien X_{n1}, \dots, X_{nn} unabhängige Zufallsvariablen mit $\mathbb{E}X_{nk} = 0$ und $\text{Var} X_{nk} = \sigma_{nk}^2$, wobei $\sum_{k=1}^n \sigma_{nk}^2 = 1$. Es gelte außerdem die folgende Ljapunow-Bedingung: Es existiert ein $\delta > 0$ mit

$$(13.7.3) \quad \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}|X_{nk}|^{2+\delta} = 0.$$

Dann gilt:

$$X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

BEWEIS. Wir zeigen, dass die Ljapunow-Bedingung (13.7.3) die Lindeberg-Bedingung (13.7.1) impliziert. Sei also (13.7.3) erfüllt. Sei $\varepsilon > 0$ beliebig. Dann gilt

$$L_n(\varepsilon) = \sum_{k=1}^n \mathbb{E} [X_{nk}^2 \mathbb{1}_{|X_{nk}| > \varepsilon}] \leq \sum_{k=1}^n \mathbb{E} \left[\frac{|X_{nk}|^{2+\delta}}{\varepsilon^\delta} \mathbb{1}_{|X_{nk}| > \varepsilon} \right] \leq \frac{1}{\varepsilon^\delta} \sum_{k=1}^n \mathbb{E} [|X_{nk}|^{2+\delta}] \xrightarrow[n \rightarrow \infty]{} 0.$$

Somit gilt die Lindeberg-Bedingung (13.7.1). □

BEISPIEL 13.7.11. Wir betrachten ein n -faches Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p_n , für die $\lim_{n \rightarrow \infty} np_n(1 - p_n) = \infty$ gilt. Dann gilt:

$$(13.7.4) \quad \frac{S_n - np_n}{\sqrt{np_n(1 - p_n)}} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

BEMERKUNG 13.7.12. Dies ist beispielsweise anwendbar, wenn $p_n = \frac{1}{\sqrt{n}}$, $p_n = 1 - \frac{1}{\sqrt{n}}$ oder wenn $p_n = p \in (0, 1)$ konstant ist. Zum Vergleich: Der Poisson-Grenzwertsatz gilt, wenn $\lim_{n \rightarrow \infty} np_n = \lambda \in (0, \infty)$.

BEWEIS VON (13.7.4). Den Satz von de Moivre-Laplace können wir nicht anwenden, denn p_n ist im Allgemeinen nicht konstant. Wir werden stattdessen den Satz von Lindeberg anwenden. Wir betrachten unabhängige Zufallsvariablen $Y_{n1}, \dots, Y_{nn} \sim \text{Bin}(1, p_n)$, d.h.

$$\mathbb{P}[Y_{nk} = 1] = p_n, \quad \mathbb{P}[Y_{nk} = 0] = 1 - p_n.$$

Für die Summe dieser Zufallsvariablen gilt dann

$$S_n := Y_{n1} + \dots + Y_{nn} \sim \text{Bin}(n, p_n).$$

Wir betrachten auch die normierten Zufallsvariablen

$$X_{nk} := \frac{Y_{nk} - p_n}{\sqrt{np_n(1 - p_n)}}.$$

Es gilt $\mathbb{E}X_{nk} = 0$ und $\sum_{k=1}^n \text{Var} X_{nk} = 1$. Für die Summe dieser Zufallsvariablen X_{nk} gilt dann

$$S_n^* := X_{n1} + \dots + X_{nn} = \frac{S_n - np_n}{\sqrt{np_n(1 - p_n)}}.$$

Wir zeigen, dass S_n^* in Verteilung gegen die Standardnormalverteilung konvergiert, indem wir die Lindeberg-Bedingung überprüfen. Sei $\varepsilon > 0$ vorgegeben. Wir können die Zufallsvariable X_{nk} wie folgt abschätzen:

$$|X_{nk}| \leq \frac{1}{\sqrt{np_n(1 - p_n)}}.$$

Somit ist $|X_{nk}| < \varepsilon$ für hinreichend großes n , denn $\lim_{n \rightarrow \infty} np_n(1 - p_n) = \infty$. Es gilt also $L_n(\varepsilon) = \sum_{k=1}^n \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}| > \varepsilon}] = 0$ für hinreichend großes n . Somit ist die Lindeberg-Bedingung erfüllt. Nun darf der zentrale Grenzwertsatz von Lindeberg benutzt werden. Er besagt, dass

$$S_n^* \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

□

BEISPIEL 13.7.13. Seien X_1, X_2, \dots unabhängige (aber nicht identisch verteilte) Zufallsvariablen mit

$$\mathbb{P}[X_n = n] = \mathbb{P}[X_n = -n] = \frac{1}{2}.$$

Definiere die Partialsummen

$$S_n = X_1 + \dots + X_n = \pm 1 \pm 2 \pm \dots \pm n.$$

Wir behaupten nun, dass

$$\sqrt{3} \cdot \frac{S_n}{n^{3/2}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

BEWEIS. Den zentralen Grenzwertsatz können wir nicht anwenden, da die Zufallsvariablen nicht identisch verteilt sind. Stattdessen muss man den Satz von Lindeberg oder den Satz von Ljapunow anwenden. Wir entscheiden uns für den Satz von Ljapunow. Zunächst bemerken wir, dass $\mathbb{E}X_k = 0$ und $\text{Var } X_k = \mathbb{E}[X_k^2] = k^2$. Die Varianz der Summe S_n ist dann

$$D_n^2 := \text{Var } S_n = \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6} \sim \frac{n^3}{3}, \quad n \rightarrow \infty.$$

Daraus folgt, dass

$$D_n \sim \frac{n^{3/2}}{\sqrt{3}}, \quad n \rightarrow \infty.$$

Um nun den zentralen Grenzwertsatz von Ljapunow anwenden zu können, definieren wir uns folgende Zufallsvariablen:

$$X_{nk} = \frac{X_k}{D_n} \text{ für } k = 1, \dots, n.$$

Wir erkennen, dass

$$\mathbb{E}X_{nk} = 0, \quad \sum_{k=1}^n \text{Var } X_{nk} = \frac{1}{D_n^2} \sum_{k=1}^n \text{Var } X_k = 1.$$

Nun zeigen wir, dass die Ljapunow-Bedingung mit $\delta = 1$ erfüllt ist:

$$\sum_{k=1}^n \mathbb{E}[|X_{nk}|^3] = \sum_{k=1}^n \left(\frac{k}{D_n} \right)^3 = \frac{1}{D_n^3} \sum_{k=1}^n k^3 \xrightarrow[n \rightarrow \infty]{} 0,$$

denn $\sum_{k=1}^n k^3 \sim \frac{n^4}{4}$ und $D_n^3 \sim \frac{n^{9/2}}{3\sqrt{3}}$ für $n \rightarrow \infty$. Der Satz Ljapunow besagt dann, dass

$$\frac{1}{D_n} \sum_{k=1}^n X_k = \sum_{k=1}^n X_{nk} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

Also ist die Behauptung bewiesen, da $D_n \sim \frac{n^{3/2}}{\sqrt{3}}$ für $n \rightarrow \infty$ ist.

KAPITEL 14

Irrfahrt

14.1. Berechnung einer Ruinwahrscheinlichkeit

Das folgende sogenannte *Ruinproblem* wurde (in einer etwas anderen Formulierung) von Pascal in einem Brief an Fermat im Jahr 1656 gestellt.

BEISPIEL 14.1.1. Man stelle sich folgendes Spiel vor. Am Anfang des Spiels besitzt man ein Startkapital von k Euro, wobei $k \in \mathbb{N}_0$. In jeder Spielrunde kann man entweder mit Wahrscheinlichkeit p ein Euro gewinnen (wodurch sich das Kapital um 1 vergrößert) oder mit Wahrscheinlichkeit q ein Euro verlieren (wodurch sich das Kapital um 1 verkleinert). Hierbei gilt, dass $p, q \in (0, 1)$ und $p + q = 1$. Die Spielrunden seien unabhängig voneinander. Erreicht das Kapital zu einem Zeitpunkt den Wert 0, so sagen wir, dass ein *Ruin* eingetreten ist. Dieses Ereignis bezeichnen wir mit R . Nach dem Ruin wird das Spiel beendet. Erreicht das Kapital zu einem Zeitpunkt einen zuvor vorgegebenen Wert $M \geq k$ (der *Zielkapital* genannt wird), so steigt man aus dem Spiel aus. Dieses Ereignis bezeichnen wir mit G . Die Spielrunden werden wiederholt, bis eines der beiden Ereignisse R oder G eintritt. Man bestimme die Wahrscheinlichkeiten der Ereignisse R und G .

LÖSUNG. SCHRITT 0. Es sei p_i die Wahrscheinlichkeit von Ruin in einem Spiel, das mit Startkapital $k = i$ startet. In der ersten Spielrunde kann man entweder 1 Euro gewinnen (Wahrscheinlichkeit = p) oder 1 Euro verlieren (Wahrscheinlichkeit = q). Hat man 1 Euro gewonnen, so hat man nach der ersten Runde ein Kapital von $i + 1$ Euro und die Wahrscheinlichkeit, dass irgendwann später der Ruin eintreten wird, ist p_{i+1} . Hat man 1 Euro verloren, so hat man nach der ersten Runde ein Kapital von $i - 1$ Euro und die Wahrscheinlichkeit, dass irgendwann später der Ruin eintritt wird, ist gleich p_{i-1} . Nach dem Satz der totalen Wahrscheinlichkeit erhalten wir, dass

$$(14.1.1) \quad p_i = p \cdot p_{i+1} + q \cdot p_{i-1}, \quad i = 1, \dots, M - 1.$$

Außerdem gelten trivialerweise die "Randbedingungen" $p_0 = 1$ und $p_M = 0$. Wir haben also ein lineares Gleichungssystem mit $M + 1$ unbekanntem, nämlich p_0, \dots, p_M , und $M + 1$ Gleichungen. Dieses Gleichungssystem werden wir nun lösen.

SCHRITT 1. Es gilt

$$(14.1.2) \quad p_i = p \cdot p_i + q \cdot p_i, \quad i = 1, \dots, M - 1,$$

denn $p + q = 1$. Zieht man nun die Gleichungen (14.1.1) und (14.1.2) voneinander ab, so erhält man

$$p(p_{i+1} - p_i) + q(p_{i-1} - p_i) = 0.$$

Führt man die Notation $d_i := p_i - p_{i+1}$ ein, so erhält man

$$d_i = \frac{q}{p} d_{i-1}, \quad i = 1, \dots, M-1.$$

Benutzt man diese Gleichung iterativ, so erhält man

$$(14.1.3) \quad d_i = \frac{q}{p} d_{i-1} = \left(\frac{q}{p}\right)^2 d_{i-2} = \dots = \left(\frac{q}{p}\right)^i d_0, \quad i = 1, \dots, M-1.$$

SCHRITT 2. Nun berechnen wir d_0 . Dazu setzen wir die Randbedingungen $p_0 = 1$ und $p_M = 0$ ein:

$$(14.1.4) \quad 1 = p_0 - p_M = \sum_{i=0}^{M-1} d_i = d_0 \sum_{i=0}^{M-1} \left(\frac{q}{p}\right)^i.$$

Wir müssen die Summe dieser geometrischen Folge berechnen. Hierzu unterscheiden wir zwei Fälle:

FALL A. Es gelte $p = q = 1/2$, also $q/p = 1$. Dann ergibt sich aus (14.1.4), dass $1 = d_0 \cdot M$, also $d_0 = 1/M$. Mit (14.1.3) folgt, dass $d_i = 1/M$ und somit

$$p_i = p_i - p_M = d_i + \dots + d_{M-1} = \frac{M-i}{M}, \quad i = 0, \dots, M.$$

Somit ist p_i eine lineare Funktion von i , die zwischen den Werten $p_0 = 1$ und $p_M = 0$ interpoliert.

FALL B. Es gelte $p \neq q$, also $r := q/p \neq 1$. Dann ergibt sich aus (14.1.4), dass $1 = d_0 \cdot \frac{1-r^M}{1-r}$. Daraus folgt, dass $d_0 = \frac{1-r}{1-r^M}$. Mit (14.1.3) erhalten wir, dass

$$d_i = r^i d_0 = \frac{r^i - r^{i+1}}{1 - r^M}, \quad i = 0, \dots, M-1.$$

Daraus erhalten wir schließlich

$$p_i = p_i - p_M = \sum_{j=i}^{M-1} d_j = \sum_{j=i}^{M-1} \frac{r^j - r^{j+1}}{1 - r^M} = \frac{r^i - r^M}{1 - r^M}, \quad i = 0, \dots, M.$$

Wir haben also die Ruinwahrscheinlichkeit bestimmt und folgendes Ergebnis erhalten: Bei Startkapital k ist die Ruinwahrscheinlichkeit gleich

$$(14.1.5) \quad p_k = \begin{cases} \frac{M-k}{M}, & \text{falls } p = q = \frac{1}{2}, \\ \frac{r^k - r^M}{1 - r^M}, & \text{falls } r := \frac{q}{p} \neq 1, \end{cases} \quad k = 0, \dots, M.$$

Es sei bemerkt, dass der erste Fall als der Grenzwert des zweiten Falls angesehen werden kann, denn

$$\lim_{r \rightarrow 1} \frac{r^k - r^M}{1 - r^M} = \frac{M-k}{M}.$$

Völlig analog kann man ein Gleichungssystem zur Bestimmung der Wahrscheinlichkeit des Ereignisses G aufstellen und lösen. Es gibt aber einen schnelleren Weg, die Wahrscheinlichkeit von G zu berechnen.

SCHRITT 3. Damit das Ereignis R eintritt, muss man k Euro verlieren, bevor man $M - k$ Euro gewonnen hat. Dabei ist die Wahrscheinlichkeit, ein Euro in einer Runde zu verlieren, gleich q . Damit das Ereignis G eintritt, muss man $M - k$ Euro gewinnen, bevor man k Euro verloren hat. Dabei ist die Wahrscheinlichkeit, ein Euro in einer Runde zu gewinnen, gleich p . Um die Wahrscheinlichkeit von G zu bestimmen, müssen wir also in der Formel für die Wahrscheinlichkeit von R folgende Substitutionen machen:

$$p \leftrightarrow q, \quad r \leftrightarrow 1/r, \quad k \leftrightarrow M - k.$$

Für die Wahrscheinlichkeit q_k von G bei Startkapital k ergibt sich die folgende Formel:

$$q_k = \begin{cases} \frac{M-(M-k)}{M} = \frac{k}{M}, & \text{falls } p = q = \frac{1}{2}, \\ \frac{\frac{1}{r^{M-k}} - \frac{1}{r^M}}{1 - \frac{1}{r^M}} = \frac{1-r^k}{1-r^M}, & \text{falls } p \neq q. \end{cases}$$

Schaut man sich nun die Formeln für p_k und q_k an, so erkennt man, dass $p_k + q_k = 1$. Das Spiel wird also mit Wahrscheinlichkeit 1 in einer endlichen Zeit beendet. Die Wahrscheinlichkeit, dass das Spiel ewig dauert, ist somit 0.

BEISPIEL 14.1.2. Nun nehmen wir an, dass das Zielkapital $M = +\infty$ ist. Man spielt also, bis man ruiniert ist (bzw. ewig, wenn der Ruin nicht eintritt). Wir berechnen die Ruinwahrscheinlichkeit p_k bei Startkapital k . Dazu lassen wir $M \rightarrow \infty$ in (14.1.5):

$$p_k = \begin{cases} 1, & \text{falls } p = q = 1/2, \\ 1, & \text{falls } p < q, \\ r^k, & \text{falls } p > q. \end{cases}$$

Im Fall $p < q$ ist das Spiel *unvorteilhaft* (man verliert mit einer größeren Wahrscheinlichkeit, als man gewinnt). Deshalb ist es keine Überraschung, dass in diesem Fall der Ruin mit Wahrscheinlichkeit 1 eintritt. Der Ruin tritt aber auch dann mit Wahrscheinlichkeit 1 ein, wenn das Spiel *fair* ist (d.h. im Fall $p = q = 1/2$). Im Fall $p > q$ ist das Spiel *vorteilhaft* (man gewinnt mit einer größeren Wahrscheinlichkeit, als man verliert). In diesem Fall ist die Ruinwahrscheinlichkeit gleich r^k . Man beachte, dass bei einem Startkapital, das gegen $+\infty$ strebt, die Ruinwahrscheinlichkeit gegen 0 geht.

14.2. Rückkehr der Irrfahrt zum Ursprung

DEFINITION 14.2.1. Seien X_1, X_2, \dots unabhängige identisch verteilte Zufallsvariablen mit

$$\mathbb{P}[X_i = 1] = p \text{ und } \mathbb{P}[X_i = -1] = q,$$

wobei $p + q = 1$ und $p, q \in (0, 1)$. Dann ist eine *Irrfahrt* die Folge S_0, S_1, S_2, \dots mit

$$S_0 = 0, \quad S_n = X_1 + \dots + X_n, \quad n \in \mathbb{N}.$$

Die Irrfahrt heißt *symmetrisch*, wenn $p = q = 1/2$.

BEMERKUNG 14.2.2. Eine Irrfahrt ist beispielsweise ein Modell für ein Glücksspiel oder ein diffundierendes Teilchens.

Sei nun Z der erste Zeitpunkt, zu dem die Irrfahrt zum Ursprung zurückkehrt, d.h.

$$Z = \min\{n \in \mathbb{N} : S_n = 0\}.$$

Wenn es keine Rückkehr zum Ursprung gibt, so definieren wir $Z = \infty$. Wie ist nun die Zufallsvariable Z verteilt? Wir definieren die Zähldichte von Z :

$$f_n = \mathbb{P}[Z = n], \quad n \in \mathbb{N}.$$

SATZ 14.2.3. Für die Zähldichte des ersten Rückkehrzeitpunktes gilt

$$(14.2.1) \quad f_{2n} = \frac{1}{2n-1} \binom{2n}{n} p^n q^n, \quad f_{2n+1} = 0.$$

BEWEIS.

SCHRITT 1. Unser Ziel ist es, die Zähldichte f_n zu berechnen. Dazu betrachten wir eine verwandte Größe, nämlich

$$u_n = \mathbb{P}[S_n = 0], \quad n \in \mathbb{N}_0.$$

Dabei ist $u_0 = 1$. Hier unterscheiden sich f_n und u_n dadurch, dass bei f_n die *erste* Rückkehr zum Ursprung betrachtet wird, bei u_n hingegen *irgendeine* Rückkehr zum Ursprung.

Trivialerweise ist es nicht möglich, nach einer ungeraden Anzahl von Schritten zum Ursprung zurückzukehren. Daher gilt $u_{2n+1} = 0$. Bei gerader Anzahl von Schritten muss die Irrfahrt genau n Schritte nach oben und n Schritte nach unten machen, um zum Zeitpunkt $2n$ wieder bei 0 anzukommen. Daher gilt (wegen der Binomialverteilung)

$$(14.2.2) \quad u_{2n} = \mathbb{P}[S_{2n} = 0] = \binom{2n}{n} p^n q^n, \quad u_{2n+1} = 0.$$

SCHRITT 2. Nun stellen wir einen Zusammenhang zwischen u_n und f_n her. Damit die Irrfahrt zum Zeitpunkt n zum Ursprung zurückkehrt, muss die *erste* Rückkehr zum Ursprung zu einem Zeitpunkt $k = 1, \dots, n$ geschehen. Wenn die Irrfahrt zum Zeitpunkt k zum ersten Mal zum Ursprung zurückkehrt, so ist der Fall für f_k eingetreten. Von hier an kann die Irrfahrt beliebig weiter laufen mit der einzigen Bedingung, dass sie zum Zeitpunkt n wieder zum Ursprung zurückkehrt. Dabei darf sie zwischen k und n zum Ursprung zurückkehren, muss aber nicht. Damit erhalten wir zwei unabhängige Ereignisse f_n und u_{n-k} für die beiden Zeitabschnitte. Für u_n gilt somit die Formel:

$$(14.2.3) \quad u_n = \sum_{k=1}^n f_k \cdot u_{n-k}.$$

Diese Gleichung werden wir nun benutzen, um f_k zu bestimmen.

SCHRITT 3. Die Faltung ist schwierig zu handhaben, daher arbeiten wir stattdessen bevorzugt mit erzeugenden Funktionen $u(s)$ und $f(s)$, die folgendermaßen definiert werden:

$$u(s) = \sum_{n=0}^{\infty} u_n s^n, \quad f(s) = \sum_{n=1}^{\infty} f_n s^n, \quad |s| \leq 1.$$

Beide Funktionen sind wohldefiniert für $|s| \leq 1$, denn $0 \leq u_n \leq 1$ und $0 \leq f_n \leq 1$. Nun kann man (14.2.3) benutzen, um $u(s)$ in Abhängigkeit von $f(s)$ darzustellen:

$$\begin{aligned}
 u(s) &= \sum_{n=0}^{\infty} u_n s^n \\
 &= 1 + \sum_{n=1}^{\infty} \left(\sum_{k=1}^n f_k \cdot u_{n-k} \right) s^n \\
 &= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^n (f_k s^k) \cdot (u_{n-k} s^{n-k}) \\
 &= 1 + \left(\sum_{n=1}^{\infty} f_n s^n \right) \cdot \left(\sum_{l=0}^{\infty} u_l s^l \right) \\
 &= 1 + f(s) \cdot u(s).
 \end{aligned}$$

Man sieht, dass wir durch die Anwendung der erzeugenden Funktionen die Faltung in ein Produkt umgewandelt haben. Also gilt die Formel

$$f(s) = \frac{u(s) - 1}{u(s)}, \quad s \in [-1, 1].$$

SCHRITT 4. Um nun $f(s)$ bestimmen zu können, benötigen wir zunächst $u(s)$:

$$u(s) = \sum_{n=0}^{\infty} u_n s^n = \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n \cdot s^{2n} = \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} (-4pqs^2)^n = \frac{1}{\sqrt{1-4pqs^2}}.$$

Hier haben wir verwendet, dass

$$\begin{aligned}
 \binom{-\frac{1}{2}}{n} &= \frac{-\frac{1}{2} \cdot (-\frac{1}{2} - 1) \cdot (-\frac{1}{2} - 2) \cdot \dots \cdot (-\frac{1}{2} - n + 1)}{n!} \\
 &= (-1)^n \frac{1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)}{2^n n!} \\
 &= (-1)^n \frac{\binom{2n}{n}}{4^n}.
 \end{aligned}$$

Danach haben wir die Newton-Formel $(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$ mit $\alpha = -1/2$ benutzt.

SCHRITT 5. Damit gilt

$$f(s) = \frac{u(s) - 1}{u(s)} = \frac{\frac{1}{\sqrt{1-4pqs^2}} - 1}{\frac{1}{\sqrt{1-4pqs^2}}} = 1 - \sqrt{1-4pqs^2}.$$

Um nun die Folge f_n zu berechnen, müssen wir die Funktion $f(s)$ in eine Taylor-Reihe entwickeln. Wir benutzen wieder die Newton-Formel $(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$, aber diesmal mit $\alpha = 1/2$:

$$f(s) = - \sum_{n=1}^{\infty} \binom{\frac{1}{2}}{n} (-4pqs^2)^n = \sum_{n=1}^{\infty} \frac{1}{2n-1} \binom{2n}{n} p^n q^n s^{2n}.$$

Dabei haben wir die Formel $-(-4)^n \binom{\frac{1}{2}}{n} = \frac{1}{2n-1} \binom{2n}{n}$ benutzt (Beweis: Übung). Also gilt

$$f_{2n} = \frac{1}{2n-1} \binom{2n}{n} p^n q^n = \frac{u_{2n}}{2n-1}, \quad f_{2n+1} = 0.$$

Die Formel für f_{2n+1} ist übrigens offensichtlich, denn eine Rückkehr zum Ursprung zu einem ungeraden Zeitpunkt ist nicht möglich. Die Formel für f_{2n} ist aber nicht trivial. \square

BEISPIEL 14.2.4. Wir definieren das Ereignis

$$A = \{\exists n \in \mathbb{N} : S_n = 0\} = \text{“Es gibt eine Rückkehr der Irrfahrt zum Ursprung”}.$$

Nun behaupten wir, dass für die Wahrscheinlichkeit dieses Ereignisses gilt:

$$\mathbb{P}[A] = 1 - |2p - 1|.$$

BEWEIS. Es gibt genau dann eine Rückkehr zum Ursprung, wenn $Z \neq \infty$. Somit gilt

$$\mathbb{P}[A] = \mathbb{P}[Z \neq \infty] = f_1 + f_2 + \dots = f(1).$$

Mit der Formel für $f(s)$ erhalten wir

$$f(1) = 1 - \sqrt{1 - 4pq} = 1 - \sqrt{(2p - 1)^2} = 1 - |2p - 1|.$$

dabei haben wir benutzt, dass $q = 1 - p$. \square

BEMERKUNG 14.2.5. Im Fall $p = 1/2$ ist die Wahrscheinlichkeit, dass es eine Rückkehr zum Ursprung gibt, gleich 1. Da die Irrfahrt nach einer Rückkehr zum Ursprung wieder neu (und unabhängig von der Vergangenheit) startet, wird sie mit Wahrscheinlichkeit 1 wieder zum Ursprung zurückkehren, usw. Mit anderen Worten, die Irrfahrt kehrt mit Wahrscheinlichkeit 1 sogar unendlich oft zum Ursprung zurück. Man sagt, dass die Irrfahrt für $p = 1/2$ *rekurrent* ist. Für $p \neq 1/2$ ist die Irrfahrt *transient*, d.h. sie kehrt mit Wahrscheinlichkeit 1 nur endlich oft zum Ursprung zurück.

Im Fall $p = 1/2$ findet eine Rückkehr zum Ursprung mit Wahrscheinlichkeit 1 statt. Wie lange muss man im Durchschnitt auf die erste Rückkehr warten? Die Antwort auf diese Frage fällt etwas unerwartet aus.

SATZ 14.2.6. Sei $p = 1/2$ und Z der Zeitpunkt der ersten Rückkehr der Irrfahrt zum Ursprung. Dann gilt $\mathbb{E}Z = +\infty$.

BEWEIS. Die erzeugende Funktion von Z ist

$$g_Z(s) = f(s) = 1 - \sqrt{1 - s^2}.$$

Daraus ergibt sich dann für den Erwartungswert $\mathbb{E}Z = g'_Z(1) = +\infty$. \square

Im nächsten Satz werden wir die Wahrscheinlichkeiten u_n und f_n approximativ (für großes n) berechnen. Dazu benötigen wir zunächst folgende Notation:

DEFINITION 14.2.7. Zwei Folgen a_1, a_2, \dots und b_1, b_2, \dots heißen *asymptotisch äquivalent*, wenn

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1.$$

BEZEICHNUNG: $a_n \sim b_n$ für $n \rightarrow \infty$.

SATZ 14.2.8. Für $p = 1/2$ gilt

$$u_{2n} \sim \frac{1}{\sqrt{\pi n}} \text{ und } f_{2n} \sim \frac{1}{2\sqrt{\pi n^3}} \text{ für } n \rightarrow \infty.$$

LÖSUNG. Zum Beweis benötigen wir die Stirling-Formel:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad n \rightarrow \infty.$$

Damit erhalten wir

$$u_{2n} = \binom{2n}{n} \frac{1}{2^{2n}} = \frac{(2n)!}{(n!)^2 \cdot 2^{2n}} \sim \frac{\sqrt{2\pi \cdot 2n} \cdot \left(\frac{2n}{e}\right)^{2n}}{(\sqrt{2\pi n})^2 \cdot \left(\frac{n}{e}\right)^{2n} \cdot 2^{2n}} = \frac{1}{\sqrt{\pi n}}.$$

Damit gilt auch $f_{2n} = \frac{u_{2n}}{2^{2n-1}} \sim \frac{1}{2n} \cdot \frac{1}{\sqrt{\pi n}} = \frac{1}{2\sqrt{\pi n^3}}$. □

BEMERKUNG 14.2.9. Mit dem obigen Satz kann man Satz 14.2.6 auf eine andere Weise beweisen. Es gilt nämlich

$$\mathbb{E}Z = \sum_{n=1}^{\infty} f_{2n} \cdot 2n = \infty, \quad \text{da } f_{2n} \cdot 2n \sim \frac{1}{\sqrt{\pi n}} \text{ und } \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty.$$

BEMERKUNG 14.2.10. Sei K_{2n} eine Zufallsvariable, welche angibt, wie oft die Irrfahrt im Zeitintervall $[1, 2n]$ zum Ursprung zurückkehrt, d.h.

$$K_{2n} = \sum_{k=1}^n \mathbb{1}_{S_{2k}=0} = \#\{k : S_{2k} = 0, 1 \leq k \leq n\}.$$

Wir werden nun den Erwartungswert von K_{2n} bestimmen. Man könnte zunächst vermuten, dass der Erwartungswert für großes n approximativ proportional zu $2n$ sein sollte, da man erwartet, dass bei einem beispielsweise doppelt so großen Zeitintervall auch doppelt so viele Rückkehrzeitpunkte auftauchen. Diese Überlegung ist falsch, wie wir nun rechnerisch zeigen:

$$\mathbb{E}K_{2n} = \mathbb{E} \left[\sum_{k=1}^n \mathbb{1}_{S_{2k}=0} \right] = \sum_{k=1}^n \mathbb{P}[S_{2k} = 0] \sim \frac{1}{\sqrt{\pi}} \sum_{k=1}^n \frac{1}{\sqrt{k}} \sim 2\sqrt{\frac{n}{\pi}}, \quad n \rightarrow \infty.$$

Der vorletzte Schritt gilt wegen Satz 14.2.8. Zu zeigen, dass \sum und \sim vertauscht werden können, ist eine nichttriviale Übung. Also ist $\mathbb{E}K_{2n}$ approximativ proportional zu \sqrt{n} und nicht zu n .

Dieses Ergebnis kann man sich folgendermaßen veranschaulichen. Wir haben gezeigt, dass die Irrfahrt mit Wahrscheinlichkeit 1 unendlich oft zum Ursprung zurückkehrt. In diesem Sinne ist sie mit einer Sinusfunktion vergleichbar. Jedoch ist bei der Sinusfunktion die Anzahl der Rückkehrpunkte proportional zu n . Es ist also besser, die Irrfahrt nicht mit einer normalen Sinusfunktion zu vergleichen, sondern mit einer Sinusfunktion, bei welcher der Abstand zwischen den Rückkehrpunkten immer größer wird.

14.3. Verteilung des Maximums der Irrfahrt

Sei S_0, S_1, S_2, \dots eine Irrfahrt mit $p = 1/2$. Wir bezeichnen mit M_n den maximalen Wert der Irrfahrt auf dem Zeitintervall $[0, n]$, d.h.

$$M_n = \max_{k=0, \dots, n} S_k.$$

Aus $S_0 = 0$ folgt, dass $M_n \geq 0$. Im nächsten Satz bestimmen wir die asymptotische Verteilung von M_n für $n \rightarrow \infty$.

SATZ 14.3.1. Für jedes $x > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{M_n}{\sqrt{n}} \leq x \right] = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

Mit anderen Worten:

$$\frac{M_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} |N|, \text{ wobei } N \sim N(0, 1).$$

BEWEIS. Sei $m \in \mathbb{N}$. Mit dem Spiegelungsprinzip (Bild fehlt) erhalten wir

$$\begin{aligned} (14.3.1) \quad \mathbb{P}[M_n \geq m] &= \mathbb{P}[M_n \geq m, S_n < m] + \mathbb{P}[M_n \geq m, S_n = m] + \mathbb{P}[M_n \geq m, S_n > m] \\ &= \mathbb{P}[S_n > m] + \mathbb{P}[S_n = m] + \mathbb{P}[S_n > m] \\ &= 2\mathbb{P}[S_n > m] + \mathbb{P}[S_n = m] \\ &= 2\mathbb{P}[S_n \geq m] - \mathbb{P}[S_n = m]. \end{aligned}$$

Sei jetzt $x > 0$. Die Zahl $x\sqrt{n}$ muss nicht ganzzahlig sein. Wir können aber eine ganze Zahl m_n mit $m_n - 1 \leq x\sqrt{n} < m_n$ finden. Dann gilt:

$$\begin{aligned} \mathbb{P} \left[\frac{M_n}{\sqrt{n}} > x \right] &= \mathbb{P}[M_n \geq m_n] \\ &= 2\mathbb{P}[S_n \geq m_n] - \mathbb{P}[S_n = m_n] \\ &= 2\mathbb{P} \left[\frac{S_n}{\sqrt{n}} \geq \frac{m_n}{\sqrt{n}} \right] - \mathbb{P}[S_n = m_n] \\ &\xrightarrow[n \rightarrow \infty]{} 2\mathbb{P}[N \geq x] - 0, \end{aligned}$$

wobei $N \sim N(0, 1)$ und wir den zentralen Grenzwertsatz benutzt haben. Außerdem haben wir benutzt, dass $\lim_{n \rightarrow \infty} \mathbb{P}[S_n = m_n] = 0$ ist. Hier ist der Beweis. Der Abstand zwischen $x\sqrt{n}$ und m_n ist höchstens 1. Sei $\varepsilon > 0$ fest. Somit liegt m_n immer zwischen $x\sqrt{n}$ und $(x + \varepsilon)\sqrt{n}$, zumindest wenn n groß genug ist. Also gilt:

$$\mathbb{P}[S_n = m_n] \leq \mathbb{P} \left[x \leq \frac{S_n}{\sqrt{n}} \leq x + \varepsilon \right] \xrightarrow[n \rightarrow \infty]{} \Phi(x + \varepsilon) - \Phi(x).$$

Dies gilt für jedes $\varepsilon > 0$. Daher können auch $\varepsilon \downarrow 0$ betrachten. Somit gilt wegen der Stetigkeit von Φ

$$\lim_{n \rightarrow \infty} \mathbb{P}[S_n = m_n] = 0.$$

Aus den obigen Überlegungen folgt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{M_n}{\sqrt{n}} > x \right] = 2\mathbb{P}[N \geq x] = \mathbb{P}[|N| > x].$$

Daraus folgt die Aussage des Satzes. □

14.4. Arcussinus–Gesetz

Eine weitere unerwartete Eigenschaft der Irrfahrt ist das sogenannte *Arcussinus–Gesetz*. Sei S_0, S_1, S_2, \dots eine Irrfahrt mit $p = 1/2$. Wir betrachten die Zufallsvariable T_{2n} , welche die Zeit angibt, die die Irrfahrt während des Zeitintervalls $[0, 2n]$ auf der positiven Halbachse verbringt. Das heißt,

$$T_{2n} = \sum_{j=0}^{2n-1} \mathbb{1}_{S_j > 0 \text{ oder } S_{j+1} > 0}.$$

Das Ereignis “ $S_j > 0$ oder $S_{j+1} > 0$ ” tritt genau dann ein, wenn die Irrfahrt die Zeit zwischen j und $j + 1$ auf der positiven Halbachse verbringt. Im nächsten Satz berechnen wir die Verteilung von T_{2n} .

SATZ 14.4.1. *Für die Zähldichte von T_{2n} gilt*

$$p_{2k, 2n} \stackrel{\text{def}}{=} \mathbb{P}[T_{2n} = 2k] = u_{2k} \cdot u_{2n-2k}, \quad k = 0, \dots, n.$$

BEWEIS. SCHRITT 0. Sei zuerst $k = n$. Es gilt

$$p_{2n, 2n} = \mathbb{P}[T_{2n} = 2n] = \mathbb{P}[S_1 \geq 0, \dots, S_{2n} \geq 0] = \mathbb{P}[S_1 \leq 0, \dots, S_{2n} \leq 0] = \mathbb{P}[M_n = 0],$$

wobei wir die Symmetrie der Irrfahrt ausgenutzt haben und $M_n = \max\{S_0, S_1, \dots, S_n\}$. Mit (14.3.1) erhalten wir

$$p_{2n, 2n} = 1 - \mathbb{P}[M_{2n} \geq 1] = 1 - 2\mathbb{P}[S_{2n} \geq 1] + \mathbb{P}[S_{2n} = 1] = \mathbb{P}[S_{2n} = 0] = u_{2n},$$

denn $\mathbb{P}[S_{2n} = 1] = 0$ und $2\mathbb{P}[S_{2n} \geq 1] = \mathbb{P}[|S_{2n}| \geq 1]$. Somit gilt die Formel $p_{2n, 2n} = u_{2n}u_0$. Analog erhält man im Fall $k = 0$

$$p_{0, 2n} = \mathbb{P}[T_{2n} = 0] = \mathbb{P}[S_1 \leq 0, \dots, S_{2n} \leq 0] = \mathbb{P}[M_n = 0] = u_{2n} = u_0u_{2n}.$$

SCHRITT 1. Sei nun $k = 1, \dots, n - 1$. Wir wollen die Wahrscheinlichkeit bestimmen, dass $T_{2n} = 2k$. Wir betrachten die erste Rückkehr zum Ursprung und bezeichnen den entsprechenden Zeitpunkt mit $2r$. Es gibt zwei Fälle:

FALL A. Die erste Rückkehr findet zum Zeitpunkt $2r$ statt, nachdem sich die Irrfahrt auf der positiven Halbachse aufgehalten hat. Die Wahrscheinlichkeit dieses Ereignisses ist $\frac{1}{2}f_{2r}$. Also hat die Irrfahrt bereits $2r$ Zeiteinheiten auf der positiven Halbachse verbracht und muss noch $2k - 2r$ Zeiteinheiten dort verbringen. Dabei stehen ihr insgesamt $2n - 2r$ Zeiteinheiten zu Verfügung. Es muss also $2r \leq 2k$ und $2n - 2r \geq 2k - 2r$ gelten. Wir erhalten somit den Term

$$\frac{1}{2} \sum_{r=1}^k f_{2r} p_{2k-2r, 2n-2r}.$$

FALL B. Die erste Rückkehr findet zum Zeitpunkt $2r$ statt, nachdem sich die Irrfahrt auf der negativen Halbachse aufgehalten hat. Die Wahrscheinlichkeit dieses Ereignisses ist $\frac{1}{2}f_{2r}$. Danach muss die Irrfahrt noch $2k$ Zeiteinheiten auf der nichtnegativen Halbachse verbringen.

Dafür stehen ihr $2n - 2r$ Zeiteinheiten zur Verfügung. Es muss also $2k \leq 2n - 2r$ gelten. Wir erhalten somit den Term

$$\frac{1}{2} \sum_{r=1}^{n-k} f_{2r} p_{2k, 2n-2r}.$$

Beide Fälle zusammengenommen ergeben die Identität

$$(14.4.1) \quad p_{2k, 2n} = \frac{1}{2} \sum_{r=1}^k f_{2r} p_{2k-2r, 2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} p_{2k, 2n-2r}.$$

Nun beenden wir den Beweis mittels Induktion.

Induktionsbasis: Für die Fälle $n = 1, k = 0$ und $n = 1, k = 1$ lässt sich die Aussage leicht durch Abzählen der Pfade überprüfen:

$$p_{0,2} = p_{2,2} = \frac{1}{2}, \quad u_0 = 1, \quad u_2 = \frac{1}{2}.$$

Induktionsannahme: Es gelte

$$(14.4.2) \quad p_{2k, 2m} = u_{2k} \cdot u_{2m-2k} \quad \text{für alle } m = 1, \dots, n-1 \text{ und } k = 1, \dots, m-1.$$

Nun ist zu zeigen, dass diese Formel auch für $m = n$ und beliebiges $k = 1, \dots, n-1$ gilt. Mit (14.4.1) und (14.4.2) erhalten wir

$$\begin{aligned} p_{2k, 2n} &= \frac{1}{2} \sum_{r=1}^k f_{2r} \cdot u_{2k-2r} \cdot u_{2n-2k} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} \cdot u_{2k} \cdot u_{2n-2k-2r} \\ &= \frac{1}{2} u_{2n-2k} \sum_{r=1}^k f_{2r} \cdot u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{n-k} f_{2r} \cdot u_{2n-2k-2r} \\ &= u_{2k} \cdot u_{2n-2k}. \end{aligned}$$

Dabei haben wir im letzten Schritt die Identitäten

$$\sum_{r=1}^k f_{2r} u_{2k-2r} = u_{2k} \quad \text{und} \quad \sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r} = u_{2n-2k},$$

siehe (14.2.3), benutzt. Somit gilt $p_{2k, 2n} = u_{2k} \cdot u_{2n-2k}$ und die Induktion ist beendet. \square

SATZ 14.4.2 (Arcussinus-Gesetz). Die Zufallsvariable $\frac{T_{2n}}{2n}$ konvergiert für $n \rightarrow \infty$ in Verteilung gegen eine Zufallsvariable T mit der Dichte

$$f_T(y) = \frac{1}{\pi \sqrt{y(1-y)}}, \quad y \in [0, 1].$$

BEMERKUNG 14.4.3. Hierbei ist $\frac{T_{2n}}{2n}$ ist der Anteil der Zeit, die die Irrfahrt auf der positiven Halbachse verbringt. Der Satz heißt Arcussinus-Gesetz (und die Zufallsvariable T heißt Arcussinus-verteilt), da die Verteilungsfunktion von T die folgende Gestalt hat:

$$F_T(y) = \frac{2}{\pi} \arcsin \sqrt{y}, \quad y \in [0, 1].$$

Die Dichte $f_T(y)$ erreicht ihr Minimum bei $t = 1/2$. Für $t \rightarrow 0$ und $t \rightarrow 1$ wird sie unendlich. Es ist demnach unwahrscheinlich, dass die Irrfahrt auf der positiven und negativen Halbachse ungefähr gleich viel Zeit verbringt. Viel wahrscheinlicher ist es, dass die Irrfahrt entweder fast die ganze Zeit auf der positiven, oder fast die ganze Zeit auf der negativen Halbachse verbringt.

BEWEIS VON SATZ 14.4.2. Hier soll nur eine Beweisidee vorgestellt werden. Wir haben in Satz 14.4.1 und Satz 14.2.8 gezeigt, dass

$$\mathbb{P}[T_{2n} = 2k] = p_{2k,2n} = u_{2k} \cdot u_{2n-2k}, \quad u_{2k} = \mathbb{P}[S_{2k} = 0] \sim \frac{1}{\sqrt{\pi k}}, \quad k \rightarrow \infty.$$

Es seien a und b zwei Zahlen mit $0 < a < b < 1$. Wir erhalten, dass

$$\begin{aligned} \mathbb{P}\left[a < \frac{T_{2n}}{2n} \leq b\right] &= \sum_{k=[na]+1}^{[nb]} u_{2k} \cdot u_{2n-2k} \\ &\sim \frac{1}{n} \sum_{k=[na]+1}^{[nb]} \frac{1}{\pi \sqrt{\frac{k}{n}(1-\frac{k}{n})}} \quad (\text{Riemann-Summe}) \\ &\xrightarrow{n \rightarrow \infty} \int_a^b \frac{dy}{\pi \sqrt{y(1-y)}}. \end{aligned}$$

Dabei ist die Begründung des Übergangs von der ersten zur zweiten Zeile eine Übung. \square

Außer Satz 14.4.2 gibt es mindestens zwei weitere Arcussinus-Gesetze, die wir ohne Beweis angeben. In den beiden Sätzen ist S_0, S_1, S_2, \dots eine Irrfahrt mit $p = 1/2$ und T ist eine Arcussinus-verteilte Zufallsvariable.

SATZ 14.4.4 (Arcussinus-Gesetz für die letzte Nullstelle). *Es sei $K_{2n} = \max\{k : S_k = 0, 0 \leq k \leq 2n\}$ die letzte Nullstelle der Irrfahrt im Zeitintervall $[0, 2n]$. Dann gilt*

$$\frac{K_{2n}}{2n} \xrightarrow[n \rightarrow \infty]{d} T.$$

SATZ 14.4.5 (Arcussinus-Gesetz für die Position des Maximums). *Es sei $R_{2n} = \max\{k : S_k = M_k, 0 \leq k \leq 2n\}$ die Position des letzten Maximums der Irrfahrt im Zeitintervall $[0, 2n]$. Dann gilt*

$$\frac{R_{2n}}{2n} \xrightarrow[n \rightarrow \infty]{d} T.$$

14.5. Gesetz vom iterierten Logarithmus

Seien X_1, X_2, \dots unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = 0$ und $\text{Var } X_k = 1$. Definiere

$$S_n = X_1 + \dots + X_n, \quad n \in \mathbb{N}.$$

Wir haben folgende Aussagen über die Wachstumsgeschwindigkeit der Folge S_1, S_2, \dots bewiesen.

1. Das starke Gesetz der großen Zahlen behauptet, dass

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

Die Folge S_n wächst also wesentlich langsamer als n .

2. Der zentrale Grenzwertsatz behauptet, dass

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N, \quad N \sim N(0, 1).$$

Der Wert S_n ist also mit großer Wahrscheinlichkeit ungefähr von der Größenordnung \sqrt{n} .

3. In Satz 14.3.1 haben wir gezeigt, dass

$$\frac{\max\{S_0, S_1, \dots, S_n\}}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} |N|, \quad N \sim N(0, 1).$$

Wir haben nur den Fall $\mathbb{P}[X_k = \pm 1] = 1/2$ betrachtet, die Aussage gilt aber allgemein. Das Maximum von S_0, S_1, \dots, S_n ist also mit großer Wahrscheinlichkeit von der Größenordnung \sqrt{n} .

Welche ist nun die richtige Geschwindigkeit, mit der die Folge S_1, S_2, S_3, \dots wächst? Diese Frage beantwortet der folgende Satz.

SATZ 14.5.1 (Gesetz vom iterierten Logarithmus). *Es seien X_1, X_2, \dots unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = 0$ und $\text{Var } X_k = 1$. Definiere $S_n = X_1 + \dots + X_n$, $n \in \mathbb{N}$. Dann gilt*

$$\mathbb{P} \left[\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log(\log n)}} = 1 \right] = \mathbb{P} \left[\liminf_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log(\log n)}} = -1 \right] = 1.$$

OHNE BEWEIS.

BEISPIEL 14.5.2. Aus dem Gesetz vom iterierten Logarithmus folgt, dass für alle $\varepsilon > 0$ mit Wahrscheinlichkeit 1 gilt:

$$\lim_{n \rightarrow \infty} \frac{S_n}{n^{\frac{1}{2} + \varepsilon}} = 0 \quad \text{jedoch} \quad \limsup_{n \rightarrow \infty} \frac{S_n}{n^{\frac{1}{2} - \varepsilon}} = +\infty.$$

Die Folge S_n wächst also langsamer als $n^{\frac{1}{2} + \varepsilon}$, aber nicht langsamer als $n^{\frac{1}{2} - \varepsilon}$.