

WESTFÄLISCHE  
WILHELMS-UNIVERSITÄT  
MÜNSTER

---

ANSCHAULICHE  
WAHRSCHEINLICHKEITSTHEORIE

---

SKRIPT  
PROF. DR. ZAKHAR KABLUCHKO

WWU MÜNSTER  
INSTITUT FÜR MATHEMATISCHE STOCHASTIK

## Inhaltsverzeichnis

Vorwort	1
Literatur	1
Kapitel 1. Grundbegriffe der Wahrscheinlichkeitstheorie	2
1.1. Ausgänge und deren Wahrscheinlichkeiten	2
1.2. Produktexperimente	5
1.3. Ereignisse und deren Wahrscheinlichkeiten	7
1.4. Verknüpfungen von Ereignissen	8
1.5. Was sind Wahrscheinlichkeiten?	13
1.6. Zufallsexperimente mit unendlicher Grundmenge	16
1.7. Zufallsvariablen	18
1.8. Erwartungswert	22
1.9. Unabhängigkeit???	25
Kapitel 2. Kombinatorik und diskrete Verteilungen	26
2.1. Geburtstagsproblem	26
2.2. Vier Urnenmodelle	28
2.3. Hypergeometrische Verteilung	35
2.4. Binomialverteilung und Multinomialverteilung	37
2.5. Uniformverteilung	40
2.6. Bernoulli-Experimente und die Binomialverteilung	40
2.7. Poisson-Verteilung	43
2.8. Geometrische Verteilung	48
2.9. Negative Binomialverteilung	51
Kapitel 3. Geometrische Wahrscheinlichkeiten	54
3.1. Uniforme Verteilung	54
3.2. Das Buffon'sche Nadelproblem	57
3.3. Das Bertrand'sche Paradoxon	61
Kapitel 4. Axiome der Wahrscheinlichkeitstheorie und Maßtheorie	64
4.1. Axiome der Wahrscheinlichkeitstheorie	64
4.2. Existenz nichtmessbarer Mengen	66
4.3. Vorüberlegungen zur Maßtheorie	67
4.4. Algebren	68
4.5. $\sigma$ -Algebren	70
4.6. Die von einem Mengensystem erzeugte $\sigma$ -Algebra	71
4.7. Borel- $\sigma$ -Algebra	72
4.8. Maße	73

4.9.	Wahrscheinlichkeitsmaße	74
4.10.	Eigenschaften der Wahrscheinlichkeit	75
4.11.	Die Siebformel	78
4.12.	Eindeutigkeit der Maßfortsetzung	82
4.13.	Limes superior und Limes inferior für Folgen von Mengen	87
4.14.	Das Lemma von Borel–Cantelli	89
Kapitel 5.	Unabhängigkeit und bedingte Wahrscheinlichkeiten	93
5.1.	Unabhängigkeit von zwei Ereignissen	93
5.2.	Unabhängigkeit von mehreren Ereignissen	95
5.3.	Unabhängigkeit von Familien von Ereignissen und der Blockungssatz	97
5.4.	Bedingte Wahrscheinlichkeiten	99
5.5.	Unabhängigkeit von Zufallsvariablen	103
5.6.	Produktexperimente	105
Kapitel 6.	Zufallsvariablen: Die allgemeine Definition	107
6.1.	Zufallsvariablen	107
6.2.	Zufallsvektoren	109
6.3.	Verteilungsfunktion einer Zufallsvariable	110
6.4.	Definition und Eigenschaften des Erwartungswerts	116
6.5.	Diskrete und absolut stetige Verteilungen	120
6.6.	Beispiele von absolut stetigen Verteilungen	122
6.7.	Singuläre Verteilungen	131
6.8.	Zerlegungssatz von Lebesgue	134
6.9.	Verteilungsfunktion eines Zufallsvektors	136
6.10.	Diskrete und absolut stetige Zufallsvektoren	137
6.11.	Randverteilungen eines Zufallsvektors	138
6.12.	Unabhängigkeit und Produktformeln	140
6.13.	Transformationsformel für die Dichte	141
6.14.	Faltungsformeln	142
6.15.	Transformationsformel für den Erwartungswert	147
6.16.	Multiplikatивität des Erwartungswerts	149
Kapitel 7.	Varianz und Kovarianz	153
7.1.	Varianz	153
7.2.	Kovarianz und Korrelationskoeffizient	157
Kapitel 8.	Gesetz der großen Zahlen	163
8.1.	Zwei Beispiele	163
8.2.	Konvergenz in Wahrscheinlichkeit und $L^2$ -Konvergenz	164
8.3.	Ungleichungen von Markow und Tschebyschew	165
8.4.	Schwaches Gesetz der großen Zahlen	166
8.5.	Fast sichere Konvergenz	167
8.6.	Starkes Gesetz der großen Zahlen: Erste Version	170
8.7.	Starkes Gesetz der großen Zahlen: Zweite Version	173
8.8.	Der Fall eines unendlichen Erwartungswerts	179
8.9.	Anwendungen des Gesetzes der großen Zahlen	180

Kapitel 9. Ungleichungen	187
9.1. Jensen-Ungleichung	187
9.2. Ljapunow-Ungleichung	188
9.3. Young-Ungleichung	189
9.4. Hölder-Ungleichung	189
9.5. Minkowski-Ungleichung	190
9.6. $L^p$ -Räume und $L^p$ -Konvergenz	191
Kapitel 10. Analytische Methoden	194
10.1. Erzeugende Funktion	194
10.2. Summen mit einer zufälligen Anzahl von Summanden	198
10.3. Verzweigungsprozesse	199
10.4. Momenterzeugende Funktion	202
10.5. Charakteristische Funktion (Fourier-Transformierte)	204
Kapitel 11. Der zentrale Grenzwertsatz	214
11.1. Einführung	214
11.2. Konvergenz in Verteilung	216
11.3. Eine Charakterisierung der Konvergenz in Verteilung	220
11.4. Straffheit und der Satz von Helly	223
11.5. Stetigkeitssatz von Lévy	229
11.6. Der zentrale Grenzwertsatz	231
11.7. Beweis des zentralen Grenzwertsatzes	235
11.8. Sätze von Lindeberg und Ljapunow	236
Kapitel 12. Irrfahrt	244
12.1. Berechnung einer Ruinwahrscheinlichkeit	244
12.2. Rückkehr der Irrfahrt zum Ursprung	246
12.3. Verteilung des Maximums der Irrfahrt	251
12.4. Arcussinus-Gesetz	252
12.5. Gesetz vom iterierten Logarithmus	255



## Vorwort

Dieses Skript ist aus den Stochastik-Vorlesungen entstanden, die ich an den Universitäten Ulm und Münster gehalten habe. Für die Hilfe bei der Erstellung einiger Abschnitte der früheren L<sup>A</sup>T<sub>E</sub>X-Versionen dieses Skriptes bedanke ich mich bei Uli Armbruster, Linda Bolay, Melanie Herz und Anita Kollwitz.

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.
7. W. Linde. *Stochastik für das Lehramt*. De Gruyter.
8. M. Löwe, H. Knöpfel. *Stochastik: Struktur im Zufall*. De Gruyter.

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. Ausgänge und deren Wahrscheinlichkeiten

In der Stochastik betrachten wir Zufallsexperimente, d.h. Experimente mit einem unsicheren, zufälligen Verlauf. Alle möglichen Ausgänge eines Zufallsexperiments fassen wir zu einer Menge zusammen. Diese Menge bezeichnen wir mit  $\Omega$  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:

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

Ist die Münze **fair** (oder **symmetrisch**), so sind diese Ausgänge gleichwahrscheinlich, d.h. es gilt

$$p(K) = p(Z) = \frac{1}{2}.$$

Dabei schreiben wir  $p(\omega)$  für die Wahrscheinlichkeit des Ausgangs  $\omega$ .

**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

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

Bei einem fairen (oder symmetrischen) Würfel sind alle Ausgänge gleichwahrscheinlich, d.h. es gilt

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$

Bei einem unsymmetrischen Würfel könnten die Wahrscheinlichkeiten  $p(1), \dots, p(6)$  beliebige Werte annehmen, es sollten aber stets die beiden Bedingungen erfüllt sein:

- Die Wahrscheinlichkeiten sind nicht-negativ, d.h.  $p(i) \geq 0$  für alle  $i$ .
- Die Wahrscheinlichkeiten summieren sich zu 1, d.h.  $p(1) + \dots + p(6) = 1$ .

**Beispiel 1.1.3.** Nun werfen wir eine Münze und einen Würfel gleichzeitig. Es gibt 2 Möglichkeiten für die Münze und 6 Möglichkeiten für den Würfel. Bei der Aufstellung der Grundmenge gehen wir systematisch vor indem wir zuerst alle Möglichkeiten auflisten, bei denen die Münze “Kopf” zeigt, und dann alle Möglichkeiten, bei denen die Münze “Zahl” zeigt. Insgesamt erhalten wir die folgende Grundmenge:

$$\Omega = \{K1, K2, K3, K4, K5, K6, Z1, Z2, Z3, Z4, Z5, Z6\}.$$

Die Anzahl der möglichen Ausgänge in diesem Experiment ist  $2 \cdot 6 = 12$  (und nicht  $2 + 6 = 8$ ), da wir jede Möglichkeit für die Münze mit jeder Möglichkeit für den Würfel kombinieren können. Experimente von diesem Typ (z.B. solche, bei denen mehrere Gegenstände unabhängig voneinander geworfen werden) werden Produktexperimente genannt. Sind beide Gegenstände symmetrisch, so sind alle 12 Ausgänge gleichwahrscheinlich. Da sich Wahrscheinlichkeiten aller Ausgänge zu 1 summieren müssen, erhalten wir

$$p(\omega) = \frac{1}{12} \quad \text{für alle } \omega \in \Omega.$$

Die **Grundmenge** eines Experiments ist eine Menge, deren Elemente alle möglichen Ausgänge des Experiments sind. Die Grundmenge wird stets mit  $\Omega = \{\omega_1, \omega_2, \dots\}$  bezeichnet, während die einzelnen Ausgänge mit  $\omega_1, \omega_2, \dots$  bezeichnet werden.

Wir bezeichnen mit  $\#\Omega$  die Anzahl der Elemente in  $\Omega$ , also die Anzahl der Ausgänge.

**Beispiel 1.1.4** (Unterscheidbarkeit der Gegenstände). Nun betrachten wir ein Experiment, bei dem **zwei** Münzen gleichzeitig geworfen werden. Die Grundmenge besteht in diesem Fall aus 4 Elementen:

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

und es gilt  $\#\Omega = 4$ . Man beachte, dass  $(K, Z)$  und  $(Z, K)$  verschiedene Ausgänge sind! Das kann man wie folgt begründen. Auch wenn die beiden Münzen am Anfang völlig identisch aussehen mögen, kann man sie immer färben und dadurch **unterscheidbar** machen. Durch das Färben der Münzen ändert sich am Verlauf des Experiments (z.B. an den komplizierten Differentialgleichungen, die die Bewegung der Münzen beschreiben) offenbar nichts. Stellen wir uns also vor, dass wir eine schwarze und eine weiße Münze werfen. Dann können die beiden Ausgänge  $(K, Z)$  und  $(Z, K)$  wie folgt beschrieben werden:

$(K, Z)$  : die schwarze Münze zeigt “Kopf” und die weiße Münze zeigt “Zahl”;

$(Z, K)$  : die schwarze Münze zeigt “Zahl” und die weiße Münze zeigt “Kopf”.

Nun ist es klar, dass es sich um zwei unterschiedliche Ausgänge handelt.

**Beispiel 1.1.5** (Wahl der Grundmenge). Bei der Modellierung der Zufallsexperimente sollte man stets darauf achten, dass die Definition der Ausgänge eine **möglichst vollständige Beschreibung** dessen beinhalten muss, was im Experiment beobachtet wird. Um dies zu verdeutlichen, betrachten wir ein Experiment, bei dem **drei** Münzen geworfen werden. Man könnte nun versuchen, eine Grundmenge  $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\}$  zu wählen, die aus den folgenden vier Ausgängen besteht:

$\omega_0$  := “keine Münze zeigt Kopf”,

$\omega_1$  := “eine Münze zeigt Kopf”,

$\omega_2$  := “zwei Münzen zeigen Kopf”,

$\omega_3$  := “alle drei Münzen zeigen Kopf”.



Diese Wahl ist allerdings schlecht und sollte vermieden werden, denn sie beschreibt das beobachtete Ergebnis nicht vollständig: bei  $\omega_1$  und  $\omega_2$  gibt man zwar an, *wieviele* Münzen “Kopf” zeigen, nicht aber um *welche* Münzen es sich handelt. Dadurch wird ein Teil der Information über das Ergebnis des Experiments verheimlicht. In der richtigen Modellierung dieses Experiments sollte man die **ganze** Wahrheit sagen indem man vollständig beschreibt, welche Münzen welche Symbole gezeigt haben. Die richtige Grundmenge besteht aus 8 Elementen:

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

Es gilt also  $\#\Omega = 8$ . Sind die Münzen fair, so haben alle 8 Ausgänge die gleiche Wahrscheinlichkeit  $1/8$ . Würde aber jemand dennoch hartnäckig auf der Modellierung mit vier Ausgängen  $\omega_0, \omega_1, \omega_2, \omega_3$  bestehen, so müsste er zugeben, dass in seiner Modellierung die “Ausgänge” (und zwar auch bei fairen Münzen) nicht gleichwahrscheinlich sind! Aus Erfahrung weiß man nämlich, dass die “Ausgänge”  $\omega_0$  und  $\omega_3$  viel seltener beobachtet werden, als die “Ausgänge”  $\omega_1$  und  $\omega_2$ . Um dies zu erklären, bemerken wir, dass man in der richtigen Modellierung des Experiments  $\omega_1$  in drei Ausgänge  $(K, Z, Z), (Z, K, Z), (Z, Z, K)$  und  $\omega_2$  ebenfalls in drei Ausgänge  $(K, K, Z), (K, Z, K), (Z, K, K)$  aufsplitten muss, während  $\omega_0$  und  $\omega_3$  den Ausgängen  $(Z, Z, Z)$  und  $(K, K, K)$  entsprechen. Für die Wahrscheinlichkeiten im Modell mit 4 Ausgängen gilt also

$$p(\omega_0) = p(\omega_3) = \frac{1}{8}, \quad p(\omega_1) = p(\omega_2) = \frac{3}{8}.$$

Im Prinzip ist also auch eine Modellierung mit 4 Ausgängen nicht falsch, vorausgesetzt, dass man die Wahrscheinlichkeiten der Ausgänge korrekt identifiziert.

**Beispiel 1.1.6.** In einer Familie gibt es zwei Kinder. Wie groß ist die Wahrscheinlichkeit, dass die Kinder das gleiche Geschlecht haben? Dabei gehen wir davon aus, dass Jungen und Mädchen mit gleicher Wahrscheinlichkeit  $1/2$  geboren werden.<sup>1</sup>

**Erste falsche Lösung.** Es gibt zwei mögliche Ausgänge:

$$\begin{aligned} \omega_1 &= \text{“die Kinder haben das gleiche Geschlecht”}, \\ \omega_2 &= \text{“die Kinder haben verschiedene Geschlechter”}. \end{aligned}$$

Also gilt  $p(\omega_1) = p(\omega_2) = 1/2$ .

Diese Lösung ist falsch, denn sie erklärt nicht, **warum** die beiden Ausgänge die gleiche Wahrscheinlichkeit haben. Würde man der gleichen Logik folgen, so müsste man zugeben, dass die Wahrscheinlichkeit, auf dem Nachhauseweg einem sprechenden Elephanten zu begegnen, ebenfalls  $1/2$  ist, weil es in diesem Experiment 2 Ausgänge gibt: entweder begegnet man einem sprechenden Elephanten oder nicht. Das ist aber absurd.

**Zweite falsche Lösung.** Es gibt drei mögliche Ausgänge:

$$\begin{aligned} \omega_1 &= \text{“beide Kinder sind Jungen”}, \\ \omega_2 &= \text{“beide Kinder sind Mädchen”}, \\ \omega_3 &= \text{“Kinder haben verschiedene Geschlechter”}. \end{aligned}$$

---

<sup>1</sup>Was eine Vereinfachung ist, denn in Wirklichkeit ist die Wahrscheinlichkeit eines Jungen etwas höher und liegt approximativ bei 51%.

Hieraus ergibt sich die Grundmenge  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ . Geht man davon aus, dass alle drei Ausgänge die gleiche Wahrscheinlichkeit haben, so ergibt sich das Ergebnis  $1/3$ . Diese Lösung ist allerdings falsch, denn sie behandelt Kinder als ununterscheidbare Objekte.

**Richtige Lösung.** Die Kinder sind unterscheidbar (man kann vom “ersten Kind” und “zweiten Kind” sprechen). Es gibt folgende Ausgänge:

- $JJ$  : “beide Kinder sind Jungen”,
- $MM$  : “beide Kinder sind Mädchen”,
- $JM$  : “erstes Kind ist Junge und zweites Kind ist Mädchen”,
- $MJ$  : “erstes Kind ist Mädchen und zweites Kind ist Junge”.

Man beachte, dass  $MJ$  und  $JM$  zwei verschiedene Ausgänge sind, genauso wie in Beispiel 1.1.4. Die Grundmenge ist

$$\Omega = \{JJ, MM, JM, MJ\}$$

und besteht aus 4 Ausgängen. Jeder Ausgang hat die gleiche Wahrscheinlichkeit  $1/4$ , d.h.

$$p(JJ) = p(MM) = p(JM) = p(MJ) = \frac{1}{4}.$$

Wir interessieren uns für das Ereignis “Kinder haben verschiedene Geschlechter”, das aus den Ausgängen  $JM$  und  $MJ$  besteht. Für die gesuchte Wahrscheinlichkeit ergibt sich also

$$\mathbb{P}[\text{Kinder haben verschiedene Geschlechter}] = p(JM) + p(MJ) = \frac{1}{2}. \quad (1.1.1)$$

Das Ergebnis  $1/2$  ist also doch richtig, jedoch ist die Begründung, die in der ersten Lösung präsentiert wurde, falsch.  $\square$

## 1.2. Produktexperimente

Wir haben bereits einige Experimente gesehen, die aus mehreren unabhängigen Telexperimenten bestehen, in welchem Fall wir von einem **Produktexperiment** sprechen.

**Beispiel 1.2.1.** Wir betrachten ein Experiment, bei dem  $n$  Münzen geworfen werden. Wir erinnern den Leser daran, dass die Münzen immer unterscheidbar sind (also mit  $1, 2, \dots, n$  nummeriert werden können) und dass man bei der Aufstellung der Grundmenge darauf achten muss, dass das beobachtete Ergebnis vollständig beschrieben wird. Eine solche vollständige Beschreibung erhält man, indem man für jede Münze das von dieser Münze gezeigte Symbol angibt. Ein Ausgang dieses Experiments ist also eine aus den Buchstaben  $K$  und  $Z$  bestehende Folge der Länge  $n$ . Die Grundmenge besteht aus allen solchen Folgen und wird wie folgt bezeichnet:

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

Dabei steht  $a_k$  für das Symbol, das die  $k$ -te Münze gezeigt hat. Für  $n = 5$  Münzen ist z.B.  $(K, Z, K, K, Z)$  ein möglicher Ausgang des Experiments.

Wieviele Elemente hat  $\Omega$ ? Da es für jede einzelne Münze zwei Möglichkeiten  $K$  und  $Z$  gibt, und da die Möglichkeiten für verschiedene Münzen beliebig kombinierbar sind (es gibt keine

verbotenen Kombinationen), ergibt sich für die Gesamtzahl der Ausgänge die Formel

$$\#\Omega = \underbrace{2 \cdot \dots \cdot 2}_{n \text{ Mal}} = 2^n.$$

In diesem und in ähnlichen Experimenten ist es egal, ob man die Münzen gleichzeitig oder nacheinander wirft. Man kann auch eine Münze nehmen und diese  $n$  Mal werfen. Die Grundmenge ändert sich dadurch nicht.

**Beispiel 1.2.2.** Das Werfen von  $n$  Würfeln ist ebenfalls ein Produktexperiment mit der Grundmenge

$$\Omega = \{1, \dots, 6\}^n \stackrel{\text{def}}{=} \{(a_1, \dots, a_n) : a_1, \dots, a_n \in \{1, \dots, 6\}\},$$

die aus  $6^n$  Ausgängen besteht.

**Beispiel 1.2.3.** Es seien drei Gegenstände gegeben: eine Münze, ein Würfel und ein Tetraeder. Für das Werfen der Münze ist die Grundmenge  $E_1 = \{K, Z\}$ , für das Würfeln  $E_2 = \{1, \dots, 6\}$ , und für das Werfen des Tetraeders  $E_3 = \{1, 2, 3, 4\}$ . Nun betrachten wir ein Gesamtexperiment, bei dem **alle drei Gegenstände** (egal ob gleichzeitig oder nacheinander) geworfen werden. Ein Ausgang des Gesamtexperiments ist ein Tupel der Form  $(a_1, a_2, a_3)$ , wobei die Komponente  $a_1 \in \{K, Z\}$  angibt, welches Symbol die Münze gezeigt hat, die Komponente  $a_2 \in \{1, \dots, 6\}$  die Seite des Würfels und  $a_3 \in \{1, 2, 3, 4\}$  die Seite des Tetraeders. Da man die Seiten der verschiedenen Gegenstände beliebig kombinieren kann, besteht die Grundmenge aus

$$\#\Omega = 2 \cdot 6 \cdot 4 = 48$$

Elementen, die wir hier nicht auflisten werden.

Bei einem **Produktexperiment** werden  $n$  Teilerperimente mit Grundmengen  $E_1, \dots, E_n$  *unabhängig* voneinander ausgeführt. Die Ausgänge des Produktexperiments sind Tupel der Form  $(e_1, \dots, e_n)$ , wobei die Komponente  $e_k$  den Ausgang des  $k$ -ten Teilerperiments angibt und somit ein Element von  $E_k$  sein muss. Die Grundmenge des Produktexperiments ist 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 des Gesamtexperiments wird nach der Produktregel berechnet:

$$\#\Omega = (\#E_1) \cdot \dots \cdot (\#E_n)$$

Die Unabhängigkeit der Teilerperimente bedeutet, dass sie einander nicht beeinflussen dürfen. Die nächsten zwei Beispiele sollen verdeutlichen, wie eine mögliche Beeinflussung aussehen könnte.

**Beispiel 1.2.4.** Eine Person besitze 5 verschiedene Hosen, 4 Hemden und 3 Hüte. Insgesamt gibt es  $5 \cdot 4 \cdot 3 = 60$  Anziehungsmöglichkeiten, wobei man aber auch solche Kombinationen zulässt, bei denen die Person komisch aussieht. Würde man Bedingungen aufstellen, die das komische Aussehen verhindern (etwa, dass die Hose und der Hut die gleiche Farbe haben

sollen), so würde man die verschiedenen Teilerperimente abhängig machen, was zu einer Verkleinerung der Anzahl der Anziehungsmöglichkeiten führen würde.

**Beispiel 1.2.5.** Wir ziehen aus einem Kartenspiel mit 52 Karten zwei Karten gleichzeitig. Für jede der beiden Karten hat man zwar 52 Möglichkeiten, diese sind allerdings nicht beliebig kombinierbar. Ist z.B. die erste Karte ein Kreuzass, so kann die zweite Karte unmöglich auch ein Kreuzass sein. Es handelt sich also nicht um ein Produktexperiment und die Anzahl der Ausgänge ist nicht  $52^2$ .

### 1.3. Ereignisse und deren Wahrscheinlichkeiten

Vereinfacht gesprochen ist ein Ereignis eine Aussage über den Ausgang des Experiments, die entweder wahr oder falsch sein kann. Nachdem das Experiment ausgeführt wird, kann man überprüfen, ob diese Aussage zutrifft oder nicht. Im ersten Fall sagen wir, dass das Ereignis eingetreten ist.

**Beispiel 1.3.1.** Wir werfen einen Würfel und betrachten beispielhaft das folgende Ereignis  $A$ : “Es wird eine gerade Zahl gewürfelt”. Die Grundmenge ist  $\Omega = \{1, \dots, 6\}$ . Das Ereignis  $A$  tritt genau dann ein, wenn eine der Zahlen 2, 4, 6 gewürfelt wird. Wir können also  $A$  mit der entsprechenden Teilmenge von  $\Omega$  identifizieren:

$$A = \text{“eine gerade Zahl wird gewürfelt”} = \{2, 4, 6\}.$$

Bei einem fairen Würfel hat jeder Ausgang die gleiche Wahrscheinlichkeit  $1/6$ . Da sich das Ereignis  $A$  aus 3 Ausgängen zusammensetzt, erhalten wir für die Wahrscheinlichkeit von  $A$

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

Wir können jeder Aussage über das beobachtete Ergebnis des Experiments die Teilmenge der Ausgänge zuordnen, bei denen diese Aussage zutrifft. Wir kommen also zur folgenden Definition:

**Definition 1.3.2.** Ein **Ereignis** ist eine Teilmenge der Grundmenge  $\Omega$ .

Fürs Erste werden wir die Wahrscheinlichkeit eines Ereignisses als die Summe der Wahrscheinlichkeiten der darin enthaltenen Ausgänge definieren:

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

Dabei benutzen wir die Notation  $\mathbb{P}[A]$  für die Wahrscheinlichkeit eines Ereignisses  $A \subset \Omega$ , während  $p(\omega)$  die Wahrscheinlichkeit des Ausgangs  $\omega \in \Omega$  bezeichnet.

**Beispiel 1.3.3.** Nun würfeln wir zweimal. Die Grundmenge ist  $\Omega = \{1, \dots, 6\}^2$  mit  $\#\Omega = 36$ . Wir betrachten das Ereignis

$$B := \text{“die Summe der Augenzahlen ist 10”}.$$

Dieses Ereignis kann durch 3 Kombinationen realisiert werden, nämlich

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

Hier ist zu beachten, dass es sich bei  $(6, 4)$  und  $(4, 6)$  um verschiedene Ausgänge handelt, denn die Würfel sind unterscheidbar. Da in diesem Experiment alle Ausgänge die gleiche Wahrscheinlichkeit  $1/36$  haben und da sich  $B$  aus 3 Ausgängen zusammensetzt, gilt für die Wahrscheinlichkeit von  $B$ , dass

$$\mathbb{P}[B] = 3 \cdot \frac{1}{36} = \frac{1}{12}.$$

Man beachte außerdem, dass “die Augensumme” an sich kein Ereignis ist. Vielmehr ist eine Aussage über mögliche Werte der Augensumme ein Ereignis, etwa “die Augensumme ist mindestens 7” oder “die Augensumme ist ungerade”.

**Beispiel 1.3.4.** Die folgenden “trivialen” Ereignisse sind erwähnenswert:

- unmögliches Ereignis, welches nie eintritt,  $A = \emptyset$ . Es gilt  $\mathbb{P}[\emptyset] = 0$ .
- sicheres Ereignis, welches immer eintritt,  $A = \Omega$ . Es gilt  $\mathbb{P}[\Omega] = 1$ .

**Beispiel 1.3.5.** Ein *Elementarereignis* ist ein aus nur einem Ausgang bestehendes Ereignis, also  $A = \{\omega\}$  mit  $\omega \in \Omega$ . Jedes Ereignis setzt sich somit aus Elementarereignissen zusammen.

**Aufgabe 1.3.6.** Wieviele Ereignisse gibt es insgesamt bei einem Experiment mit  $n$  Ausgängen?

## 1.4. Verknüpfungen von Ereignissen

Ein Ereignis ist eine Aussage über den Ausgang des Experiments. Jede Aussage kann verneint werden, was zum Begriff des Gegenereignisses führt.

Das **Komplement** oder das **Gegenereignis** von  $A$  besteht aus allen Ausgängen, die nicht in  $A$  enthalten sind:

$$A^c = \{\omega \in \Omega : \omega \notin A\} = \text{“}A \text{ tritt nicht ein”}.$$

**Beispiel 1.4.1.** Beim werfen eines Würfels ist das Komplement des Ereignisses

$$A := \text{“die Augenzahl ist gerade”} = \{2, 4, 6\}$$

das Ereignis

$$A^c = \text{“die Augenzahl ist ungerade”} = \{1, 3, 5\}.$$

**Beispiel 1.4.2.** Wenn ein Experte der Meinung ist, dass im kommenden Jahr mit Wahrscheinlichkeit 70% eine Wirtschaftskrise ausbricht, dann muss er auch der Meinung sein, dass die Wahrscheinlichkeit für das Nichtausbrechen der besagten Wirtschaftskrise bei 30% liegt. Diese Überlegung kann man wie folgt verallgemeinern.

Für die Wahrscheinlichkeit des Komplements gilt

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

Bei einem Laplace-Experiment (???????) kann die Wahrscheinlichkeit jedes Ereignisses durch Abzählen der Ausgänge bestimmt werden. Allerdings kann das bei einer großen Grundmenge aufwendig werden. Ein Trick, der beim Abzählen der Ausgänge oft hilft, ist der Übergang zum Gegenereignis.

**Beispiel 1.4.3** (Gegenereignis betrachten). Es werden 10 faire Münzen geworfen. Man bestimme die Wahrscheinlichkeit des Ereignisses

$$A := \text{“mindestens eine Münze zeigt Kopf”}.$$

**Lösung.** Die Grundmenge ist hier  $\Omega = \{K, Z\}^{10}$  mit  $\#\Omega = 2^{10}$ . Das Ereignis  $A$  besteht aus sehr vielen Ausgängen, die man nicht so einfach abzählen kann. Allerdings ist das Komplement des Ereignisses  $A$  ganz einfach:

$$A^c = \text{“keine Münze zeigt Kopf”} = \text{“alle Münzen zeigen Zahl”}.$$

Somit besteht  $A^c$  aus einem einzigen Ausgang  $(Z, Z, \dots, Z)$  und es gilt  $\#(A^c) = 1$ . Daraus ergibt sich die Anzahl der Ausgänge in  $A$ :

$$\#A = \#\Omega - \#(A^c) = 1023.$$

Also gilt für die Wahrscheinlichkeit von  $A$ :

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega} = \frac{1023}{1024}. \quad \square$$

**Euler-Venn-Diagramme.** Im Folgenden wird es sehr hilfreich sein, sich die Grundmenge  $\Omega$  als ein Rechteck und die Ereignisse als Gebiete (z.B. Kreise) in diesem Rechteck vorzustellen. So kann man sich z.B. ein Ereignis und sein Komplement vorstellen:

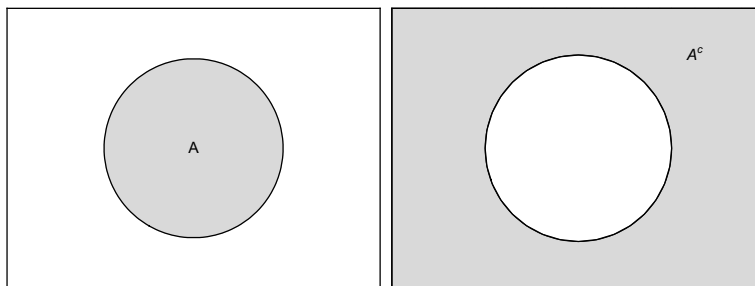


ABBILDUNG 1. Ereignis und sein Komplement

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

Die **Vereinigung** von  $A$  und  $B$  ist das Ereignis, das aus allen Ausgängen besteht, die in mindestens einem der beiden Ereignisse enthalten sind:

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\} = \text{“}A \text{ tritt ein oder } B \text{ tritt ein”}.$$

Der **Schnitt** von  $A$  und  $B$  ist das Ereignis, das aus allen Ausgängen besteht, die sowohl in  $A$  als auch in  $B$  enthalten sind:

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ und } \omega \in B\} = \text{“}A \text{ tritt ein und } B \text{ tritt ein“}.$$

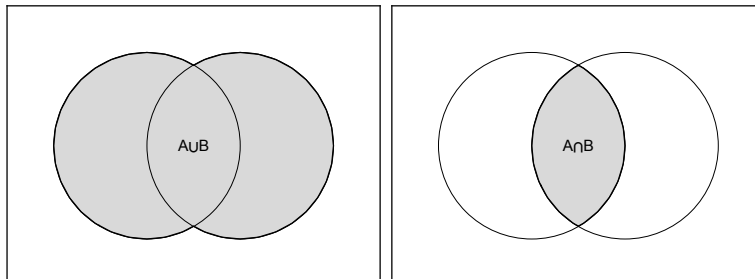


ABBILDUNG 2. Vereinigung und Schnitt von zwei Ereignissen.

**Beispiel 1.4.4.** Wir werfen zwei Würfel. Die Grundmenge ist  $\Omega = \{1, \dots, 6\}^2$  und besteht aus 36 Elementen. Betrachte folgende Ereignisse:

$$A := \text{“erster Würfel zeigt eine 6”} = \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\},$$

$$B := \text{“zweiter Würfel zeigt eine 6”} = \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6)\}.$$

Die beiden Ereignisse haben nur ein gemeinsames Element  $(6, 6)$ , deshalb gilt für den Schnitt dieser Ereignisse

$$A \cap B = \text{“beide Würfel zeigen 6”} = \{(6, 6)\}.$$

Die Vereinigung der beiden Ereignisse ist

$$A \cup B = \text{“mindestens einmal eine 6 gewürfelt”}$$

$$= \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6), (1, 6), (2, 6), (3, 6), (4, 6), (5, 6)\}.$$

Man beachte, dass der Ausgang  $(6, 6)$ , der in den beiden Ereignissen liegt, nur einmal aufgelistet wird. Es gilt  $\#(A \cup B) = 11$ .

Die **Mengendifferenz**  $A \setminus B$  besteht aus allen Ausgängen, die in  $A$  jedoch nicht in  $B$  enthalten sind:

$$A \setminus B = A \cap B^c = \text{“}A \text{ tritt ein, aber } B \text{ tritt nicht ein“}.$$

Analog kann man auch die Mengendifferenz  $B \setminus A$  definieren:

$$B \setminus A = B \cap A^c = \text{“}B \text{ tritt ein, aber } A \text{ tritt nicht ein“}.$$

Man beachte, dass die Mengendifferenz nicht kommutativ ist:  $A \setminus B \neq B \setminus A$ .

Die **symmetrische Differenz** der Ereignisse  $A$  und  $B$  besteht aus allen Ausgängen, die in **genau einem** der beiden Ereignisse enthalten sind:

$$A \triangle B = (A \setminus B) \cup (B \setminus A) = \text{“}A \text{ tritt ein oder } B \text{ tritt ein, aber nicht beide gleichzeitig“}.$$

Man kann die symmetrische Differenz auch als  $A\Delta B = (A \cup B) \setminus (A \cap B)$  darstellen.

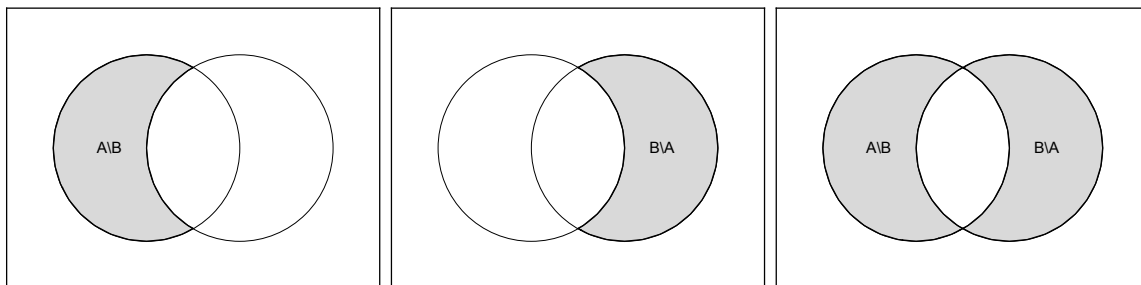


ABBILDUNG 3.  $A \setminus B$ ,  $B \setminus A$ ,  $A \Delta B$

**Beispiel 1.4.5.** Wir betrachten weiterhin die in Beispiel 1.4.4 eingeführten Ereignisse. Die symmetrische Differenz von  $A$  und  $B$  ist die Vereinigung, aus der das gemeinsame Element  $(6, 6)$  entfernt werden muss:

$$\begin{aligned} A\Delta B &= \text{“genau einmal eine 6 gewürfelt”} \\ &= \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (1, 6), (2, 6), (3, 6), (4, 6), (5, 6)\}. \end{aligned}$$

Es gilt also  $\#(A\Delta B) = 10$ .

**Aufgabe 1.4.6.** Seien  $A$ ,  $B$ ,  $C$  und  $D$  Ereignisse in einer Grundmenge  $\Omega$ . Wie lauten die folgenden Ereignisse in Mengenschreibweise?

- Unter den vier Ereignissen tritt nur das Ereignis  $C$  ein.
- Es treten genau zwei der vier Ereignisse ein.
- Es treten mindestens drei der vier Ereignisse ein.
- Es tritt höchstens eines der vier Ereignisse ein.

Zwei Ereignisse  $A$  und  $B$  heißen **disjunkt**, falls  $A \cap B = \emptyset$ . Disjunkt sind also Ereignisse, die nicht gleichzeitig eintreten können.

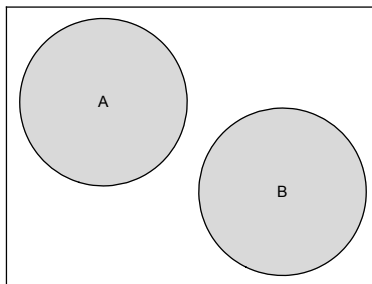


ABBILDUNG 4.  $A$  und  $B$  sind disjunkt

**Beispiel 1.4.7.** Folgende Paare von Ereignissen sind immer disjunkt:



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

Wir schreiben  $A \subset B$  und sagen, dass  $A$  ein **Teilereignis** von  $B$  ist, falls alle Elemente von  $A$  auch in  $B$  enthalten sind.

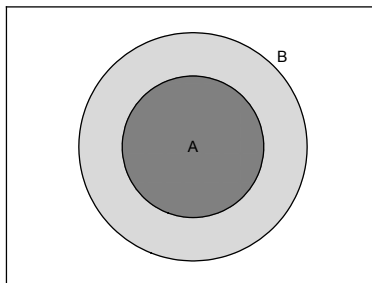


ABBILDUNG 5.  $A$  ist Teilereignis von  $B$

**Beispiel 1.4.8.** Betrachte ein Experiment, bei dem eine zufällige Person aus einer Population ausgewählt wird. Betrachte weiterhin die Ereignisse

$$C := \text{“Die ausgewählte Person besitzt eine Katze”},$$

$$D := \text{“Die ausgewählte Person besitzt ein Haustier”}.$$

Dann gilt offenbar  $C \subset D$ .

Eine sehr wichtige Rolle werden im Folgenden die de Morgan’schen Regeln spielen, die das Komplement einer Vereinigung bzw. eines Schnitts von Ereignissen beschreiben.

**De Morgan’sche Regeln.** Für beliebige Ereignisse  $A_1, A_2, \dots \subset \Omega$  gilt

$$(A_1 \cup A_2 \cup \dots)^c = A_1^c \cap A_2^c \cap \dots,$$

$$(A_1 \cap A_2 \cap \dots)^c = A_1^c \cup A_2^c \cup \dots$$

**Beispiel 1.4.9.** In einem Raum befinden sich  $n$  Personen. Wir betrachten die Aussage “Alle anwesenden Personen besitzen Führerschein”. Was ist die Verneinung dieser Aussage? Die Antwort “Keine der anwesenden Personen besitzt Führerschein” ist falsch! Die richtige Antwort ist “Nicht alle Personen haben einen Führerschein”, was man auch als “Mindestens eine Person hat keinen Führerschein” formulieren kann. Wir definieren nun für jede Person  $i \in \{1, \dots, n\}$  das Ereignis

$$A_i := \text{“die } i\text{-te Person besitzt Führerschein”}.$$

Die Aussage “Alle anwesenden Personen haben Führerschein” entspricht dann dem *Schnitt* dieser Ereignisse, also  $A_1 \cap \dots \cap A_n$ . Die Verneinung der Aussage lautet “Mindestens eine Person hat keinen Führerschein”, was der *Vereinigung*  $A_1^c \cup \dots \cup A_n^c$  entspricht. Also gilt

$$(A_1 \cap \dots \cap A_n)^c = A_1^c \cup \dots \cup A_n^c$$

in Übereinstimmung mit der zweiten Regel von de Morgan. Wir überlassen es dem Leser als Übung, eine ähnliche Interpretation für die erste Regel zu finden.

Zum Schluss dieses Abschnitts erwähnen wir noch einige Identitäten für Ereignisse, die im Folgenden oft benutzt werden.

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:  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ .
- (4) Assoziativgesetze:  $A \cap (B \cap C) = (A \cap B) \cap C$  und  $A \cup (B \cup C) = (A \cup B) \cup C$ .
- (5) Kommutativgesetze:  $A \cap B = B \cap A$  und  $A \cup B = B \cup A$ .

Es besteht eine gewisse Analogie zwischen den mengentheoretischen Operationen  $\cup$  und  $\cap$  und den arithmetischen Operationen  $+$  und  $\cdot$ . Das Analogon des zweiten Distributivgesetzes ist die Identität  $a \cdot (b + c) = a \cdot b + a \cdot c$ . Interessanterweise erstreckt sich diese Analogie nicht auf das erste Distributivgesetz, denn  $a + (b \cdot c) \neq (a + b) \cdot (a + c)$ .

### 1.5. Was sind Wahrscheinlichkeiten?

Die in der Überschrift gestellte Frage ist nicht einfach zu beantworten, denn Wahrscheinlichkeit ist eben ein Grundbegriff und lässt sich genauso schwierig wie andere Grundbegriffe der Mathematik (wie etwa Punkt, Zahl oder Menge) definieren. Wir unternehmen trotzdem einen Versuch, diese Frage zu beantworten.

Vereinfacht gesprochen ist die Wahrscheinlichkeit eines Ereignisses ein Maß dafür, wie oft dieses Ereignis eintritt. Warum sagt man z.B. dass die Wahrscheinlichkeit, dass eine faire Münze “Kopf” zeigt, gleich 50% oder  $1/2$  ist? Es gibt mehrere Erklärungen.

**Frequentistischer Wahrscheinlichkeitsbegriff.** Wir werfen eine faire Münze  $n$  Mal und bezeichnen mit  $f(n)$  die Anzahl der Würfe, in denen die Münze “Kopf” gezeigt hat. Aus Erfahrung weiß man, dass bei einer steigenden Anzahl der Würfe die relative Häufigkeit  $f(n)/n$  gegen  $1/2$  konvergiert:

$$\lim_{n \rightarrow \infty} \frac{f(n)}{n} = \frac{1}{2}.$$

Diese Behauptung wurde empirisch überprüft. Zum Beispiel haben die Mathematiker Buffon und Pearson mit der Münze experimentiert:

- Buffon:  $n = 4040$  Münzwürfe, davon 2048 “Kopf”.
- Pearson:  $n = 24000$  Münzwürfe, davon 12012 “Kopf”.

In beiden Fällen zeigte die Münze ungefähr in 50% aller Würfe “Kopf”. Man kann den Münzwurf sehr leicht auf dem Rechner simulieren. In meiner Simulation ergaben sich folgende Ergebnisse:

- $n = 10^5$  Münzwürfe, davon 50106 “Kopf”. Die relative Häufigkeit ist 0.50106.
- $n = 10^7$  Münzwürfe, davon 4999965 “Kopf”. Die relative Häufigkeit ist 0.4999965.

Bei noch mehr Münzwürfen würde sich die relative Häufigkeit immer mehr  $1/2$  annähern.

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

$$\mathbb{P}[A] := \lim_{n \rightarrow \infty} \frac{f(n)}{n}.$$

Die Zahl  $f(n)/n$  ist die *relative Häufigkeit* des Eintretens von  $A$  in  $n$  Experimenten. Der Limes der relativen Häufigkeit für  $n \rightarrow \infty$  heißt die *Wahrscheinlichkeit* von  $A$  und wird mit  $\mathbb{P}[A]$  bezeichnet. Aus der obigen Interpretation folgt, dass Wahrscheinlichkeit stets Werte zwischen 0 und 1 annimmt, wobei die beiden Randwerte 0 und 1 zugelassen sind:

$$0 \leq \mathbb{P}[A] \leq 1.$$

Allerdings bleibt die Frage offen, warum der Limes von  $f(n)/n$  existiert. Auf der einen Seite weiß man aus Erfahrung, dass sich  $f(n)/n$  bei großem  $n$  einem Wert zu nähern scheint (weshalb die obige Behauptung als **empirisches** Gesetz der großen Zahlen bezeichnet wird). Auf der anderen Seite ist es in der Praxis unmöglich, **unendlich viele** Wiederholungen eines Experiments durchzuführen, weshalb die Existenz des Limes niemals empirisch nachgewiesen werden kann. Auf diese Fragen werden wir im Kapitel “Gesetze der großen Zahlen” eingehen.

**Laplace’scher Wahrscheinlichkeitsbegriff.** Die Münze hat zwei Seiten. Bei einer symmetrischen (oder fairen) Münze sollten diese Seiten die gleiche Wahrscheinlichkeit haben. Also hat jede Seite eine Wahrscheinlichkeit von  $1/2$ .

Bei diesem und vielen anderen Experimenten kann man aus Symmetriegründen davon ausgehen, dass alle Ausgänge gleichwahrscheinlich sind. Solche Experimente heißen Laplace-Experimente.

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

$$p(\omega) = \frac{1}{n}.$$

Die Wahrscheinlichkeit eines Ereignisses  $A \subset \Omega$  ist dann

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

also die Anzahl der “günstigen” Ausgänge geteilt durch Anzahl aller Ausgänge.

In Laplace-Experimenten können Wahrscheinlichkeiten aller Ereignisse zumindest theoretisch durch Abzählen der Ausgänge berechnet werden. Zahlreiche Beispiele von solchen Experimenten haben wir bereits gesehen.

**Beispiel 1.5.1.** Ein fairer (d.h. symmetrischer) Würfel werde zweimal gerollt. Man bestimme die Wahrscheinlichkeit des Ereignisses

$$A := \text{“Die Augensumme der beiden Würfe ist gleich 8”}.$$

**Lösung.** Die Grundmenge ist  $\Omega = \{1, \dots, 6\}^2$  mit  $\#\Omega = 6^2 = 36$ . Das Ereignis  $A$  besteht aus folgenden Ausgängen:

$$A = \{(6, 2), (5, 3), (4, 4), (3, 5), (2, 6)\}.$$

Man beachte, dass die Ausgänge  $(6, 2)$  und  $(2, 6)$  unterschiedlich sind und dass der Ausgang  $(4, 4)$  nur einmal gezählt wird. Die Wahrscheinlichkeit von  $A$  ist

$$\mathbb{P}[A] = \frac{\#A}{\#\Omega} = \frac{5}{36}. \quad \square$$

Nicht jedes Experiment ist ein Laplace-Experiment. Zum Beispiel ist das Werfen einer Reißzwecke kein Laplace-Experiment, denn es ist aus der Erfahrung bekannt, dass sich die Wahrscheinlichkeit einer seitlichen Landung von der Wahrscheinlichkeit einer Landung auf dem Kopf unterscheidet. Die Bestimmung der Blutgruppe einer Person ist ebenfalls kein Laplace-Experiment, da die Blutgruppen nicht gleichwahrscheinlich sind. Weitere Beispiele entstehen beim Werfen von unfairen (=unsymmetrischen) Würfeln oder Münzen.

**Beispiel 1.5.2.** Wir betrachten ein zweistufiges Experiment, das nach den folgenden Regeln verläuft. Zuerst wird eine faire Münze geworfen. Zeigt diese Münze “Kopf”, so wird sie noch einmal geworfen, zeigt die Münze Zahl, so werfen wir einen fairen Würfel. Die Grundmenge dieses Experiments ist

$$\Omega = \{KK, KZ, Z1, Z2, Z3, Z4, Z5, Z6\}.$$

Diese Ausgänge sind allerdings nicht gleichwahrscheinlich! In der Tat, da die Münze beim ersten Wurf mit Wahrscheinlichkeit  $1/2$  “Kopf” und mit der gleichen Wahrscheinlichkeit “Zahl” zeigt, sollten die Wahrscheinlichkeiten der beiden Ereignisse

$$A := \{KK, KZ\} \text{ und } B := \{Z1, Z2, Z3, Z4, Z5, Z6\}$$

jeweils  $1/2$  betragen, also

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

Da der in der zweiten Stufe des Experiments benutzte Gegenstand symmetrisch ist, müssen alle Ausgänge *innerhalb* jedes der beiden Ereignisse  $A$  und  $B$  gleichwahrscheinlich sein.

Die Wahrscheinlichkeiten der einzelnen Ausgänge sind also in diesem Experiment wie folgt gegeben:

$$p(KK) = p(KZ) = \frac{1}{4}, \quad p(Z1) = p(Z2) = p(Z3) = p(Z4) = p(Z5) = p(Z6) = \frac{1}{12}.$$

Man beachte, dass sich die Wahrscheinlichkeiten zu 1 summieren.

**Bayes'scher Wahrscheinlichkeitsbegriff.** In dieser Interpretation wird Wahrscheinlichkeit als ein Grad persönlicher Überzeugung definiert. So könnte z.B. ein Experte meinen, dass im nächsten Jahr mit Wahrscheinlichkeit 70% eine Weltwirtschaftskrise ausbricht. Aus frequentistischer Sicht hat eine solche Aussage wenig Sinn, denn in der vorliegenden Situation kann man die Beobachtung nicht  $n \rightarrow \infty$  Mal unter identischen Bedingungen wiederholen.

**Aufgabe 1.5.3.** Man wirft zwei faire Würfel. Für welches  $k$  ist die Wahrscheinlichkeit des Ereignisses "Die Augensumme ist  $k$ " am größten?

**Aufgabe 1.5.4.** Ein fairer Würfel werde zweimal geworfen. Bestimmen Sie die Wahrscheinlichkeiten der folgenden Ereignisse:

- (a) "Die Augenzahl in jedem Wurf ist kleiner als 4".
- (b) "Es wird mindestens eine 6 gewürfelt".
- (c) "Die Augenzahl im ersten Wurf ist echt größer als die im zweiten Wurf".

**Aufgabe 1.5.5.** Ein fairer Würfel werde dreimal geworfen. Wie groß ist die Wahrscheinlichkeit, dass die Augensumme der drei Würfe gleich 12 ist?

**Aufgabe 1.5.6.** Man wirft  $n$  faire Würfel. Man bestimme die Wahrscheinlichkeit, dass alle Augenzahlen gleich sind.

**Aufgabe 1.5.7.** Von drei Personen habe jede einen kompletten Satz an Cent-Münzen in der Tasche, also jeweils ein 1-, 2-, 5-, 10-, 20- und 50-Cent-Stück. Jeder zieht zufällig eine Münze.

- (a) Geben Sie eine geeignete Grundmenge für dieses Experiment an.
- (b) Mit welcher Wahrscheinlichkeit ergeben die drei gezogenen Münzen zusammen mindestens 40 Cent?

*Hinweis:* Vermeiden Sie bitte Lösungen, in denen Sie dutzende Fälle oder vielleicht sogar sämtliche Ergebnisse auflisten und abzählen. Versuchen Sie also, die Komplexität argumentativ zu reduzieren.

## 1.6. Zufallsexperimente mit unendlicher Grundmenge

In den vorangegangenen Abschnitten haben wir viele Beispiele von Zufallsexperimenten mit einer *endlichen* Grundmenge  $\Omega = \{\omega_1, \dots, \omega_n\}$  gesehen. Die Wahrscheinlichkeiten der Ausgänge können dabei beliebige Zahlen  $p_1, \dots, p_n$  sein, solange die folgenden Bedingungen gelten:

- $p_i \geq 0$  für alle  $i \in \{1, \dots, n\}$  und
- $p_1 + \dots + p_n = 1$ .

**Beispiel 1.6.1.** Wir werfen einen (unsymmetrischen) Gegenstand mit  $n$  Seiten. Die Wahrscheinlichkeit, dass der Gegenstand auf der  $i$ -ten Seite landet sei mit  $p_i$  bezeichnet. Dann

können  $p_1, \dots, p_n$  im Prinzip beliebige Zahlen sein, solange die beiden oben formulierten Bedingungen erfüllt sind.

Man kann sich aber auch Zufallsexperimente mit einer *abzählbar unendlichen*<sup>2</sup> Grundmenge  $\Omega = \{\omega_1, \omega_2, \dots\}$  vorstellen. Die Wahrscheinlichkeiten der einzelnen Ausgänge erfüllen dann die Bedingungen

- $p_i \geq 0$  für alle  $i \in \mathbb{N}$  und
- $p_1 + p_2 + \dots = 1$ .

Ein solches Experiment kann aber unmöglich ein Laplace-Experiment sein! Hätten nämlich alle Ausgänge  $\omega_i$  die gleiche strikt positive Wahrscheinlichkeit  $a > 0$ , so würden sich die Wahrscheinlichkeiten aller Ausgänge auf  $\infty$  summieren. Ein Widerspruch, denn die Summe sollte 1 sein. Wäre die Wahrscheinlichkeit von jedem Ausgangs gleich 0, so würde die Summe aller Wahrscheinlichkeiten ebenfalls 0 sein. Wieder ein Widerspruch, denn die Summe sollte 1 sein.

**Beispiel 1.6.2** (Würfeln bis zur ersten 6: Abzählbar unendliche Grundmenge). Ein Spieler rollt einen fairen Würfel so lange, bis zum ersten Mal eine 6 erscheint. Danach hört er auf. Prinzipiell könnte dies beliebig lange dauern. Als Ausgang des Experiments betrachten wir die Anzahl der Würfe. Daher ist hier die Grundmenge

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

Es sei bemerkt, dass bei dieser Wahl von  $\Omega$  das beobachtete Ergebnis nicht vollständig beschrieben wird, weshalb diese Modellierung nicht empfehlenswert ist. Es wird nämlich nicht verraten, welche Zahlen (und in welcher Reihenfolge) vor der ersten 6 erschienen sind. Insbesondere handelt es sich um kein Laplace-Experiment.

Wir werden nun die Wahrscheinlichkeiten der einzelnen Ausgänge bestimmen. Zu diesem Zweck stellen wir uns vor, dass jeder Mensch auf der Erde das oben beschriebene Experiment durchführt, also bis zur ersten 6 würfelt. Ungefähr ein Sechstel der Weltbevölkerung wird bereits in der ersten Runde eine 6 würfeln. Somit gilt  $p_1 = \frac{1}{6}$ . Es verbleiben  $\frac{5}{6}$  der Weltbevölkerung, die es in der ersten Runde nicht geschafft haben. Von diesen  $\frac{5}{6}$  wird  $\frac{1}{6}$  in der zweiten Runde eine Sechs würfeln, somit gilt  $p_2 = \frac{5}{6} \cdot \frac{1}{6}$ . Es verbleiben  $\frac{5}{6} \cdot \frac{5}{6}$  der Weltbevölkerung, die es auch in der zweiten Runde nicht geschafft haben. Von diesen wird ein Sechstel eine 6 in der dritten Runde würfeln, also gilt  $p_3 = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$ . Indem wir weiter auf diese Weise argumentieren, erhalten wir die Formel

$$p_n = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{n-1}, \quad \text{falls } n \in \{1, 2, \dots\}, \quad \text{und } p_\infty = 0.$$

Im Kapitel über die geometrische Verteilung werden wir diese Formeln noch einmal herleiten.

---

<sup>2</sup>Eine unendliche Menge  $A$  heißt abzählbar, wenn es eine Bijektion zwischen dieser Menge und der Menge der natürlichen Zahlen  $\{1, 2, \dots\}$  gibt. Ansonsten heißt  $A$  überabzählbar. Eine abzählbare Menge kann man als eine Folge  $A = \{a_1, a_2, \dots\}$  darstellen. Abzählbar sind z.B. die Menge der natürlichen Zahlen  $\mathbb{N}$ , die Menge der ganzen Zahlen  $\mathbb{Z}$ , die Menge der rationalen Zahlen  $\mathbb{Q}$ . Überabzählbar ist die Menge der reellen Zahlen  $\mathbb{R}$  oder auch jedes nicht-leere Intervall  $[a, b]$

Zufallsexperimente mit einer endlichen oder abzählbar unendlichen Grundmenge heißen **diskret**.

Alle Experimente, die wir bislang gesehen haben, sind diskret. Weitere Beispiele von diskreten Zufallsexperimenten werden wir im Kapitel “Kombinatorik” sehen.

Im Folgenden werden wir uns aber auch intensiv mit Experimenten beschäftigen, die eine **überabzählbar unendliche** Grundmenge haben. An dieser Stelle erwähnen wir nur zwei solche Experimente, weitere Beispiele werden wir im Kapitel “Geometrische Wahrscheinlichkeiten” sehen.

**Beispiel 1.6.3** (Unendlicher Münzwurf: Überabzählbar unendliche Grundmenge). Wir betrachten ein Experiment, bei dem eine faire Münze **unendlich oft** geworfen wird. Alternativ kann man gleichzeitig unendlich viele Münzen  $M_1, M_2, \dots$  werfen. Zugegebenermaßen lässt sich ein solches Experiment nicht praktisch durchführen, es sollte aber kein Problem sein, sich dieses Experiment vorzustellen. Die Grundmenge besteht aus allen unendlichen Folgen der Symbole  $K$  und  $Z$ , d.h.

$$\Omega = \{K, Z\}^\infty = \{(a_1, a_2, a_3, \dots) : a_i \in \{K, Z\} \text{ für alle } i \in \mathbb{N}\}.$$

So ist etwa  $(K, K, K, K, K, K, \dots)$  ein möglicher Ausgang dieses Experiments. Wie groß ist die Wahrscheinlichkeit dieses Ausgangs? Diese sollte auf jeden Fall kleiner sein, als die Wahrscheinlichkeit, dass die Münze in den ersten 100 Würfeln “Kopf” zeigt, also kleiner als  $1/2^{100}$ . Da wir 100 durch eine beliebig große Zahl ersetzen können, muss die Wahrscheinlichkeit des Ausgangs gleich 0 sein. Analog überzeugt man sich, dass die gleiche Schlussfolgerung für jeden Ausgang gilt:

$$p(\omega) = 0 \text{ für alle } \omega \in \Omega.$$

Auf der anderen Seite sollte  $\mathbb{P}[\Omega] = 1$  sein. Also gilt die Formel  $\mathbb{P}[\Omega] = \sum_{\omega \in \Omega} p(\omega)$  in unserer Situation offenbar nicht! Somit kann man die Wahrscheinlichkeit eines **überabzählbar unendlichen** Ereignisses nicht als die Summe der Wahrscheinlichkeiten der darin enthaltenen Ausgänge berechnen. Diese Beobachtung wird bei der Wahl der Axiome der Wahrscheinlichkeitstheorie eine wichtige Rolle spielen.

**Beispiel 1.6.4.** Die tägliche Entwicklung eines Aktienindex kann man als ein Zufallsexperiment auffassen, dessen Ausgänge Funktionen sind. Die Grundmenge  $\Omega$  ist also die Menge aller Funktionen  $\omega : [0, T] \rightarrow \mathbb{R}$  und somit überabzählbar. Auch hier erscheint plausibel, dass für jede konkrete Funktion  $\omega$ , die Wahrscheinlichkeit *exakt* diese Funktion zu beobachten, gleich Null ist, also  $p(\omega) = 0$  für alle  $\omega \in \Omega$ .

## 1.7. Zufallsvariablen

Als Ergebnis eines Zufallsexperiments ergibt sich oft eine Zahl. Diese vom Zufall abhängige Zahl bezeichnen wir als Zufallsvariable.

**Beispiel 1.7.1.** Beim zweimaligen Würfeln kann man folgende Zufallsvariablen betrachten: die Augensumme, die größere Augenzahl, die kleinere Augenzahl, die Differenz der Augenzahlen,  $\dots$ . All diese Größen sind gewisse Funktionen der beiden Augenzahlen. Deshalb führen wir die folgende Definition ein:

**Definition 1.7.2.** Eine **Zufallsvariable** ist eine auf der Grundmenge definierte Funktion  $X : \Omega \rightarrow \mathbb{R}$ .

Für einen Ausgang  $\omega \in \Omega$  heißt  $X(\omega)$  der Wert von  $X$  zum Ausgang  $\omega$ .

**Beispiel 1.7.3.** Wir betrachten das zweimalige Würfeln mit einem fairen Würfel. Die Grundmenge dieses Experiments ist

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

Wir können z.B. die 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. Die Zufallsvariablen  $X_1$  und  $X_2$  modellieren somit die erste und die zweite Augenzahl. Nun können wir auch die Augensumme  $S := X_1 + X_2$  definieren. Als Funktion auf  $\Omega$  ist  $S$  durch die folgende Vorschrift definiert:

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

Wird z.B. der Ausgang  $(2, 6)$  gewürfelt, so nimmt die Zufallsvariable  $S$  den Wert  $S(2, 6) = 8$  an. Ein weiteres Beispiel ist die Zufallsvariable  $M := \max(X_1, X_2)$ , die die größere der beiden Augenzahlen beschreibt. Als Funktion auf  $\Omega$  ist  $M$  durch die folgende Vorschrift definiert:

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

Wird z.B. der Ausgang  $(2, 6)$  gewürfelt, so nimmt die Zufallsvariable  $M$  den Wert  $M(2, 6) = 6$  an.

**Beispiel 1.7.4.** Es sollte stets eine klare Trennlinie zwischen den Begriffen “Ereignis” und “Zufallsvariable” gezogen werden. Für ein Ereignis gibt es nur zwei Möglichkeiten: es tritt ein oder es tritt nicht ein. Für eine Zufallsvariable gibt es hingegen typischerweise mehrere Möglichkeiten, nämlich alle Zahlenwerte, die diese annehmen kann. Betrachten wir beispielhaft wieder das zweimalige Würfeln. Dann ist die “Augensumme”  $S$  an sich kein Ereignis, sondern eine Zufallsvariable, die die Werte  $2, 3, \dots, 12$  annehmen kann. Hingegen ist eine *Aussage* über die Werte der Augensumme (oder eine *Bedingung* an diese Werte) ein Ereignis. Z.B. können wir das Ereignis “die Augensumme ist mindestens 10” betrachten. Für dieses benutzen wir dann die Notation

$$\{S \geq 9\} = \{(a_1, a_2) : a_1 + a_2 \geq 9\} = \{(6, 6), (6, 5), (6, 4), (5, 6), (5, 5), (4, 6)\}.$$

Analog ist “die größere der beiden Augenzahlen” eine Zufallsvariable (die wir mit  $M$  bezeichnet haben) jedoch kein Ereignis. Ein Ereignis entsteht, wenn wir eine Bedingung an die Werte von  $M$  stellen, etwa “die größere der beiden Augenzahlen ist gleich 3”. Für dieses Ereignis benutzen wir die Notation

$$\{M = 3\} = \{(a_1, a_2) : \max(a_1, a_2) = 3\} = \{(3, 1), (3, 2), (3, 3), (1, 3), (2, 3)\}.$$

**Beispiel 1.7.5** (Indikatorvariablen). Wir können jedem Ereignis eine Zufallsvariable zuordnen, die den Wert 1 annimmt, wenn dieses Ereignis eintritt, und den Wert 0 sonst.



**Definition 1.7.6.** Sei  $A \subset \Omega$  ein Ereignis. Die **Indikatorvariable** von  $A$  ist die Zufallsvariable  $\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$ , die wie folgt definiert wird:

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

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

- $\mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$ .
- $\mathbb{1}_{A \cup B} = \max\{\mathbb{1}_A, \mathbb{1}_B\} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \cdot \mathbb{1}_B$ .
- $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B$  falls  $A \cap B = \emptyset$ .
- $\mathbb{1}_{A \Delta B} = \mathbb{1}_A + \mathbb{1}_B \pmod 2$  für beliebige  $A$  und  $B$ .
- $\mathbb{1}_A + \mathbb{1}_{A^c} = 1$  und  $\mathbb{1}_A \cdot \mathbb{1}_{A^c} = 0$ .

Den Nachweis überlassen wir dem Leser als Übung.

Die **Verteilung** einer Zufallsvariable ist die Angabe der Werte, die diese Zufallsvariable annehmen kann, sowie der Wahrscheinlichkeiten, mit denen diese Werte angenommen werden.

**Beispiel 1.7.7.** Beim einmaligen Würfeln können wir die Verteilung der Zufallsvariable  $X :=$  "Augenzahl" beschreiben, indem wir sagen, dass diese Zufallsvariable Werte  $1, 2, \dots, 6$  mit der gleichen Wahrscheinlichkeit  $1/6$  annimmt:

$$\mathbb{P}[X = 1] = \mathbb{P}[X = 2] = \dots = \mathbb{P}[X = 6] = \frac{1}{6}.$$

Am Einfachsten beschreibt man die Verteilung mithilfe der sogenannten Zähl-dichte.

**Definition 1.7.8.** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine beliebige Zufallsvariable. Die **Zähl-dichte** von  $X$  ist die Funktion  $p_X : \mathbb{R} \rightarrow [0, 1]$  mit

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

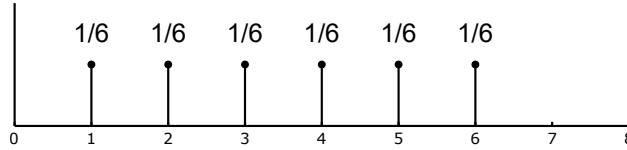
Mit anderen Worten ist  $p_X(y)$  die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  den Wert  $y$  annimmt.

**Beispiel 1.7.9.** Für die Augenzahl beim Rollen eines fairen Würfels ist die Zähl-dichte gegeben durch

$$p(y) = \begin{cases} 1/6, & \text{falls } y \in \{1, \dots, 6\}, \\ 0, & \text{sonst.} \end{cases}$$

Der Graph der Zähl-dichte sieht wie folgt aus:

**Beispiel 1.7.10.** Es sei  $S$  die Augensumme beim Würfeln mit 2 fairen Würfeln. Wir bestimmen die Zähl-dichte von  $S$ .



**Lösung.** Die Grundmenge dieses Experiments ist  $\Omega = \{1, \dots, 6\}^2$  mit  $\#\Omega = 36$ . Die Menge der Werte, die  $S$  annehmen kann, ist  $\{2, 3, \dots, 12\}$ . Nun bestimmen wir die Zähldichte von  $S$ :

$$p_S(2) = \mathbb{P}[S = 2] = \frac{1}{36},$$

$$p_S(3) = \mathbb{P}[S = 3] = \frac{2}{36},$$

⋮

$$p_S(7) = \mathbb{P}[S = 7] = \frac{6}{36},$$

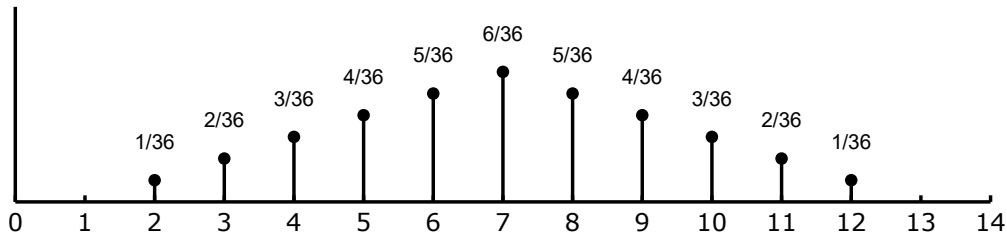
⋮

$$p_S(11) = \mathbb{P}[S = 11] = \frac{2}{36},$$

$$p_S(12) = \mathbb{P}[S = 12] = \frac{1}{36}.$$

Zusammenfassend gilt

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



**Beispiel 1.7.11.** Sei  $A \subset \Omega$  ein Ereignis und  $\mathbb{1}_A$  die Indikatorvariable von  $A$ . Für die Zähldichte dieser Indikatorvariable gilt

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

**Bemerkung 1.7.12.** Für die Zähldichte einer beliebigen Zufallsvariable  $Z$  gelten stets die folgenden zwei Eigenschaften:

- $p_Z(y) \in [0, 1]$  für alle  $y \in \mathbb{R}$ .

- $\sum_{y \in \mathbb{R}} p_Z(y) = 1.$

**Aufgabe 1.7.13.** Man wirft zwei faire Würfel und bezeichnet mit  $X_1$  und  $X_2$  die Augenzahlen. Bestimmen Sie die Zähldichten der folgenden Zufallsvariablen:

- (a) Die größere Augenzahl  $\max(X_1, X_2).$
- (b) Die kleinere Augenzahl  $\min(X_1, X_2).$
- (c) Die Differenz der Augenzahlen  $X_1 - X_2.$
- (d) Der Betrag der Differenz der Augenzahlen  $|X_1 - X_2|.$

## 1.8. Erwartungswert

**Beispiel 1.8.1.** Wir betrachten ein Glücksspiel, bei dem man mit Wahrscheinlichkeit  $1/5$  eine Summe von 100 Euro gewinnt und mit Wahrscheinlichkeit  $4/5$  einen Betrag von 30 Euro verliert. Wieviel gewinnt oder verliert man im Durchschnitt? Ein Laie könnte wie folgt argumentieren. In 5 Runden eines solchen Spiels gewinnt man “im Durchschnitt” einmal und verliert “im Durchschnitt” viermal. Also gewinnt man in fünf Runden “im Durchschnitt”  $1 \cdot 100 - 4 \cdot 30 = -20$  Euro. In einer Runde gewinnt man somit durchschnittlich  $-20/5 = -4$  Euro.

Diese Überlegung können wir zur Definition des Erwartungswerts ausbauen, der den “durchschnittlichen Wert” einer Zufallsvariable darstellt. Wir betrachten eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  mit  $n$  Werten  $y_1, y_2, \dots, y_n.$  Diese kann z.B. den Gewinn in einem Glücksspiel modellieren. Die Wahrscheinlichkeit, dass  $X$  den Wert  $y_i$  annimmt, sei mit  $p_i := \mathbb{P}[X = y_i]$  bezeichnet:

$$\begin{array}{l} \text{Werte von } X : \quad y_1 \quad y_2 \quad \dots \quad y_n \\ \text{Wahrscheinlichkeiten : } \quad p_1 \quad p_2 \quad \dots \quad p_n. \end{array}$$

**Definition 1.8.2.** Der **Erwartungswert** von  $X$  ist die Zahl

$$\mathbb{E}X \stackrel{\text{def}}{=} p_1 y_1 + \dots + p_n y_n = \sum_{i=1}^n p_i y_i. \tag{1.8.1}$$

Der Erwartungswert ist also eine gewichtete Summe der Werte der Zufallsvariable, wobei jeder Wert mit seiner Wahrscheinlichkeit gewichtet wird. Je “wahrscheinlicher” ein Wert ist, umso stärker wird er berücksichtigt.

**Beispiel 1.8.3.** Wir werfen einmal einen fairen Würfel und bezeichnen mit  $X$  die Augenzahl. Die möglichen Werte von  $X$  sind 1, 2, 3, 4, 5, 6, und jeder Wert hat Wahrscheinlichkeit  $\frac{1}{6}.$  Somit gilt für den Erwartungswert von  $X:$

$$\mathbb{E}X = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{1}{6}(1 + \dots + 6) = 3.5.$$

Viele Anfänger definieren den Erwartungswert als “den wahrscheinlichsten Wert einer Zufallsvariable”. Dieses Beispiel zeigt, dass diese Definition falsch ist: der Wert 3.5 wird gar nicht angenommen und somit sicherlich nicht der wahrscheinlichste Wert.

**Beispiel 1.8.4.** Sei  $A \subset \Omega$  ein Ereignis und  $\mathbb{1}_A$  die Indikatorvariable von  $A$ . Diese nimmt den Wert 1 genau dann an, wenn das Ereignis eintritt. Ansonsten nimmt sie den Wert 0 an. Demnach gilt:

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

Der Erwartungswert einer Indikatorvariable ist somit die Wahrscheinlichkeit des entsprechenden Ereignisses.

**Satz 1.8.5** (Linearität des Erwartungswerts).

(a) Seien  $X, Y : \Omega \rightarrow \mathbb{R}$  Zufallsvariablen. Dann gilt

$$\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y. \quad (1.8.2)$$

(b) Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable und  $a \in \mathbb{R}$  eine Konstante. Dann gilt

$$\mathbb{E}[aX] = a\mathbb{E}X. \quad (1.8.3)$$

**Beweis von (a).** Wir bezeichnen die Werte von  $X$  mit  $x_1, \dots, x_n$ . Die Werte von  $Y$  seien analog mit  $y_1, \dots, y_m$  bezeichnet. Es sei  $A_i$  ein Ereignis, das genau dann eintritt, wenn  $X$  den Wert  $x_i$  annimmt, d.h.

$$A_i = \{X = x_i\} = \{\omega \in \Omega : X(\omega) = x_i\}, \quad i = 1, \dots, n.$$

Die Ereignisse  $A_1, \dots, A_n$  sind disjunkt, d.h.  $A_i \cap A_j = \emptyset$  für alle  $i \neq j$ . Außerdem ist die Vereinigung dieser Ereignisse  $\Omega$ . Der Erwartungswert von  $X$  ist definitionsgemäß

$$\mathbb{E}X = \sum_{i=1}^n x_i \mathbb{P}[A_i].$$

Analog verfahren wir mit der Zufallsvariable  $Y$ . Sei  $B_j$  das Ereignis, das genau dann eintritt, wenn  $Y$  den Wert  $y_j$  annimmt, d.h.

$$B_j = \{Y = y_j\} = \{\omega \in \Omega : Y(\omega) = y_j\}, \quad j = 1, \dots, m.$$

Die Ereignisse  $B_1, \dots, B_m$  sind disjunkt, d.h.  $B_i \cap B_j = \emptyset$  für alle  $i \neq j$ . Die Vereinigung dieser Ereignisse ist  $\Omega$ . Der Erwartungswert von  $Y$  ist somit

$$\mathbb{E}Y = \sum_{j=1}^m y_j \mathbb{P}[B_j].$$

Wir haben also zwei disjunkte Zerlegungen von  $\Omega$ , nämlich  $\Omega = A_1 \cup \dots \cup A_n$  und  $\Omega = B_1 \cup \dots \cup B_m$ . Es folgt, dass auch die Zerlegung

$$\Omega = \bigcup_{i=1}^n \bigcup_{j=1}^m (A_i \cap B_j)$$

disjunkt ist.

Nun schauen wir uns die Zufallsvariable  $X + Y$  an. Auf dem Ereignis  $A_i \cap B_j$  nimmt  $X$  den Wert  $p_i$  und  $Y$  den Wert  $q_j$  an. Somit nimmt die Zufallsvariable  $X + Y$  den Wert  $p_i + q_j$  an.

Für den Erwartungswert von  $X + Y$  erhalten wir

$$\begin{aligned}\mathbb{E}[X + Y] &= \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) \mathbb{P}[A_i \cap B_j] \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i \mathbb{P}[A_i \cap B_j] + \sum_{i=1}^n \sum_{j=1}^m y_j \mathbb{P}[A_i \cap B_j] \\ &= \sum_{i=1}^n x_i \sum_{j=1}^m \mathbb{P}[A_i \cap B_j] + \sum_{j=1}^m y_j \sum_{i=1}^n \mathbb{P}[A_i \cap B_j].\end{aligned}$$

Nun bilden die Ereignisse  $A_i \cap B_1, A_i \cap B_2, \dots, A_i \cap B_m$  eine disjunkte Zerlegung des Ereignisses  $A_i$ . Es gilt also

$$\sum_{j=1}^m \mathbb{P}[A_i \cap B_j] = \mathbb{P}[A_i].$$

Analog bilden die Ereignisse  $A_1 \cap B_j, A_2 \cap B_j, \dots, A_n \cap B_j$  eine disjunkte Zerlegung des Ereignisses  $B_j$ . Also gilt

$$\sum_{i=1}^n \mathbb{P}[A_i \cap B_j] = \mathbb{P}[B_j].$$

Setzen wir beide Identitäten in die Formel für  $\mathbb{E}[X + Y]$  ein, so erhalten wir

$$\mathbb{E}[X + Y] = \sum_{i=1}^n x_i \mathbb{P}[A_i] + \sum_{j=1}^m y_j \mathbb{P}[B_j] = \mathbb{E}X + \mathbb{E}Y,$$

was den Beweis abschließt. □

**Beweis von (b).** Übung. □

**Bemerkung 1.8.6.** 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 1.8.7.** 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  $S$ :

$$\mathbb{E}S = \mathbb{E}X_1 + \dots + \mathbb{E}X_n = n \cdot \mathbb{E}X_1 = 3.5 \cdot n.$$

**Aufgabe 1.8.8.** Betrachten Sie das folgende Würfelspiel: Man darf sich aus einem großen Vorrat an fairen Würfeln beliebig viele herausgreifen, um sie dann zu werfen. Ist eine Sechs unter den geworfenen Augenzahlen, so gewinnen Sie nichts, andernfalls bekommen Sie die geworfene Augensumme in Euro ausgezahlt. Es sei  $X$  die Zufallsvariable, die Ihren zufälligen Gewinn in diesem Spiel beschreibt.

- (1) Bestimmen Sie den erwarteten Gewinn  $a_n = \mathbb{E}X$  in Abhängigkeit von der Anzahl  $n$  der herausgegriffenen Würfel.
- (2) Wieviele Würfel würden Sie nehmen, um den erwarteten Gewinn zu maximieren?

\*\*\*\*\* Verschieben!!!

**Beispiel 1.8.9** (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:

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

Die Anzahl der Richtigen ist dann  $S = X_1 + \dots + X_6$ . Mit Satz 1.8.5 gilt für den Erwartungswert von  $S$ :

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

### 1.9. Unabhängigkeit???

Zwei Gegenstände werden geworfen...

## Kombinatorik und diskrete Verteilungen

Kombinatorik ist die Kunst des Abzählens von Kombinationen von Objekten. Wir beginnen mit einem Beispiel.

### 2.1. Geburtstagsproblem

**Beispiel 2.1.1** (Geburtstagsproblem). In einem Raum befinden sich  $k = 200$  Studenten. Wie groß ist die Wahrscheinlichkeit, dass es eine “Kollision” der Geburtstage gibt, d.h. dass es im Raum mindestens 2 Studenten gibt, die am gleichen Tag Geburtstag feiern?

**Lösung.** Zuerst beschreiben wir die Grundmenge. Es gibt  $n = 365$  Tage im Jahr (Schaltjahre werden vernachlässigt), die wir mit  $1, \dots, n$  nummerieren. Das uns interessierende Zufallsexperiment besteht darin, dass  $k = 200$  Personen ihre Geburtstage auf einen Zettel schreiben. Die Ausgänge dieses Experiments sind Listen  $(a_1, \dots, a_k)$ , wobei die Komponente  $a_i \in \{1, \dots, n\}$  den Geburtstag der  $i$ -ten Person angibt. Die Grundmenge ist die Menge aller möglichen Listen, also

$$\Omega = \{1, \dots, n\}^k = \{(a_1, \dots, a_k) : a_i \in \{1, \dots, 365\} \text{ für alle } i = 1, \dots, k\}.$$

Die Anzahl der Elemente in  $\Omega$  ist  $\#\Omega = n^k$ . Es sei bemerkt, dass die Geburtstage der verschiedenen Personen unabhängig voneinander sind, so dass es sich um ein Produktexperiment handelt.

Wir gehen in dieser Lösung davon aus, dass alle Tage im Jahr als Geburtstage gleichwahrscheinlich sind, so dass es sich um ein Laplace-Experiment handelt und alle  $n^k$  Ausgänge gleichwahrscheinlich sind.

Nun definieren wir das uns interessierende Ereignis

$$A = \text{“mindestens zwei Studenten haben am selben Tag Geburtstag”}.$$

Das direkte Abzählen der Elemente in  $A$  ist schwierig, deshalb betrachten wir das Gegenereignis

$$A^c = \text{“keine zwei Studenten haben am selben Tag Geburtstag”}.$$

Äquivalent kann man  $A^c$  auch so beschreiben:

$$A^c = \text{“alle Studenten haben Geburtstage an verschiedenen Tagen”}.$$

Um die Anzahl der Ausgänge im Ereignis  $A^c$  zu berechnen, gehen wir wie folgt vor. Wir möchten die Anzahl der Möglichkeiten bestimmen, für alle Studenten Geburtstage auszuwählen, so dass keine zwei Geburtstage gleich sind. Der erste Student hat für seinen Geburtstag  $n$  Möglichkeiten, der zweite hat  $n - 1$  Möglichkeiten, usw. Der letzte,  $k$ -te Student,





Hashfunktion gleichwahrscheinlich sind, so zeigt die obige Aufgabe, dass für  $k = 2^{\ell/2}$  die Attacke mit einer Wahrscheinlichkeit von  $\approx 1 - e^{-1/2} \approx 0.39$  erfolgreich ist. Bei nicht uniform verteilten Werten der Hashfunktion ist die Erfolgswahrscheinlichkeit noch größer, s. Bemerkung 2.1.3. Der Wert von  $\ell$  sollte so groß sein, das eine solche Attacke in einer vernünftigen Rechenzeit nicht möglich ist.

## 2.2. Vier Urnenmodelle

Viele Probleme der Wahrscheinlichkeitstheorie lassen sich auf sogenannte Urnenmodelle zurückführen. Es sei eine Urne mit  $n$  Bällen gegeben. Die Bälle seien mit  $1, \dots, n$  beschriftet. Wir betrachten ein Zufallsexperiment, bei dem  $k$  Mal jeweils ein Ball aus der Urne gezogen und seine Nummer notiert wird. Beim Ziehen der Bälle gibt es 2 Möglichkeiten:

- die Bälle werden mit/ohne Zurücklegen gezogen.

Beim Ziehen mit Zurücklegen wird nach jeder Ziehung der gezogene Ball zurück in die Urne gelegt. Dabei kann also ein Ball mehrmals aus der Urne gezogen werden. Beim Ziehen ohne Zurücklegen wird jeder Ball höchstens einmal gezogen.

Beim Notieren der Nummern der Bälle gibt es ebenfalls 2 Möglichkeiten:

- die Nummern werden mit/ohne Berücksichtigung der Reihenfolge notiert.

Dabei bedeutet “Notieren mit Reihenfolge”, dass zwei Ausgänge des Experiments auch dann als unterschiedlich angesehen werden, wenn sie sich nur durch die Reihenfolge der gezogenen Bälle unterscheiden.

Insgesamt ergeben sich  $2 \times 2 = 4$  Urnenmodelle, die wir nun einzeln betrachten werden.

### Modell 1: Ziehen mit Zurücklegen und mit Berücksichtigung der Reihenfolge.

Bei diesem Modell ziehen wir  $k$ -Mal aus einer Urne mit  $n$  Bällen. Nach jeder Ziehung wird der gezogene Ball zurück in die Urne gelegt. Die Nummern der gezogenen Bälle werden als eine Liste  $(a_1, \dots, a_k)$  notiert, wobei  $a_i$  für die Nummer des bei der  $i$ -ten Ziehung gezogenen Balls steht. Die Grundmenge kann somit wie folgt dargestellt werden:

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

Es gilt also  $\#\Omega = n^k$ . Man beachte:

- In der Liste sind Wiederholungen möglich, d.h. es ist möglich, dass  $a_i = a_j$  (wir ziehen nämlich mit Zurücklegen).
- Zwei Listen, die sich durch die Reihenfolge unterscheiden, entsprechen verschiedenen Ausgängen, etwa  $(5, 6, 2) \neq (2, 5, 6)$  (wir notieren mit Reihenfolge).

Alternativ können wir den Ausgang des Experiments durch eine Funktion  $f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$  kodieren, wobei  $f(i)$  für die Nummer des bei der  $i$ -ten Ziehung gezogenen Balls steht. Die Grundmenge ist die Menge aller Funktionen:

$$\Omega = \{f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}\}.$$

Bei diesem Modell handelt es sich um ein Produktexperiment: das Ziehen eines Balls aus einer Urne mit  $n$  Bällen wird  $k$  Mal unter gleichen Bedingungen wiederholt.

**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.** Ebenso kann das  $k$ -malige Werfen einer fairen Münze als ein Urnenmodell mit 2 Bällen angesehen werden.

**Beispiel 2.2.3.**  $k$  Personen nennen ihre Geburtstage. Die möglichen Geburtstage können als  $n = 365$  Bälle dargestellt werden.

**Modell 2: Ziehen ohne Zurücklegen und mit Berücksichtigung der Reihenfolge.** “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. Die Ausgänge sind also Listen  $(a_1, \dots, a_k)$ , in denen es keine Wiederholungen gibt. Die Grundmenge kann wie folgt dargestellt werden:

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

Alternativ können wir die Ausgänge durch Funktionen  $f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$  kodieren, die jeden Wert höchstens einmal annehmen. Die Grundmenge ist also

$$\Omega = \{f : \{1, \dots, k\} \rightarrow \{1, \dots, n\} : f \text{ is injektiv}\}.$$

Die Anzahl der Ausgänge kann wie folgt bestimmt werden. Für den ersten gezogenen Ball gibt es  $n$  Möglichkeiten, für den zweiten  $n - 1$ , für den dritten  $n - 2$ , usw. Für die letzte Ziehung gibt es  $n - k + 1$  Möglichkeiten. Somit gilt

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

Diese Überlegung funktioniert nur für  $k \leq n$ . Für  $k > n$  ist das Experiment nicht möglich: wir können nicht mehr Bälle ziehen, als es in der Urne gibt.

**Definition 2.2.4.** Die Zahl  $n(n - 1) \cdot \dots \cdot (n - k + 1)$  heißt die *fallende Faktorielle* und wird mit  $(n)_k$  bezeichnet.

**Permutationen.** Im Spezialfall wenn  $k = n$  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 begründen: 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 Zurücklegen und ohne Berücksichtigung der Reihenfolge.** Am Einfachsten stellt man sich vor, dass man  $k$  Bälle aus einer Urne mit  $n$  Bällen mit einem Griff zieht. Das Ergebnis ist dann eine  $k$ -elementige Teilmenge  $A = \{a_1, \dots, a_k\}$  von  $\{1, \dots, n\}$ . Man beachte, dass die Elemente  $a_i$  paarweise verschieden sind (da wir ohne Zurücklegen ziehen) und dass zwei Ausgänge, die sich nur durch die Reihenfolge unterscheiden, als gleich gelten (da wir die Reihenfolge der Ziehungen nicht berücksichtigen). Zum Beispiel sind folgende Ausgänge gleich:

$$\{5, 6, 3\} = \{5, 3, 6\} = \{6, 3, 5\} = \{6, 5, 3\} = \{3, 5, 6\} = \{3, 6, 5\}.$$

Die Grundmenge kann wie folgt dargestellt werden:

$$\Omega = \{A \subset \{1, \dots, n\} : \#A = k\}.$$

Die Anzahl der Elemente in  $\Omega$  kann wie folgt bestimmt werden. Zuerst können wir wie in Modell 2, also *mit* Reihenfolge und ohne Zurücklegen ziehen. Es gäbe dann  $(n)_k$  Ausgänge. Nun müssen wir aber die Reihenfolge vergessen. Das heißt, wir müssen Ausgänge, die sich nur durch Permutationen unterscheiden, identifizieren (z.B. müssen die 6 Permutationen von  $(5, 6, 3)$  zu einem Ausgang identifiziert werden). Da man  $k!$  Permutationen von  $k$  Elementen hat, werden jeweils  $k!$  Ausgänge zu einem Ausgang zusammengefasst. Es gilt also

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

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

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

**Beispiel 2.2.6.** Aus einer Klasse mit 20 Schülern sollen 10 Schüler für die Teilnahme an einer Reise ausgewählt werden. Dafür gibt es  $\binom{20}{10}$  Möglichkeiten.

**Beispiel 2.2.7** (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

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

Beide Lösungen liefern also das gleiche Ergebnis. □

**Beispiel 2.2.8** (Fortsetzung). Im Jahre 1995 wurde beim Lotto “6 aus 49” nach 40 Jahren und 3016 Ziehungen eine Gewinnreihe zum zweiten Mal gezogen. Handelt es sich um eine Sensation?

**Lösung.** Um diese Frage zu beantworten, betrachten wir das folgende Experiment: es wird 3016 Mal eine 6-elementige Teilmenge (Gewinnreihe) aus einer 49-elementigen Menge zufällig gezogen. Uns interessiert die Wahrscheinlichkeit des Ereignisses

$A :=$  “Bei mindestens 2 der 3016 Ziehungen wurde die gleiche 6-elementige Menge gezogen”.

Nun kann man genauso wie beim Geburtstagsproblem vorgehen. Das Komplement von  $A$  lautet

$A^c :=$  “Bei den 3016 Ziehungen wurden verschiedene 6-elementige Mengen gezogen”.

Die Wahrscheinlichkeit von  $A^c$  ist

$$\mathbb{P}[A^c] = \frac{n(n-1) \dots (n-k+1)}{n^k},$$

wobei  $k = 3016$  die Anzahl der Ziehungen und  $n = \binom{49}{6}$  die Anzahl der Gewinnreihen ist. Es folgt, dass

$$\mathbb{P}[A] = 1 - \frac{n(n-1) \dots (n-k+1)}{n^k} \approx 0.278.$$

Diese Wahrscheinlichkeit ist ziemlich groß. Es handelt sich also um keine große Sensation. □

**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 Sitzmöglichkeiten 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.  $\square$

**Aufgabe 2.2.10.** Aus einem Kartenspiel mit 52 Karten (darunter 4 Asse) werden 4 Karten ohne Zurücklegen zufällig gezogen. Wie groß ist die Wahrscheinlichkeit, dass alle 4 gezogenen Karten Asse sind?

**Aufgabe 2.2.11.** Aus einem Kartenspiel mit 52 Karten (darunter jeweils 13 in jeder der 4 Farben) werden 4 Karten ohne Zurücklegen zufällig gezogen. Wie groß ist die Wahrscheinlichkeit, dass alle 4 Karten die gleiche Farbe haben?

**Modell 4: Ziehen ohne Reihenfolge und mit Zurücklegen.** **Muss überarbeitet werden...** 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  $\Omega$  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 2.2.12.**  $k$  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 2.2.13.** 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.  $\square$

**Aufgabe 2.2.14.** Ein Topf enthalte drei 1-Euro-Münzen und drei 2-Euro-Münzen. Es werden genau drei Münzen ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass in der Summe genau 4 Euro gezogen werden?

**Aufgabe 2.2.15.** Fünf Studierende fahren mit einem Auto in den Urlaub. Nur zwei von ihnen haben einen Führerschein. Wie viele legale Sitzmöglichkeiten gibt es, wenn der Wagen genau 5 Plätze hat? (Plätze sind unterscheidbar).

**Aufgabe 2.2.16.** Sie würfeln 6-mal mit einem fairen Würfel. Wie groß ist die Wahrscheinlichkeit, dass alle 6 Augenzahlen paarweise verschieden sind?

**Aufgabe 2.2.17.** Sie würfeln 18-mal mit einem fairen Würfel. Mit welcher Wahrscheinlichkeit kommt jede mögliche Augenzahl genau dreimal vor?

**Aufgabe 2.2.18.** In einer Reihe stehen 50 Stühle. Es sollen 10 Stühle so ausgewählt werden, dass keine zwei davon nebeneinander stehen. Wieviele Möglichkeiten gibt es, solch eine Auswahl zu treffen?

**Aufgabe 2.2.19.** Was ist wahrscheinlicher: in 4 Würfeln mit einem fairen Würfel mindestens eine 6 zu erzielen, oder in 24 Würfeln mit jeweils 2 fairen Würfeln mindestens eine Doppelsechs (d.h. einen Sechserpasch) zu erzielen?

**Aufgabe 2.2.20.** Zwölf Politiker, darunter auch Frau  $X$  und Herr  $Y$ , nehmen an einem runden Tisch völlig zufällig Platz. Wie groß ist die Wahrscheinlichkeit, dass Frau  $X$  und Herr  $Y$  nebeneinander sitzen?

**Aufgabe 2.2.21.** In einem Topf liegen gut vermischt  $n$  Lose, darunter 1 Gewinnlos und  $n - 1$  Nieten. In einer Warteschlange stehen  $n$  Personen, darunter auch Sie, die eine nach

der anderen die Lose (ohne Zurücklegen) aus dem Topf ziehen. Sie wollen das Gewinnlos ziehen. Was ist besser: als erste oder als letzte Person in der Warteschlange zu stehen?

**Eigenschaften der Binomialkoeffizienten.** Wir erinnern, dass der Binomialkoeffizient

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

die Anzahl der  $k$ -elementigen Teilmengen einer  $n$ -elementigen Menge angibt. Die Binomialkoeffizienten lassen sich im sogenannten **Pascal'schen Dreieck** darstellen:

$n = 0$				1				
$n = 1$				1	1			
$n = 2$			1	2	1			
$n = 3$		1	3	3	1			
$n = 4$		1	4	6	4	1		
$n = 5$	1	5	10	10	5	1		
$n = 6$	1	6	15	20	15	6	1	

In der  $n$ -ten Reihe stehen die Zahlen  $\binom{n}{1}, \binom{n}{2}, \dots, \binom{n}{n-1}, \binom{n}{n}$ . Man merkt, dass das Pascal'sche Dreieck symmetrisch bezüglich der vertikalen Achse ist:

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

**BEWEIS.** Jeder  $k$ -elementigen Teilmenge von  $\{1, \dots, n\}$  ihr Komplement zugeordnet werden kann, das  $n - k$  Elemente besitzt. Dies stellt eine Bijektion zwischen der Menge der  $k$ -elementigen Teilmengen und der Menge der  $(n - k)$ -elementigen Teilmengen, woraus folgt, dass die beiden Mengen gleich viele Elemente haben.  $\square$

Außerdem merkt man, dass im Pascal'schen Dreieck jede Zahl mit der Summe der zwei darüberstehenden Zahlen übereinstimmt:

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

**BEWEIS.** Einerseits ist die Anzahl der  $k$ -elementigen Teilmengen von  $\{1, \dots, n\}$  gleich  $\binom{n}{k}$ . Andererseits kann diese Anzahl wie folgt bestimmt werden. Die  $k$ -elementigen Teilmengen teilen sich in zwei disjunkte Klassen auf:

- Teilmengen, die das Element  $n$  beinhalten. Außerdem muss eine solche Teilmenge  $k - 1$  Elemente aus  $\{1, \dots, n - 1\}$  beinhalten, also gibt es insgesamt  $\binom{n-1}{k-1}$  solche Teilmengen.
- Teilmengen, die das Element  $n$  nicht beinhalten. Eine solche Teilmenge muss  $k$  Elemente aus der Menge  $\{1, \dots, n - 1\}$  beinhalten. Also gibt es  $\binom{n-1}{k}$  solche Teilmengen.

Insgesamt gibt es also  $\binom{n-1}{k-1} + \binom{n-1}{k}$   $k$ -elementige Teilmengen von  $\{1, \dots, n\}$ . Diese Zahl muss mit  $\binom{n}{k}$  übereinstimmen.  $\square$

Die Binomialkoeffizienten tauchen im binomischen Lehrsatz auf:

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

**Aufgabe 2.2.22.** Zeigen Sie: Die  $n$ -te Zeilensumme im Pascal'schen Dreieck ist  $2^n$ , d.h.

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

Zeigen Sie, dass die entsprechende alternierende Summe verschwindet:

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

**Aufgabe 2.2.23.** Es sei  $f^{(n)}$  die  $n$ -te Ableitung der Funktion  $f$ . Zeigen Sie, dass

$$(fg)^{(n)} = \sum_{k=0}^n \binom{n}{k} f^{(k)} g^{(n-k)}.$$

**Aufgabe 2.2.24.** Zeigen Sie, dass

$$\sin(nx) = \binom{n}{1} (\cos x)^{n-1} \sin x - \binom{n}{3} (\cos x)^{n-3} (\sin x)^3 + \binom{n}{5} (\cos x)^{n-5} (\sin x)^5 - \dots,$$

$$\cos(nx) = \cos^n x - \binom{n}{2} (\cos x)^{n-2} (\sin x)^2 + \binom{n}{4} (\cos x)^{n-4} (\sin x)^4 - \dots$$

### 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  $\Omega$  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

$$A = \text{“man hat genau 3 Nummern richtig getippt”}.$$

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

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

**Aufgabe 2.3.5.** Beim Lotto werden in einer Ziehung 6 Zahlen aus der Menge  $\{1, 2, \dots, 49\}$  ohne Zurücklegen gezogen. Jemand behauptet: bei der nächsten Ziehung der Lottozahlen ist es wahrscheinlicher, mindestens eine Zahl aus der letzten Ziehung zu erhalten als ausschließlich neue Zahlen. Stimmt die Behauptung?

## 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  $k_2$ , ..., 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

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

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

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

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 (mit jeweils 11 Schülern) 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. Unterscheidbarkeit könnte z.B. dadurch entstehen, 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.

Sind die 3 Mannschaften *ununterscheidbar* (z.B. wenn sie alle in derselben Liga spielen sollen), 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

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

**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

$A =$  “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

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

Nun werden wir verschiedene Beispiele von diskreten Zufallsvariablen betrachten.

## 2.5. Uniformverteilung

**Definition 2.5.1.** Eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt *gleichverteilt* (oder *uniformverteilt*, oder *Laplace-verteilt*) auf einer endlichen Menge  $\{y_1, \dots, y_n\} \subset \mathbb{R}$ , wenn

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

Eine Zufallsvariable ist also gleichverteilt, wenn sie  $n$  Werte annehmen kann und die Wahrscheinlichkeiten dieser  $n$  Werte gleich sind. 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 2.5.2.** Definition 2.5.1 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.

## 2.6. Bernoulli-Experimente und die Binomialverteilung

**Definition 2.6.1.** 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 2.6.2.** Eine Zufallsvariable  $X$  heißt *Bernoulli-verteilt* mit Parameter  $p \in [0, 1]$ , falls

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

Wir schreiben dann  $X \sim \text{Bern}(p)$ .

**Definition 2.6.3.** Ein *n-faches Bernoulli-Experiment* ist ein Bernoulli-Experiment, das  $n$ -mal unabhängig voneinander ausgeführt wurde.

**Beispiel 2.6.4.** Wir können eine (faire oder unfaire) Münze  $n$ -mal werfen und zum Beispiel “Kopf” als “Erfolg” auffassen.

**Beispiel 2.6.5.** 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  $p = \frac{1}{6}$ .

**Beispiel 2.6.6.** Wir stellen  $n$  zufälligen Personen jeweils eine Frage, die man mit “Ja” oder “Nein” beantworten kann.

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

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{für alle } k = 0, 1, \dots, n. \quad (2.6.1)$$

**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 (2.6.1).  $\square$

**Definition 2.6.8.** Eine Zufallsvariable  $X$ , die (2.6.1) erfüllt, heißt *binomialverteilt* mit Parametern  $n \in \mathbb{N}$  und  $p \in [0, 1]$ . Wir schreiben dann  $X \sim \text{Bin}(n, p)$ .

**Bemerkung 2.6.9.** 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’.

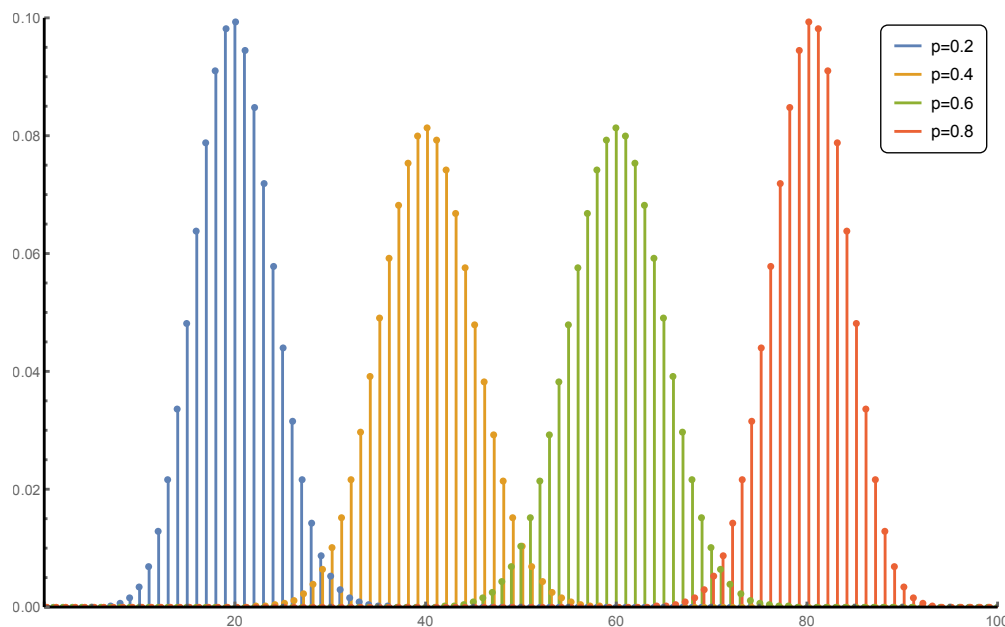


ABBILDUNG 1. Zähl-dichten der Binomialverteilungen mit  $n = 100$  und verschiedenen Werten von  $p$ .

**Satz 2.6.10.** 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  $\mathbb{E}X = \mathbb{E}X_1 + \dots + \mathbb{E}X_n = np$ . □

**Aufgabe 2.6.11.** Bestimmen Sie  $\sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}$ .

**Aufgabe 2.6.12.** Man würfelt mit einem fairen Würfel. Was ist wahrscheinlicher: mindestens 2 Sechsen in 12 Würfeln zu erzielen, oder mindestens 3 Sechsen in 18 Würfeln zu erzielen?

**Aufgabe 2.6.13.** Das Blut einer sehr großen Anzahl von Personen soll auf einen Krankheitserreger untersucht werden, wobei man annimmt, dass die Personen unabhängig voneinander mit jeweils der gleichen Wahrscheinlichkeit  $p \in (0, 1)$  infiziert sind. Für die Untersuchung stehen zwei Methoden zur Auswahl:

- (M1) Jede Probe wird einzeln untersucht.
- (M2) Die Hälften von je  $k$  Proben werden zusammengeschüttet und dann untersucht. Ein negatives Ergebnis bedeutet, dass der Erreger in keiner der Proben vorhanden ist. Bei einem positiven Testergebnis muss jede Probe einzeln überprüft werden.

Es soll nun der optimale Wert von  $k$  bestimmt werden.

- (a) Bestimmen Sie die erwartete Anzahl der nötigen Untersuchungen für jede der beiden Methoden.
- (b) Wann ist die Methode M2 vorzuziehen (Ungleichung für  $k$ )?
- (c) Bestimmen Sie für  $p = 0.001$  den optimalen Wert von  $k$  und die in diesem Fall erwartete Anzahl von Untersuchungen pro Person.

*Bemerkung.* Die Methode wurde während des zweiten Weltkrieges tatsächlich verwendet.

## 2.7. Poisson-Verteilung

**Definition 2.7.1.** Eine Zufallsvariable  $X$  hat *Poisson-Verteilung* mit Parameter (oder *Intensität*)  $\lambda > 0$ , wenn

$$\mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!} \text{ für alle } k = 0, 1, 2, \dots \quad (2.7.1)$$

Wir schreiben dann  $X \sim \text{Poi}(\lambda)$ .

**Bemerkung 2.7.2.** Die Wahrscheinlichkeiten in (2.7.1) 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 2.7.3.** 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 die Indexverschiebung  $m := k - 1$  durchgeführt.  $\square$

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

**Satz 2.7.4** (Poisson-Grenzwertsatz). Sei  $p_n \in (0, 1)$  eine Folge mit

$$\lim_{n \rightarrow \infty} np_n = \lambda \in (0, \infty). \quad (2.7.2)$$

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

$$\lim_{n \rightarrow \infty} \mathbb{P}[S_n = k] = e^{-\lambda} \frac{\lambda^k}{k!}. \quad (2.7.3)$$

**Beispiel 2.7.5.** 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 2.7.4.** 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 (2.7.2). Außerdem haben wir die folgende Formel benutzt:

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

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

**Beispiel 2.7.6.** 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 (2.7.1) 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

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

□

**Aufgabe 2.7.7.**  $10^7$  Personen spielen Lotto (6 aus 49). Jede Person tippt zufällig auf genau eine Kombination aus 6 Zahlen, wobei alle Kombinationen gleich wahrscheinlich seien. Die Tipps der verschiedenen Personen seien stochastisch unabhängig voneinander und es gebe keine Lieblingsnummern. Berechnen Sie die Wahrscheinlichkeit, dass mindestens eine Person 6 Richtige hat

- (1) exakt.
- (2) mit Hilfe der Poisson-Approximation.

**Beispiel 2.7.8.** Wie ist die Anzahl der Tore in einem Fussballspiel verteilt? Um diese Frage zu beantworten, stellen wir uns vor, dass ein Fussballspiel in eine große Zahl  $n$  kleiner Zeitabschnitte aufgeteilt ist. In jedem Zeitabschnitt kann ein Tor mit einer kleinen Wahrscheinlichkeit  $p$  fallen. Nehmen wir an, dass verschiedene Zeitabschnitte stochastisch unabhängig sind, so suggeriert der Poisson-Grenzwertsatz, dass die Anzahl der Tore im ganzen Spiel Poisson-verteilt sein muss. Der Parameter  $\lambda$  kann dabei als mittlere Anzahl der Tore pro Spiel interpretiert werden. Eine solche theoretische Überlegung muss natürlich an den realen Daten getestet werden. In der Saison 2018/2019 wurden  $T = 973$  Tore in  $N = 306$  Spielen geschossen. Die mittlere Anzahl der Tore pro Spiel betrug  $\lambda = T/N = 3.17$ . Können wir anhand dieser Daten die Anzahl der torlosen Spiele in dieser Saison schätzen? Dazu nehmen wir an, dass die Anzahl der Tore Poisson-verteilt mit Parameter  $\lambda = 3.17$  ist. Dann ist die Wahrscheinlichkeit, dass ein Spiel torlos ausgeht,  $e^{-\lambda}$ . Die erwartete Anzahl der torlosen Spiele ist dann  $Ne^{-\lambda} = 12.72$ . In Wirklichkeit gingen 17 Spiele torlos aus. Unser Ergebnis war also gar nicht so schlecht.

Nun beweisen wir eine verbesserte Version des Poisson-Grenzwertsatzes, in der wir Bernoulli-Variablen mit verschiedenen Erfolgswahrscheinlichkeiten  $p_1, \dots, p_n$  zulassen und eine Abschätzung an die Konvergenzgeschwindigkeit geben.

**Satz 2.7.9** (Quantitativer Poisson-Satz). Seien  $X_1, \dots, X_n$  unabhängige Zufallsvariablen mit  $X_i \sim \text{Bern}(p_i)$ , d. h.  $\mathbb{P}[X_i = 1] = p_i$  und  $\mathbb{P}[X_i = 0] = 1 - p_i$ . Dann gilt für  $S := X_1 + \dots + X_n$  und  $\lambda := p_1 + \dots + p_n$  die Abschätzung

$$\sum_{k=0}^{\infty} \left| \mathbb{P}[S = k] - e^{-\lambda} \frac{\lambda^k}{k!} \right| \leq 2 \sum_{i=1}^n p_i^2.$$

**Beispiel 2.7.10.** Für  $p_1 = \dots = p_n = \lambda/n$  gilt  $2 \sum_{i=1}^n p_i^2 = 2n \cdot \frac{\lambda^2}{n^2} = \frac{2\lambda^2}{n}$  und wir erhalten die Abschätzung

$$\sum_{k=0}^{\infty} \left| \mathbb{P}[S = k] - e^{-\lambda} \frac{\lambda^k}{k!} \right| \leq \frac{2\lambda^2}{n}.$$

Es sei bemerkt, dass die rechte Seite gegen 0 für  $n \rightarrow \infty$  konvergiert.

BEWEIS VON SATZ 2.7.9. In diesem Beweis werden wir die sogenannte Coupling-Methode verwenden, bei der sowohl die Bernoulli-Variablen als auch die approximierende Poisson-Variablen auf einem gemeinsamen Wahrscheinlichkeitsraum definiert werden, wodurch ein ‘‘Coupling’’ zwischen den Variablen hergestellt wird.

Zuerst halten wir ein  $i \in \{1, \dots, n\}$  fest und stellen ein ‘‘Coupling’’ zwischen  $X_i \sim \text{Bern}(p_i)$  und einer Poisson-Variablen  $Y_i \sim \text{Poi}(p_i)$  her. Dabei wollen wir erreichen, dass  $X_i = Y_i$  mit ‘‘großer’’ Wahrscheinlichkeit gilt. Der wohl einfachste Wahrscheinlichkeitsraum, auf dem man eine  $\text{Poi}(\lambda_i)$ -verteilte Zufallsvariable definieren kann, ist die Grundmenge  $\{0, 1, \dots\}$  mit der Wahrscheinlichkeitsfunktion

$$q_i(k) = e^{-\lambda_i} \frac{\lambda_i^k}{k!}, \quad k = 0, 1, \dots$$

Damit wir auf demselben Wahrscheinlichkeitsraum auch eine  $\text{Bern}(p_i)$ -verteilte Zufallsvariable definieren können, brauchen wir zwei Ereignisse mit Wahrscheinlichkeiten  $p_i$  und  $1 - p_i$ . Deshalb werden wir den Ausgang 0 (der eine Wahrscheinlichkeit von  $e^{-p_i}$  hat) in zwei verschiedene Ausgänge aufspalten, die wir mit  $0'$  und  $0''$  bezeichnen und denen wir Wahrscheinlichkeiten  $1 - p_i$  und  $e^{-p_i} - (1 - p_i)$  zuordnen.<sup>1</sup> Wir betrachten also die Grundmenge  $\Omega_i := \{0', 0'', 1, 2, \dots\}$  versehen mit der folgenden Wahrscheinlichkeitsfunktion  $p_i : \Omega \rightarrow [0, 1]$ :

$$p_i(0') = 1 - p_i, \quad p_i(0'') = e^{-p_i} - (1 - p_i), \quad p_i(k) = e^{-p_i} \frac{p_i^k}{k!}, \quad k \in \{1, 2, \dots\}.$$

Es sei bemerkt, dass sich die Wahrscheinlichkeiten tatsächlich zu 1 summieren.

Damit wir nun diese Konstruktion für alle  $i = 1, \dots, n$  auf einem gemeinsamen Wahrscheinlichkeitsraum durchführen können, betrachten wir den Produktraum  $\Omega = \Omega_1 \times \dots \times \Omega_n$  mit der  $\sigma$ -Algebra  $\mathcal{F} := \mathcal{P}(\Omega)$  und der Wahrscheinlichkeitsfunktion

$$p(a_1, \dots, a_n) := p_1(a_1) \cdot \dots \cdot p_n(a_n), \quad (a_1, \dots, a_n) \in \Omega.$$

Nun definieren wir die Zufallsvariablen  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  wie folgt:

$$X_i(a_1, \dots, a_n) := \begin{cases} 0, & \text{falls } a_i = 0', \\ 1, & \text{sonst.} \end{cases}$$

Da  $X_i$  nur von der Koordinate  $a_i$  abhängt und wegen der Produktstruktur der Funktion  $p$  sind  $X_1, \dots, X_n$  unabhängig. Es gilt außerdem

$$\mathbb{P}[X_i = 0] = p_i(0') = 1 - p_i, \quad \mathbb{P}[X_i = 1] = 1 - (1 - p_i) = p_i.$$

Somit ist  $X_i$  Bernoulli-verteilt mit Parameter  $p_i$ . Die approximierenden Zufallsvariablen  $Y_1, \dots, Y_n : \Omega \rightarrow \mathbb{R}$  werden wie folgt definiert:

$$Y_i(a_1, \dots, a_n) = \begin{cases} a_i, & \text{falls } a_i \in \{1, 2, \dots\}, \\ 0, & \text{falls } a_i = 0' \text{ oder } a_i = 0''. \end{cases}$$

Die Zufallsvariablen  $Y_i$  sind unabhängig und es gilt  $Y_i \sim \text{Poi}(p_i)$ , denn

$$\mathbb{P}[Y_i = 0] = p_i(0') + p_i(0'') = e^{-p_i}, \quad \mathbb{P}[Y_i = k] = p_i(k) = e^{-p_i} \frac{p_i^k}{k!}, \quad k \in \mathbb{N}.$$

<sup>1</sup>Die zweite Wahrscheinlichkeit ist größer als 0, denn es gilt die Ungleichung  $e^x \geq 1 + x$  für alle  $x \in \mathbb{R}$  (Übung).

Somit haben wir die Zufallsvariablen  $X_1, \dots, X_n, Y_1, \dots, Y_n$  auf einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  konstruiert, was man als ein Coupling der beiden Familien bezeichnet.

Nun behaupten wir, dass  $Y_i$  eine Approximation von  $X_i$  ist in dem Sinne, dass das Ereignis  $\{X_i = Y_i\}$  eine "große" Wahrscheinlichkeit hat. Dieses Ereignis tritt genau dann ein, wenn  $a_i \in \{0', 1\}$  ist. Es gilt somit

$$\mathbb{P}[X_i \neq Y_i] = 1 - p_i(0') - p_i(1) = 1 - (1 - p_i) - e^{-p_i} p_i = p_i(1 - e^{-p_i}) \leq p_i^2,$$

wobei wir wieder die Ungleichung  $1 - e^{-x} \leq x$  benutzt haben. Für die Zufallsvariablen  $S := X_1 + \dots + X_n$  und  $Y := Y_1 + \dots + Y_n \sim \text{Poi}(p_1 + \dots + p_n)$  gilt

$$\begin{aligned} \sum_{k=0}^{\infty} |\mathbb{P}[S = k] - \mathbb{P}[Y = k]| &= \sum_{k=0}^{\infty} |\mathbb{P}[S = k, Y \neq k] - \mathbb{P}[Y = k, S \neq k]| \\ &\leq \sum_{k=0}^{\infty} (\mathbb{P}[S = k, Y \neq k] + \mathbb{P}[Y = k, S \neq k]) \\ &= 2\mathbb{P}[S \neq Y]. \end{aligned}$$

Damit das Ereignis  $\{S \neq Y\}$  eintritt, muss mindestens eines der Ereignisse  $\{X_1 \neq Y_1\}, \dots, \{X_n \neq Y_n\}$  eintreten. Indem wir die Subadditivität der Wahrscheinlichkeit benutzen, erhalten wir

$$\sum_{k=0}^{\infty} |\mathbb{P}[S = k] - \mathbb{P}[Y = k]| \leq 2\mathbb{P}[S \neq Y] \leq 2 \sum_{i=1}^n \mathbb{P}[X_i \neq Y_i] \leq 2 \sum_{i=1}^n p_i^2.$$

Der Satz ist bewiesen. □

## 2.8. 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 2.8.1.** 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 2.8.2.** Die Wartezeit auf  $T$  auf den ersten “Erfolg” in einem unendlich oft wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit  $p$  ist wie folgt verteilt:

$$\mathbb{P}[T = k] = (1 - p)^{k-1}p \text{ für alle } k = 1, 2, \dots \quad (2.8.1)$$

**Bemerkung 2.8.3.** Eine Zufallsvariable  $T$ , die (2.8.1) mit einem  $p \in (0, 1]$  erfüllt, heißt *geometrisch verteilt* mit Parameter  $p$ .

**Notation 2.8.4.**  $T \sim \text{Geo}(p)$ .

**Bemerkung 2.8.5.** Die Wahrscheinlichkeiten in (2.8.1) 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 2.8.2.** 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 2.8.6.** 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 2.8.7.** 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 2.8.8** (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 2.8.9** (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. □

**Aufgabe 2.8.10.** Man würfelt mit einem fairen Würfel und stoppt nach der ersten 6. Wie groß ist die Wahrscheinlichkeit, dass die Anzahl der Würfe gerade ist?

**Aufgabe 2.8.11.** Man würfelt mit einem fairen Würfel bis man jede der 6 Augenzahlen mindestens einmal gewürfelt hat. Sei  $T$  die Anzahl der Würfe. Bestimmen Sie  $\mathbb{E}T$ .

**Aufgabe 2.8.12.** Zwei Personen werfen gleichzeitig je eine faire Münze und wiederholen dies, bis jemand zum ersten Mal “Kopf” geworfen hat. Hat der Gegner gleichzeitig “Zahl”

geworfen, so hat derjenige gewonnen, der “Kopf” geworfen hat, sonst gibt es Unentschieden. Mit welcher Wahrscheinlichkeit tritt ein Unentschieden auf?

**Aufgabe 2.8.13.** Bestimmen Sie für  $|q| < 1$  die Summe  $\sum_{k=1}^{\infty} k^3 q^k$ .

## 2.9. 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 2.9.1.** 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 2.9.2.** 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:

$$\mathbb{P}[T_r = k] = \binom{k-1}{r-1} p^r (1-p)^{k-r} \text{ für alle } k \in \{r, r+1, \dots\}. \quad (2.9.1)$$

**Bemerkung 2.9.3.** 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 2.9.4.** Eine Zufallsvariable  $T_r$ , die (2.9.1) mit einem  $r \in \mathbb{N}$  und einem  $p \in (0, 1]$  erfüllt, heißt *negativ binomialverteilt* mit Parametern  $r$  und  $p$ . Wir benutzen die Notation  $T_r \sim \text{NB}(r, p)$ .

**Beispiel 2.9.5.** Die geometrische Verteilung ist ein Spezialfall der negativen Binomialverteilung:  $\text{Geo}(p) = \text{NB}(1, p)$ .

**Bemerkung 2.9.6.** Negative Binomialverteilung wird auch Pascal- oder Polya-Verteilung genannt.

**Beweis von Satz 2.9.2.** 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:

$$\underbrace{\begin{matrix} ??? \dots \dots ? 1, \\ 123 \dots \dots k-1 k \end{matrix}}_{r-1 \text{ "Erfolge"}}$$

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 2.9.7.** Wir zeigen noch, dass die Summe aller Wahrscheinlichkeiten in (2.9.1) gleich 1 ist:

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

Eigentlich folgt das aus Satz 2.9.2. 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  $\alpha$  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:

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

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  $\alpha$  einen negativen Wert eingesetzt haben.

**Satz 2.9.8.** Für  $T_r \sim \text{NB}(r, p)$  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.  $\square$

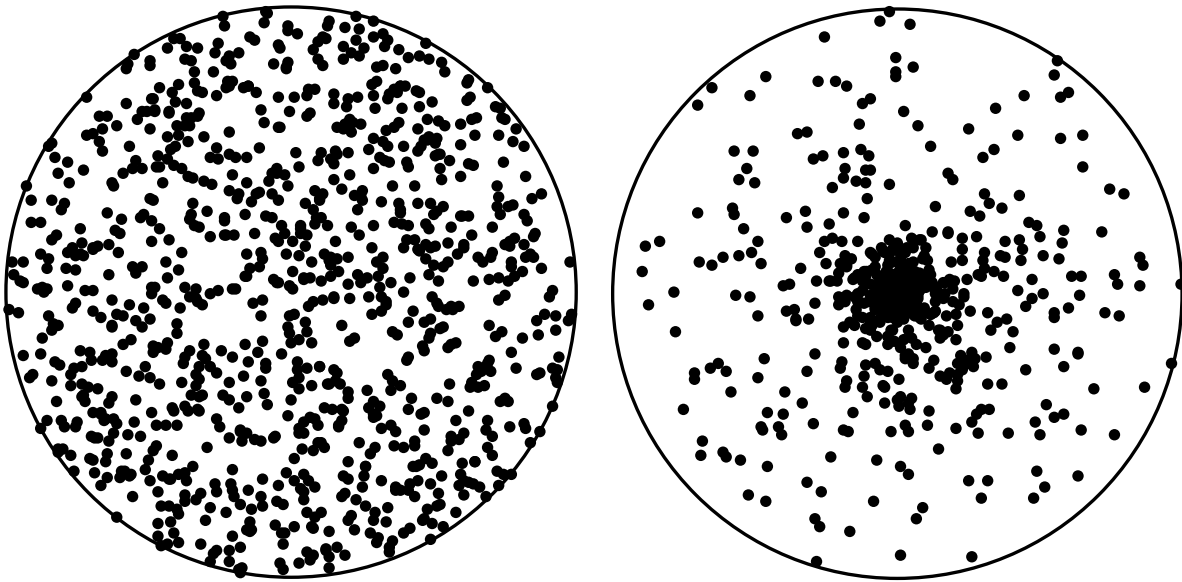
## KAPITEL 3

### Geometrische Wahrscheinlichkeiten

In diesem Kapitel werden wir eine Reihe von Beispielen betrachten, in denen Punkte, Strecken, Dreiecke oder andere geometrische Objekte in der Ebene oder im Raum zufällig konstruiert werden. Die Grundmengen in diesen Experimenten sind **überabzählbar**, d.h. man kann die Ausgänge nicht als eine Folge aufschreiben.

#### 3.1. Uniforme Verteilung

Stellen wir uns zwei Dartspieler vor, die den Mittelpunkt jeweils einer Dartscheibe treffen wollen. Zur Vereinfachung nehmen wir an, dass beide Spieler so gut seien, dass sie die Dartscheibe immer treffen. Nach jeweils 1000 Würfeln sehen die beiden Dartscheiben wie folgt aus:



Die Ergebnisse des ersten Spielers (links) entsprechen dem, was man intuitiv als “uniforme Verteilung” oder “Gleichverteilung” der Punkte auf dem Kreis bezeichnen würde. Beim zweiten Spieler (der offenbar besser wirft) sind die Punkte nicht uniform verteilt.

Nun versuchen wir, ein stochastisches Modell für *einen* Wurf des *ersten* Spielers aufzustellen. Die Grundmenge  $\Omega$  ist offenbar die Dartscheibe, also ein Kreis in der Ebene. Für jeden Punkt im Kreis ist die Wahrscheinlichkeit, diesen Punkt exakt zu treffen, gleich 0. Also gilt für jeden Ausgang  $\omega$

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

Die gleiche Formel gilt aber auch für den zweiten Werfer, auch wenn seine Würfe völlig anders verteilt sind! Wir stellen also fest, dass die Angabe der Wahrscheinlichkeiten aller *Ausgänge* eines Experiments im Allgemeinen noch keine zufriedenstellende Beschreibung des Experiments darstellt. Der Ausweg besteht darin, die Wahrscheinlichkeiten aller *Ereignisse*, also aller Teilmengen des Kreises, anzugeben. Beim ersten Werfer gehen wir davon aus, dass die Wahrscheinlichkeit, eine Teilmenge  $A$  zu treffen, proportional zum Flächeninhalt von  $A$  sein muss. Es gilt also

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

wobei  $\lambda(A)$  den Flächeninhalt von  $A$  bezeichnet. Es sei bemerkt, dass wir durch  $\lambda(\Omega)$  teilen, um sicherzustellen, dass  $\mathbb{P}[\Omega] = 1$  ist. Diese Formel ist die Definition der uniformen Verteilung. Sie ist offenbar analog zum Laplace-Ansatz, bei dem wir gefordert haben, dass

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

wobei  $\#A$  für die Anzahl der Elemente in  $A$  steht. Im Beispiel mit der Dartscheibe ergibt der Laplace-Ansatz allerdings keinen Sinn, denn  $\#\Omega = \infty$  und  $\#A$  kann ebenfalls unendlich sein, wodurch sich die Unbestimmtheit  $\infty/\infty$  ergeben würde.

**Beispiel 3.1.1.** Man berechne die Wahrscheinlichkeit, dass bei einem Wurf des ersten Spielers der Abstand vom Pfeil zum Mittelpunkt der Dartscheibe kleiner ist, als der Abstand zwischen dem Pfeil und dem Rand.

**Lösung.** Es sei  $R$  der Radius der Dartscheibe. Damit das besagte Ereignis eintritt, muss der Werfer einen Kreis vom Radius  $R/2$  treffen, der also das uns interessierende Ereignis  $A$  darstellt. Gehen wir nun von der Annahme der Gleichverteilung aus, so muss

$$\mathbb{P}[A] = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{\pi(R/2)^2}{\pi R^2} = \frac{1}{4}$$

gelten. □

Im Folgenden werden wir noch einige Beispiele betrachten, bei denen wir von der Gleichverteilungsannahme ausgehen. Ob eine solche Annahme gerechtfertigt ist, muss natürlich jedes Mal hinterfragt werden.

**Beispiel 3.1.2.** 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  $y \in [0, 1]$ . Das Paar  $(x, y)$  stellt den Ausgang des uns interessierenden Experiments dar. Als Grundmenge

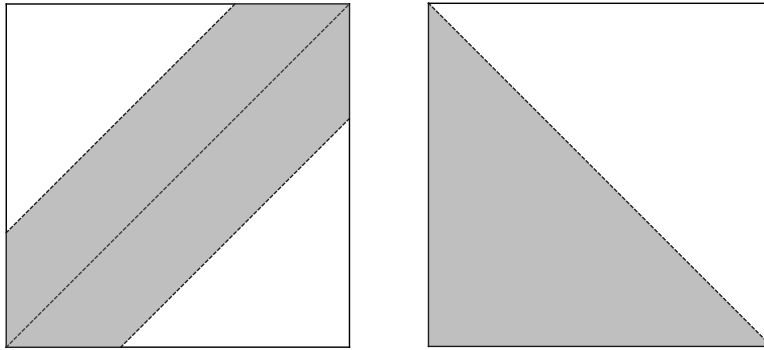


ABBILDUNG 1. Links: Skizze zu Beispiel 3.1.2. Die schattierte Fläche stellt das Ereignis  $A = \text{“Freunde treffen sich”}$  dar. Rechts: Skizze zu Beispiel 3.1.3. Die schattierte Fläche stellt das Ereignis  $A = \{x + y < 1\}$  dar.

können wir somit das Einheitsquadrat betrachten:

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

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

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

welche auf Abbildung 1 (links) zu sehen ist. Gehen wir nun davon aus, dass der Punkt  $(x, y)$  gleichverteilt auf  $[0, 1]^2$  ist, so 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 haben wir benutzt, dass das Komplement von  $A$  aus zwei rechtwinkligen Dreiecken mit Flächeninhalt von jeweils  $\frac{2}{9}$  besteht.  $\square$

**Beispiel 3.1.3.** 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:

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

Es handelt sich um ein Dreieck, das auf Abbildung 1 (rechts) zu sehen ist. Von einem idealen Zufallsgenerator erwartet man, dass der Punkt  $(x, y)$  uniformverteilt auf  $[0, 1]^2$  ist. Gehen

wir von dieser Annahme aus, so erhalten wir

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

**Aufgabe 3.1.4.** Ein Zufallsgenerator erzeugt zwei Zufallszahlen  $X$  und  $Y$ , die auf dem Intervall  $[0, 1]$  uniformverteilt und unabhängig voneinander sind. Für ein vorgegebenes  $a \in [0, 2]$  bestimmen Sie die Wahrscheinlichkeit, dass  $X + Y \leq a$ .

**Aufgabe 3.1.5.** Ein Zufallsgenerator erzeugt zwei Zufallszahlen  $X$  und  $Y$ , die auf dem Intervall  $[0, 1]$  uniformverteilt und unabhängig voneinander sind. Bestimmen Sie die Wahrscheinlichkeit, dass  $X > Y^2$ .

**Aufgabe 3.1.6.** An einem Stab der Länge 1 werden zufällig, unabhängig voneinander und uniformverteilt zwei Stellen  $X$  und  $Y$  markiert. An diesen Stellen wird der Stab durchgesägt. Mit welcher Wahrscheinlichkeit lässt sich aus den so gewonnenen Stücken ein Dreieck bilden? *Hinweis:* Damit man aus drei Strecken der Längen  $a, b, c$  ein Dreieck bilden kann, ist es notwendig, dass  $a + b > c$ . Diese Bedingung ist aber nicht hinreichend.

**Aufgabe 3.1.7.** An einem Stab der Länge 1 wird zuerst eine Stelle  $X$  zufällig und uniform verteilt markiert. Der Stab wird dann an dieser Stelle durchgesägt. Danach wird das linke Teilstück (das Länge  $X$  hat) an einer zufälligen und uniform verteilten Stelle markiert und noch einmal durchgesägt. Mit welcher Wahrscheinlichkeit lässt sich aus den so gewonnenen Stücken ein Dreieck bilden?

**Aufgabe 3.1.8.** In einer Wand befindet sich ein äußerlich nicht sichtbares Drahtgeflecht aus 4 mm starkem Draht, das Rechtecke mit den Seitenlängen 50 mm und 80 mm (gemessen von Drahtmitte zu Drahtmitte) bildet. An einer rein zufällig ausgewählten Stelle wird mit einem Bohrer ein Loch mit 10 mm Durchmesser in die Wand gebohrt. Mit welcher Wahrscheinlichkeit wird dabei das Drahtgeflecht getroffen?

**Aufgabe 3.1.9.** In einem Quadrat  $ABCD$  wird zufällig und uniform verteilt ein Punkt  $X$  ausgewählt. Bestimmen Sie die Wahrscheinlichkeit, dass das Dreieck  $ABX$  spitzwinklig ist.

### 3.2. Das Buffon'sche Nadelproblem

Das folgende Problem wurde 1777 von Comte de Buffon, einem französischen Naturforscher, vorgeschlagen. Auf ein liniertes Papier wird "völlig zufällig" eine Nadel geworfen. Der Abstand zwischen aufeinanderfolgenden parallelen Geraden sei o.E.d.A. gleich 1. Die Länge der Nadel bezeichnen wir mit  $\ell$  und nehmen an, dass  $\ell < 1$ , so dass die Nadel höchstens eine Gerade kreuzen kann. Gefragt wird nach der Wahrscheinlichkeit des Ereignisses

$A :=$  "Die Nadel kreuzt eine Gerade".

Wir behaupten, dass

$$\mathbb{P}[A] = \frac{2}{\pi}\ell.$$

**Abschweifung.** Überraschenderweise enthält diese Formel die Kreiszahl  $\pi$ . Dadurch kann sie als Methode zur approximativen Berechnung von  $\pi$  verwendet werden. Dazu lassen wir  $N$  Nadeln auf das Papier zufällig fallen, s. Abbildung 2. Die Anzahl der Nadeln, die eine

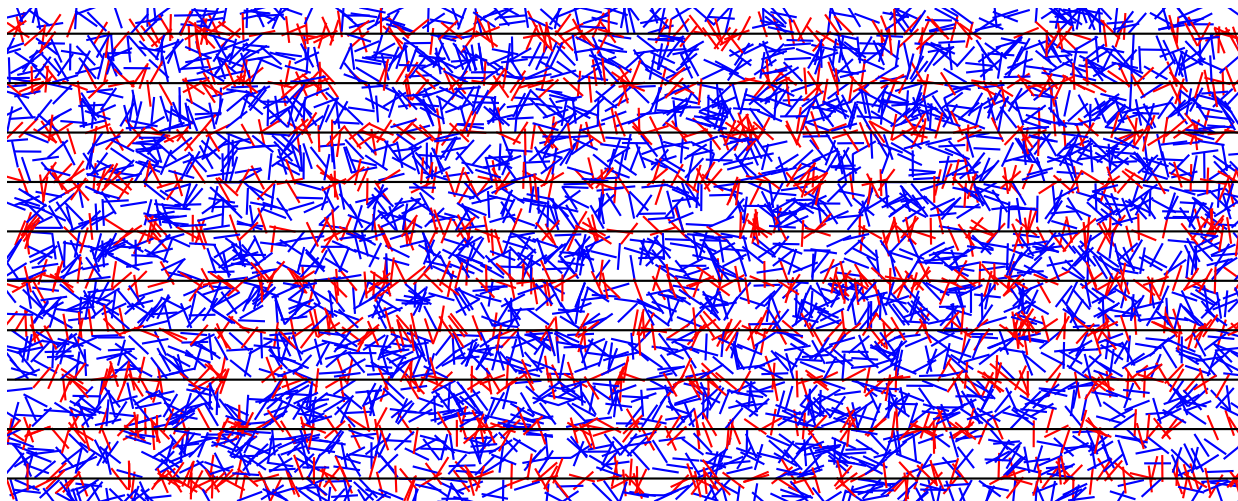


ABBILDUNG 2. Buffon'sches Problem mit 5000 Nadeln. Nadeln, die eine Gerade kreuzen, werden in rot gezeigt. Alle anderen Nadeln werden in blau gezeigt.

Gerade kreuzen, sei mit  $n$  bezeichnet. Bei einem großen  $N$  sollte wegen der frequentistischen Interpretation der Wahrscheinlichkeit

$$\frac{n}{N} \approx \frac{2}{\pi} \ell$$

gelten, woraus sich eine approximative Formel für die Zahl  $\pi$  ergibt:

$$\pi \approx \frac{N}{n} \cdot 2\ell.$$

In einer Computersimulation ließen wir  $N = 10^7$  Nadeln der Länge  $\ell = 1/2$  auf das Papier fallen, davon kreuzten  $k = 3182613$  eine Gerade. Das entspricht der Approximation  $\pi \approx 3.14207$ . Die Konvergenzgeschwindigkeit dieser Methode kann mit dem zentralen Grenzwertsatz (der später behandelt wird) abgeschätzt werden und ist sehr langsam im Vergleich zu anderen, deutlich effizienteren Methoden zur Berechnung von  $\pi$ .

**Lösung des Buffon'schen Problems.** Beim Nadelproblem ist die Wahl der Grundmenge nicht ganz trivial. Ausgänge des Experiments sind mögliche Positionen der Nadel. Die Position der Nadel kann z.B. durch die Angabe ihres Mittelpunktes und ihrer Richtung angegeben werden. Jeder Punkt in der Ebene kommt als ein möglicher Mittelpunkt in Frage. Eine Uniformverteilung auf der Ebene existiert allerdings nicht, da der Flächeninhalt der Ebene unendlich ist, was die Definition der uniformen Verteilung absurd macht.

Wir werden deshalb anders vorgehen. Wir betrachten den "Korridor" zwischen zwei aufeinanderfolgenden Geraden  $L_1$  und  $L_2$ , in dem der Mittelpunkt der Nadel landet, und beschreiben die Position des Mittelpunktes der Nadel durch die Angabe des vertikalen Abstandes  $y$  zu  $L_1$ , siehe Abbildung 3. Dabei denken wir uns die Geraden  $L_1$  und  $L_2$  als horizontal, so dass  $L_1$  unter  $L_2$  liegt. Die horizontale Koordinate des Mittelpunktes der Nadel spielt für das Kreuzen oder Nichtkreuzen keine Rolle, weshalb wir sie in der Modellierung gar nicht berücksichtigen. Die möglichen Werte von  $y$  liegen zwischen 0 und 1. Außerdem bezeichnen

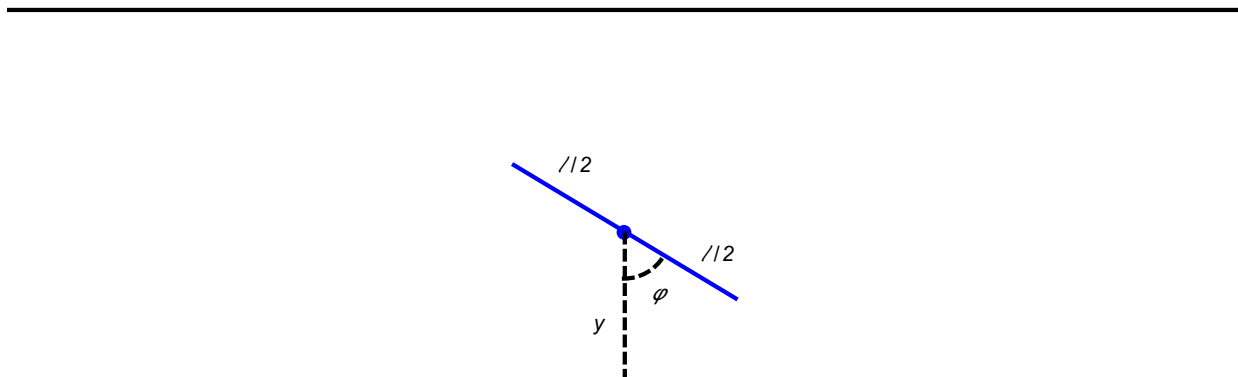


ABBILDUNG 3. Modellierung im Buffon'schen Problem.

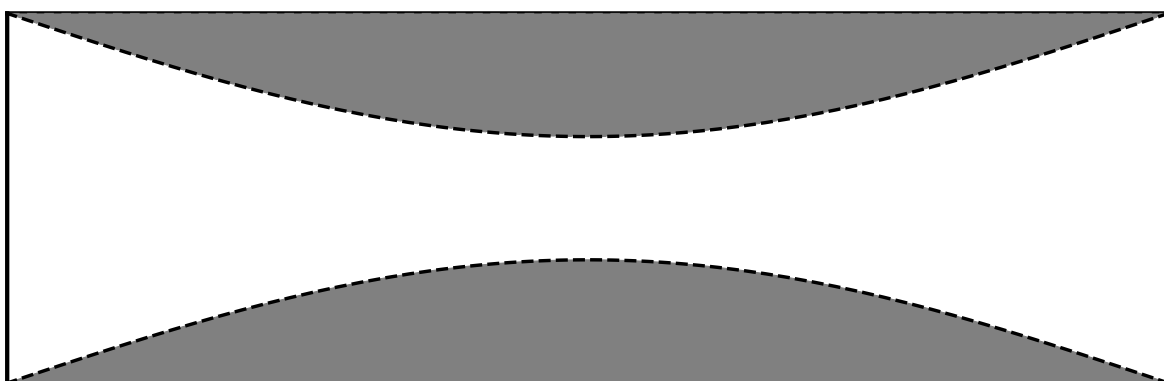


ABBILDUNG 4. Grundmenge im Buffon'schen Problem.

wir mit  $\varphi$  den Winkel zwischen der vertikalen Richtung und der Nadel. Die möglichen Werte von  $\varphi$  liegen zwischen  $-\frac{\pi}{2}$  und  $\frac{\pi}{2}$ .

Die Grundmenge des Experiments ist also das Rechteck

$$\Omega = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times (0, 1) = \left\{(\varphi, y) : \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), y \in (0, 1)\right\}.$$

Wir werden in dieser Lösung davon ausgehen, dass der Punkt  $(\varphi, y)$  uniformverteilt auf  $\Omega$  ist. Würde man die Nadel aus einer kleinen Höhe fallen lassen, so wäre diese Annahme sicherlich falsch. Bei einer Höhe, die sehr groß im Vergleich zum Abstand zwischen den aufeinanderfolgenden Linien ist, erscheint die Annahme der Uniformverteilung plausibel.

Wir beschreiben nun das uns interessierende Ereignis  $A :=$  "die Nadel kreuzt eine Gerade" als Teilmenge von  $\Omega$ . Dazu stellen wir fest, dass die Nadel die untere Gerade  $L_1$  genau dann kreuzt, wenn

$$\frac{\ell}{2} \cos \varphi > y.$$



Analog überzeugt man sich, dass die obere Gerade  $L_2$  genau dann gekreuzt wird, wenn

$$\frac{\ell}{2} \cos \varphi > 1 - y.$$

Das Ereignis  $A$  entspricht also der folgenden Teilmenge von  $\Omega$ :

$$A = \left\{ \left( -\frac{\pi}{2}, \frac{\pi}{2} \right) \times (0, 1) : \frac{\ell}{2} \cos \varphi > y \text{ oder } \frac{\ell}{2} \cos \varphi > 1 - y \right\}.$$

Die Menge  $A$  ist auf Abbildung 4 zu sehen. Sie besteht aus zwei Komponenten, die sich wegen der Annahme  $\ell < 1$  nicht schneiden. Die Fläche von  $A$  ist

$$\lambda(A) = 2 \int_{-\pi/2}^{\pi/2} \frac{\ell}{2} \cos \varphi \, d\varphi = \ell \int_{-\pi/2}^{\pi/2} \cos \varphi \, d\varphi = 2\ell.$$

Für die Wahrscheinlichkeit von  $A$  erhalten wir die Formel

$$\mathbb{P}[A] = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{2\ell}{\pi},$$

wobei wir  $\lambda(\Omega) = \pi$  benutzt haben. □

**Abschweifung.** Wir werden nun eine zweite Lösung des Buffon-Problems vorstellen. Diese ist deutlich eleganter<sup>1</sup> als die erste, benutzt jedoch den Begriff des Erwartungswerts, der erst später eingeführt werden soll. Für eine Nadel der Länge  $\ell > 0$  bezeichnen wir mit  $E(\ell)$  die erwartete Anzahl der Geraden, die diese kreuzt. Im Spezialfall wenn  $\ell < 1$  kann höchstens eine Gerade gekreuzt werden, so dass  $E(\ell)$  mit der Wahrscheinlichkeit übereinstimmt, dass eine Gerade gekreuzt wird. Zerlegen wir eine Nadel der Länge  $\ell = x + y$  in zwei Teilnadeln der Längen  $x$  und  $y$ , so entspricht die erwartete Anzahl der Kreuzungen der Gesamtnadel der Summe der erwarteten Kreuzungszahlen der beiden Teilnadeln:

$$E(x + y) = E(x) + E(y).$$

Das gilt für beliebige  $x > 0$  und  $y > 0$ . Somit erfüllt die Funktion  $E(\ell)$  die Cauchy-Funktionalgleichung. Außerdem ist die Funktion  $E(x)$  monoton nichtfallend. Aus der Beschreibung der Lösungen der Cauchy-Funktionalgleichung folgt, dass  $E(\ell)$  eine lineare Funktion sein muss, d.h.

$$E(\ell) = \ell E(1)$$

für alle  $\ell > 0$ .

Es bleibt nur noch, den Koeffizienten  $c := E(1)$  zu bestimmen. Zu diesem Zweck betrachten wir zuerst eine polygonale Nadel, die aus Teilen der Längen  $\ell_1, \dots, \ell_n$  besteht. Für diese ist die erwartete Anzahl an Kreuzungen offenbar  $c\ell_1 + \dots + c\ell_n$ , also  $c$  multipliziert mit der Länge der Nadel. Indem wir nun eine beliebig gekrümmte Nadel durch polygonale Nadeln approximieren, kommen wir zum Schluss, dass die erwartete Anzahl an Kreuzungen für eine solche Nadel der Länge  $\ell$  gleich  $c\ell$  ist. Nun betrachten wir eine Nadel, die die Form eines Kreises vom Radius  $1/2$  hat. Für diese ist die erwartete Anzahl der Kreuzungen besonders einfach zu bestimmen, denn eine solche Nadel kreuzt mit Wahrscheinlichkeit 1 genau eine

<sup>1</sup>Diese Lösung stammt aus "Das Buch der Beweise" von Martin Aigner und Günter Ziegler, Kapitel 26

Gerade. Auf der anderen Seite ist diese Anzahl  $c\pi/2$ , denn die Länge des Kreises ist  $\pi/2$ . Es gilt also  $c\pi/2 = 1$ , woraus sich  $c = 2/\pi$  ergibt! Wir haben somit gezeigt, dass

$$E(\ell) = \frac{2}{\pi}\ell.$$

Dieses Ergebnis gilt für eine beliebig gekrümmte Nadel der Länge  $\ell$ . Für eine Nadel der Länge  $\ell < 1$  ist höchstens eine Kreuzung möglich und das Ergebnis bedeutet, dass die Wahrscheinlichkeit einer Kreuzung  $\frac{2}{\pi}\ell$  ist.

**Bemerkung 3.2.1.** Für eine Nadel der Länge  $\ell > 1$  schneiden sich die beiden Komponenten von  $A$ . Die Formel für  $\mathbb{P}[A]$  ist nicht mehr so einfach und kann im Buch der Beweise, Kapitel 26, gefunden werden.

### 3.3. Das Bertrand'sche Paradoxon

In einem Kreis mit Radius 1 wählen wir zufällig eine Sehne und fragen nach der Wahrscheinlichkeit, dass die Sehne mindestens so lang wie die Seite des einbeschriebenen gleichseitigen Dreiecks ist. Bezeichnen wir die Länge der Sehne mit  $L$ , so interessiert uns das Ereignis  $A := \{L > \sqrt{3}/2\}$ . Es stellt sich heraus, dass dieses harmlos klingende Problem keine eindeutige Lösung besitzt, weshalb es als "Paradoxon" bezeichnet wird. Die entscheidende Schwierigkeit besteht darin, dass es unklar ist, was man unter einer "zufälligen Sehne" verstehen soll. Es können mehrere natürliche Modelle für die Wahl der zufälligen Sehne vorgeschlagen werden, die zu verschiedenen Wahrscheinlichkeiten für das uns interessierende Ereignis führen.

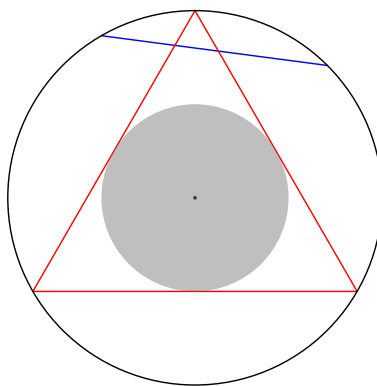


ABBILDUNG 5. Zum Paradoxon von Bertrand: Eine zufällige Sehne und ein einbeschriebenes gleichseitiges Dreieck.

**Modell 1.** Wir konstruieren eine Sehne indem wir ihren Mittelpunkt konstruieren. Wir wählen nämlich einen Punkt  $P = (x, y)$  zufällig und uniform verteilt im Einheitskreis und betrachten dann die Sehne, deren Mittelpunkt  $P$  ist. Eine solche Sehne ist eindeutig definiert, es sei denn  $P$  stimmt mit dem Mittelpunkt des Kreises überein, was allerdings Wahrscheinlichkeit 0 hat und vernachlässigt werden kann. Die Grundmenge ist also der Einheitskreis

$$\Omega = \{(x, y) : x^2 + y^2 \leq 1\}.$$

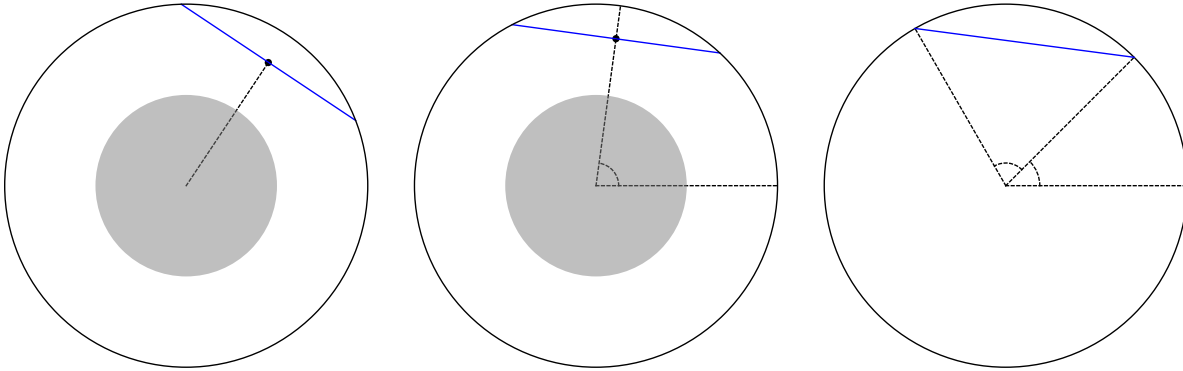


ABBILDUNG 6. Die drei Modelle für die zufällige Sehne aus dem Paradoxon von Bertrand.

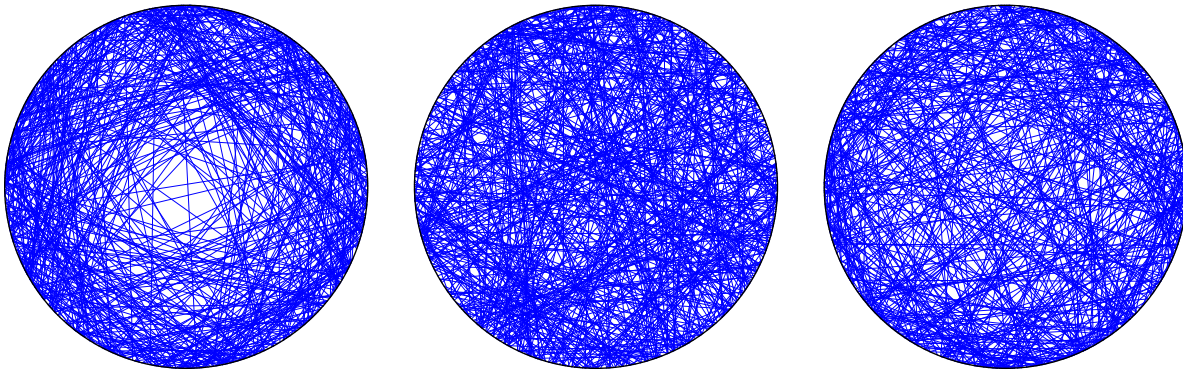


ABBILDUNG 7. Die drei Modelle aus dem Paradoxon von Bertrand. Jedes Bild zeigt 500 Sehnen, die gemäß dem entsprechenden Modell erzeugt wurden.

Da jedes einbeschriebene gleichseitige Dreieck die senkrecht auf den Seiten stehenden Radien halbiert, tritt das Ereignis  $L > \sqrt{3}/2$  genau dann ein, wenn  $P$  innerhalb des Kreises vom Radius  $1/2$  um den Ursprung liegt, d.h.

$$A = \{(x, y) : x^2 + y^2 \leq 1/2\}.$$

Für die Wahrscheinlichkeit von  $A$  erhalten wir

$$\mathbb{P}[A] = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{\pi/4}{\pi} = \frac{1}{4}.$$

**Modell 2.** Wir konstruieren die Sehne indem wir den Radius, der sie senkrecht in ihrem Mittelpunkt schneidet, sowie den Abstand  $R$  dieses Schnittpunktes vom Kreismittelpunkt angeben. Der Radius kann eindeutig durch den Winkel  $\varphi \in [0, 2\pi)$  parametrisiert werden, und der Abstand zum Schnittpunkt durch die Zahl  $r \in (0, 1)$ . Wir werden sowohl den Abstand als auch den Winkel uniform verteilt auf dem entsprechenden Intervall auswählen. Der Wahrscheinlichkeitsraum ist also das ‘‘Rechteck’’

$$\Omega = [0, 2\pi) \times (0, 1).$$

Das gesuchte Ereignis  $A$  tritt genau dann ein, wenn  $r < 1/2$ , also

$$A = [0, 2\pi) \times (0, 1/2).$$

Für die Wahrscheinlichkeit von  $A$  folgt, dass

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

**Modell 3.** Hier parametrisieren wir die Sehne durch den Radius zum ersten (gegen den Uhrzeigersinn) Punkt der Sehne, sowie durch den Öffnungswinkel. Der Radius wird durch einen Winkel  $\varphi \in [0, 2\pi)$  parametrisiert, während der Öffnungswinkel Winkel  $\psi$  Werte im Intervall  $(0, \pi)$  annehmen kann. Das Paar  $(\varphi, \psi)$  soll uniform verteilt auf dem Rechteck

$$\Omega := [0, 2\pi) \times (0, \pi)$$

ausgewählt werden. Die Sehne ist länger als die Seite des einbeschriebenen gleichseitigen Dreiecks genau dann, wenn der Öffnungswinkel größer als  $2\pi/3$  ist. Somit gilt

$$A = [0, 2\pi) \times \left(\frac{2\pi}{3}, \pi\right).$$

Für die Wahrscheinlichkeit von  $A$  erhalten wir diesmal den Wert

$$\mathbb{P}[A] = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{2\pi \cdot \frac{\pi}{3}}{2\pi \cdot \pi} = \frac{1}{3}.$$

Die obigen Lösungen unterscheiden sich dadurch, dass wir drei verschiedene Definitionen der zufälligen Sehne (bzw. drei verschiedene “Verteilungen” auf der Menge aller Sehnen) benutzt haben. Alle drei Modelle scheinen mehr oder weniger natürlich. Das Fazit von Bertrand lautet: “Keine der drei Lösungen ist falsch, allerdings ist auch keine ganz richtig. Die Frage ist schlecht gestellt”.

Ähnliche Schwierigkeiten treten beim sogenannten Sylvester-Vierpunkteproblem auf. Dieses fragt nach der Wahrscheinlichkeit, dass 4 zufällige Punkte in einer Ebene ein Viereck bilden (eine andere Möglichkeit wäre es, dass einer der Punkte innerhalb des von den drei anderen Punkten aufgespannten Dreiecks liegt). Auch hier ist es nicht klar, was man unter einer zufälligen Wahl von 4 Punkten verstehen soll. Man könnte die Punkte uniform verteilt in einem Quadrat, einem Dreieck oder einem Kreis auswählen. Man kann zeigen, dass die entsprechenden Wahrscheinlichkeiten  $\frac{25}{36}$ ,  $\frac{2}{3}$  und  $1 - \frac{35}{12\pi^2}$  sind. Es sollte aber hervorgehoben werden, dass es keine Uniformverteilung in der ganzen Ebene gibt.

## Axiome der Wahrscheinlichkeitstheorie und Maßtheorie

Dieses Kapitel bedarf einer Überarbeitung. Es kann in zwei Kapitel aufgespalten werden: “Das Volumenproblem” und “Axiome der Wtheorie”. Gute Referenz zu Lebesgue-Integralen (mit der Erklärung der Caratheodory-Konstruktion) ist Kurtz, Swartz: Theories of Integration: Integrals of Riemann, Lebesgue,...

### 4.1. Axiome der Wahrscheinlichkeitstheorie

Auf dem zweiten internationalen Mathematikerkongress, das im Jahre 1900 in Paris stattfand, stellte der deutsche Mathematiker David Hilbert eine Liste aus 23 Problemen, die die Entwicklung der Mathematik im 20. Jahrhundert bestimmen sollten. Das sechste Problem aus seiner Liste lautete “Mathematische Behandlung der Axiome der Physik” und fragte insbesondere nach der axiomatischen Behandlung der Mechanik und der Wahrscheinlichkeitstheorie. Eine Axiomatisierung der gesamten Physik ist bis heute nicht in Sicht. Hier werden wir uns mit der Axiomatisierung der Wahrscheinlichkeitsrechnung befassen, die im Jahre 1933 dem russischen Mathematiker Andrei Kolmogorov gelang.

Unter einer axiomatischen Behandlung der Wahrscheinlichkeitsrechnung versteht man Folgendes. Wir wollen eine möglichst knappe Liste einfacher Eigenschaften (oder Axiome) aufstellen, die für jedes Zufallsexperiment gelten. Diese Eigenschaften sollen “offensichtlich” sein. Alle weiteren Eigenschaften des Zufalls sollen dann als logische Konsequenzen der Axiome hergeleitet werden können.

Zuallererst müssen wir festlegen, welche Daten für eine eindeutige Beschreibung eines Zufallsexperiments nötig sind. Auf jeden Fall wird die Grundmenge  $\Omega$  benötigt, die eine beliebige nichtleere Menge sein darf.

**Schritt 1: Diskrete Wahrscheinlichkeitsräume.** Ist  $\Omega$  höchstens abzählbar, so können wir jedem Ausgang  $\omega \in \Omega$  seine Wahrscheinlichkeit  $p(\omega)$  zuordnen. Dabei ist  $p : \Omega \rightarrow [0, 1]$  eine Funktion mit

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

Das Paar  $(\Omega, p)$  beschreibt das Zufallsexperiment eindeutig in dem Sinne, dass wir die Wahrscheinlichkeiten aller Ereignisse  $A \subset \Omega$  ausrechnen können, nämlich

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

Experimente mit höchstens abzählbarer Grundmenge heißen diskret.

**Schritt 2: Naive Axiome.** Ist allerdings  $\Omega$  überabzählbar, so gilt typischerweise  $p(\omega) = 0$  für alle  $\omega \in \Omega$ . Das haben wir z.B. im Kapitel über geometrische Wahrscheinlichkeiten gesehen. Durch die Angabe von  $p(\omega)$  wird das Experiment nicht eindeutig beschrieben und wir

brauchen mehr Daten. Der Ausweg besteht darin, die Wahrscheinlichkeiten aller *Ereignisse* (und nicht nur aller Ausgänge) anzugeben.

Mit  $\mathcal{P}(\Omega)$  bezeichnen wir die *Potenzmenge* von  $\Omega$ , also die Menge aller Teilmengen von  $\Omega$ .

Wir bezeichnen mit  $\mathbb{P}[A]$  die Wahrscheinlichkeit des Ereignisses  $A$ .

**Beispiel 4.1.1.** Ist  $\Omega$  endlich mit  $n$  Elementen, so besteht die Potenzmenge  $\mathcal{P}(\Omega)$  aus  $2^n$  Elementen.

**Aufgabe 4.1.2.** Benennen Sie alle Elemente von  $\mathcal{P}(\emptyset)$ .

**Naive Axiome der Wahrscheinlichkeitstheorie.** Jedes Zufallsexperiment wird durch ein Paar  $(\Omega, \mathbb{P})$  beschrieben, wobei  $\Omega$  eine beliebige nichtleere Menge ist und  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  eine Funktion, die den folgenden Anforderungen genügt:

- (a) Normiertheit:  $\mathbb{P}[\Omega] = 1$ .
- (b)  $\sigma$ -Additivität: Für beliebige paarweise disjunkte Mengen  $A_1, A_2, \dots \subset \Omega$  gilt

$$\mathbb{P}[A_1 \cup A_2 \cup \dots] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \dots$$

Leider können diese Axiome nicht als Grundlage für die Wahrscheinlichkeitstheorie dienen, was wir In Abschnitt????? zeigen werden.

### Schritt 3: Axiome von Kolmogorov.

**Axiome von Kolmogorov.** Jedes Zufallsexperiment wird durch ein Tripel  $(\Omega, \mathcal{F}, \mathbb{P})$  beschrieben. Dabei haben die einzelnen Komponenten des Tripels die folgenden Eigenschaften:

- $\Omega$  ist eine beliebige nichtleere Menge.
- $\mathcal{F} \subset \mathcal{P}(\Omega)$  ist eine  $\sigma$ -**Algebra**, d.h. ein System von Teilmengen von  $\Omega$  mit folgenden Eigenschaften:
  - $\Omega \in \mathcal{F}$ ;
  - Für jedes  $A \in \mathcal{F}$  gilt auch  $A^c \in \mathcal{F}$ ;
  - Für beliebige Mengen  $A_1, A_2, \dots \in \mathcal{F}$  gilt auch  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ .

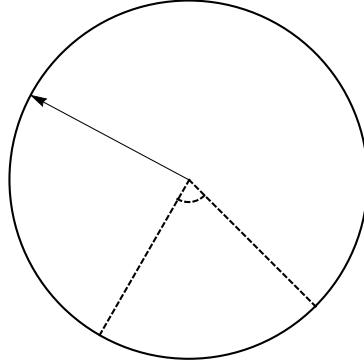
Gehört eine Menge  $A$  zu  $\mathcal{F}$ , so nennen wir  $A$  ein **Ereignis** oder eine **messbare Menge**.

- $\mathbb{P}$  ist ein **Wahrscheinlichkeitsmaß** auf  $(\Omega, \mathcal{F})$ , d.h.  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  ist eine Funktion mit den folgenden Eigenschaften:
  - (a) Normiertheit:  $\mathbb{P}[\Omega] = 1$ .
  - (b)  $\sigma$ -Additivität: Für beliebige **paarweise disjunkte Ereignisse**  $A_1, A_2, \dots \in \mathcal{F}$  gilt

$$\mathbb{P}[A_1 \cup A_2 \cup \dots] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \dots$$

## 4.2. Existenz nichtmessbarer Mengen

Wir betrachten dazu ein Zufallsexperiment, bei dem ein Glücksrad gedreht wird. Nach mehreren Runden stoppt der Zeiger und zeigt in eine zufällige Richtung. Wir möchten ein mathematisches Modell für dieses Experiment aufstellen.



Als Grundmenge können wir den Einheitskreis in der Ebene wählen:

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

Jeder Menge  $A \subset \Omega$  ordnen wir die Wahrscheinlichkeit  $\mathbb{P}[A]$  zu, dass der Zeiger auf einen Punkt aus dieser Menge zeigt. Es erscheint plausibel, dass die Funktion  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  die folgenden Eigenschaften haben muss:

- (a) Normiertheit:  $\mathbb{P}[\Omega] = 1$ .
- (b)  $\sigma$ -Additivität: Für beliebige paarweise disjunkte Mengen  $A_1, A_2, \dots \subset \Omega$  gilt

$$\mathbb{P}[A_1 \cup A_2 \cup \dots] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \dots$$

- (c) Rotationsinvarianz: Die Wahrscheinlichkeit einer beliebigen Menge  $A \subset \Omega$  ändert sich nicht, wenn man die Menge um einen beliebigen Winkel dreht.

Der nächste Satz kommt wie ein Donnerschlag aus heiterem Himmel.

**Satz 4.2.1** (Vitali, 1905). Eine Funktion  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ , die den Anforderungen (a), (b), (c) genügt, existiert nicht.

**BEWEIS.** Wir sagen, dass zwei Punkte  $x$  und  $y$  auf dem Einheitskreis äquivalent sind, wenn der Öffnungswinkel zwischen diesen Punkten ein *rationales* Vielfaches von  $2\pi$  ist. Das heißt, wir definieren  $x \sim y$ , falls es ein  $q \in \mathbb{Q}$  gibt mit  $x = T_{2\pi q}y$ , wobei  $T_{2\pi q} : \Omega \rightarrow \Omega$  eine Rotation um den Winkel  $2\pi q$  bezeichnet. Man rechnet leicht nach, dass  $\sim$  eine Äquivalenzrelation ist. Somit zerfällt der Kreis  $\Omega$  in eine Vereinigung von disjunkten Äquivalenzklassen. Dabei besteht die Äquivalenzklasse von  $z$  aus allen Punkten der Form  $T_{2\pi q}z$ ,  $q \in \mathbb{Q}$ . Nun benutzen wir das Auswahlaxiom: Aus jeder Äquivalenzklasse wählen wir genau einen Repräsentanten aus. Sei  $V$  die aus diesen Repräsentanten gebildete Menge  $V$ . Der Schnitt von  $V$  mit jeder Äquivalenzklasse besteht aus exakt einem Punkt. Rotieren wir die Menge  $V$  um alle beliebigen Winkel  $2\pi q$  mit  $q \in \mathbb{Q}$ , so erhalten wir disjunkte Mengen, die wir mit  $V_{2\pi q}$  bezeichnen

und die eine disjunkte Zerlegung von  $\Omega$  bilden:

$$\Omega = \bigcup_{q \in \mathbb{Q}} V_{2\pi q}, \quad V_{2\pi q} \cap V_{2\pi r} = \emptyset \text{ für alle } q, r \in \mathbb{Q} \text{ mit } q \neq r.$$

Wegen Eigenschaft (c) haben all diese Mengen die gleiche Wahrscheinlichkeit, die wir mit  $p := \mathbb{P}[V_{2\pi q}]$  bezeichnen. Wegen Eigenschaften (a) und (b) gilt dann

$$1 = \mathbb{P}[\Omega] = \sum_{q \in \mathbb{Q}} \mathbb{P}[V_{2\pi q}] = \sum_{q \in \mathbb{Q}} p.$$

Nun betrachten wir zwei Fälle:

FALL 1:  $p > 0$ . Dann gilt  $\sum_{q \in \mathbb{Q}} p = \infty$ , ein Widerspruch.

FALL 2:  $p = 0$ . Dann gilt  $\sum_{q \in \mathbb{Q}} p = 0$ , wieder ein Widerspruch.  $\square$

Das obige Gegenbeispiel von Vitali ist der Anfang einer Reihe von paradoxen Ergebnissen, die in der ersten Hälfte des 20. Jahrhunderts entdeckt wurden. Das wohl überraschendste dieser Resultate ist das Banach-Tarski-Paradoxon (1924). Dieses behauptet, dass man eine Kugel in 6 disjunkte Teilmengen zerlegen und aus diesen dann mithilfe von starren Rotationen 2 Kugeln zusammensetzen kann, die genauso groß wie die ursprüngliche Kugel sind. Für mehr Einzelheiten und weitere überraschende Ergebnisse verweisen wir auf das Buch von Stan Wagon "The Banach-Tarski Paradox".

### 4.3. Vorüberlegungen zur Maßtheorie

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

**Beispiel 4.3.1** (Verteilung der Ladung oder der Masse). Man stelle sich eine positive elektrische Ladung, die sich über eine Menge  $\Omega$  verteilt hat. Dabei kann  $\Omega$  zum Beispiel ein Gebiet im drei- oder zweidimensionalen Raum sein. Ist  $A$  eine Teilmenge von  $\Omega$ , so kann man mit  $\mu(A)$  die Gesamtladung bezeichnen, die sich in der Menge  $A$  aufhält. 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  $\sigma$ -Additivität, gilt: Für beliebige paarweise disjunkte Teilmengen  $A_1, A_2, \dots \subset \Omega$  gilt

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

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

**Beispiel 4.3.2** (Wahrscheinlichkeit). Man stelle sich ein Zufallsexperiment mit Grundmenge  $\Omega$  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  $\sigma$ -Additivität gilt: Für beliebige paarweise disjunkte Teilmengen  $A_1, A_2, \dots \subset \Omega$  gilt

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



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

**Beispiel 4.3.3** (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  $\sigma$ -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.

#### 4.4. Algebren

**Definition 4.4.1.** Eine Teilmenge der Potenzmenge  $\mathcal{P}(\Omega)$  (d.h. eine Menge, deren Elemente Teilmengen von  $\Omega$  sind) werden wir als eine *Mengenfamilie* oder ein *Mengensystem* bezeichnen.

**Beispiel 4.4.2.** Ist  $\Omega = \mathbb{R}^2$  die Ebene, so können wir z.B. das Mengensystem aller Kreise oder das Mengensystem aller Quadrate betrachten. Bezeichnen wir mit  $\mathcal{Q}$  das Mengensystem aller Quadrate und mit  $\mathcal{R}$  das Mengensystem aller Rechtecke, so gilt  $\mathcal{Q} \subset \mathcal{R}$ .

Eine Algebra ist ein Mengensystem mit der Eigenschaft, dass man dieses Mengensystem nicht verlässt, wenn man beliebige mengentheoretischen Operationen auf *endlich* viele Mengen aus diesem Mengensystem anwendet.

**Definition 4.4.3.** Eine Mengenfamilie  $\mathcal{F} \subset \mathcal{P}(\Omega)$  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 4.4.4.** Folgende Mengenfamilien sind Algebren:

- $\mathcal{F} = \{\emptyset, \Omega\}$ .

- $\mathcal{F} = \mathcal{P}(\Omega)$ .

**Beispiel 4.4.5.** Sei  $\Omega = \{1, 2, 3\}$ . Folgende Mengenfamilie ist eine Algebra:

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

**Beispiel 4.4.6.** 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 4.4.7.** 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:

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

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 einer beliebigen Reihenfolge an, so erhält man wieder ein Element aus der Algebra. Einige Spezialfälle dieser Aussage werden im folgenden Satz bewiesen.

**Satz 4.4.8.** Sei  $\mathcal{F} \subset \mathcal{P}(\Omega)$  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.

□

## 4.5. $\sigma$ -Algebren

**Definition 4.5.1.** Eine Mengenfamilie  $\mathcal{F} \subset \mathcal{P}(\Omega)$  heißt  $\sigma$ -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)  $A_1, A_2, \dots \in \mathcal{F} \Rightarrow A_1 \cup A_2 \cup \dots \in \mathcal{F}$ . ( $\sigma$ -Vereinigungsstabilität).

**Beispiel 4.5.2.** Folgende Mengenfamilien sind  $\sigma$ -Algebren:

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

**Satz 4.5.3.** Ist  $\mathcal{F}$  eine  $\sigma$ -Algebra, so ist  $\mathcal{F}$  auch eine Algebra.

**Beweis.** Sei  $\mathcal{F}$  eine  $\sigma$ -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  $\sigma$ -Algebren gilt auch  $\emptyset = \Omega^c \in \mathcal{F}$ . Aus Eigenschaft (3) der  $\sigma$ -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 4.5.4.** Die Umkehrung von Satz 4.5.3 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  $\sigma$ -Algebra.

**Beispiel 4.5.5.** Die Mengenfamilie  $\mathcal{G}$  aus Beispiel 4.4.7 ist eine Algebra, aber keine  $\sigma$ -Algebra.

**Beispiel 4.5.6.** Sei  $\Omega = \mathbb{R}$ . Die folgende Mengenfamilie ist eine  $\sigma$ -Algebra (und somit auch eine Algebra):

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

**Satz 4.5.7.** Sei  $\mathcal{F}$  eine  $\sigma$ -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  $\sigma$ -Algebra ist  $\sigma$ -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  $\sigma$ -Algebra auf eine Teilmenge einschränken kann.

**Satz 4.5.8.** Sei  $\mathcal{F} \subset \mathcal{P}(\Omega)$  eine  $\sigma$ -Algebra. Sei  $A \subset \Omega$  nichtleer. Definiere die folgende Mengenfamilie:

$$\mathcal{F}_A := \{A \cap B : B \in \mathcal{F}\} \subset \mathcal{P}(A).$$

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

**Beweis.** Wir beweisen nur die  $\sigma$ -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  $\sigma$ -Algebra ist, erhalten wir, dass  $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$ . Es folgt aus der Definition von  $\mathcal{F}_A$ , dass  $A \cap \left(\bigcup_{n=1}^{\infty} B_n\right) \in \mathcal{F}_A$ . Somit gilt  $\bigcup_{n=1}^{\infty} C_n \in \mathcal{F}_A$ .  $\square$

#### 4.6. Die von einem Mengensystem erzeugte $\sigma$ -Algebra

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

**Satz 4.6.1.** Sei  $I$  eine beliebige nichtleere Menge. Für jedes  $i \in I$  sei  $\mathcal{F}_i \subset \mathcal{P}(\Omega)$  eine  $\sigma$ -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  $\sigma$ -Algebra.

**Beweis.** Wir zeigen nur, dass  $\mathcal{F}$   $\sigma$ -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  $\sigma$ -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 4.6.2.** Sei  $\mathcal{E} \subset \mathcal{P}(\Omega)$  eine beliebige nichtleere Mengenfamilie. Die Mengenfamilie

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

heißt die *von  $\mathcal{E}$  erzeugte  $\sigma$ -Algebra*. Der Schnitt wird über alle  $\sigma$ -Algebren  $\mathcal{F} \subset \mathcal{P}(\Omega)$ , die  $\mathcal{E}$  enthalten, genommen.

**Bemerkung 4.6.3.** Aus Satz 4.6.1 folgt, dass  $\sigma(\mathcal{E})$  tatsächlich eine  $\sigma$ -Algebra ist.

**Bemerkung 4.6.4.** Eine äquivalente Beschreibung:  $\sigma(\mathcal{E})$  ist die kleinste  $\sigma$ -Algebra, die alle Mengen aus  $\mathcal{E}$  enthält. Dabei heißt “die kleinste” folgendes: ist  $\mathcal{F} \subset \mathcal{P}(\Omega)$  irgendeine  $\sigma$ -Algebra, die  $\mathcal{E}$  enthält, so gilt  $\sigma(\mathcal{E}) \subset \mathcal{F}$ .

**Aufgabe 4.6.5.** Sei  $\Omega = \mathbb{R}$  und  $\mathcal{E} = \{\{\omega\} : \omega \in \mathbb{R}\}$  das System aller einelementigen Teilmengen von  $\mathbb{R}$ . Zeigen Sie: Die von  $\mathcal{E}$  erzeugte  $\sigma$ -Algebra besteht aus allen höchstens abzählbaren Teilmengen von  $\mathbb{R}$  und deren Komplementen.

## 4.7. Borel- $\sigma$ -Algebra

**Definition 4.7.1.** 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  $\sigma$ -Algebra  $\mathcal{B} := \sigma(\mathcal{E}) \subset \mathcal{P}(\mathbb{R})$  heißt die *Borel- $\sigma$ -Algebra* auf  $\mathbb{R}$ . Elemente von  $\mathcal{B}$  heißen *Borel-Mengen*.

**Bemerkung 4.7.2.** Man kann zeigen, dass die Borel- $\sigma$ -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 4.7.3.** Sei  $\Omega = \mathbb{R}^d$  und  $\mathcal{E}$  die Familie aller "Orthanten" 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  $\sigma$ -Algebra  $\mathcal{B}^d := \sigma(\mathcal{E}) \subset \mathcal{P}(\mathbb{R}^d)$  heißt die *Borel- $\sigma$ -Algebra* auf  $\mathbb{R}^d$ . Elemente von  $\mathcal{B}^d$  heißen *Borel-Mengen*.

**Bemerkung 4.7.4.** Man kann zeigen, dass die Borel- $\sigma$ -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 4.7.5.** 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  $\mathcal{P}(\mathbb{R}^d)$  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.

Borel-Hierarchie. Muss ergänzt werden...

#### 4.8. Maße

**Definition 4.8.1.** Sei  $\Omega$  eine nichtleere Menge und  $\mathcal{F} \subset \mathcal{P}(\Omega)$  eine  $\sigma$ -Algebra. Dann heißt das Paar  $(\Omega, \mathcal{F})$  ein *Messraum*. Mengen (oder Ereignisse)  $A \subset \Omega$  mit  $A \in \mathcal{F}$  heißen *messbar*.

**Definition 4.8.2.** Sei  $(\Omega, \mathcal{F})$  ein Messraum. Ein *Maß*  $\mu$  auf  $(\Omega, \mathcal{F})$  ist eine Funktion  $\mu : \mathcal{F} \rightarrow [0, +\infty]$  mit der folgenden Eigenschaft, die  $\sigma$ -*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 4.8.3.** 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  $\mu$  eine Verteilung der positiven elektrischen Ladung auf  $\Omega$  beschreibt, bei der sich die Ladung nur in den Punkten  $\omega_1, \omega_2, \dots$  sammelt, wobei die Ladung des Punktes  $\omega_i$  gleich  $m_i$  ist.

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

**Definition 4.8.4.** 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  $\lambda$  kein Maß auf  $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$ : sie ist nicht  $\sigma$ -additiv und sogar nicht additiv.

**Satz 4.8.5** (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  $\lambda$  nur auf die  $\sigma$ -Algebra der Borel-Mengen einschränken, wird sie  $\sigma$ -additiv.

**Satz 4.8.6** (Satz von Lebesgue).  $\lambda$  ist ein Maß auf  $(\mathbb{R}^d, \mathcal{B}^d)$ .

#### 4.9. Wahrscheinlichkeitsmaße

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

**Definition 4.9.1.** Sei  $(\Omega, \mathcal{F})$  ein Messraum. Ein *Wahrscheinlichkeitsmaß*  $\mathbb{P}$  auf  $(\Omega, \mathcal{F})$  ist eine Funktion  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  mit folgenden zwei Eigenschaften:

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

$$\mathbb{P}[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]. \quad (4.9.1)$$

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

**Beispiel 4.9.2** (Diskrete Wahrscheinlichkeitsräume). Zuerst betrachten wir ein Zufallsexperiment mit einer *höchstens abzählbaren* Grundmenge  $\Omega = \{\omega_1, \omega_2, \dots\}$ . Die Wahrscheinlichkeiten der einzelnen Ausgänge bezeichnen wir mit  $p(\omega_1), p(\omega_2), \dots$ . Offenbar ist  $p : \Omega \rightarrow [0, 1]$  eine Funktion mit der Eigenschaft

$$\sum_n p(\omega_n) = 1.$$

Nun wollen wir unserem Zufallsexperiment einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  zuordnen. Wir definieren  $\mathcal{F} := \mathcal{P}(\Omega)$  und das Wahrscheinlichkeitsmaß  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  wie folgt:

$$\mathbb{P}[A] := \sum_{\omega \in A} p(\omega), \quad \text{für alle } A \subset \Omega.$$

Wahrscheinlichkeitsräume  $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ , die auf diese Weise mit einem endlichen oder abzählbaren  $\Omega$  konstruiert wurden, heißen *diskrete Wahrscheinlichkeitsräume*.

**Beispiel 4.9.3** (Geometrische Wahrscheinlichkeiten). Sei  $\Omega \subset \mathbb{R}^d$  eine Borel-Menge mit Lebesgue-Maß  $\lambda(\Omega)$ , das weder 0 noch  $\infty$  ist. Wir betrachten ein Zufallsexperiment, bei dem ein zufälliger, “uniformverteilter” Punkt  $S$  in der Menge  $\Omega$  ausgewählt wird. Solche

Experimente haben wir z.B. in Kapitel 3 betrachtet. Der Wahrscheinlichkeitsraum, der dieses Experiment beschreibt, kann wie folgt konstruiert werden.

Als Grundmenge dieses Experiments können wir  $\Omega$  betrachten. Als  $\sigma$ -Algebra auf  $\Omega$  wählen wir die Einschränkung

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

der Borel- $\sigma$ -Algebra  $\mathcal{B}^d$  auf  $\Omega$ ; siehe Satz 4.5.8. Die Wahrscheinlichkeit, dass der zufällige Punkt  $S$  in eine Menge  $A \in \mathcal{B}_\Omega^d$  fällt, ist

$$\mathbb{P}[A] = \frac{\lambda(A)}{\lambda(\Omega)}, \quad A \in \mathcal{B}_\Omega^d, \quad (4.9.2)$$

was wir als Definition der Uniformverteilung auffassen. Das so konstruierte Tripel  $(\Omega, \mathcal{B}_\Omega^d, \mathbb{P})$  ist ein Modell für das oben beschriebene Zufallsexperiment.

Es sei hervorgehoben, dass diese Konstruktion für  $\lambda(\Omega) = 0$  oder  $\lambda(\Omega) = \infty$  nicht möglich ist, weil wir in diesen Fällen durch 0 oder durch  $\infty$  teilen müssten, was zu Unbestimmtheiten führen könnte. Insbesondere ist die Uniformverteilung auf ganz  $\mathbb{R}^d$  nicht wohldefiniert.

Im obigen Experiment ist die Wahrscheinlichkeit jedes einzelnen Punktes gleich 0:

$$\mathbb{P}[\{\omega\}] = 0 \quad \text{für alle } \omega \in \Omega.$$

Überdies ist sogar die Wahrscheinlichkeit jeder abzählbaren Menge gleich 0:

$$\mathbb{P}[A] = 0 \quad \text{falls } A = \{\omega_1, \omega_2, \dots\} \text{ abzählbar.}$$

**Bemerkung 4.9.4** (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  $\sigma$ -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.

#### 4.10. Eigenschaften der Wahrscheinlichkeit

Wir leiten nun einige Formeln, die in jedem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  gelten.

**Lemma 4.10.1.** Unmögliches Ereignis hat Wahrscheinlichkeit 0. Das heißt,  $\mathbb{P}[\emptyset] = 0$ .

**Beweis.** Setze  $A_1 = A_2 = \dots = \emptyset$  in (4.9.1). Es ergibt sich  $\mathbb{P}[\emptyset] = \mathbb{P}[\emptyset] + \mathbb{P}[\emptyset] + \dots$ . Das kann nur für  $\mathbb{P}[\emptyset] = 0$  gelten.  $\square$

Eines der Axiome der Wahrscheinlichkeitstheorie postuliert die  $\sigma$ -Additivität: Die Wahrscheinlichkeit einer *abzählbar unendlichen* Vereinigung ist die Summe der Wahrscheinlichkeiten der einzelnen Ereignisse. Eine ähnliche Eigenschaft für *endliche* Vereinigungen (die Additivität genannt wird) ergibt sich daraus als Konsequenz.



**Lemma 4.10.2** (Additivität). Für beliebiges  $n \in \mathbb{N}$  und beliebige paarweise disjunkte Ereignisse  $A_1, \dots, A_n \in \mathcal{F}$  gilt

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

**Beweis.** Setze  $A_{n+1} = A_{n+2} = \dots = \emptyset$  in (4.9.1) und benutze, dass  $\mathbb{P}[\emptyset] = 0$ .  $\square$

Nun berechnen wir die Wahrscheinlichkeit des Gegenereignisses.

**Lemma 4.10.3.** Für jedes Ereignis  $A \in \mathcal{F}$  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 folgt  $1 = \mathbb{P}[\Omega] = \mathbb{P}[A \cup A^c] = \mathbb{P}[A] + \mathbb{P}[A^c]$ . Durch Umstellen ergibt sich die Behauptung.  $\square$

**Lemma 4.10.4.** Für beliebige Ereignisse  $A, B \in \mathcal{F}$  gilt

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

**Beweis.** Das Ereignis  $A$  ist eine disjunkte Vereinigung der Ereignisse  $A \cap B$  und  $A \setminus B$ . Mit der Additivität folgt  $\mathbb{P}[A] = \mathbb{P}[A \cap B] + \mathbb{P}[A \setminus B]$ . Die Behauptung ergibt sich durch Umstellen.  $\square$

**Lemma 4.10.5.** Für beliebige Ereignisse  $A, B \in \mathcal{F}$  (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 4.10.4 folgt  $\mathbb{P}[A] - \mathbb{P}[A \cap B] + \mathbb{P}[B] = \mathbb{P}[A \cup B]$  und die Behauptung ergibt sich durch Umstellen.  $\square$

**Lemma 4.10.6** (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 4.10.7** (Subadditivität). Für beliebige Ereignisse  $A, B \in \mathcal{F}$  gilt

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

**Beweis.** Folgt aus Lemma 4.10.5, denn  $\mathbb{P}[A \cap B] \geq 0$ .  $\square$

**Lemma 4.10.8** (Subadditivität). Für beliebige Ereignisse  $A_1, \dots, A_n \in \mathcal{F}$  gilt:

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

**Beweis.** Folgt aus Lemma 4.10.7 mit Induktion.  $\square$

**Satz 4.10.9** ( $\sigma$ -Subadditivität). Für beliebige Ereignisse  $A_1, A_2, \dots \in \mathcal{F}$  gilt

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

**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  $\sigma$ -Algebra ist, sind die Mengen  $B_1, B_2, \dots$  messbar. Die Mengen  $B_1, B_2, \dots$  sind disjunkt und es gilt  $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ . Aus der  $\sigma$ -Additivität von  $\mathbb{P}$  folgt, dass

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

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 4.10.10.** 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 4.10.11.** Eine 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  $\sigma$ -Subadditivität. Die zweite Aussage: de Morgan'sche Regel.  $\square$

**Beispiel 4.10.12.** Wir betrachten das Experiment, bei dem eine zufällige, uniform verteilte Zahl im Intervall  $[0, 1]$  ausgewählt wird. Die Grundmenge ist  $\Omega = [0, 1]$ , die  $\sigma$ -Algebra  $\mathcal{F}$  besteht aus allen Borel-Teilmengen von  $[0, 1]$ , und  $\mathbb{P}$  ist das Lebesgue-Maß. Dann ist z.B. jedes Ereignis der Form  $\{\omega\}$  ein Nullereignis. Auch jede abzählbare Menge ist ein Nullereignis, z.B. die Menge der rationalen Zahlen. Eine überabzählbare Vereinigung von Nullereignissen muss allerdings kein Nullereignis sein, denn wir können  $[0, 1]$  als Vereinigung der einzelnen Punkte darstellen.

**Satz 4.10.13** (Stetigkeit der Wahrscheinlichkeit). Seien  $A_1, A_2, \dots \in \mathcal{F}$  Ereignisse. Dann gelten die folgenden zwei Aussagen.

- (a) Ist die Folge  $A_1 \subset A_2 \subset \dots$  aufsteigend, so gilt  $\mathbb{P}[\bigcup_{i=1}^{\infty} A_i] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n]$ .  
 (b) Ist die Folge  $A_1 \supset A_2 \supset \dots$  fallend, so gilt  $\mathbb{P}[\bigcap_{i=1}^{\infty} A_i] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n]$ .

**Beweis von (a).** Es gelte  $A_1 \subset A_2 \subset \dots$ . Definiere  $A_0 = \emptyset$  und

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad \dots, \quad B_n = A_n \setminus A_{n-1}, \quad \dots$$

Dann sind die Mengen  $B_1, B_2, \dots$  messbar, disjunkt und es gilt

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

Aus der  $\sigma$ -Additivität der Wahrscheinlichkeit folgt, dass

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \mathbb{P}\left[\bigcup_{i=1}^{\infty} B_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[B_i] = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}[B_i].$$

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 (b).** Es gelte  $A_1 \supset A_2 \supset \dots$ . Mit den Regeln von de Morgan erhalten wir

$$\mathbb{P}\left[\bigcap_{i=1}^{\infty} A_i\right] = 1 - \mathbb{P}\left[\left(\bigcap_{i=1}^{\infty} A_i\right)^c\right] = 1 - \mathbb{P}\left[\bigcup_{i=1}^{\infty} (A_i^c)\right].$$

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:

$$1 - \mathbb{P}\left[\bigcup_{i=1}^{\infty} (A_i^c)\right] = 1 - \lim_{n \rightarrow \infty} \mathbb{P}[A_n^c] = 1 - \lim_{n \rightarrow \infty} (1 - \mathbb{P}[A_n]) = \lim_{n \rightarrow \infty} \mathbb{P}[A_n].$$

□

#### 4.11. Die Siebformel

Wie wir bereits wissen, gilt für beliebige Ereignisse  $A$  und  $B$  die Formel

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

Gibt es eine analoge Formel zur Berechnung der Wahrscheinlichkeit von  $A \cup B \cup C$  für drei Ereignisse  $A, B, C \in \mathcal{F}$ ? Bei disjunkten Ereignissen würden wir einfach die Summe  $\mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C]$  bilden. Bei sich überschneidenden Ereignissen würden in dieser Summe einige Ausgänge mehrfach gezählt. Also können wir versuchen,  $\mathbb{P}[A \cap B] + \mathbb{P}[B \cap C] + \mathbb{P}[A \cap C]$  abzuziehen. Dies ist allerdings nicht das Ende der Lösung, denn alle Ausgänge in  $A \cap B \cap C$  wurden zuerst dreimal gezählt und dann dreimal abgezogen. Wir müssen also noch  $\mathbb{P}[A \cap B \cap C]$  dazuaddieren und erhalten die Formel

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

Für vier Ereignisse  $A, B, C, D$  kann man ebenfalls eine ähnliche Formel zur Berechnung von  $\mathbb{P}[A \cup B \cup C \cup D]$  aufstellen. Zuerst bildet man die Summe der Wahrscheinlichkeiten der einzelnen Ereignisse, dann subtrahiert man Wahrscheinlichkeiten aller Zweierschnitte, dann addiert man die Wahrscheinlichkeiten aller Dreierschnitte und schließlich subtrahiert man die Wahrscheinlichkeit des Viererschnitts:

$$\begin{aligned} \mathbb{P}[A \cup B \cup C \cup D] &= \mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] + \mathbb{P}[D] \\ &\quad - \mathbb{P}[A \cap B] - \mathbb{P}[A \cap C] - \mathbb{P}[A \cap D] - \mathbb{P}[B \cap C] - \mathbb{P}[B \cap D] - \mathbb{P}[C \cap D] \\ &\quad + \mathbb{P}[A \cap B \cap C] + \mathbb{P}[A \cap C \cap D] + \mathbb{P}[A \cap B \cap D] + \mathbb{P}[B \cap C \cap D] \\ &\quad - \mathbb{P}[A \cap B \cap C \cap D]. \end{aligned}$$

Nun formulieren wir die allgemeine Identität für eine beliebige Anzahl  $n$  von Ereignissen.

**Satz 4.11.1** (Siebformel oder Einschluss-Ausschluss-Identität). Für beliebige  $n$  Ereignisse  $A_1, A_2, \dots, A_n \in \mathcal{F}$  gilt

$$\mathbb{P}[A_1 \cup \dots \cup A_n] = S_1 - S_2 + S_3 - \dots + (-1)^{n-1} S_n,$$

wobei  $S_1, \dots, S_n$  wie folgt definiert werden:

$$S_r = \sum_{1 \leq i_1 < \dots < i_r \leq n} \mathbb{P}[A_{i_1} \cap \dots \cap A_{i_r}].$$

Vollständig ausgeschrieben, sieht die Siebformel wie folgt aus:

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

Es ist möglich, die Siebformel per Induktion zu beweisen, was allerdings langwierig und nicht besonders elegant ist. Ein kurzer Beweis der Siebformel wird später gegeben, nachdem wir den Begriff des Erwartungswertes eingeführt haben. ????

**Beispiel 4.11.2.** Es werden  $n$  Briefe rein zufällig in  $n$  dafür vorgesehene, adressierte Umschläge gesteckt. Dabei wird darauf geachtet, dass in jeden Umschlag genau ein Brief gesteckt wird. Wir bestimmen die Wahrscheinlichkeit des Ereignisses

$$C_n = \text{“mindestens ein Brief gelangt in den richtigen Umschlag”}.$$

**Lösung.** Die Ausgänge des Experiments sind alle möglichen *bijektiven* Funktionen  $f : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  (d.h. Permutationen). Dabei beschreibt  $f(k)$  den Umschlag, in den der  $k$ -te Brief gesteckt wurde. Für die Anzahl der Elemente in  $\Omega$  gilt also

$$\#\Omega = n!$$

Da die Briefe “rein zufällig” in die Umschläge gesteckt werden, gehen wir von der Laplace-Annahme aus:

$$\mathbb{P}[B] = \frac{\#B}{\#\Omega} \quad \text{für jedes } B \subset \Omega.$$

Für jedes  $k \in \{1, \dots, n\}$  definieren wir das Ereignis

$$A_k = \{f \in \Omega : f(k) = k\} = \text{“der } k\text{-te Brief gelangt in den richtigen Umschlag”}.$$

Uns interessiert das Ereignis

$$C_n = \text{“mindestens ein Brief gelangt in den richtigen Umschlag”} = A_1 \cup \dots \cup A_n.$$

Leider sind die Ereignisse  $A_1, \dots, A_n$  nicht disjunkt, denn es können mehrere Briefe in den jeweils richtigen Umschlag gelangen. Also benutzen wir zur Bestimmung von  $\mathbb{P}[A]$  die Siebformel. Für diese benötigen wir die Wahrscheinlichkeiten aller möglichen Schnitte von  $A_1, \dots, A_n$ .

Zuerst bestimmen wir exemplarisch die Anzahl der Ausgänge in  $A_1$ . Für  $f \in A_1$  gilt  $f(1) = 1$ , während die Werte  $f(2), \dots, f(n)$  eine beliebige Permutation von  $2, \dots, n$  bilden dürfen. Es gilt also

$$\#A_1 = 1 \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1 = (n-1)!$$

Für die Wahrscheinlichkeit von  $A_1$  folgt, dass

$$\mathbb{P}[A_1] = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Diese Formel ist nicht überraschend: Für den ersten Brief gibt es  $n$  gleichwahrscheinliche Umschläge, in die er gesteckt werden kann, und nur einer dieser Umschläge führt zum Eintreten von  $A_1$ . Analog zeigt man, dass

$$\mathbb{P}[A_k] = \frac{1}{n} \quad \text{für alle } k = 1, \dots, n.$$

Nun bestimmen wir die Wahrscheinlichkeit des Ereignisses  $A_{i_1} \cap \dots \cap A_{i_r}$  mit  $1 \leq i_1 < \dots < i_r \leq n$ . Diese kommen in der Siebformel vor. Wir betrachten exemplarisch das Ereignis

$$\begin{aligned} A_1 \cap \dots \cap A_r &= \{f \in \Omega : f(1) = 1, \dots, f(r) = r\} \\ &= \text{“Briefe } 1, \dots, r \text{ werden in den jeweils richtigen Umschlag gesteckt”}. \end{aligned}$$

Da die Zuordnung der Briefe  $1, \dots, r$  eindeutig festgelegt ist, während die restlichen Briefe  $r+1, \dots, n$  völlig frei auf die Umschläge  $r+1, \dots, n$  verteilt werden dürfen, was einer Permutation von  $n-r$  Elementen entspricht, gilt für die Anzahl der Elemente in  $A_1 \cap \dots \cap A_r$  die Formel

$$\#(A_1 \cap \dots \cap A_r) = (n-r)!$$

Für die Wahrscheinlichkeit dieses Ereignisses erhalten wir

$$\mathbb{P}[A_1 \cap \dots \cap A_r] = \frac{(n-r)!}{n!}.$$

Analoge Überlegungen zeigen, dass für beliebige Indizes  $1 \leq i_1 < \dots < i_r \leq n$

$$\mathbb{P}[A_{i_1} \cap \dots \cap A_{i_r}] = \frac{(n-r)!}{n!}.$$

Nun können wir die Siebformel anwenden:

$$\mathbb{P}[C_n] = \sum_{r=1}^n (-1)^{r-1} S_r.$$

Dabei ist  $S_r$  gegeben durch

$$S_r = \sum_{1 \leq i_1 < \dots < i_r \leq n} \mathbb{P}[A_{i_1} \cap \dots \cap A_{i_r}] = \binom{n}{r} \cdot \frac{(n-r)!}{n!} = \frac{1}{r!},$$

wobei wir benutzt haben, dass alle Ereignisse  $\mathbb{P}[A_{i_1} \cap \dots \cap A_{i_r}]$  (bei fixem  $r$ ) die gleiche Wahrscheinlichkeit haben. Somit ergibt sich, dass

$$\mathbb{P}[C_n] = \sum_{r=1}^n \frac{(-1)^{r-1}}{r!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots + (-1)^{n-1} \frac{1}{n!}, \quad (4.11.1)$$

was die Lösung abschließt. □

**Bemerkung 4.11.3.** Was passiert mit der Wahrscheinlichkeit des Ereignisses  $C_n$  bei einer sehr großen Anzahl an Briefen, also für  $n \rightarrow \infty$ ? Für jeden konkreten Brief wird die Wahrscheinlichkeit, in den richtigen Umschlag gesteckt zu werden, klein, die Anzahl der Briefe wird aber groß. Inwieweit gleichen sich diese Effekte aus? Wir können diese Frage beantworten, indem wir den Grenzwert von  $\mathbb{P}[C_n]$  für  $n \rightarrow \infty$  ausrechnen. Durch Grenzwertbildung in (4.11.1) ergibt sich

$$\lim_{n \rightarrow \infty} \mathbb{P}[C_n] = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots,$$

wobei die Reihe nun unendlich viele Terme enthält. Um die Summe der Reihe zu bestimmen, benutzen wir die Taylor-Reihe

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Mit  $x = -1$  ergibt sich die Formel

$$\frac{1}{e} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots$$

Es folgt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}[C_n] = 1 - \frac{1}{e} \approx 0.632121.$$

**Aufgabe 4.11.4** (Anzahl der Fixpunkte in einer zufälligen Permutation). In der Situation von Beispiel 4.11.2 bestimmen Sie für ein vorgegebenes  $k \in \{0, \dots, n\}$  die Wahrscheinlichkeit, dass genau  $k$  Briefe in den jeweils richtigen Umschlag gelangen. Bestimmen Sie den Grenzwert dieser Wahrscheinlichkeit für ein festes  $k \in \mathbb{N}$  und  $n \rightarrow \infty$ .

**Aufgabe 4.11.5** (Anzahl der Fixpunkte in einer zufälligen Abbildung). Es werden  $n$  Briefe unabhängig voneinander zufällig auf  $n$  Briefkästen verteilt. Alle Briefkästen seien dabei gleichwahrscheinlich und ein Briefkasten kann mehrere Briefe bekommen.

- (a) Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Briefkasten leer bleibt?
- (b) Verwenden Sie Aufgabenteil (a) um die folgende Identität zu beweisen:

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

**Aufgabe 4.11.6.** 1000 Personen schreiben Ihre Geburtstage auf einen Zettel. Es sei  $X$  die Anzahl der Tage im Jahr, die nicht in der Liste vorkommen. Bestimmen Sie  $\mathbb{P}[X = m]$  für alle  $m \in \{0, 1, \dots, 364\}$ .

**Aufgabe 4.11.7.** Auf dem Nachhauseweg aus dem Kindergarten zieht jedes der  $n$  Kinder ein zufälliges linkes und ein zufälliges rechtes Schuh an.

- (a) Zeigen Sie: Die Wahrscheinlichkeit des Ereignisses, dass bei keinem der Kinder beide Schuhe richtig sind, ist

$$\sum_{k=0}^n (-1)^k \frac{(n-k)!}{k!n!}.$$

- (b) Zeigen Sie: Die Wahrscheinlichkeit des Ereignisses, dass bei allen Kindern beide Schuhe falsch sind, ist

$$\left( \sum_{k=2}^n \frac{(-1)^k}{k!} \right)^2.$$

**Aufgabe 4.11.8.** Die Möbius-Funktion  $\mu : \mathbb{N} \rightarrow \{0, +1, -1\}$  ist definiert durch die Vorschrift

$$\mu(n) = \begin{cases} 0, & \text{falls } n \text{ durch eine Quadratzahl teilbar ist,} \\ (-1)^k, & \text{falls } n = p_1 \dots p_k \text{ mit verschiedenen Primzahlen } p_1, \dots, p_k. \end{cases}$$

- (a) Zeigen Sie: Für jede Zahl  $n \in \mathbb{N}$  gilt

$$\sum_{d:d|n} \mu(d) = \begin{cases} 1, & \text{falls } n = 1, \\ 0, & \text{sonst,} \end{cases}$$

wobei die Summe über alle Teiler  $d$  von  $n$  gebildet wird.

- (b) Die Riemann'sche Zeta-Funktion ist definiert durch  $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$  für  $s > 1$ . Zeigen Sie, dass

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}.$$

## 4.12. Eindeutigkeit der Maßfortsetzung

Das Lebesgue-Maß wurde definiert indem es zuerst für alle Rechtecke definiert und dann auf Borel-Mengen fortgesetzt wurde. Ist diese Fortsetzung eindeutig? In diesem Kapitel werden wir uns die folgende allgemeine Frage der Eindeutigkeit der Maßfortsetzung stellen. Sei  $\Omega$  eine Grundmenge und  $\mathcal{E} \subset \mathcal{P}(\Omega)$  ein Mengensystem auf  $\Omega$ . Wir bezeichnen mit  $\mathcal{F} = \sigma(\mathcal{E})$  die von  $\mathcal{E}$  erzeugte  $\sigma$ -Algebra.

**Frage:** Es seien  $\mu$  und  $\nu$  zwei Wahrscheinlichkeitsmaße auf  $(\Omega, \mathcal{F})$ , die auf dem Mengensystem  $\mathcal{E}$  übereinstimmen, d.h. es gelte

$$\mu(A) = \nu(A) \text{ für alle } A \in \mathcal{E}.$$

Gilt dann  $\mu(A) = \nu(A)$  sogar für alle  $A \in \mathcal{F}$ ?

Die Antwort ist im Allgemeinen negativ, wie das folgende Beispiel zeigt.

**Beispiel 4.12.1.** Sei  $\Omega = \{1, 2, 3, 4\}$  und betrachte das Mengensystem  $\mathcal{E} = \{\{1, 2\}, \{2, 3\}\}$ . Man überzeugt sich leicht, dass  $\sigma(\mathcal{E}) = \mathcal{P}(\Omega)$ . Nun betrachten wir die durch die folgende Vorschrift eindeutig festgelegten Wahrscheinlichkeitsmaße  $\mu$  und  $\nu$  auf  $(\Omega, \mathcal{P}(\Omega))$ :

$$\begin{aligned}\mu(\{1\}) &= \mu(\{2\}) = \mu(\{3\}) = \mu(\{4\}) = \frac{1}{4}, \\ \nu(\{1\}) &= \nu(\{3\}) = \frac{1}{2}, \quad \nu(\{2\}) = \nu(\{4\}) = 0.\end{aligned}$$

Somit ist  $\mu$  die Gleichverteilung auf  $\Omega$ , während  $\nu$  die Gleichverteilung auf  $\{1, 3\}$  ist. Man sieht leicht, dass

$$\mu(\{1, 2\}) = \mu(\{2, 3\}) = \frac{1}{2} = \nu(\{1, 2\}) = \nu(\{2, 3\}).$$

Somit stimmen die beiden Wahrscheinlichkeitsmaße auf dem Erzeuger  $\mathcal{E}$  überein. Dabei folgt aus der Definition von  $\mu$  und  $\nu$ , dass es sich um unterschiedliche Maße handelt.

Damit die Frage nach der Eindeutigkeit der Maßfortsetzung positiv beantwortet werden kann, muss man eine zusätzliche Bedingung an das Mengensystem  $\mathcal{E}$  stellen.

**Definition 4.12.2.** Das Mengensystem  $\mathcal{E} \subset \mathcal{P}(\Omega)$  heißt **schnittstabil** oder  **$\cap$ -stabil**, wenn es mit beliebigen zwei Teilmengen auch deren Schnitt enthält:

$$\forall A, B \in \mathcal{E} \Rightarrow A \cap B \in \mathcal{E}.$$

**Beispiel 4.12.3.** Sei  $\Omega = \mathbb{R}^d$ . Die Familie der Orthanten der Form  $(-\infty, a_1] \times \dots \times (-\infty, a_d]$  mit  $a_1, \dots, a_d \in \mathbb{R}$  ist schnittstabil. Ebenso schnittstabil ist die Familie der Quader  $[a_1, b_1] \times \dots \times [a_d, b_d]$ , wenn man sie um die leere Menge  $\emptyset$  erweitert. Beide Mengensysteme erzeugen die Borel- $\sigma$ -Algebra  $\mathcal{B}^d$  und werden deshalb als schnittstabile Erzeuger von  $\mathcal{B}^d$  bezeichnet.

Der nächste Satz behauptet, dass zwei auf einem schnittstabilen Erzeuger übereinstimmende Wahrscheinlichkeitsmaße gleich sein müssen.

**Satz 4.12.4** (Satz über die Eindeutigkeit der Maßfortsetzung). Seien  $\mu, \nu$  zwei Wahrscheinlichkeitsmaße auf  $(\Omega, \mathcal{F})$ , wobei  $\mathcal{F} = \sigma(\mathcal{E})$  und  $\mathcal{E} \subset \mathcal{F}$  ein schnittstabiles Mengensystem sei. Es gelte außerdem

$$\mu(A) = \nu(A) \text{ für alle } A \in \mathcal{E}.$$

Dann gilt sogar  $\mu(A) = \nu(A)$  für alle  $A \in \mathcal{F}$ , d.h.  $\mu = \nu$ .

**Beispiel 4.12.5.** Seien  $\mu$  und  $\nu$  zwei Wahrscheinlichkeitsmaße auf  $(\mathbb{R}, \mathcal{B})$  mit  $\mu([a, b]) = \nu([a, b])$  für jedes Intervall  $[a, b]$ . Da das System der Intervalle der Form  $[a, b]$  (erweitert um die leere Menge  $\emptyset$ ) schnittstabil ist und die Borel- $\sigma$ -Algebra  $\mathcal{B}$  erzeugt, folgt, dass  $\mu(A) = \nu(A)$  für jede Borel-Menge  $A \subset \mathbb{R}$ .



Für den Beweis von Satz 4.12.4 werden einige Vorüberlegungen benötigt. Wir werden versuchen, das Mengensystem, auf dem die Wahrscheinlichkeitsmaße  $\mu$  und  $\nu$  bereinstimmen, um weitere Mengen zu erweitern. Gilt z.B.  $\mu(A_n) = \nu(A_n)$  für alle  $n \in \mathbb{N}$  und **disjunkte** Mengen  $A_1, A_2, \dots$ , so können wir wegen der  $\sigma$ -Additivität schließen, dass

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

Diese Überlegung schlägt allerdings fehl, wenn  $A_1, A_2, \dots$  nicht disjunkt sind. Das motiviert die nächste Definition.

**Definition 4.12.6.** Ein Mengensystem  $\mathcal{D} \subset \mathcal{P}(\Omega)$  heißt ein **Dynkin-System** (oder ein  **$\lambda$ -System**, falls

- (1)  $\Omega \in \mathcal{D}$ ;
- (2)  $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$ .
- (3) Für jede Folge  $A_1, A_2, \dots \in \mathcal{D}$  von **paarweise disjunkten** Mengen aus  $\mathcal{D}$  gilt auch  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$ .

Diese Definition unterscheidet sich von der Definition einer  $\sigma$ -Algebra nur dadurch, dass die Abgeschlossenheit bezüglich abzählbaren Vereinigungen abgeschwächt wird, indem sie nur für disjunkte Vereinigungen gefordert wird. Eine  $\sigma$ -Algebra ist somit auch ein Dynkin-System. Die Umkehrung gilt aber nicht, wie die folgende Aufgabe zeigt.

**Aufgabe 4.12.7.** Sei  $\Omega = \{1, 2, 3, 4\}$ . Zeigen Sie, dass  $\mathcal{D} := \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}, \{2, 3\}, \{1, 4\}\}$  ein Dynkin-System, aber keine  $\sigma$ -Algebra definiert.

Genauso wie für  $\sigma$ -Algebren zeigt man, dass der Schnitt von beliebig vielen Dynkin-Systemen wieder ein Dynkin-System ist:

**Satz 4.12.8.** Sei  $I$  eine beliebige nichtleere Menge. Für jedes  $i \in I$  sei  $\mathcal{D}_i \subset \mathcal{P}(\Omega)$  ein Dynkin-System. Dann ist auch die Mengenfamilie

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

ein Dynkin-System.

BEWEIS. Übung. □

Dieser Satz ermöglicht es uns, das von einem Mengensystem  $\mathcal{E}$  erzeugte Dynkin-System als den Schnitt aller Dynkin-Systeme, die  $\mathcal{E}$  enthalten, zu definieren.

**Definition 4.12.9.** Sei  $\mathcal{E} \subset \mathcal{P}(\Omega)$  eine beliebige nichtleere Mengenfamilie. Die Mengenfamilie

$$\mathcal{D}(\mathcal{E}) := \bigcap_{\substack{\mathcal{D}: \mathcal{E} \subset \mathcal{D} \subset \mathcal{P}(\Omega) \\ \mathcal{D} \text{ ist Dynkin-System}}} \mathcal{D}$$

heißt das von  $\mathcal{E}$  erzeugte Dynkin-System.

Mit anderen Worten ist  $\mathcal{D}(\mathcal{E})$  das minimale Dynkin-System, das  $\mathcal{E}$  enthält. Den wichtigsten Schritt im Beweis von Satz 4.12.4 bildet das folgende

**Lemma 4.12.10** (Dynkin-Lemma). Sei  $\mathcal{E} \subset \mathcal{P}(\Omega)$  ein schnittstabiles Mengensystem. Dann ist das von  $\mathcal{E}$  erzeugte Dynkin-System  $\mathcal{D}(\mathcal{E})$  sogar eine  $\sigma$ -Algebra.

BEWEIS VON LEMMA 4.12.10. SCHRITT 1. Wir zeigen, dass  $\mathcal{D} := \mathcal{D}(\mathcal{E})$  schnittstabil ist. Sei dazu  $B \in \mathcal{D}$ . Definiere das Mengensystem

$$\mathcal{D}_B := \{A \in \mathcal{D} : A \cap B \in \mathcal{D}\}.$$

Unser Ziel ist zu zeigen, dass  $\mathcal{D}_B = \mathcal{D}$ . Zuerst bemerken wir, dass  $\mathcal{D}_B$  ein Dynkin-System ist. Dies kann wie folgt begründet werden.

- (1)  $\Omega \in \mathcal{D}_B$ , denn  $\Omega \cap B = B \in \mathcal{D}$ .
- (2) Sei  $A \in \mathcal{D}_B$ . Wir zeigen, dass  $A^c \in \mathcal{D}_B$ . Gegeben ist, dass  $A \cap B \in \mathcal{D}$ . Daraus folgt, dass

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

Somit gilt  $A^c \in \mathcal{D}_B$ .

- (3) Seien  $A_1, A_2, \dots \in \mathcal{D}_B$  disjunkt. Wir zeigen, dass  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}_B$ . Gegeben ist, dass  $A_i \cap B \in \mathcal{D}$  für alle  $i \in \mathbb{N}$ . Daraus ergibt sich

$$\left( \bigcup_{i=1}^{\infty} A_i \right) \cap B = \bigcup_{i=1}^{\infty} (A_i \cap B) \in \mathcal{D},$$

denn  $A_1 \cap B, A_2 \cap B, \dots$  sind disjunkt und in  $\mathcal{D}$ . Wir haben gezeigt, dass  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}_B$ .

Also ist  $\mathcal{D}_B$  ein Dynkin-System für jedes  $B \in \mathcal{D}$ . Sei nun  $E \in \mathcal{E}$ . Dann gilt  $\mathcal{E} \subset \mathcal{D}_E$ , denn  $\mathcal{E}$  ist schnittstabil. Somit ist  $\mathcal{D}_E$  ein Dynkin-System, das  $\mathcal{E}$  enthält. Auf der anderen Seite ist  $\mathcal{D}$  das *minimale* Dynkin-System, das  $\mathcal{E}$  enthält. Es folgt, dass  $\mathcal{D}_E \supset \mathcal{D}$ , und zwar für alle  $E \in \mathcal{E}$ . Für beliebige  $E \in \mathcal{E}$  und  $A \in \mathcal{D}$  gilt also  $A \cap E \in \mathcal{D}$ . Um die gewünschte Schnittstabilität von  $\mathcal{D}$  zu beweisen, müssen wir noch zeigen, dass die Bedingung  $E \in \mathcal{E}$  durch die schwächere Bedingung  $E \in \mathcal{D}$  ersetzt werden kann.

Indem wir  $A$  in  $B$  umbenennen, können wir das Obige wie folgt formulieren: Für alle  $E \in \mathcal{E}$  und  $B \in \mathcal{D}$  gilt  $B \cap E \in \mathcal{D}$ . Mit anderen Worten, gilt  $\mathcal{E} \subset \mathcal{D}_B$ . Somit ist  $\mathcal{D}_B$  ein Dynkin-System, das  $\mathcal{E}$  enthält. Auf der anderen Seite ist  $\mathcal{D}$  das *minimale* Dynkin-System, das  $\mathcal{E}$  enthält. Es folgt, dass  $\mathcal{D}_B \supset \mathcal{D}$ . Auf der anderen Seite gilt definitionsgemäß  $\mathcal{D}_B \subset \mathcal{D}$ . Es ergibt sich, dass  $\mathcal{D}_B = \mathcal{D}$ , und zwar für alle  $B \in \mathcal{D}$ . Somit ist  $\mathcal{D}$  schnittstabil.

SCHRITT 2. Wir zeigen, dass  $\mathcal{D}$  eine  $\sigma$ -Algebra ist. Seien dazu  $A_1, A_2, \dots \in \mathcal{D}$  nicht unbedingt disjunkte Mengen. Wir wollen zeigen, dass  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$ . Dazu schreiben wir

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \cap A_1^c) \cup (A_3 \cap A_1^c \cap A_2^c) \cup \dots = \bigcup_{n=1}^{\infty} (A_n \cap A_1^c \cap \dots \cap A_{n-1}^c).$$

Die Vereinigung auf der rechten Seite ist disjunkt. Außerdem gilt  $A_n \cap A_1^c \cap \dots \cap A_{n-1}^c \in \mathcal{D}$ , denn  $\mathcal{D}$  ist schnittstabil (was wir in Schritt 1 gezeigt haben) und komplementstabil (als Dynkin-System). Aus der Dynkin-Eigenschaft von  $\mathcal{D}$  folgt nun, dass  $\cup_{i=1}^{\infty} A_i \in \mathcal{D}$ . Wir haben gezeigt, dass  $\mathcal{D}$  eine  $\sigma$ -Algebra ist.  $\square$

BEWEIS VON SATZ 4.12.4. Seien  $\mu, \nu$  zwei Wahrscheinlichkeitsmaße auf  $(\Omega, \mathcal{F})$ , wobei  $\mathcal{F} = \sigma(\mathcal{E})$  und der Erzeuger  $\mathcal{E}$  schnittstabil sei. Betrachte das Mengensystem

$$\mathcal{D}_0 := \{A \in \mathcal{F} : \mu(A) = \nu(A)\}.$$

Zu zeigen ist, dass  $\mathcal{D}_0 = \mathcal{F}$ .

Zuerst zeigen wir, dass  $\mathcal{D}_0$  ein Dynkin-System ist.

- (1)  $\Omega \in \mathcal{D}_0$ , denn  $\mu(\Omega) = \nu(\Omega) = 1$ .
- (2) Aus  $\mu(A) = \nu(A)$  folgt, dass  $\mu(A^c) = 1 - \mu(A) = 1 - \nu(A) = \nu(A^c)$ . Somit gilt auch  $A^c \in \mathcal{D}_0$ .
- (3) Seien  $A_1, A_2, \dots$  **disjunkte** Mengen mit  $\mu(A_i) = \nu(A_i)$  für alle  $i \in \mathbb{N}$ . Aus der  $\sigma$ -Additivität von  $\mu$  und  $\nu$  folgt, dass

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

Also gilt auch  $\cup_{i=1}^{\infty} A_i \in \mathcal{D}_0$ .

Wir haben gezeigt, dass  $\mathcal{D}_0$  ein Dynkin-System ist, das zudem nach Voraussetzung  $\mathcal{E}$  enthält. Auf der anderen Seite ist  $\mathcal{D} = \mathcal{D}(\mathcal{E})$  das *minimale* Dynkin-System, das  $\mathcal{E}$  enthält. Es folgt, dass  $\mathcal{D}(\mathcal{E}) \subset \mathcal{D}_0$ . Das Dynkin-Lemma behauptet aber, dass  $\mathcal{D}(\mathcal{E})$  sogar eine  $\sigma$ -Algebra ist, woraus folgt, dass  $\mathcal{F} = \sigma(\mathcal{E}) \subset \mathcal{D}(\mathcal{E}) \subset \mathcal{D}_0$ . Wir haben gezeigt, dass  $\mathcal{D} = \mathcal{F}$ , also stimmen die Wahrscheinlichkeitsmaße  $\mu$  und  $\nu$  auf  $\mathcal{F}$  überein.  $\square$

Zum Schluss erwähnen wir eine Erweiterung von Satz 4.12.4 auf unendliche Maße.

**Aufgabe 4.12.11.** Seien  $\mu, \nu$  zwei unendliche Maße auf einem Messraum  $(\Omega, \mathcal{F})$ , wobei  $\mathcal{F} = \sigma(\mathcal{E})$  und  $\mathcal{E} \subset \mathcal{F}$  ein schnittstabiles Mengensystem sei. Es gelte  $\mu(A) = \nu(A)$  für alle  $A \in \mathcal{E}$ . Außerdem gebe es eine Darstellung  $\Omega = \cup_{i=1}^{\infty} E_i$  mit  $E_i \in \mathcal{E}$  und  $\mu(E_i) < \infty$  für alle  $i \in \mathbb{N}$ . Zeigen Sie, dass dann  $\mu = \nu$ .

**Aufgabe 4.12.12.** Seien  $\mu$  ein Maß auf  $(\mathbb{R}, \mathcal{B})$  mit  $\mu([a, b]) = b - a$  für jedes Intervall  $[a, b]$ . Zeigen Sie:  $\mu$  ist das Lebesgue-Maß.

**Aufgabe 4.12.13.** Konstruieren Sie zwei verschiedene  $\sigma$ -endliche Maße  $\mu$  und  $\nu$  auf  $(\mathbb{R}, \mathcal{B})$ , so dass  $\mu([a, b]) = \nu([a, b])$  für jedes Intervall  $[a, b]$ . Somit kann man in der vorherigen Aufgabe auf die Bedingung  $E_i \in \mathcal{E}$  nicht verzichten.

### 4.13. Limes superior und Limes inferior für Folgen von Mengen

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

**Definition 4.13.1.** Seien  $A_1, A_2, \dots \subset \Omega$ . Dann ist  $\limsup_{n \rightarrow \infty} A_n$  eine Teilmenge von  $\Omega$ , 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  $\limsup A_n$  wie folgt beschreiben:

$\limsup A_n =$  "Das Ereignis  $A_i$  tritt für unendliche viele  $i \in \mathbb{N}$  ein".

**Definition 4.13.2.** Seien  $A_1, A_2, \dots \subset \Omega$ . Dann ist  $\liminf_{n \rightarrow \infty} A_n$  eine Teilmenge von  $\Omega$ , die wie folgt definiert wird:

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

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

$\liminf A_n =$  "Das Ereignis  $A_i$  tritt für alle bis auf endlich viele Werte von  $i \in \mathbb{N}$  ein".

**Beispiel 4.13.3.** 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 4.13.4.** Es gilt:

- (1)  $\liminf A_n \subset \limsup A_n$ .
- (2)  $(\liminf A_n)^c = \limsup(A_n^c)$  und  $(\limsup A_n)^c = \liminf(A_n^c)$ .

**Bemerkung 4.13.5.** Sind  $A_1, A_2, \dots \in \mathcal{F}$  und ist  $\mathcal{F}$  eine  $\sigma$ -Algebra, so gilt auch  $\limsup A_n \in \mathcal{F}$  und  $\liminf A_n \in \mathcal{F}$ .

**Bemerkung 4.13.6.** Die Indikatorfunktion von  $\limsup A_n$  ist  $\limsup$  von Indikatorfunktionen:

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

Analog für  $\liminf$ :

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

**Aufgabe 4.13.7.** Seien  $B$  und  $C$  zwei Ereignisse und betrachte die Folge

$$A_n = \begin{cases} B, & \text{falls } n \text{ gerade ist,} \\ C, & \text{falls } n \text{ ungerade ist.} \end{cases}$$

Zeigen Sie, dass  $\limsup_{n \rightarrow \infty} A_n = B \cup C$  und  $\liminf_{n \rightarrow \infty} A_n = B \cap C$ .

#### 4.14. Das Lemma von Borel–Cantelli

Es seien  $A_1, A_2, \dots$  Ereignisse. Das Lemma von Borel–Cantelli beantwortet die Frage nach der Wahrscheinlichkeit des Ereignisses

$$\limsup A_n = \text{“Es treten unendlich viele } A_n \text{ ein”}.$$

**Lemma 4.14.1** (Borel–Cantelli). Es sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und betrachte Ereignisse  $A_1, A_2, \dots \in \mathcal{F}$ .

(1) Ist  $\sum_{i=1}^{\infty} \mathbb{P}[A_i] < \infty$ , so gilt  $\mathbb{P}[\limsup A_n] = 0$ . Mit anderen Worten gilt

$$\mathbb{P}[\text{“Es treten unendlich viele } A_n \text{ ein”}] = 0.$$

(2) Ist  $\sum_{i=1}^{\infty} \mathbb{P}[A_i] = \infty$  und sind die Ereignisse  $A_1, A_2, \dots$  unabhängig, so gilt  $\mathbb{P}[\limsup A_n] = 1$ . Mit anderen Worten gilt

$$\mathbb{P}[\text{“Es treten unendlich viele } A_n \text{ ein”}] = 1.$$

**Beweis des Lemmas von Borel–Cantelli, Teil 1.** Definiere für jedes  $k \in \mathbb{N}$  die Menge  $B_k = \cup_{n \geq k} A_n$ . Dann sind  $B_1, B_2, \dots$  messbar als abzählbare Vereinigungen messbarer Mengen und es gilt  $B_1 \supset B_2 \supset \dots$ . Mit der Definition von  $\limsup$  und mit dem Stetigkeitsatz 4.10.13 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  $\sigma$ -Subadditivität (Satz 4.10.9):

$$\lim_{k \rightarrow \infty} \mathbb{P}[B_k] \leq \lim_{k \rightarrow \infty} \sum_{n \geq k} \mathbb{P}[A_n] = 0.$$

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  $A_1, A_2, \dots$  unabhängige Ereignisse mit  $p_n = \mathbb{P}[A_n]$  und  $p_1 + p_2 + \dots = \infty$ . Für  $k, m \in \mathbb{N}$  definiere das Ereignis

$$B_{k,m} = \cup_{n=k}^{k+m} A_n.$$

Die Mengen  $B_{k,m}$  sind messbar als endliche Vereinigungen messbarer Mengen. Sei nun  $k \in \mathbb{N}$  fest. Es gilt dann

$$B_{k,1} \subset B_{k,2} \subset \dots \text{ und } \cup_{m=1}^{\infty} B_{k,m} = \cup_{n=k}^{\infty} A_n \stackrel{\text{def}}{=} B_k.$$

Im Folgenden wollen wir zeigen, dass  $\mathbb{P}[B_k] = 1$ . Aus der Stetigkeit der Wahrscheinlichkeit (Satz 4.10.13) folgt, 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: (Beweis?????)

$$\mathbb{P}[B_k] \geq \liminf_{m \rightarrow \infty} \left( 1 - \prod_{n=k}^{k+m} e^{-p_n} \right) = \liminf_{m \rightarrow \infty} (1 - e^{-(p_k + p_{k+1} + \dots + p_{k+m})}) = 1,$$

wobei der letzte Schritt aus  $p_k + p_{k+1} + \dots = \infty$  folgt. Wir haben somit 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 4.10.11), 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.  $\square$

**Beispiel 4.14.2.** 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.** Alle Ereignisse  $A_i$  haben die gleiche Wahrscheinlichkeit  $\mathbb{P}[A_i] = p > 0$ . 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 mit  $p > 0$  unendlich viele Erfolge erzielt, ist 1. Dies ist im Einklang mit der Intuition: Eine Münze unendlich oft geworfene Münze zeigt “Kopf” in unendlich vielen Würfen mit Wahrscheinlichkeit 1. Analog zeigt man, dass

$$\mathbb{P}[\limsup(A_n^c)] = 1.$$

In Worten: Die Wahrscheinlichkeit, dass man in einem unendlich oft wiederholten Bernoulli-Experiment mit  $p < 1$  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 ebenfalls im Einklang mit der Intuition: Die Wahrscheinlichkeit, dass die Münze beim unendlichen Münzwurf ab irgendwann nur noch “Kopf” zeigt, ist 0. Analog gilt

$$\mathbb{P}[\liminf(A_n^c)] = \mathbb{P}[\text{“Ab irgendwann nur noch Misserfolge”}] = 0.$$

Das obige Beispiel hat eine interessante Interpretation.

**Beispiel 4.14.3** (Affe und Schreibmaschine). Ein Affe bekommt eine Schreibmaschine und beginnt, zufällig auf Tasten einzuschlagen. Dadurch entsteht ein zuerst sinnlos erscheinender unendlich langer Text:

5LWjY Z 3v k DQ nqz ux hpH kU r V 9 CjzbB 3J GX8 8i h iE bTaY Il33iBp  
 ABC 4 kc Kl PyY 4u C mK5 NF LM k qnhCwhcnry J5wJ u4zDTA O6w  
 6y t8T0V fR0AKgd jRD H OQo xD Kky Q Kn hYU V ISORU WO Nt K4  
 a HOGO 36E v LXkgr rb4epri.....

Wir behaupten nun, dass dieser Text mit Wahrscheinlichkeit 1 einen beliebigen zuvor vorgegebenen Satz, z.B. “Cogito ergo sum” enthält, und zwar sogar unendlich oft! Um dies zu begründen, nehmen wir an, dass die einzelnen Tastenanschläge stochastisch unabhängig voneinander erfolgen und dass jede Taste eine positive Wahrscheinlichkeit hat, betätigt zu werden. Die Wahrscheinlichkeit, dass bereits die ersten 15 Tastenanschläge den erwünschten Satz liefern, ergibt sich als Produkt der Wahrscheinlichkeiten der einzelnen Buchstaben C,o,g,... Diese Wahrscheinlichkeit  $p$  ist zwar klein, aber positiv. Hat es bei den ersten 15 Tastenanschlägen nicht geklappt, so kann der Affe hoffen, in den nächsten 15 Tastenanschlägen einen Erfolg zu erzielen. Insgesamt hat der Affe unendlich viele Versuche, jeder Versuch besteht aus 15 zufälligen Tastenanschlägen und hat eine positive Erfolgswahrscheinlichkeit. Es folgt aus dem Lemma von Borel-Cantelli, dass es in dieser Serie mit Wahrscheinlichkeit 1 unendlich viele Erfolge geben wird. Diese Überlegung kann man auf jeden endlichen Text anwenden.

Die Wahrscheinlichkeit, dass der vom Affen produzierte Text jede endliche Buchstabenfolge unendlich oft enthält, ist 1.

Wir werden dieses Beispiel noch einmal beim Satz von Borel über normale Zahlen aufgreifen.

**Beispiel 4.14.4.** Es seien  $A_1, A_2, \dots$  unabhängige Ereignisse mit

$$\mathbb{P}[A_n] = \frac{1}{n^\alpha},$$





## Unabhängigkeit und bedingte Wahrscheinlichkeiten

### 5.1. Unabhängigkeit von zwei Ereignissen

**Beispiel 5.1.1.** Nehmen wir an, dass  $1/6$  aller Bewohner von Münster Studenten sind. Außerdem nehmen wir an, dass  $1/7$  aller Bewohner von Münster an einem Dienstag geboren sind. Definieren wir die Ereignisse

$A :=$  “ein zufälliger Bewohner von Münster ist ein Student”,

$B :=$  “ein zufälliger Bewohner von Münster wurde an einem Dienstag geboren”,

so gilt

$$\mathbb{P}[A] = \frac{1}{6}, \quad \mathbb{P}[B] = \frac{1}{7}.$$

Es erscheint plausibel, dass die beiden Eigenschaften “unabhängig” voneinander sind. Mathematisch bedeutet das, dass der Anteil der am Dienstag geborenen unter allen Bewohnern von Münster genauso groß ist, wie unter allen Münsteraner Studenten. Somit können wir die Wahrscheinlichkeit, dass ein zufälliger Bewohner von Münster ein am Dienstag geborener Student ist, wie folgt berechnen:

$$\mathbb{P}[A \cap B] = \frac{1}{6} \cdot \frac{1}{7} = \frac{1}{42}.$$

Dieses Beispiel führt zur folgenden Definition.

**Definition 5.1.2.** Seien  $(\Omega, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $A, B \in \mathcal{F}$  zwei Ereignisse. Dann heißen  $A$  und  $B$  *unabhängig*, wenn

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]. \quad (5.1.1)$$

Ansonsten heißen die Ereignisse  $A$  und  $B$  *abhängig*.

**Beispiel 5.1.3.** 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? Auf den ersten Blick scheint es so zu sein, dass beide Ereignisse vom Ergebnis der Wurfes abhängen, was für eine Abhängigkeit

spricht. Wir werden uns aber von dieser Überlegung nicht in die Irre führen lassen und einfach nachrechnen, ob Bedingung (5.1.1) gilt. 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.

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 5.1.4.** Das sichere Ereignis  $\Omega$  ist von jedem Ereignis  $A$  unabhängig, denn  $\mathbb{P}[\Omega] = 1$  und somit

$$\mathbb{P}[\Omega \cap A] = \mathbb{P}[A] = \mathbb{P}[A] \cdot \mathbb{P}[\Omega].$$

Das unmögliche Ereignis  $\emptyset$  ist ebenfalls von jedem Ereignis  $A$  unabhängig, denn  $\mathbb{P}[\emptyset] = 0$  und somit

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

Die beiden Behauptungen sind nicht überraschend. Zum Beispiel tritt das sichere Ereignis  $\Omega$  immer ein, unabhängig davon, ob irgend ein anderes Ereignis  $A$  eintritt oder nicht.

Auch der nächste Satz erscheint natürlich.

**Satz 5.1.5.** Es seien  $A, B \in \mathcal{F}$  zwei unabhängige Ereignisse. Dann gilt:

- (i)  $A$  und  $B^c$  sind unabhängig.
- (ii)  $A^c$  und  $B$  sind unabhängig.
- (iii)  $A^c$  und  $B^c$  sind unabhängig.

**Beweis.** Wir zeigen beispielhaft, 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.  $\square$

**Bemerkung 5.1.6.** Disjunktheit und Unabhängigkeit sind völlig verschiedene Begriffe. Zwei Eigenschaften sind disjunkt, wenn sie einander ausschließen. Das Ausschließen ist aber eine Art von Abhängigkeit! Wir werden diese Überlegung nun bestätigen indem wir zeigen, dass disjunkte Ereignisse immer abhängig sind, abgesehen vom degenerierten Fall in dem eines der Ereignisse Wahrscheinlichkeit 0 besitzt. Seien  $A$  und  $B$  disjunkte Ereignisse (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.

## 5.2. Unabhängigkeit von mehreren Ereignissen

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

**Definition 5.2.1.** Eine Familie  $A_1, A_2, \dots \subset \Omega$  von Ereignissen heißt *unabhängig*, falls

$$\mathbb{P}[A_{i_1} \cap \dots \cap A_{i_k}] = \mathbb{P}[A_{i_1}] \cdot \dots \cdot \mathbb{P}[A_{i_k}] \quad (5.2.1)$$

für alle  $k \in \mathbb{N}$  und alle Indizes  $i_1 < \dots < i_k$ .

Man beachte, dass in der obigen Definition die Gültigkeit der Produktformel (5.2.1) nicht nur für alle Paare von Ereignissen, sondern auch für alle Tripel, Quadrupel, usw. gefordert wird. Würde man die Gültigkeit der Produktformel nur für alle Paare von Ereignissen fordern, so würde man einen den Begriff der *paarweisen Unabhängigkeit* erhalten. Wir werden zeigen, dass die beiden Begriffe *nicht* äquivalent sind.

**Definition 5.2.2.** Ereignisse  $A_1, A_2, \dots \subset \Omega$  heißen *paarweise unabhängig*, falls

$$\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i] \cdot \mathbb{P}[A_j] \quad (5.2.2)$$

für alle  $i, j \in \mathbb{N}$  mit  $i \neq j$ .

Unabhängige Ereignisse sind paarweise unabhängig. Das folgt direkt aus den beiden Definitionen. Wir wollen nun an einem Beispiel zeigen, dass die Umkehrung im Allgemeinen falsch ist.

**Beispiel 5.2.3** (Drei Ereignisse, die paarweise unabhängig aber nicht unabhängig sind). Wir betrachten das Würfeln mit 3 fairen Würfeln. Die Würfel muss man sich wie immer als unterscheidbar vorstellen. Die Grundmenge dieses Experiments ist  $\Omega = \{1, \dots, 6\}^3$  mit  $\#\Omega = 6^3$ . Wir bezeichnen mit  $X_1, X_2, X_3$  die Zahlen, die diese drei Würfel zeigen. Also sind  $X_1, X_2, X_3 : \Omega \rightarrow \mathbb{R}$  Funktionen mit

$$X_i(a_1, a_2, a_3) = a_i = \text{“Augenzahl, die der } i\text{-te Würfel zeigt”}, \quad i \in \{1, 2, 3\},$$

wobei  $(a_1, a_2, a_3) \in \{1, \dots, 6\}^3$ . Wir betrachten die folgenden drei Ereignisse  $A, B$  und  $C$ :

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

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

$$C = \{X_3 = X_1\} = \text{“der dritte und der erste Würfel zeigen die gleiche Augenzahl”}.$$

Wir behaupten, dass diese drei Ereignisse paarweise unabhängig aber nicht unabhängig sind. Als Teilmengen von  $\Omega$  können wir diese Ereignisse wie folgt darstellen:

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

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

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

Somit gilt  $\#A = \#B = \#C = 6^2$  und folglich

$$\mathbb{P}[A] = \mathbb{P}[B] = \mathbb{P}[C] = \frac{1}{6}.$$

Wir zeigen, dass die Ereignisse  $A, B, C$  paarweise unabhängig sind. Diese Behauptung mag auf den ersten Blick überraschend erscheinen, denn sowohl  $A$  als auch  $B$  hängen von der zweiten Augenzahl  $X_2$  ab. Nichtsdestotrotz sind die Ereignisse  $A$  und  $B$  unabhängig, wie wir gleich zeigen werden. Es gilt

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

und somit

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

Analog überzeugt man sich von der Gültigkeit der Produktformel für  $\mathbb{P}[B \cap C]$  und  $\mathbb{P}[C \cap A]$ . Somit sind die Ereignisse  $A, B, C$  paarweise unabhängig.

Wir zeigen nun, dass die Familie  $A, B, C$  nicht unabhängig ist. Der Schnitt der drei Ereignisse ist gegeben durch

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

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

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

Also ist die Familie  $A, B, C$  nicht unabhängig.

**Bemerkung 5.2.4.** Von nun an wird nur noch der Begriff der Unabhängigkeit benutzt. Die paarweise Unabhängigkeit spielt keine Rolle in der Zukunft.

**Satz 5.2.5.** Es seien  $A, B, C \subset \Omega$  unabhängige Ereignisse. Dann gilt

- (i)  $A$  ist unabhängig von  $B \cup C$ .
- (ii)  $A$  ist unabhängig von  $B \cap C$ .

**Beweis.** Übung. □

**Beispiel 5.2.6.** Seien  $A_1, A_2, A_3$  drei unabhängige Ereignisse mit  $\mathbb{P}[A_1] = p_1$ ,  $\mathbb{P}[A_2] = p_2$  und  $\mathbb{P}[A_3] = p_3$ . Wir bestimmen die Wahrscheinlichkeit von  $A_1 \cup A_2 \cup A_3$ .

**Lösung.** Mit der Siebformel gilt

$$\mathbb{P}[A_1 \cup A_2 \cup A_3] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \mathbb{P}[A_3] - \mathbb{P}[A_1 \cap A_2] - \mathbb{P}[A_2 \cap A_3] - \mathbb{P}[A_3 \cap A_1] + \mathbb{P}[A_1 \cap A_2 \cap A_3].$$

Wegen der Unabhängigkeit der Ereignisse können wir die Wahrscheinlichkeiten der Schnitte als Produkte der Einzelwahrscheinlichkeiten schreiben:

$$\mathbb{P}[A_1 \cup A_2 \cup A_3] = p_1 + p_2 + p_3 - p_1p_2 - p_2p_3 - p_3p_1 + p_1p_2p_3.$$

□

**Aufgabe 5.2.7.** Seien  $A_1, A_2, A_3, A_4$  unabhängige Ereignisse mit Wahrscheinlichkeiten  $p_1, p_2, p_3, p_4$ . Bestimmen Sie die Wahrscheinlichkeit, dass genau zwei dieser Ereignisse eintreten.

**Aufgabe 5.2.8.** Seien  $A, B, C$  unabhängige Ereignisse.

- (a) Zeigen Sie, dass  $A$  unabhängig von  $B \setminus C$  ist.
- (b) Zeigen Sie, dass  $A$  unabhängig von  $B \Delta C$  ist.

### 5.3. Unabhängigkeit von Familien von Ereignissen und der Blockungssatz

Aus der Unabhängigkeit der Ereignisse  $A, B, C$  folgt die Unabhängigkeit von  $A$  und  $B \cup C$ . Wir wollen nun diese Aussage verallgemeinern indem wir größere Familien von Ereignissen und beliebige mengentheoretische Operationen zulassen.

**Beispiel 5.3.1.** Seien  $A_1, \dots, A_7$  unabhängige Ereignisse. Wir können diese Ereignisse in disjunkte “Blöcke” zerlegen, z.B.  $\{A_1, A_5, A_7\}$  und  $\{A_2, A_3, A_4, A_6\}$ . Danach können wir auf die Ereignisse in jedem Block beliebige mengentheoretische Operationen anwenden. Z.B. können wir aus dem ersten “Block” das Ereignis  $A_1 \cap (A_5 \cup A_7^c)$ , und aus dem zweiten Block das Ereignis  $A_2 \cup (A_3 \setminus A_4) \cup (A_4 \cap A_6)^c$  bilden. Die Intuition suggeriert, dass diese Ereignisse unabhängig sein sollten. Dabei ist die Tatsache entscheidend, dass es keine Überschneidungen zwischen den Blöcken geben darf. So sind etwa die Ereignisse  $A_1 \cap A_2$  und  $A_1 \cap A_3$  im Allgemeinen abhängig, da sie beide von  $A_1$  “abhängen” (obwohl es auch Ausnahmesituationen geben kann, wo auch diese Ereignisse unabhängig sind, siehe Beispiel 5.2.3).

Im Folgenden werden wir den sogenannten Blockungssatz beweisen, das die obige Intuition bestätigt. Zuerst bedarf es allerdings einiger vorbereitender Definitionen. Wir halten einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$  fest und erinnern den Leser daran, dass eine Familie von Ereignissen eine Teilmenge von  $\mathcal{F}$  ist. Elemente einer solchen Familie sind Ereignisse.

**Definition 5.3.2.** Ein Ereignis  $A \in \mathcal{F}$  ist unabhängig von einer Familie  $\mathcal{B} \subset \mathcal{F}$ , wenn für jedes  $B \in \mathcal{B}$  die Ereignisse  $A$  und  $B$  unabhängig sind.

**Beispiel 5.3.3.** Sie werfen eine Münze und definieren das Ereignis  $A :=$  “Münze zeigt Kopf”. Eine andere Person würfelt mit 20 Würfeln und betrachtet alle möglichen Ereignisse, die mit diesen Würfeln zusammenhängen, z.B.

$$B_1 := \text{“der erste Würfel zeigt 6”}, \quad B_2 := \text{“die Augensumme aller Würfe ist } < 100\text{”}, \quad \dots$$

Dann ist  $A$  von der Familie  $\mathcal{B} := \{B_1, B_2, \dots\}$  unabhängig. Es sei hervorgehoben, dass dabei Abhängigkeiten innerhalb der Familie  $\mathcal{B}$  zugelassen sind.

**Satz 5.3.4.** Sei  $A \in \mathcal{F}$  ein von einer  $\cap$ -stabilen Familie  $\mathcal{B}$  unabhängiges Ereignis. Dann ist  $A$  sogar von der  $\sigma$ -Algebra  $\sigma(\mathcal{B})$  unabhängig.

**Beweis.** Wir betrachten die Familie  $\mathcal{D} := \{D \in \mathcal{F} : A \text{ und } D \text{ sind unabhängig}\}$ . Laut Voraussetzung gilt  $\mathcal{B} \subset \mathcal{D}$ . Außerdem überzeugt man sich leicht, dass  $\mathcal{D}$  ein Dynkin-System

ist (Übung). Aus dem Dynkin-Lemma folgt unter Berücksichtigung der  $\cap$ -Stabilität von  $\mathcal{B}$ , dass  $\sigma(\mathcal{B}) \subset \mathcal{D}$ .  $\square$

Die nächste Aufgabe zeigt, dass man auf die Bedingung der  $\cap$ -Stabilität nicht verzichten kann.

**Aufgabe 5.3.5.** Konstruieren Sie drei Ereignisse  $A, B, C$ , so dass  $A$  von  $\{B, C\}$  unabhängig ist, allerdings die Ereignisse  $A$  und  $B \cap C$  abhängig sind.

Nun werden wir die vorherige Definition ein Stück weiter verallgemeinern.

**Definition 5.3.6.** Seien  $\mathcal{A} \subset \mathcal{F}$  und  $\mathcal{B} \subset \mathcal{F}$  zwei Familien von Ereignissen. Wir sagen, dass  $\mathcal{A}$  von  $\mathcal{B}$  unabhängig ist, wenn jedes Ereignis  $A \in \mathcal{A}$  von jedem Ereignis  $B \in \mathcal{B}$  unabhängig ist.

Dabei sind Abhängigkeiten innerhalb der Familien  $\mathcal{A}$  und  $\mathcal{B}$  zugelassen.

**Beispiel 5.3.7.** Eine Person wirft 20 Münzen und fasst alle möglichen Ereignisse, die mit diesen Münzen zusammenhängen, zu einer Familie  $\mathcal{A}$  zusammen. Eine andere Person wirft 10 Würfel und fasst alle möglichen Ereignisse, die mit diesen Würfeln zusammenhängen, zu einer Familie  $\mathcal{B}$  zusammen. Dann ist  $\mathcal{A}$  unabhängig von  $\mathcal{B}$ .

**Satz 5.3.8.** Die Familien  $\mathcal{A}$  und  $\mathcal{B}$  seien unabhängig voneinander und  $\cap$ -stabil. Dann sind auch die von diesen Familien erzeugten  $\sigma$ -Algebren  $\sigma(\mathcal{A})$  und  $\sigma(\mathcal{B})$  unabhängig voneinander.

**Beweis.** Sei  $A \in \mathcal{A}$  ein Ereignis. Eine Anwendung von Satz 5.3.4 auf dieses Ereignis und die Familie  $\mathcal{B}$  ergibt, dass  $A$  unabhängig von  $\sigma(\mathcal{B})$  ist. Dies gilt für jedes  $A \in \mathcal{A}$ , also ist  $\mathcal{A}$  unabhängig von  $\sigma(\mathcal{B})$ . Sei nun  $B \in \sigma(\mathcal{B})$  ein Ereignis. Eine Anwendung von Satz 5.3.4 auf dieses Ereignis und die Familie  $\mathcal{A}$  ergibt, dass  $B$  von  $\sigma(\mathcal{A})$  unabhängig ist. Dies gilt für jedes  $B \in \sigma(\mathcal{B})$ . Also sind  $\sigma(\mathcal{A})$  und  $\sigma(\mathcal{B})$  unabhängig.  $\square$

**Satz 5.3.9** (Blockungssatz mit zwei Blöcken). Es sei  $\mathcal{A}$  eine Familie von unabhängigen Ereignissen und  $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$  eine disjunkte Zerlegung von  $\mathcal{A}$  in zwei Teilfamilien. Dann sind die Familien  $\sigma(\mathcal{B})$  und  $\sigma(\mathcal{C})$  unabhängig voneinander.

**Beweis.** Die Familien  $\mathcal{B}$  und  $\mathcal{C}$  müssen nicht  $\cap$ -stabil sein. Wir erweitern sie deshalb zu folgenden  $\cap$ -stabilen Familien:

$$\begin{aligned}\bar{\mathcal{B}} &:= \{B_{i_1} \cap \dots \cap B_{i_n} : n \in \mathbb{N}, B_{i_1}, \dots, B_{i_n} \in \mathcal{B}\}, \\ \bar{\mathcal{C}} &:= \{C_{j_1} \cap \dots \cap C_{j_m} : m \in \mathbb{N}, C_{j_1}, \dots, C_{j_m} \in \mathcal{C}\}.\end{aligned}$$

Dann sind  $\bar{\mathcal{B}}$  und  $\bar{\mathcal{C}}$   $\cap$ -stabil. Außerdem ist  $\bar{\mathcal{B}}$  unabhängig von  $\bar{\mathcal{C}}$  (Übung). Mit Satz 5.3.8 folgt, dass  $\sigma(\bar{\mathcal{B}})$  unabhängig von  $\sigma(\bar{\mathcal{C}})$  ist. Da aber  $\sigma(\bar{\mathcal{B}}) = \sigma(\mathcal{B})$  und  $\sigma(\bar{\mathcal{C}}) = \sigma(\mathcal{C})$ , ergibt sich die Behauptung des Satzes.  $\square$

**Beispiel 5.3.10.** Es sei  $A_1, A_2, \dots$  eine Folge unabhängiger Ereignisse. Dann sind die  $\sigma$ -Algebren  $\sigma(A_1, A_3, A_5, \dots)$  und  $\sigma(A_2, A_4, A_6, \dots)$  unabhängig voneinander. Z.B. sind die Ereignisse  $A_1 \cap A_3 \cap A_5 \cap \dots$  und  $A_2 \cap A_4 \cap A_6 \cap \dots$  unabhängig.

Man kann den Blockungssatz auf Zerlegungen in mehr als zwei Teilfamilien erweitern. Zuerst müssen wir definieren, was es heißt, dass mehrere Familien von Ereignissen unabhängig voneinander sind.

**Definition 5.3.11.** Familien von Ereignissen  $\mathcal{A}_1, \mathcal{A}_2, \dots$  heißen unabhängig voneinander, wenn für jedes  $n \in \mathbb{N}$  und jede beliebige Auswahl  $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$  die Ereignisse  $A_1, \dots, A_n$  unabhängig sind.

**Satz 5.3.12** (Blockungssatz). Es sei  $\mathcal{A}$  eine Familie von unabhängigen Ereignissen und  $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \dots$  eine disjunkte Zerlegung von  $\mathcal{A}$  in Teilfamilien. Dann sind die Familien  $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \dots$  unabhängig voneinander.

**Beweis.** Übung. □

**Beispiel 5.3.13.** Seien  $A_1, A_2, \dots$  unabhängige Ereignisse. Dann sind die  $\sigma$ -Algebren  $\sigma(A_1, A_2), \sigma(A_3, A_4), \sigma(A_5, A_6), \dots$  unabhängig voneinander. Z.B. sind die Ereignisse  $A_1 \cup A_2, A_3 \cup A_4, A_5 \cup A_6, \dots$  unabhängig.

Um die Notation zu vereinfachen, haben wir den Blockungssatz nur für abzählbar viele Blöcke formuliert. Der Satz bleibt für eine beliebige Mächtigkeit der Menge der Blöcke gültig.

## 5.4. Bedingte Wahrscheinlichkeiten

**Beispiel 5.4.1.** Stellen wir uns vor, dass jemand mit 2 fairen Würfeln würfelt. Die Grundmenge ist  $\Omega = \{1, \dots, 6\}^2$ . Wir betrachten zwei Ereignisse:

$$A = \text{“erster Würfel zeigt 6”} = \{(6, 1), (6, 2), \dots, (6, 6)\},$$

$$B = \text{“Augensumme = 10”} = \{(6, 4), (5, 5), (4, 6)\}.$$

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

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}. \quad (5.4.1)$$

**Bemerkung 5.4.3.** Beachte:  $A|B$  ist kein Ereignis, sondern lediglich eine Notation für eine neue Art von Wahrscheinlichkeit.

**Satz 5.4.4.** 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 (5.4.1). 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 5.4.5.** Aus (5.4.1) sieht man, dass  $\mathbb{P}[A|B]$  und  $\mathbb{P}[B|A]$  im Allgemeinen nicht gleich sein müssen.

**Satz 5.4.6** (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 5.4.7** (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 5.4.8** (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  $i = 1, \dots, n$ :

$$\mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|B_i] \cdot \mathbb{P}[B_i]}{\sum_{k=1}^n \mathbb{P}[A|B_k] \cdot \mathbb{P}[B_k]}. \quad (5.4.2)$$

**Beweis.** Wir wenden die Definition der bedingten Wahrscheinlichkeit (5.4.1) 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 (5.4.2). Die zweite Hälfte folgt aus dem Satz von der totalen Wahrscheinlichkeit. □

**Beispiel 5.4.9** (Fortsetzung von Beispiel 5.4.7). 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 bereits eingetreten. Ausgestattet mit dieser Information wollen wir wissen, mit welcher (bedingten!) Wahrscheinlichkeit das Ereignis  $B_1$  eintritt. Gefragt wird also nach der bedingten Wahrscheinlichkeit  $\mathbb{P}[B_1|A]$ . Die Bayes-Formel (5.4.2) 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:

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

**Aufgabe 5.4.10.** In einem Topf liegen 10 Münzen: davon 7 fair und 3 gezinkt. Eine gezinkte Münze zeigt “Kopf” mit Wahrscheinlichkeit 0.8. Es wird rein zufällig eine Münze aus dem Topf entnommen und geworfen. Mit welcher Wahrscheinlichkeit zeigt sie “Kopf”?

**Aufgabe 5.4.11.** In einem Topf liegen 10 identisch aussehende Münzen: davon 7 fair und 3 gezinkt. Eine gezinkte Münze zeigt “Kopf” mit Wahrscheinlichkeit 0.8. Es wird rein zufällig eine Münze aus dem Topf entnommen und fünf Mal geworfen. In vier Würfeln zeigt sie “Kopf”, in einem Wurf “Zahl”. Mit welcher (bedingten) Wahrscheinlichkeit ist diese Münze gezinkt?

**Aufgabe 5.4.12.** Wie groß ist der Anteil der Familien mit zwei Mädchen unter allen Familien mit zwei Kindern, von denen mindestens ein Kind ein Mädchen ist? (Die Wahrscheinlichkeit, dass ein Kind ein Mädchen ist, kann als  $1/2$  angenommen werden. Geschlechter von verschiedenen Kindern sind stochastisch unabhängig. Ignorieren Sie weitere Einflüsse, wie z.B. unterschiedliche Sterbewahrscheinlichkeiten).

**Aufgabe 5.4.13.** In Ihrem E-Mail Postfach befinden sich zwei Arten von Mails: Spam und Nicht-Spam. Erfahrungsgemäß seien (leider) 80% Ihrer Mails Spam. In einer Spam-Mail kommt das Wort “money” mit Wahrscheinlichkeit 0.6 vor und in einer Nicht-Spam-Mail mit Wahrscheinlichkeit 0.1.

- (1) Sie bekommen eine E-Mail. Bestimmen Sie die Wahrscheinlichkeit, dass sie das Wort “money” enthält.
- (2) Sie bekommen eine E-Mail. Mit welcher Wahrscheinlichkeit ist es eine das Wort “money” enthaltende Spam-Mail?
- (3) Sie bekommen eine E-Mail, die das Wort “money” enthält. Mit welcher Wahrscheinlichkeit ist diese Mail Spam?

**Aufgabe 5.4.14.** Seien  $A$  und  $B$  zwei Ereignisse mit  $\mathbb{P}[A|B] = \mathbb{P}[B|A]$  und  $\mathbb{P}[A \cap B] \neq 0$ . Beweisen Sie, dass dann  $\mathbb{P}[A] = \mathbb{P}[B]$ .

**Beispiel 5.4.15** (Ziegenproblem). **Muss noch ergänzt werden...**

**Aufgabe 5.4.16.** In einer Fernsehshow stehen Sie vor 5 geschlossenen Türen 1, 2, 3, 4, 5. Hinter einer (zufällig ausgewählten) Tür steht ein Auto, hinter den anderen 4 sind Ziegen. Sie wollen das Auto gewinnen. Sie zeigen auf die erste Tür. Diese Tür bleibt aber vorerst

verschlossen. Der Moderator öffnet 2 der 4 verbliebenen Türen und es stellt sich heraus, dass hinter den beiden Türen Ziegen waren. Sie dürfen nun bei Ihrer Wahl bleiben und Tür 1 öffnen oder eine der zwei verbliebenen Türen öffnen. Was ist besser?

- (a) Lösen Sie diese Aufgabe unter der Annahme, dass der Moderator weiß, was hinter allen Türen versteckt ist, und dass er Ihnen helfen will, indem er nur Türen öffnet, hinter denen Ziegen sind. D.h. der Moderator wählt die beiden zu öffnenden Türen unter allen Türen, hinter denen Ziegen sind, zufällig aus.
- (b) Lösen Sie die gleiche Aufgabe unter der Annahme, dass der Moderator nicht weiß, was hinter den Türen ist, und dass er rein zufällig zwei Türen öffnet. (Falls der Moderator die Tür mit dem Auto geöffnet hätte, hätten Sie verloren. Diese Situation ist jedoch laut Aufgabenstellung nicht eingetreten).
- (c) Lösen Sie die gleiche Aufgabe unter der Annahme, dass der Moderator immer 2 Türen öffnet, die auf die von Ihnen ausgewählte Tür folgen. (Da Sie auf Tür 1 gezeigt haben, öffnet er Türen 2 und 3. Hätten Sie z.B. auf 3 gezeigt, dann hätte er 4 und 1 geöffnet).

*Hinweis:* Wenn Sie sich für einen Türwechsel entscheiden, heißt es nicht, dass beide noch nicht gewählten Türen geöffnet werden. Es wird nur eine Tür geöffnet, und zwar die, auf die Sie zeigen.

**Aufgabe 5.4.17.** Anton, Bernd und Charles sind Insassen in einem Gefängnis, das zu voll ist, um weitere Straftäter aufzunehmen. Weil die Reststrafe der Drei gering ist und sie sich vorbildlich verhalten, soll einer von ihnen frühzeitig entlassen werden. Der Gefängnischef lost auf eine faire Art und Weise eine der drei Personen aus, welche dann in der nächsten Woche frühzeitig entlassen werden soll. Das Ergebnis teilt der Chef nur dem Gefängniswächter mit, damit dieser alle Vorbereitungen treffen kann. Die drei Gefangenen Anton, Bernd und Charles haben von ihrer möglichen Entlassung gehört und können es nicht abwarten, zu erfahren, wer ausgelost wurde. Deshalb versucht Anton, vom Wächter den Namen zu erfahren. Weil der Wächter sich weigert, die Information frühzeitig auszulplaudern, schlägt Anton folgendes vor: „Nenne mir nach der folgenden Regel den Namen einer Person, die im Gefängnis bleiben muss:

- wenn Bernd frei kommt, nenne mir Charles.
- wenn Charles frei kommt, nenne mir Bernd.
- wenn ich frei komme, wirf eine faire Münze und nenne mir bei ‚Kopf‘ Bernd und bei ‚Zahl‘ Charles.“

Der Wächter denkt nach und nennt ihm den Namen Charles. Glücklicherweise geht Anton davon aus, dass seine Chancen auf eine frühzeitige Entlassung von  $\frac{1}{3}$  auf  $\frac{1}{2}$  gestiegen sind. Es bleiben ja nur er und Bernd übrig.

Hat Anton Recht? Wie groß ist mit den vorhandenen Informationen die bedingte Wahrscheinlichkeit, dass Anton freikommt? Wie groß ist mit den vorhandenen Informationen die bedingte Wahrscheinlichkeit, dass Bernd freikommt?

## 5.5. Unabhängigkeit von Zufallsvariablen

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

**Definition 5.5.1.** 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:

$$\mathbb{P}[X_1 = y_1, \dots, X_n = y_n] = \mathbb{P}[X_1 = y_1] \cdot \dots \cdot \mathbb{P}[X_n = y_n]. \quad (5.5.1)$$

Diese Definition lässt sich in folgender äquivalenter Form darstellen.

**Satz 5.5.2.** Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum  $(\Omega, \mathbb{P})$ . 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:

$$\mathbb{P}[X_1 \in B_1, \dots, X_n \in B_n] = \mathbb{P}[X_1 \in B_1] \cdot \dots \cdot \mathbb{P}[X_n \in B_n]. \quad (5.5.2)$$

- (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 5.5.3.** 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 5.5.4.** 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  $\Omega$  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, was natürlich im Einklang mit dem gesunden Verstand steht. 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 5.5.5.** In diesem Beispiel betrachten wir die Augensumme  $S = X_1 + \dots + X_n$  beim  $n$ -maligen Würfeln mit einem fairen Würfel. Wir zeigen, dass die Zufallsvariablen  $S$  und  $X_1$  abhängig sind. Zuerst eine intuitive Überlegung. Wissen wir, dass  $X_1$  einen "großen" Wert angenommen hat, so erhöht das die Chancen, dass auch die ganze Summe  $S$  "groß" sein wird. Das ist ein klarer Fall von stochastischer Abhängigkeit. Um diese Überlegung mit einem Beweis zu untermauern, werden wir zeigen, dass für die Ereignisse  $\{X_1 = 1\}$  und  $\{S = 6n\}$  die Produktformel nicht gilt. Die Wahrscheinlichkeiten der beiden Ereignisse sind

strikt positiv, nämlich

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

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

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

Die Produktformel gilt also nicht und somit sind die Zufallsvariablen  $S$  und  $X_1$  abhängig.

**Bemerkung 5.5.6.** 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.

## 5.6. Produktexperimente

**Dieser Abschnitt muss überarbeitet werden...** 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  $(\Omega, \mathbb{P})$  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 5.6.1.** 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

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

## Zufallsvariablen: Die allgemeine Definition

### 6.1. 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 6.1.1.** 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 6.1.2.** 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 6.1.3.** 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 6.1.2 benötigen wir eine Hilfsaussage.



**Proposition 6.1.4.** Seien  $(\Omega, \mathcal{F})$  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 6.1.5.** Mit anderen Worten: Um zu zeigen, dass die Urbilder aller Mengen aus einer  $\sigma$ -Algebra messbar sind, reicht es zu zeigen, dass die Urbilder aller Mengen aus einem Erzeuger dieser  $\sigma$ -Algebra messbar sind.

**Beweis von Proposition 6.1.4.** 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  $\sigma$ -Algebra ist. Außerdem gilt  $\mathcal{E} \subset \mathcal{A}$  laut Voraussetzung. Die Familie  $\mathcal{A}$  ist also eine  $\sigma$ -Algebra, die  $\mathcal{E}$  enthält. Die von  $\mathcal{E}$  erzeugte  $\sigma$ -Algebra  $\sigma(\mathcal{E})$  ist (laut Definition der erzeugten  $\sigma$ -Algebra) die *kleinste*  $\sigma$ -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  $\sigma$ -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  $\sigma$ -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  $\mathcal{A}$   $\sigma$ -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  $\sigma$ -Algebra und somit  $\sigma$ -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  $\sigma$ -Algebra ist. □

**Beweis von Satz 6.1.2.** 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 6.1.4 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- $\sigma$ -Algebra.  $\square$

**Beispiel 6.1.6.** 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  $\sigma$ -Algebra. Somit ist  $X$  messbar.  $\square$

## 6.2. Zufallsvektoren

**Definition 6.2.1.** 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 6.2.2.** 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  $\sigma$ -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- $\sigma$ -Algebra  $\mathcal{B}^d$ . Mit Proposition 6.1.4 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 6.2.3.** 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 6.2.4.** 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 6.2.5.** 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- $\sigma$ -Algebra. Mit Proposition 6.1.4 folgt, dass auch das Urbild jeder Borel-Menge eine Borel-Menge ist. Somit ist  $f$  Borel-messbar.  $\square$

**Satz 6.2.6.** 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 6.2.2 ist  $f \circ X$  ein Zufallsvektor.  $\square$

**Korollar 6.2.7.** 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 6.2.6.  $\square$

### 6.3. Verteilungsfunktion einer Zufallsvariable

Die Familie der Borel-Teilmengen von  $\mathbb{R}$  wird mit  $\mathcal{B}$  bezeichnet.

**Definition 6.3.1.** 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 6.3.2.** 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  $\sigma$ -Algebra der Borel-Teilmengen von  $\Omega$ . Als Wahrscheinlichkeitsmaß wählen wir

$$\mathbb{P}[A] = \frac{\lambda(A)}{\pi}, \quad A \in \mathcal{F},$$

wobei  $\lambda$  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 6.3.3.** 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- $\sigma$ -Algebra  $\mathcal{B}^2$  auf  $\Omega$ . 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 6.3.4.** 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 6.3.5.** 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 6.3.6.** 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 6.3.4 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  $F_X(t_1) \leq F_X(t_2)$ . □

**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 6.3.7.** 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 6.3.8** (Eindeutigkeit der Maß-Fortsetzung). Sei  $\Omega$  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  $\sigma$ -Algebra. Seien  $\mathbb{P}_1$  und  $\mathbb{P}_2$  zwei Wahrscheinlichkeitsmaße auf  $(\Omega, \mathcal{A})$  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 6.3.9.** Mit anderen Worten: stimmen zwei Wahrscheinlichkeitsmaße auf dem Erzeuger einer  $\sigma$ -Algebra überein, so stimmen sie auch auf der ganzen  $\sigma$ -Algebra überein, wenn der Erzeuger schnittstabil ist.

**Beweis von Satz 6.3.7.** 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- $\sigma$ -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 6.3.10.** 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.  $\square$

**Beispiel 6.3.11.** 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  $\mathbb{P}[X = t] = 0$ .  $\square$

**Definition 6.3.12.** 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 6.3.13.** 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 6.3.14.** 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

$$\begin{aligned} \mathbb{P}[a \leq X \leq b] &= \mathbb{P}[X \leq b] - \mathbb{P}[X < a] = F_X(b) - \lim_{s \uparrow a} F_X(s), \\ \mathbb{P}[a \leq X < b] &= \mathbb{P}[X < b] - \mathbb{P}[X < a] = \lim_{s \uparrow b} F_X(s) - \lim_{s \uparrow a} F_X(s), \\ \mathbb{P}[a < X \leq b] &= \mathbb{P}[X \leq b] - \mathbb{P}[X \leq a] = F_X(b) - F_X(a), \\ \mathbb{P}[a < X < b] &= \mathbb{P}[X < b] - \mathbb{P}[X \leq a] = \lim_{s \uparrow b} F_X(s) - F_X(a). \end{aligned}$$



## 6.4. 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  $\Omega$  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:

$$\mathbb{E}X = \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega). \quad (6.4.1)$$

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 6.4.1.** 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

$$X^+(\omega) \stackrel{\text{def}}{=} \begin{cases} X(\omega), & \text{falls } X(\omega) \geq 0, \\ 0, & \text{falls } X(\omega) < 0, \end{cases} \quad X^-(\omega) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{falls } X(\omega) \geq 0, \\ |X(\omega)|, & \text{falls } X(\omega) < 0. \end{cases}$$

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

$$\mathbb{E}X \stackrel{\text{def}}{=} \mathbb{E}X^+ - \mathbb{E}X^-.$$

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

$$\mathbb{E}X \stackrel{\text{def}}{=} +\infty.$$

FALL 3: Gilt  $\mathbb{E}X^+ < +\infty$  und  $\mathbb{E}X^- = +\infty$ , so definieren wir

$$\mathbb{E}X \stackrel{\text{def}}{=} -\infty.$$

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 6.4.2.** 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 6.4.3.** 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 6.4.4.** 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 6.4.5.** 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 6.4.6.** 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 6.4.7.** Sind  $X$  und  $Y$  fast überall gleich und eine der Zufallsvariablen integrierbar, so ist auch die andere Zufallsvariable integrierbar und es gilt  $\mathbb{E}X = \mathbb{E}Y$ .

\*\*\*\*\*

**Definition 6.4.8.** 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:

$$\mathbb{E}X \stackrel{\text{def}}{=} \sum_i p_i y_i. \tag{6.4.2}$$

**Bemerkung 6.4.9.** Nimmt  $X$  nur endlich viele Werte an, so ist die Bedingung  $\sum_i p_i |y_i| < \infty$  erfüllt und  $X$  ist integrierbar.

**Bemerkung 6.4.10.** 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 abhängt.

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

**Beispiel 6.4.11.** Betrachte die alternierende harmonische Reihe:

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \dots = \ln 2. \quad (6.4.3)$$

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:

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \dots = \frac{3}{2} \log 2. \quad (6.4.4)$$

Beweis: Übung. *Hinweis:* Betrachten Sie zuerst die Summe der ersten  $3n$  Terme in der Reihe. Wie lauten die letzten 3 Terme in dieser Summe? Sie können benutzen, dass

$$H_n := 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} = \log n + \gamma + \varepsilon_n,$$

wobei  $\gamma$  die Euler–Mascheroni–Konstante ist und  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ .

Die Summen der Reihen (6.4.3) und (6.4.4) sind unterschiedlich. Dabei haben beide Reihen die gleichen Summanden, lediglich die Reihenfolge der Summanden ist unterschiedlich.

**Bemerkung 6.4.12.** Die Definition des Erwartungswerts kann man auch wie folgt darstellen:

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

falls die Reihe absolut konvergiert.

**Satz 6.4.13.** Sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsvariable. Dann ist  $X$  integrierbar genau dann, wenn

$$\sum_{\omega \in \Omega} p(\omega) |X(\omega)| < \infty. \quad (6.4.5)$$

Ist (6.4.5) erfüllt, so gilt

$$\mathbb{E}X = \sum_{\omega \in \Omega} p(\omega) X(\omega). \quad (6.4.6)$$

**Beweis.** Für jedes  $y \in \mathfrak{S}(X)$  definieren wir das Ereignis

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

Es gilt dann:

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

Laut Definition ist  $X$  integrierbar genau dann, wenn

$$\sum_{y \in \mathfrak{S}(X)} |y| \cdot \mathbb{P}[X = y] < \infty. \quad (6.4.7)$$

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

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

Somit sind Bedingungen (6.4.5) und (6.4.7) äquivalent. Es folgt, dass  $X$  genau dann integrierbar ist, wenn (6.4.5) gilt.

Nun nehmen wir an, dass (6.4.5) gilt. Es folgt, dass

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

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

**Aufgabe 6.4.14** (Zur harmonischen Reihe). Ein Auto ist durch einen zunächst 1 Meter langen Gummistreifen mit einer Wand verbunden. Der Gummistreifen kann sich beliebig weit ausdehnen. Eine Schnecke sitzt zunächst an dem an der Wand angebrachten Ende auf dem Streifen. Nun bewegen sich das Auto und die Schnecke abwechselnd nach den folgenden Regeln:

- Erst fährt das Auto einen Meter weiter. Dabei dehnt es den Streifen weiter. Beachten Sie, dass dadurch auch die Schnecke (relativ zum Boden) bewegt wird.
- Die Schnecke kriecht danach einen Zentimeter weiter.

Danach wird der Durchgang wiederholt. Zeigen Sie, dass die Schnecke das Auto nach endlich vielen Durchgängen einholt.

*Hinweis.* Sei  $a_n$  der Abstand (in Metern) zwischen der Schnecke und der Wand vor dem  $n$ -ten Durchgang, so dass  $a_1 = 0$ . Leiten Sie eine Formel her, die  $a_{n+1}$  durch  $a_n$  darstellt.

## 6.5. Diskrete und absolut stetige Verteilungen

**Definition 6.5.1.** 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 6.5.2.** 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 6.5.3.** 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 6.5.4.** 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 6.5.5.** 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 6.5.6.** 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- $\sigma$ -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 6.5.7.** 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: $\mathbb{E}X = \int_{\mathbb{R}} y \cdot f_X(y)dy$

## 6.6. Beispiele von absolut stetigen Verteilungen

**6.6.1. Uniformverteilung auf einem Intervall.** Eine Zufallsvariable  $X$  ist uniformverteilt auf einem Intervall  $[a, b]$ , wenn ihre Dichte auf diesem Intervall konstant ist und außerhalb des Intervalls verschwindet. Der konstante Wert der Dichte auf dem Intervall ist  $1/(b - a)$  und ergibt sich aus der Bedingung  $\int_{\mathbb{R}} f_X(y)dy = 1$ .

**Definition 6.6.1.** Eine Zufallsvariable  $X$  heißt *uniformverteilt* auf einem Intervall  $[a, b]$ , wobei  $a, b \in \mathbb{R}$  und  $a < b$ , wenn die Dichte von  $X$  durch die folgende Formel gegeben ist:

$$f_X(y) = \begin{cases} \frac{1}{b-a}, & y \in [a, b], \\ 0, & y \notin [a, b]. \end{cases}$$

Wir benutzen dann die Notation  $X \sim \text{Unif}[a, b]$ .

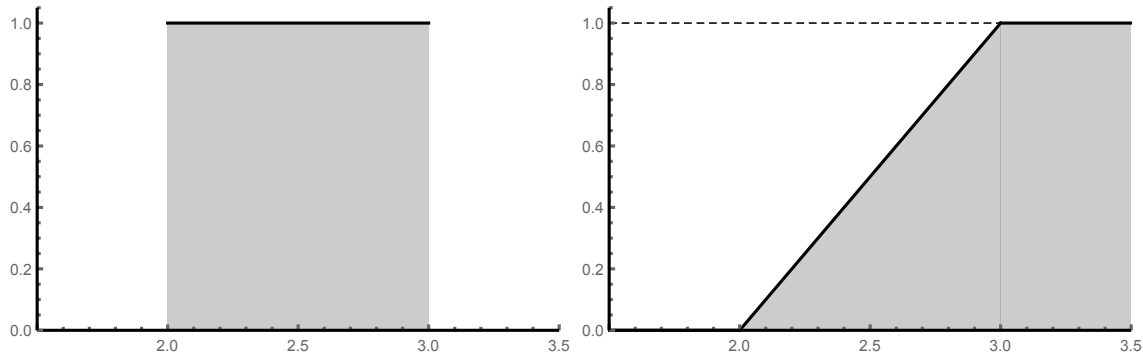


ABBILDUNG 1. Links: Dichte der Uniformverteilung auf dem Intervall  $[2, 3]$ . Rechts: Die entsprechende Verteilungsfunktion.

Die Werte außerhalb des Intervalls  $[a, b]$  können von  $X$  nicht (genauer gesagt, nur mit Wahrscheinlichkeit 0) angenommen werden, da die Dichte außerhalb von  $[a, b]$  verschwindet.

**Bemerkung 6.6.2.** Die Verteilungsfunktion einer uniformverteilten Zufallsvariable  $X \sim \text{Unif}[a, b]$  ist gegeben durch

$$F_X(t) = \int_{-\infty}^t f_X(y) dy = \begin{cases} 0, & t \leq a, \\ \frac{t-a}{b-a}, & t \in [a, b], \\ 1, & t \geq b. \end{cases}$$

Die Verteilungsfunktion ist also linear auf dem Intervall  $[a, b]$  und konstant sonst. Die Verteilungsfunktion ist überall stetig. Außerdem ist sie differenzierbar mit Ausnahme der Stellen 0 und 1. Die Ableitung der Verteilungsfunktion stimmt mit der Dichte überein, was in vollem Einklang mit dem Hauptsatz der Integralrechnung steht.

Wir bestimmen nun den Erwartungswert einer uniformverteilten Zufallsvariable. Aus Symmetriegründen sollte der Erwartungswert genau im Mittelpunkt des Intervalls  $[a, b]$  liegen, was der nächste Satz bestätigt.

**Satz 6.6.3.** Sei  $X$  eine Zufallsvariable, die auf  $[a, b]$  uniformverteilt ist. Dann gilt

$$\mathbb{E}X = \frac{a+b}{2}.$$

**Beweis.** Nach der allgemeinen Formel für den Erwartungswert gilt

$$\mathbb{E}X = \int_{\mathbb{R}} y f_X(y) dy = \int_a^b \frac{y}{b-a} dy = \frac{1}{b-a} \int_a^b y dy = \frac{1}{b-a} \left( \frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.$$

□



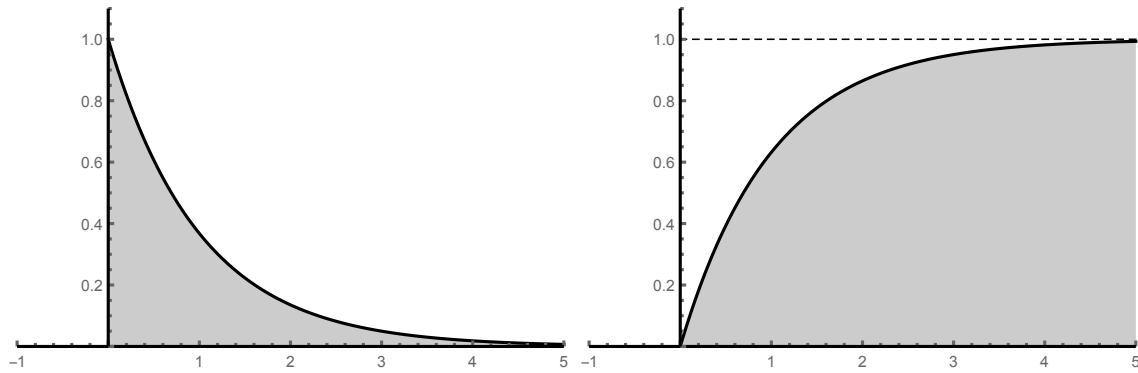


ABBILDUNG 2. Links: Dichte der Exponentialverteilung mit Parameter  $\lambda = 1$ . Rechts: Die entsprechende Verteilungsfunktion.

### 6.6.2. Exponentialverteilung.

**Definition 6.6.4.** 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:

$$f_X(y) = \begin{cases} \lambda e^{-\lambda y}, & y > 0, \\ 0, & y < 0. \end{cases}$$

Wir benutzen dann die Notation  $X \sim \text{Exp}(\lambda)$ .

Das Integral der Dichte ist gleich 1, denn  $\int_0^\infty \lambda e^{-\lambda y} dy = 1$  für  $\lambda > 0$ . Da die Dichte auf der negativen Halbachse verschwindet, kann eine exponentialverteilte Zufallsvariable nur positive Werte annehmen, d.h.  $\mathbb{P}[X > 0] = 1$ .

**Bemerkung 6.6.5.** 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  $F_X(t) = \int_{-\infty}^t f_X(s) ds = \int_{-\infty}^t 0 ds = 0$ , denn die Dichte  $f_X$  verschwindet auf der negativen Halbachse. Für  $t \geq 0$  gilt

$$F_X(t) = \int_{-\infty}^t f_X(y) dy = \int_0^t \lambda e^{-\lambda y} dy = -e^{-\lambda y} \Big|_0^t = 1 - e^{-\lambda t}.$$

□

Alternativ kann man die Exponentialverteilung durch die folgende Formel definieren:

$$\mathbb{P}[X > t] = e^{-\lambda t}, \quad t \geq 0.$$

Nun bestimmen wir den Erwartungswert der Exponentialverteilung.

**Satz 6.6.6.** Für eine mit Parameter  $\lambda > 0$  exponentialverteilte 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:

$$\mathbb{E}X = \int_0^\infty y (-e^{-\lambda y})' dy = -ye^{-\lambda y} \Big|_0^\infty + \int_0^\infty e^{-\lambda y} dy = \frac{1}{\lambda}.$$

□

Wir möchten nun eine bemerkenswerte Eigenschaft der Exponentialverteilung formulieren, die sie unter allen Verteilungen auszeichnet.

**Satz 6.6.7** (Vergessenseigenschaft der Exponentialverteilung). Sei  $X$  eine Zufallsvariable. Die folgenden zwei Eigenschaften sind äquivalent:

- (a)  $X$  ist exponentialverteilt mit einem Parameter  $\lambda > 0$ .
- (b) Für alle  $s, t > 0$  gilt:

$$\mathbb{P}[X > t + s | X > t] = \mathbb{P}[X > s].$$

**Beweis von (a)  $\Rightarrow$  (b).** 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 6.6.8** (Cauchy-Funktionalgleichung). Sei  $g : [0, \infty) \rightarrow \mathbb{R}$  eine monotone Funktion mit

$$g(t + s) = g(t) + g(s) \text{ für alle } t, s \geq 0.$$

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 (b)  $\Rightarrow$  (a) in Satz 6.6.7.** 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 6.6.8 ein  $\lambda$  mit  $g(t) = -\lambda t$  für alle  $t \geq 0$ . Dabei muss  $\lambda$  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 6.6.9** (Radioaktiver Zerfall). Bestimmte Isotope haben die Eigenschaft, spontan zu zerfallen. Solche Isotope heißen instabil oder radioaktiv. Die Lebensdauer eines instabilen Atoms ist eine Zufallsvariable, die wir mit  $X$  bezeichnen. Da man davon ausgehen muss, dass ein Atom "kein Gedächtnis" hat, muss die Lebensdauer  $X$  exponentialverteilt sein. Dabei hängt der Parameter  $\lambda$  davon ab, um welches Isotop es sich handelt. In der Physik wird neben  $\lambda$  auch ein anderer Parameter, die sogenannte Halbwertszeit benutzt. Diese ist als der Median von  $X$  definiert, also als eine Zahl  $m$  mit der Eigenschaft

$$\mathbb{P}[X > m] = \frac{1}{2}.$$

Zum Beispiel beträgt die Halbwertszeit von Cäsium-137, das durch die Reaktorunglücke von Tschernobyl und Fukushima in größeren Mengen in die Umwelt gelangte, etwa 30 Jahre. Praktisch bedeutet das, dass nach 30 Jahren aus einem Gramm Cäsium-137 nur ein halbes Gramm wird, während der Rest zerfällt und sich in andere Isotope verwandelt.

Um den Zusammenhang zwischen  $m$  und  $\lambda$  herzustellen, erinnern wir uns, dass

$$\mathbb{P}[X > m] = e^{-\lambda m} = \frac{1}{2}.$$

Löst man die Gleichung auf, so erhält man

$$m = \frac{\log(2)}{\lambda}.$$

An dieser Formel wird ersichtlich, dass sich die Halbwertszeit von der mittleren Lebensdauer  $\mathbb{E}X = \frac{1}{\lambda}$  unterscheidet.

**Beispiel 6.6.10** (Zerfallsrate). Sei  $X$  eine beliebige nichtnegative Zufallsvariable mit Dichte  $f$  und Verteilungsfunktion  $F$ . Wir können uns  $X$  z.B. als die Lebensdauer eines instabilen Atoms, eines Geräts oder einer Person denken. Wir betrachten ein sehr kleines Zeitintervall  $(t, t + dt)$  und stellen uns die folgende Frage:

Gegeben, dass das Gerät zum Zeitpunkt  $t$  noch funktioniert, wie groß ist die bedingte Wahrscheinlichkeit, dass es zwischen  $t$  und  $t + dt$  ausfällt?

Da die Länge des Zeitintervalls sehr klein ist, muss auch die Wahrscheinlichkeit sehr klein sein, vermutlich proportional zu  $dt$ . Wir können sie wie folgt berechnen:

$$\mathbb{P}[X \in (t, t + dt) | X > t] = \frac{\mathbb{P}[X \in (t, t + dt), X > t]}{\mathbb{P}[X > t]} = \frac{\mathbb{P}[X \in (t, t + dt)]}{\mathbb{P}[X > t]} \approx \frac{f_X(t)}{1 - F_X(t)} dt.$$

Wir definieren also die **Zerfallsrate** (oder die **Ausfallrate**) von  $X$  durch

$$h(t) = \frac{f(t)}{1 - F(t)}, \quad t \geq 0.$$

Bei einem instabilen Atom geht man davon aus, dass die Zerfallsrate konstant (d.h. unabhängig von  $t$ ) sein muss. Die Exponentialverteilung erfüllt diese Forderung, denn für  $X \sim \text{Exp}(\lambda)$  gilt

$$h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

**Aufgabe 6.6.11.** Seien  $X_1 \sim \text{Exp}(\lambda_1), \dots, X_n \sim \text{Exp}(\lambda_n)$  unabhängige Zufallsvariablen mit Parametern  $\lambda_1, \dots, \lambda_n > 0$ . Bestimmen Sie die Dichte von  $Z := \min(X_1, \dots, X_n)$ . Welche Verteilung hat  $Z$ ?

### 6.6.3. Normalverteilung (oder Gauß-Verteilung).

**Definition 6.6.12.** Eine Zufallsvariable  $X$  heißt *standardnormalverteilt*, wenn  $X$  absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad y \in \mathbb{R}.$$

Wir benutzen die Notation  $X \sim N(0, 1)$ .

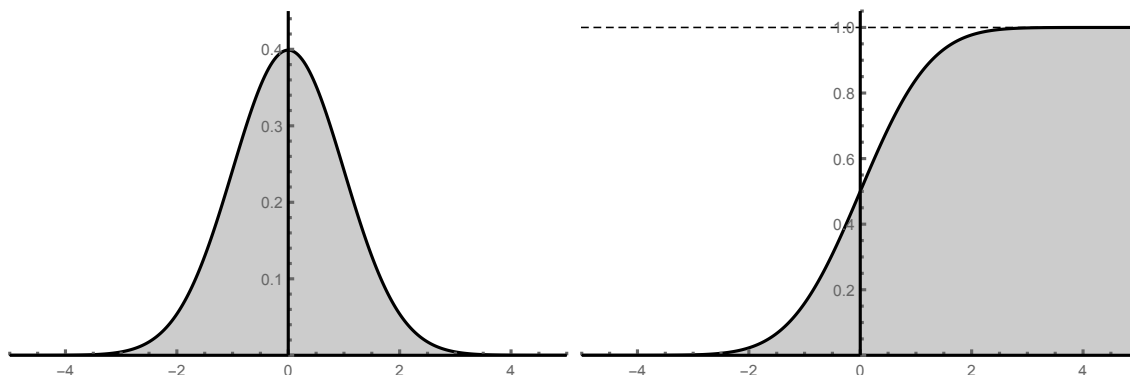


ABBILDUNG 3. Links: Dichte der Standardnormalverteilung. Rechts: Die entsprechende Verteilungsfunktion.

Die Funktion  $f_X$  ist trivialerweise nicht negativ. Aber warum ist das Integral dieser Funktion gleich 1? Das beantwortet der folgende Satz.

**Satz 6.6.13.** Es gilt  $I := \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$ .

**Beweis nach Poisson.** 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 6.6.14.** 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 6.6.15.** Aus  $X \sim N(0, 1)$  folgt für den Erwartungswert  $\mathbb{E}X = 0$ .

**Beweis.** Zuallererst bemerken wir, dass  $X$  integrierbar ist, denn  $\int_{\mathbb{R}} |y|e^{-\frac{y^2}{2}} dy < \infty$ . Für den Erwartungswert erhalten wir dann

$$\mathbb{E}X = \int_{\mathbb{R}} ye^{-\frac{y^2}{2}} dy = 0,$$

denn es wird eine ungerade Funktion integriert. □

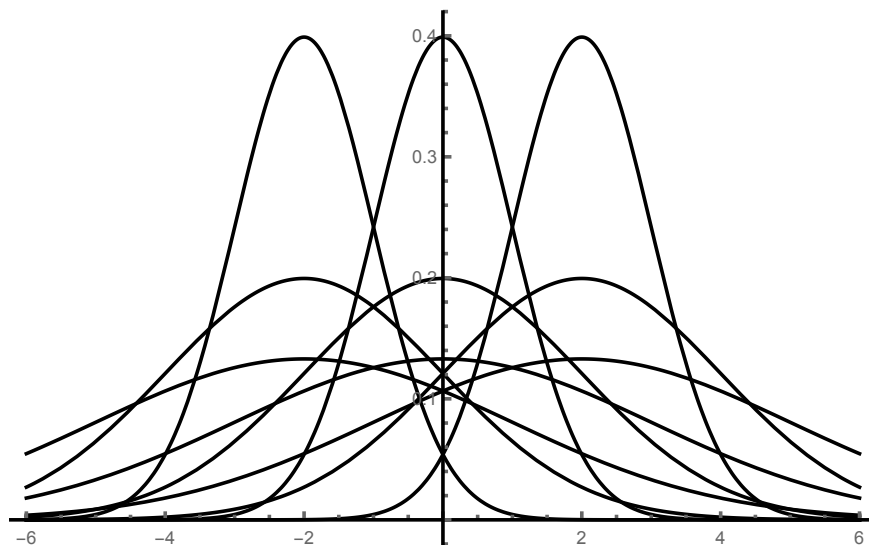


ABBILDUNG 4. Dichten der Normalverteilungen mit verschiedenen Parametern.

**Definition 6.6.16.** 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:

$$f_X(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad y \in \mathbb{R}.$$

Wir benutzen die Notation  $X \sim N(\mu, \sigma^2)$ .

**Bemerkung 6.6.17.** Ist  $Y \sim N(0, 1)$  standardnormalverteilt, so ist  $\mu + \sigma Y$  normalverteilt mit Parametern  $\mu$  und  $\sigma^2$  (Übung). Daraus folgt, dass der Erwartungswert einer  $N(\mu, \sigma^2)$ -verteilten Zufallsvariable gleich  $\mu$  ist. Später werden wir zeigen, dass der zweite Parameter  $\sigma^2$  mit der Varianz der Zufallsvariable übereinstimmt.

#### 6.6.4. Cauchy-Verteilung.

**Definition 6.6.18.** Eine Zufallsvariable  $X$  heißt *Cauchyverteilt*, wenn  $X$  absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\pi} \frac{1}{1 + y^2}, \quad y \in \mathbb{R}.$$

Wir benutzen die Notation  $X \sim \text{Cauchy}$ .

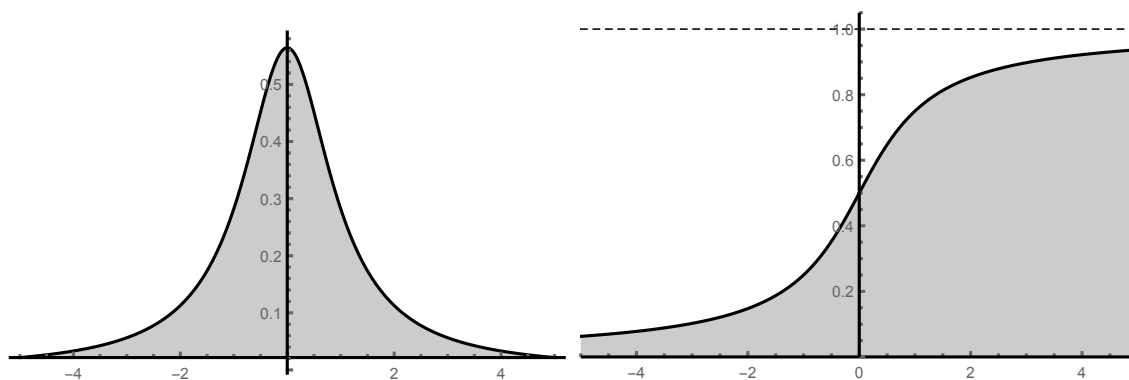


ABBILDUNG 5. Links: Dichte der standard Cauchyverteilung. Rechts: Die entsprechende Verteilungsfunktion.

**Bemerkung 6.6.19.** 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 6.6.20.** Sei  $\varphi \sim \text{Unif}[-\pi/2, \pi/2]$  gleichverteilt auf dem Intervall  $[-\pi/2, \pi/2]$  und sei  $X = \tan(\varphi)$ . Nun behaupten wir, dass  $X \sim \text{Cauchy}$ . **Muss erklärt werden, was das bedeutet.**

**Beweis.** Wir berechnen die Verteilungsfunktion von  $X$ :

$$\mathbb{P}[X \leq t] = \mathbb{P}[\varphi \leq \arctan(t)] = \frac{\arctan t - (-\frac{\pi}{2})}{\pi} = \frac{1}{2} + \frac{1}{\pi} \arctan t.$$

□

**Beispiel 6.6.21.** Sei  $X$  Cauchy-verteilt. Was ist der Erwartungswert von  $X$ ? Da die Dichte von  $X$  eine gerade Funktion ist, erscheint die Vermutung plausibel, dass der Erwartungswert von  $X$  gleich 0 ist. Dies ist jedoch nicht der Fall! Wir behaupten, dass  $X$  nicht integrierbar ist

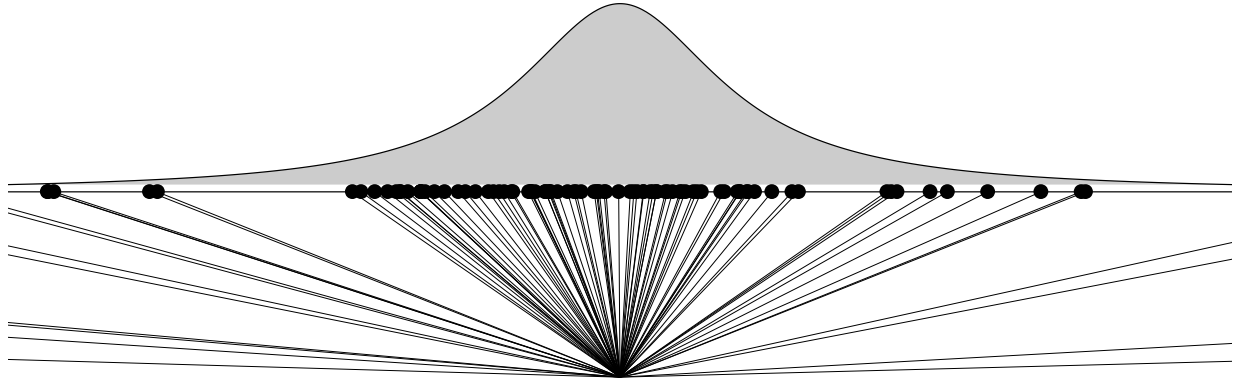


ABBILDUNG 6. Entstehung der Cauchy-Verteilung.

und somit gar keinen Erwartungswert besitzt. Wir erinnern, dass eine Zufallsvariable genau dann integrierbar ist, wenn  $\mathbb{E}X^+ < \infty$  und  $\mathbb{E}X^- < \infty$ . Also berechnen wir  $\mathbb{E}X^+$ :

$$\mathbb{E}X^+ = \mathbb{E}[X \mathbb{1}_{X \geq 0}] = \int_{\mathbb{R}} y \mathbb{1}_{y \geq 0} f_X(y) dy = \frac{2}{\pi} \int_0^{\infty} \frac{y}{1+y^2} dy = \frac{1}{\pi} \log(1+y^2) \Big|_0^{\infty} = +\infty.$$

Analog zeigt man, dass  $\mathbb{E}X^- = +\infty$ . Für den Erwartungswert ergibt sich die Unbestimmtheit  $\mathbb{E}X = (+\infty) - (+\infty)$ . Somit ist die Zufallsvariable  $X$  nicht integrierbar. Wir möchten hervorheben, dass der Erwartungswert von  $X$  weder  $\infty$ , noch  $+\infty$  oder  $-\infty$  ist! Er ist überhaupt nicht definiert.

## 6.7. 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 6.7.1.** Eine Zufallsvariable  $X$  heißt *singulär*, wenn die folgenden zwei Bedingungen gelten:

- $X$  hat keine Atome, d.h. es gilt  $\mathbb{P}[X = y] = 0$  für alle  $y \in \mathbb{R}$ .
- 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, hat dabei aber keine Atome. Nun werden wir eine singuläre Zufallsvariable konstruieren. Zuerst beschreiben wir die Nullmenge  $N$ , in der diese Zufallsvariable ihre Werte annehmen soll.





ABBILDUNG 7. Erste Schritte in der Konstruktion der Cantor-Menge.

**Beispiel 6.7.2** (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 6.7.3** (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.

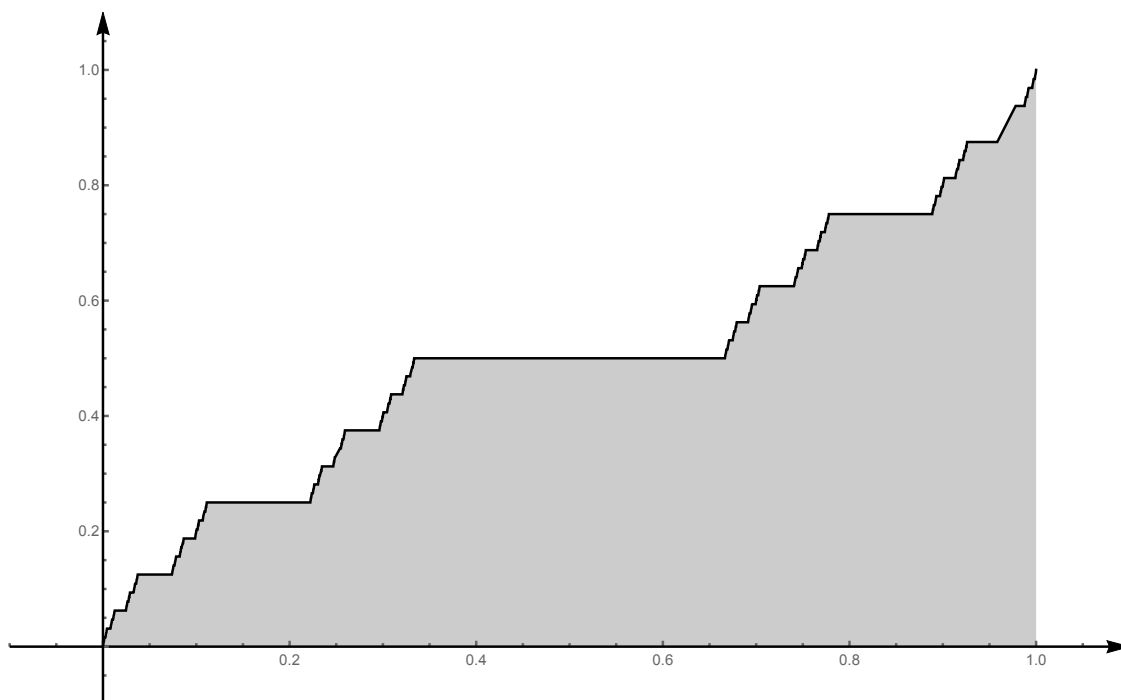


ABBILDUNG 8. Die Cantor-Funktion.

**Bemerkung 6.7.4.** 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 6.7.5.** 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.  $\square$

## 6.8. 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 6.8.1** (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 6.8.2** (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  $\sigma$ -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  $\lambda$  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, den ein einzelner Punkt ist lediglich eine Nullmenge. Es gilt also

$$\mathbb{P}[X = 0] = \frac{10}{20} = \frac{1}{2}.$$

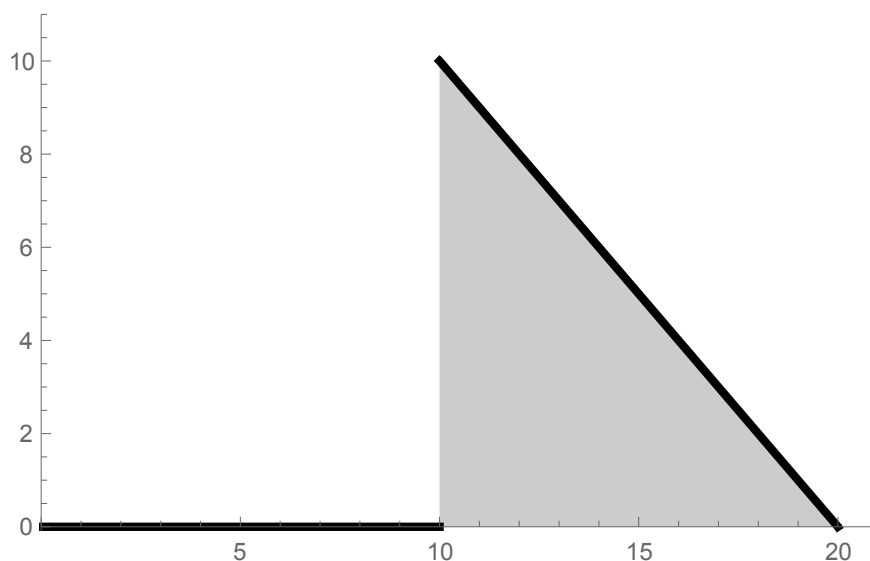


ABBILDUNG 9. Die Wartezeit an der Bahnschranke als Funktion auf der Grundmenge  $\Omega = (0, 20)$ .

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  $\frac{1}{2}$  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

$$F_X = \frac{1}{2}F_1 + \frac{1}{2}F_2.$$

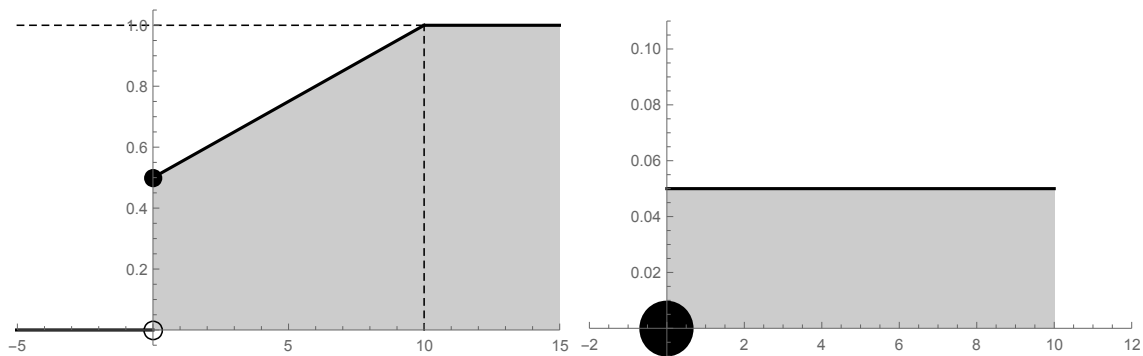


ABBILDUNG 10. Links: Verteilungsfunktion der Wartezeit an der Bahnshranke. Rechts: Versuch, die Verteilung der Wartezeit zu visualisieren. Gezeigt wird die Dichte der absolut stetigen Komponente (mit Gesamtmasse  $\frac{1}{2}$ ) sowie das Atom an der Stelle 0 (mit Masse  $\frac{1}{2}$ ).

## 6.9. Verteilungsfunktion eines Zufallsvektors

Die Familie der Borel-Teilmengen von  $\mathbb{R}^d$  wird mit  $\mathcal{B}^d$  bezeichnet.

**Definition 6.9.1.** 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 6.9.2.** 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. □

## 6.10. Diskrete und absolut stetige Zufallsvektoren

Wie im Fall von Zufallsvariablen lassen sich diskrete, absolut stetige und singuläre Vektoren definieren.

**Definition 6.10.1.** 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 6.10.2.** 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 6.10.3.** 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 6.10.4.** 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

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

### 6.11. 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 6.11.1.** 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 6.11.2.** 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 6.11.3.** 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

$$f_{X_i}(t_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(t_1, \dots, t_d) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_d \quad (6.11.1)$$

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 (6.11.1) 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 6.11.4.** 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.



## 6.12. 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 6.12.1.** 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 6.12.2.** 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 6.12.3.** 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 6.12.4.** 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  $[0, 1]^2$ .

### 6.13. 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 6.13.1.** 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

$$f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y)) \cdot |(\varphi^{-1})'(y)| \quad \text{für fast alle } y \in \varphi(I).$$

Dabei ist  $\varphi^{-1}$  die inverse Funktion von  $\varphi$ .

**Beweis.** Wegen  $\varphi'(x) \neq 0$  für alle  $x \in I$  ist  $\varphi$  entweder monoton steigend oder monoton fallend. Sei o.B.d.A.  $\varphi$  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 6.13.2.** Für monoton fallendes  $\varphi$  geht dies analog, jedoch mit  $|(\varphi^{-1})'(t)|$  anstelle von  $(\varphi^{-1})'(t)$ .

**Beispiel 6.13.3.** 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 6.13.4.** 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.

**Aufgabe 6.13.5.** Die Zufallsvariable  $X$  sei uniform verteilt auf dem Intervall  $[-2, 1]$ . Bestimmen Sie die Verteilungsfunktion und die Dichte von  $X^2$ .

## 6.14. 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 6.14.1** (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

$$p_{X_1+X_2}(z) = \sum_{y \in \mathfrak{S}(X_1)} p_{X_1}(y) \cdot p_{X_2}(z - y) \text{ für alle } z \in \mathbb{R}. \quad (6.14.1)$$

**Definition 6.14.2.** 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 6.14.3.** Zur Erinnerung:  $p_{X_1}(y) = \mathbb{P}[X_1 = y]$  ist die Zähldichte von  $X_1$  und  $\mathfrak{S}(X_1) = \{y \in \mathbb{R} : \mathbb{P}[X_1 = y] \neq 0\}$  ist die Menge der Werte, die  $X_1$  annehmen kann.

**Beweis von Satz 6.14.1.** Sei  $z \in \mathbb{R}$ . Es gilt dann

$$\begin{aligned} p_{X_1+X_2}(z) &= \mathbb{P}[X_1 + X_2 = z] \\ &= \sum_{y \in \mathfrak{S}(X_1)} \mathbb{P}[X_1 = y, X_2 = z - y] \\ &= \sum_{y \in \mathfrak{S}(X_1)} \mathbb{P}[X_1 = y] \cdot \mathbb{P}[X_2 = z - y] \\ &= \sum_{y \in \mathfrak{S}(X_1)} p_{X_1}(y) \cdot p_{X_2}(z - y). \end{aligned}$$

□

**Satz 6.14.4.** 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  $X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$ .

□

**Aufgabe 6.14.5.** 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:  $\text{Bin}(n_1, p) * \text{Bin}(n_2, p) = \text{Bin}(n_1 + n_2, p)$ .

**Aufgabe 6.14.6.** 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

**Aufgabe 6.14.7.** 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 6.14.8** (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

$$f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_{X_1}(y)f_{X_2}(z-y)dy \text{ für fast alle } z \in \mathbb{R}. \quad (6.14.2)$$

**Definition 6.14.9.** 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 6.14.10.** 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 6.14.8.** 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 6.14.11.** 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 6.14.12.** 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

$$f_{X_1+X_2}(z) = \begin{cases} ze^{-z} & z > 0, \\ 0, & z \leq 0. \end{cases}$$

**Aufgabe 6.14.13.** 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  $X_1$  und  $X_2$  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

$$p_{X_1 X_2}(z) = \sum_{y \in \mathfrak{S}(X_1) \setminus \{0\}} p_{X_1}(y) p_{X_2}\left(\frac{z}{y}\right), \quad z \neq 0.$$

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

$$f_{X_1/X_2}(z) = \int_{-\infty}^{\infty} |y| f_{X_1}(zy) f_{X_2}(y) dy.$$

**Aufgabe 6.14.14.** Die Zufallsvariablen  $X$  und  $Y$  seien beide uniformverteilt auf der Menge  $\{1, \dots, n\}$  und unabhängig. Bestimmen Sie die Verteilung von  $X + Y$  (d.h. geben Sie die möglichen Werte von  $X + Y$  zusammen mit den entsprechenden Wahrscheinlichkeiten an).

**Aufgabe 6.14.15.** Seien  $X$  und  $Y$  zwei auf dem Intervall  $(0, 1)$  uniform verteilte und unabhängige Zufallsvariablen. Bestimmen Sie die Verteilungsfunktion und die Dichte von  $X + Y$ ,  $XY$  und  $X/Y$ .

**Aufgabe 6.14.16.** Seien  $X_1 \sim \text{Geo}(p_1)$  und  $X_2 \sim \text{Geo}(p_2)$  unabhängige, mit Parametern  $p_1$  und  $p_2$  verteilte Zufallsvariablen, wobei  $p_1 \in (0, 1)$ ,  $p_2 \in (0, 1)$  und  $p_1 \neq p_2$ . Zeigen Sie, dass

$$\mathbb{P}[X_1 + X_2 = k] = \frac{p_1 \cdot p_2}{p_1 - p_2} \cdot \left( (1 - p_2)^{k-1} - (1 - p_1)^{k-1} \right) \text{ für alle } k \geq 2.$$

**Aufgabe 6.14.17.** Es seien  $U_1, U_2, \dots$  unabhängige, auf  $[0, 1]$  uniformverteilte Zufallsvariablen. Definiere

$$T(x) := \min\{n \in \mathbb{N} : U_1 + \dots + U_n > x\}, \quad 0 < x < 1.$$

Zeigen Sie:  $\mathbb{P}[T(x) > k] = x^k/k!$  für alle  $k \in \mathbb{N}$ .

## 6.15. 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 6.15.1** (Transformationsformel für den Erwartungswert). Sei  $X$  eine absolut stetige Zufallsvariable und  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  eine Borel-Funktion. Dann gilt

$$\mathbb{E}\varphi(X) = \int_{-\infty}^{\infty} \varphi(y) f_X(y) dy, \quad (6.15.1)$$

falls  $\int_{-\infty}^{\infty} |\varphi(y)| f_X(y) dy < \infty$ .

**Bemerkung 6.15.2.** 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 6.15.1.** 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 (6.15.1) 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 (6.15.1).

SCHRITT 3. Sei nun  $\varphi \geq 0$  eine beliebige, nichtnegative Borel-Funktion. Wir approximieren  $\varphi$  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  $\varphi_n$  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  $\varphi_n$  ist eine Elementarfunktion), also sind die Terme  $\int \varphi(y) f_X(y) dy$  und  $\mathbb{E}\varphi(X)$  auch gleich und die Formel (6.15.1) 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  $\varphi_+$  und  $\varphi_-$  der positive bzw. der negative Anteil von  $\varphi$  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 (6.15.1). □

**Beispiel 6.15.3.** Mit  $\varphi(y) = y$  erhalten wir die Formel

$$\mathbb{E}X = \int_{-\infty}^{\infty} yf_X(y)dy.$$

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 6.15.4.** 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 6.15.5.** Eine ähnliche Transformationsformel gilt auch für diskrete Zufallsvariablen: ist  $X$  diskret, so haben wir

$$\mathbb{E}\varphi(X) = \sum_{y \in \mathfrak{S}(X)} \varphi(y)p_X(y),$$

falls  $\sum_{y \in \mathfrak{S}(X)} |\varphi(y)|p_X(y) < \infty$ .

## 6.16. Multiplikatивität 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 6.16.1.** Die Zufallsvariablen  $X, Y : \Omega \rightarrow \mathbb{R}$  seien integrierbar und unabhängig. Dann ist auch  $XY$  integrierbar und es gilt

$$\mathbb{E}[XY] = (\mathbb{E}X) \cdot (\mathbb{E}Y). \tag{6.16.1}$$

**Satz 6.16.2.** Seien  $X, Y : \Omega \rightarrow \mathbb{R}$  unabhängige integrierbare Zufallsvariablen. Dann ist auch das Produkt  $X \cdot Y$  integrierbar und es gilt

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \quad (6.16.2)$$

\*\*\*\*\* **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  $\Omega$ . 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 6.4.13 folgt, dass  $XY$  integrierbar ist.

Für den Erwartungswert von  $X \cdot Y$  gilt nun mit Satz 6.4.13:

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

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

**Bemerkung 6.16.3.** Der soeben 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

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

\*\*\*\*\* **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  $X_n(\omega) \geq 0, \quad Y_n(\omega) \geq 0$ .
- (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 (6.16.1).

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 7

### Varianz und Kovarianz

#### 7.1. Varianz

**Definition 7.1.1.** 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 7.1.2.** 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 7.1.3.** 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 7.1.4.** 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 7.1.5.** Für jede Zufallsvariable  $X \in L^2$  gilt die Formel:

$$\text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

**Bemerkung 7.1.6.** 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 7.1.7.** 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 7.1.8.** 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:

$$\text{Var}(aX+b) = \mathbb{E}[(aX+b-\mathbb{E}(aX+b))^2] = \mathbb{E}[(aX+b-a\mathbb{E}X-b)^2] = \mathbb{E}[a^2(X-\mathbb{E}X)^2] = a^2 \text{Var} X.$$

□

**Beweis von (2).**  $\sigma(aX + b) = \sqrt{\text{Var}(aX + b)} = \sqrt{a^2 \text{Var} X} = |a|\sigma(X)$ . □

**Beispiel 7.1.9.** 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 7.1.10.** Die Varianz einer Zufallsvariable ist immer  $\geq 0$ . 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 7.1.11.** Sei  $X \sim N(\mu, \sigma^2)$  normalverteilt mit Parametern  $(\mu, \sigma^2)$ . 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:

$$\text{Var } X = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} y(-e^{-y^2/2})' dy = \frac{1}{\sqrt{2\pi}} \left( -ye^{-y^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} dy \right) = 1.$$

Sei nun  $X \sim N(\mu, \sigma^2)$  normalverteilt mit beliebigen Parametern  $(\mu, \sigma^2)$ . 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 7.1.12.** 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

$$\text{Var } X = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad \square$$

**Aufgabe 7.1.13.** Sei  $X$  eine Zufallsvariable mit der Dichte

$$f(t) = \begin{cases} ct^{-6}, & \text{falls } t > 1, \\ 0, & \text{sonst.} \end{cases}$$

Bestimmen Sie  $c$ ,  $\mathbb{P}[X > 4]$ ,  $\mathbb{E}X$  und  $\text{Var } X$ .

**Aufgabe 7.1.14.** Sei  $X \sim \text{Exp}(\lambda)$ . Zeigen Sie, dass  $\text{Var } X = 1/\lambda^2$ .

## 7.2. Kovarianz und Korrelationskoeffizient

**Definition 7.2.1.** 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 7.2.2.**  $\text{Cov}(X, X) = \text{Var } X$ .

**Bemerkung 7.2.3.** Wir zeigen, dass die Kovarianz wohldefiniert ist. Seien  $X, Y \in L^2$ . Nach Lemma 7.1.2 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

$$\mathbb{E}[|\tilde{X}\tilde{Y}|] \leq \frac{1}{2} \left( \mathbb{E}[\tilde{X}^2] + \mathbb{E}[\tilde{Y}^2] \right) < \infty.$$

Somit ist die Zufallsvariable  $\tilde{X}\tilde{Y} = (X - \mathbb{E}X)(Y - \mathbb{E}Y)$  integrierbar und die Kovarianz existiert. □

**Satz 7.2.4.** 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. □

**Bemerkung 7.2.5.** Für die Berechnung der Kovarianz kann man folgende Formeln benutzen:

(1) Für  $X, Y$  diskret:

$$\mathbb{E}[XY] = \sum_{\substack{s \in \mathfrak{S}X \\ t \in \mathfrak{S}Y}} st \cdot \mathbb{P}[X = s, Y = t].$$

(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 7.2.6.** 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. □

Für die Kovarianz gelten ähnliche Rechenregeln wie für das Skalarprodukt.

**Satz 7.2.7.** 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:

$$\text{Cov}(a_1 X_1 + \dots + a_n X_n, b_1 Y_1 + \dots + b_m Y_m) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j).$$

**Beweis.** Übungsaufgabe. □

**Satz 7.2.8** (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 7.2.9.** 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 7.2.8.** Für beliebige Konstante  $a \in \mathbb{R}$  gilt

$$0 \leq \mathbb{E}[(aX + Y)^2] = \mathbb{E}[a^2 X^2 + 2aXY + Y^2] = a^2 \mathbb{E}[X^2] + 2a \mathbb{E}[XY] + \mathbb{E}[Y^2].$$

Die rechte Seite ist eine quadratische Funktion von  $a$ . Diese Funktion ist für alle  $a$  nichtnegativ. Die Diskriminante  $D$  muss somit  $\leq 0$  sein:

$$D = 4(\mathbb{E}[XY])^2 - 4\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2] \leq 0.$$

Das beweist die Behauptung. □

**Korollar 7.2.10.** 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 7.2.11.** 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 7.2.12.** Aus Korollar 7.2.10 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  $\pm 1$  ist, wenn zwischen  $X$  und  $Y$  ein linearer Zusammenhang besteht, d.h., wenn  $Y = aX + b$ .

**Bemerkung 7.2.13.** Seien  $a, b, c, d \in \mathbb{R}$  und  $a, c > 0$ . Dann gilt:

$$\rho(aX + b, cY + d) = \rho(X, Y).$$

**Bemerkung 7.2.14.**  $\rho(X, X) = 1$ ,  $\rho(X, -X) = -1$ .

**Definition 7.2.15.** Zwei Zufallsvariablen  $X, Y \in L^2$  heißen unkorreliert, wenn  $\text{Cov}(X, Y) = 0$  (und somit auch  $\rho(X, Y) = 0$ ).

**Satz 7.2.16.** 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 7.2.17.** 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.  $\square$

**Satz 7.2.18.** 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

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X - \mathbb{E}X) + (Y - \mathbb{E}Y)]^2 \\ &= \mathbb{E}[(\tilde{X} + \tilde{Y})^2] \\ &= \mathbb{E}[\tilde{X}^2] + 2\mathbb{E}[\tilde{X}\tilde{Y}] + \mathbb{E}[\tilde{Y}^2] \\ &= \text{Var}(X) + 2 \text{Cov}(X, Y) + \text{Var}(Y). \end{aligned}$$

$\square$

**Korollar 7.2.19.** Für unabhängige Zufallsvariablen  $X, Y \in L^2$  gilt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

**Bemerkung 7.2.20.** Zum Vergleich: Für beliebige Zufallsvariablen  $X, Y \in L^1$  gilt

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

**Bemerkung 7.2.21.** 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 7.2.22.** 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 7.2.23.** 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  $\pm 1$  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 7.2.24.** 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  $\mathbb{P}[\Delta = 0] = 1$ . □

**Aufgabe 7.2.25.** Seien  $X$  und  $Y$  unabhängig und auf  $[0, 1]$  uniform verteilt. Bestimmen Sie die Kovarianz von  $U := \min(X, Y)$  und  $V := \max(X, Y)$ .

**Aufgabe 7.2.26.**  $n$  verschiedene Hüte werden rein zufällig an ihre Besitzer verteilt. (Jede Person erhält genau einen Hut). Es sei  $Z_i$  die Indikatorvariable des Ereignisses, dass die  $i$ -te Person den eigenen Hut erhält, wobei  $i \in \{1, \dots, n\}$ .

- (a) Bestimmen Sie  $\text{Cov}(Z_i, Z_j)$  für  $i \neq j$ .
- (b) Sei  $X$  die Anzahl der Personen, die den eigenen Hut erhalten. Bestimmen Sie  $\text{Var } X$ .

## Gesetz der großen Zahlen

## 8.1. Zwei Beispiele

**Beispiel 8.1.1.** Wir kippen einen Sack mit  $n$  fairen Würfeln um, wobei  $n$  sehr groß ist und bezeichnen mit  $X_1, \dots, X_n$  die Augenzahlen der einzelnen Würfel. Wie groß ist ungefähr die Augensumme  $X_1 + \dots + X_n$  aller  $n$  Würfel? Intuitiv geht man davon aus, dass ungefähr  $n/6$  Würfel eine 1 zeigen, ungefähr  $n/6$  Würfel eine 2 zeigen, usw. Für die Augensumme aller  $n$  Würfel ergibt sich die Approximation

$$X_1 + \dots + X_n \approx \frac{n}{6} \cdot 1 + \frac{n}{6} \cdot 2 + \frac{n}{6} \cdot 3 + \frac{n}{6} \cdot 4 + \frac{n}{6} \cdot 5 + \frac{n}{6} \cdot 6 = 3.5 \cdot n.$$

Diese Aussage gilt approximativ und lediglich bei “großem”  $n$ . Der von uns verwendeten Begriffe “ungefähr” und “groß” sind natürlich keine mathematischen Begriffe. Streng genommen, sollte die obige Aussage als ein Grenzwert für  $n \rightarrow \infty$  formuliert werden, nämlich

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = 3.5. \quad (8.1.1)$$

Wie kann die auf der rechten Seite aufgetauchte Zahl 3.5 interpretiert werden? Es ist der Erwartungswert für die Augenzahl eines Würfels. Für alle  $i \in \mathbb{N}$  gilt nämlich, dass

$$\mathbb{E}X_i = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6.$$

**Beispiel 8.1.2.** Wir betrachten ein  $n$ -faches Bernoulli-Experiment mit Erfolgswahrscheinlichkeit  $p$ . Dabei soll  $n$  sehr groß sein. Man kann sich z.B. vorstellen, dass ein großer Sack mit  $n$  Reißzwecken umgekippt wird, wobei das Landen einer Reißzwecke auf dem Kopf jeweils als Erfolg gilt. Die Zufallsvariable, die den Ausgang des  $i$ -ten Experiments beschreibt, sei

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

Die Anzahl der Erfolge in allen  $n$  Experimenten ist dann die Zufallsvariable  $X_1 + \dots + X_n$ . Intuitiv sollte die Anzahl der Erfolge “ungefähr”  $np$  betragen, jedenfalls wenn  $n$  “groß” ist. Es sollte also gelten, dass

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = p. \quad (8.1.2)$$

Auch hier ist der Grenzwert  $p$  nichts Anderes als der Erwartungswert jeder einzelnen Zufallsvariable  $X_i$ , denn

$$\mathbb{E}X_i = p \cdot 1 + (1 - p) \cdot 0 = p.$$

**Verallgemeinerung.** Die Formeln (8.1.1) und (8.1.2) sind Spezialfälle eines allgemeinen Naturphänomens, das als *Gesetz der großen Zahlen* bezeichnet wird. Um die obigen Beispiele zu verallgemeinern, betrachten wir eine unendliche Folge unabhängiger Zufallsvariablen



$X_1, X_2, \dots$ , die die gleiche (und ansonsten beliebige) Verteilung besitzen sollen. Dann kann man vermuten, dass

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mathbb{E}X_1,$$

falls der Erwartungswert auf der rechten Seite existiert. Das Ziel dieses Kapitels ist, diese Behauptung (das **Gesetz der großen Zahlen**) auf der Basis der Axiome der Wahrscheinlichkeitstheorie zu beweisen. Als Beispiel kann man sich vorstellen, dass die Zufallsvariable  $X_1$  den Gewinn in einer Runde eines Glücksspiels ist. Es werden nun  $n$  Runden dieses Spiels unabhängig und nach den gleichen Regeln gespielt. Dann ist  $\frac{X_1 + \dots + X_n}{n}$  der mittlere Gewinn und das Gesetz der großen Zahlen behauptet, dass der mittlere Gewinn für  $n \rightarrow \infty$  gegen  $\mathbb{E}X_1$  strebt.

In der obigen Diskussion haben wir eine heikle Frage nicht angesprochen: Was bedeutet es, dass eine Folge von Zufallsvariablen konvergiert? Bekanntlich sind Zufallsvariablen Funktionen auf  $\Omega$ . Aus der Analysis sollte bekannt sein, dass es für Funktionen viele verschiedene (nicht-äquivalente) Konvergenzbegriffe gibt, z.B. die punktweise Konvergenz und die gleichmäßige Konvergenz. Für die Wahrscheinlichkeitstheorie spielen die folgenden 4 Konvergenzarten eine besonders wichtige Rolle:

- Konvergenz in Wahrscheinlichkeit (auch stochastische Konvergenz genannt);
- Konvergenz in  $L^p$  (oder im  $p$ -ten Mittel);
- fast sichere Konvergenz;
- Konvergenz in Verteilung.

Diese Begriffe werden wir im Folgenden einführen.

## 8.2. Konvergenz in Wahrscheinlichkeit und $L^2$ -Konvergenz

**Definition 8.2.1.** 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:

$$\lim_{n \rightarrow \infty} \mathbb{P}[|Z_n - Z| > \varepsilon] = 0.$$

Wir schreiben dann  $Z_n \xrightarrow[n \rightarrow \infty]{P} Z$ .

**Definition 8.2.2.** 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

$$\lim_{n \rightarrow \infty} \mathbb{E}[(Z_n - Z)^2] = 0.$$

**Notation 8.2.3.**  $Z_n \xrightarrow[n \rightarrow \infty]{L^2} Z$ .

Im Folgenden werden wir zeigen, dass aus der  $L^2$ -Konvergenz die stochastische Konvergenz folgt, jedoch nicht umgekehrt.

### 8.3. Ungleichungen von Markow und Tschebyschew

**Lemma 8.3.1.** 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 8.3.2** (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 8.3.1:

$$\mathbb{P}[(X - \mathbb{E}X)^2 \geq a^2] \leq \frac{\mathbb{E}[(X - \mathbb{E}X)^2]}{a^2} = \frac{\text{Var } X}{a^2}.$$

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 8.3.3** (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 8.3.1, indem man dort  $f(Z)$  anstelle von  $Z$  einsetzt. □

**Beispiel 8.3.4.** 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 8.3.5.** 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  $Z_n \xrightarrow{P} Z$ . □

#### 8.4. Schwaches Gesetz der großen Zahlen

**Satz 8.4.1** (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 8.3.5 folgt nun auch die stochastische Konvergenz:

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{P} \mu. \quad \square$$

**Beispiel 8.4.2.** 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 8.4.1 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  $\infty$  geht.

**Bemerkung 8.4.3.** 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.

### 8.5. Fast sichere Konvergenz

**Definition 8.5.1.** 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 8.5.2.** Ein Ereignis  $B \in \mathcal{F}$  heißt ein *fast unmögliches* Ereignis (oder ein *Nullereignis*), wenn  $\mathbb{P}[B] = 0$ .

**Beispiel 8.5.3.** Das Ereignis  $\Omega$  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 8.5.4.** 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 8.5.5.** 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 8.5.6.** 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- $\sigma$ -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  $\sigma$ -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 8.5.7.** 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:

$$\mathbb{P} \left[ \lim_{n \rightarrow \infty} Z_n = Z \right] = 1.$$

**Notation 8.5.8.**  $Z_n \xrightarrow[n \rightarrow \infty]{f.s.} Z$  oder  $\lim_{n \rightarrow \infty} Z_n = Z$  f.s.

**Bemerkung 8.5.9.** Damit diese Definition Sinn ergibt, 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  $\sigma$ -Algebra bilden, sind auch beliebige abzählbare Schnitte und Vereinigungen dieser Mengen messbar, also auch  $A$ . Es folgt, dass  $A \in \mathcal{F}$ .

**Satz 8.5.10.** 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

$$A = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega) \right\}$$

Wahrscheinlichkeit  $\mathbb{P}[A] = 1$ . Sei  $\varepsilon > 0$ . Wir definieren Ereignisse

$$A_k = \left\{ \omega \in \Omega : \forall n \geq k : |Z_n(\omega) - Z(\omega)| \leq \varepsilon \right\}, \quad k \in \mathbb{N}.$$

Es gilt  $A \subset \bigcup_{k \in \mathbb{N}} A_k$ . Aus  $1 = \mathbb{P}[A] \leq \mathbb{P}[\bigcup_{k \in \mathbb{N}} A_k]$  folgt, dass

$$\mathbb{P}[\bigcup_{k \in \mathbb{N}} A_k] = 1.$$

Außerdem gilt  $A_1 \subset A_2 \subset A_3 \subset \dots$ . Mit der Stetigkeit des Wahrscheinlichkeitsmaßes folgt

$$\lim_{k \rightarrow \infty} \mathbb{P}[A_k] = \mathbb{P}[\bigcup_{k \in \mathbb{N}} A_k] = 1.$$

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 8.5.11.** 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  $Z_n \xrightarrow[n \rightarrow \infty]{f.s.} 0$ .

## 8.6. Starkes Gesetz der großen Zahlen: Erste Version

**Satz 8.6.1** (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  $n \in \mathbb{N}$ . Es sei außerdem die folgende Bedingung erfüllt

$$\sum_{n=1}^{\infty} \frac{\text{Var } X_n}{n^2} < \infty.$$

Für die Partialsummen  $S_n = X_1 + \dots + X_n$  gilt dann

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0.$$

**Beispiel 8.6.2.** Wir betrachten einen Spezialfall. Seien  $X_1, X_2, \dots$  unabhängige Zufallsvariablen mit einem gemeinsamen Erwartungswert  $\mathbb{E}X_n = \mu$  und einer gemeinsamen Varianz  $\text{Var } X_n = \sigma^2$ . Diese Bedingung ist z.B. dann erfüllt, wenn  $X_1, X_2, \dots$  identisch verteilt sind. 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 8.6.1 ist also anwendbar und besagt, dass  $\frac{S_n - n\mu}{n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$ , woraus folgt, dass

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \mu.$$

Diese Aussage ist stärker als das schwache Gesetz der großen Zahlen, denn die stochastische Konvergenz folgt aus der fast sicheren Konvergenz. Das erklärt die Bezeichnung starkes Gesetz der großen Zahlen.

Im Folgenden werden wir Satz 8.6.1 beweisen. Dafür benötigen wir einige Hilfsmittel.

**Satz 8.6.3** (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 8.6.4.** Die Tschebyschew-Ungleichung ist schwächer, denn damit 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 8.6.3.** 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

$$\begin{aligned} A_k &:= \{|S_1| < a, |S_2| < a, \dots, |S_{k-1}| < a, |S_k| \geq a\} \\ &= \text{“Erster Austritt aus dem Intervall } (-a, a) \text{ findet zum Zeitpunkt } k \text{ statt.”} \end{aligned}$$

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 8.6.1.** 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 8.6.1 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  $\frac{U_n}{2^n} \xrightarrow[n \rightarrow \infty]{f.s.} 0$ . □

### 8.7. 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 8.7.1** (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 8.7.2.** 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 8.7.1 benötigen wir mehrere Lemmata.

**Lemma 8.7.3.** 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:

$$A := \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\},$$

$$B := \{\omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)\}.$$

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 8.7.4.** 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 8.7.5.** 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 8.7.6 (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 8.7.1.** 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 8.7.3 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 8.6.1 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 8.7.5 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 8.6.1 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:

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}[X_n'' \neq 0] &= \sum_{n \geq 1} \mathbb{P}[|X_n| \geq n] \\ &= \sum_{n \geq 1} \mathbb{P}[|X_1| \geq n] \quad (\text{gilt, da } X_n \text{ identisch verteilt}) \\ &\leq \mathbb{E}|X_1| \quad (\text{gilt mit Lemma 8.7.4}) \\ &< \infty. \end{aligned}$$

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 8.7.6 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. □

### 8.8. 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 8.8.1.** 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:

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}[C_n] &= \sum_{n \geq 1} \mathbb{P}[|X_n| \geq n] \\ &= \sum_{n \geq 1} \mathbb{P}[|X_1| \geq n] \quad (\text{gilt, da } X_i \text{ identisch verteilt}) \\ &\geq \mathbb{E}|X_1| - 1 \quad (\text{gilt mit Lemma 8.7.4, da } |X_1| \geq 0) \\ &= \infty. \end{aligned}$$

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  $\mathbb{P}[A] = 0$ .  $\square$

**Beispiel 8.8.2.** 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 8.8.1 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.

## 8.9. Anwendungen des Gesetzes der großen Zahlen

Die nächsten 3 Beispiele sind Spezialfälle der sogenannten Monte-Carlo-Methode.

**Berechnung von  $\pi$ .** Man kann die Zahl  $\pi$  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 Betrag 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  $\pi$  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  $\varphi$  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  $\Omega$  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  $\infty$  geht.

**Beispiel 8.9.1** (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 8.9.2** (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  $2\varepsilon$  konstruiert. Da jedoch  $\varepsilon$  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  $\omega$ .

**Definition 8.9.3.** 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  $\omega$  mit Häufigkeit  $10^{-l}$  vor, und das gilt für jedes  $l$ .

**Definition 8.9.4.** 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  $\omega$  jede Ziffer mit einer Häufigkeit von  $1/10$  vorkommt.

**Beispiel 8.9.5.** Die Zahl  $\frac{1}{3} = 0,3333\dots$  ist nicht normal. Allgemein sind rationale Zahlen nicht normal, da sie eine periodische Dezimalbruchentwicklung besitzen.

**Beispiel 8.9.6.** Unbewiesene Vermutung: Die Zahlen

$$\begin{aligned}\pi &= 3.1415926535897932384626433832795028841971693993751 \dots, \\ e &= 2.7182818284590452353602874713526624977572470937000 \dots, \\ \log 2 &= 0.6931471805599453094172321214581765680755001343602 \dots\end{aligned}$$

sind normal.

**Beispiel 8.9.7.** Unbewiesene Vermutung: Jede algebraische Zahl ist normal, sofern sie nicht rational ist. Zum Beispiel ist die Zahl

$$\sqrt{2} = 1.4142135623730950488016887242096980785696718753769 \dots$$

normal.

**Beispiel 8.9.8.** 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 8.9.9** (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  $\sigma$ -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  $\Omega$ .

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  $\omega$ , 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  $\omega$  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  $\omega$ , 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 8.9.10** (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 9

# Ungleichungen

### 9.1. Jensen-Ungleichung

**Definition 9.1.1.** 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 9.1.2.** 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  $(x_0, g(x_0))$  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 9.1.3** (Jensen-Ungleichung). Sei  $g$  eine konvexe Funktion und  $X$  eine Zufallsvariable mit  $X \in L^1$ ,  $g(X) \in L^1$ . Dann gilt:

$$g(\mathbb{E}X) \leq \mathbb{E}[g(X)]. \quad (9.1.1)$$

**Beispiel 9.1.4.** 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 9.1.3.** 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:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(\mathbb{E}X)] + K_0\mathbb{E}[X - \mathbb{E}X] = g(\mathbb{E}X) + K_0(\mathbb{E}X - \mathbb{E}X) = g(\mathbb{E}X).$$

□



## 9.2. Ljapunow-Ungleichung

Ein Spezialfall der Jensen-Ungleichung ist die Ljapunow-Ungleichung.

**Satz 9.2.1** (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 9.2.2.** Sei  $p \geq 0$ . Die  $L^p$ -Norm einer Zufallsvariable  $X$  ist definiert durch

$$\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}.$$

**Bemerkung 9.2.3.** Mit dieser Notation besagt die Ljapunow-Ungleichung, dass

$$\|X\|_s \leq \|X\|_t, \quad 0 < s < t.$$

**Beispiel 9.2.4.** 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 9.2.5.** Wir betrachten den Wahrscheinlichkeitsraum  $[0, 1]$  mit der Borel- $\sigma$ -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 9.2.1.** 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. □

### 9.3. Young-Ungleichung

**Satz 9.3.1** (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 9.3.2.** 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. □

### 9.4. Hölder-Ungleichung

**Satz 9.4.1** (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 9.4.2.** 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 9.4.3.** Mit  $p = q = 2$  erhalten wir die Cauchy-Schwarzsche Ungleichung

$$\|XY\|_1 \leq \|X\|_2 \cdot \|Y\|_2. \tag{9.4.1}$$

**Beweis von Satz 9.4.1.** 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 9.4.4.** 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 9.4.5.** Betrachte den Wahrscheinlichkeitsraum  $\Omega = [0, 1]$  mit der Borel- $\sigma$ -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.

## 9.5. Minkowski-Ungleichung

**Satz 9.5.1** (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 9.5.2.** Äquivalente Schreibweise:

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Die Minkowski-Ungleichung ist also eine Dreiecksungleichung für die  $L^p$ -Norm.

**Beispiel 9.5.3.** 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 9.5.4.** Betrachte den Wahrscheinlichkeitsraum  $\Omega = [0, 1]$  mit der Borel- $\sigma$ -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 9.5.1.** 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

$$(\mathbb{E}[|X + Y|^p])^{\frac{1}{p}} \leq \|X\|_p + \|Y\|_p.$$

□

## 9.6. $L^p$ -Räume und $L^p$ -Konvergenz

**Definition 9.6.1.** 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 9.6.2.** 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 9.6.3.** 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 9.6.4.** 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 9.6.5.** 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 9.6.6.** 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 9.6.7.** 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 9.6.8.** Äquivalente Schreibweise:  $\lim_{n \rightarrow \infty} \mathbb{E}[|X - X_n|^p] = 0$ .

**Notation 9.6.9.**  $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$ .

**Bemerkung 9.6.10.** 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  $L^\infty$ -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.

### 10.1. Erzeugende Funktion

**Definition 10.1.1.** 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 10.1.2.** 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 10.1.3.** Die erzeugende Funktion ist also wohldefiniert auf dem 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 10.1.2.** 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 10.1.4.** Für jedes  $n \in \mathbb{N}_0$  gilt

$$\mathbb{P}[X = n] = \frac{g_X^{(n)}(0)}{n!}.$$

**Bemerkung 10.1.5.** 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 10.1.4.** 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  $g_X^{(n)}(0) = n! p_n$ .  $\square$

Warum betrachten wir überhaupt erzeugende Funktionen? Weil sie uns erlauben, die Faltung in ein Produkt umzuwandeln. Das wird im nächsten Satz gezeigt.

**Satz 10.1.6.** 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.  $\square$

**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 10.1.7.** 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 10.1.8.** 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:  $X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$ . □

**Beispiel 10.1.9.** Man würfelt mit einem fairen Würfel  $n = 100$  Mal. Wir bestimmen die Wahrscheinlichkeit, dass die Augensumme gleich 350 ist.

**Lösung.** Wir bezeichnen die Augenzahlen in den einzelnen Würfeln mit  $X_1, \dots, X_n$ . Dann ist jedes  $X_k$  uniform verteilt auf  $\{1, \dots, 6\}$ . Folglich ist die erzeugende Funktion von jedem  $X_k$  gegeben durch

$$g_{X_k}(t) = \frac{1}{6}(t + t^2 + t^3 + t^4 + t^5 + t^6), \quad k \in \{1, \dots, n\}.$$

Wegen der Unabhängigkeit der Würfel ist die erzeugende Funktion der Augensumme  $S := X_1 + \dots + X_n$  gegeben durch

$$g_S(t) = g_{X_1}(t) \cdot \dots \cdot g_{X_n}(t) = \frac{1}{6^n}(t + t^2 + t^3 + t^4 + t^5 + t^6)^n.$$

Es sei bemerkt, dass  $g_S(t)$  ein Polynom in  $t$  ist. Uns interessiert die Wahrscheinlichkeit des Ereignisses  $\{S = 350\}$ , also der Koeffizient von  $t^{350}$  in diesem Polynom. Obwohl es dafür keine einfache geschlossene Formel gibt und eine Rechnung per Hand zu lange dauern würde, können Programme wie MATLAB oder Mathematica den Koeffizienten ohne Schwierigkeiten ausrechnen:

$$\mathbb{P}[S = 350] = \frac{211626289699720876779325110056760077261291341544525363062928447069862398743}{9073869770834318140231809266084136396349218201013262104764888421798571409408} \approx 0.0233226.$$

□

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 10.1.10.** 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 10.1.11.** Für jedes  $n \in \mathbb{N}_0$  gilt

$$\mathbb{E}[X(X-1)\dots(X-n+1)] = g_X^{(n)}(1).$$

**Bemerkung 10.1.12.** 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 10.1.13.** Zwei Spezialfälle dieser Formel sind:

$$\begin{aligned} \mathbb{E}X &= g_X'(1), & n &= 1, \\ \mathbb{E}[X(X-1)] &= g_X''(1), & n &= 2. \end{aligned}$$

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 10.1.14.** Die erzeugende Funktion ist im Allgemeinen nur für  $|z| \leq 1$  definiert. Daher müssen wir uns die Ableitung der erzeugenden Funktion an den Stellen  $\pm 1$  als eine einseitige Ableitung vorstellen.

**Beweis von Satz 10.1.11.** 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. □

**Beispiel 10.1.15.** 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:

$$\begin{aligned}\mathbb{E}X &= g'_X(1) = \lambda, \\ \text{Var } X &= g''_X(1) + g'_X(1) - (g'_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

## 10.2. Summen mit einer zufälligen Anzahl von Summanden

**Beispiel 10.2.1.** 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 10.2.2.** 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 10.2.3.** 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 10.1.6, 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 10.2.4** (Wald-Identität). Unter den gleichen Voraussetzungen wie in Satz 10.2.2 gilt

$$\mathbb{E}[X_1 + \dots + X_N] = \mathbb{E}[N] \cdot \mathbb{E}[X_1].$$

**Beweis.** Übung. □

**Bemerkung 10.2.5.** Im Beispiel mit der Versicherung kann man den erwarteten Gesamtschaden berechnen, indem man die erwartete Anzahl an gemeldeten Schäden mit der zu erwartenden jeweiligen Schadenshöhe multipliziert.

**Aufgabe 10.2.6.** Eine Teilchenquelle emittiert  $N$  Teilchen. Dabei sei  $N$  eine mit Parameter  $\lambda > 0$  Poisson-verteilte Zufallsvariable. Jedes Teilchen wird (unabhängig von allen anderen Teilchen und unabhängig von der Anzahl der emittierten Teilchen) mit Wahrscheinlichkeit  $p$  von einem Zähler registriert (und mit Wahrscheinlichkeit  $1 - p$  nicht registriert). Es sei  $X$  die Anzahl der registrierten Teilchen. Bestimmen Sie  $\mathbb{P}[X = k]$  für alle  $k \in \{0, 1, 2, \dots\}$ . Welche Verteilung hat  $X$ ?

### 10.3. 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

$$Z_{n+1} = \sum_{i=0}^{Z_n} X_{i,n}. \quad (10.3.1)$$

Somit ist  $Z_n$  die Anzahl der Teilchen in Generation  $n$ . Wie bestimmt man nun die Verteilung von  $Z_n$ ?

**Satz 10.3.1.** 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 10.2.2 auf die Formel (10.3.1) 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 10.3.2.** Sei  $\mu := \mathbb{E}[X_{1,0}] = \sum_{k=0}^{\infty} p_k k$ . Dann gilt

$$\mathbb{E}[Z_n] = \mu^n.$$

**Bemerkung 10.3.3.** Diese Formel ist nicht überraschend. Jedes Teilchen erzeugt im Durchschnitt  $\mu$  Töchterteilchen für die nächste Generation. Jede Generation ist also im Durchschnitt  $\mu$ -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 (10.3.1) 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 10.3.4.** 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 10.3.5.** 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 10.3.6.** 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 10.3.7.** 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 10.3.8.** 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 10.3.5.** 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. □

## 10.4. Momenterzeugende Funktion

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 10.4.1.** 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 *momenterzeugende Funktion* von  $X$ .

**Beispiel 10.4.2.** Sei  $X$  eine absolut stetige Zufallsvariable mit Dichte  $f_X$ . Dann ist die momenterzeugende Funktion gegeben durch

$$m_X(t) = \int_{\mathbb{R}} e^{ty} f_X(y) dy.$$

**Aufgabe 10.4.3.** Zeigen Sie, dass  $m_X(0) = 1$ .

**Bemerkung 10.4.4.** Ein Nachteil der momenterzeugende Funktion ist, dass sie auch den Wert  $\infty$  annehmen kann. Im schlimmsten Fall kann sie sogar überall  $+\infty$  sein bis auf den trivialen Wert  $m_X(0) = 1$ . Es sei z.B.  $X$  Cauchy-verteilt, d.h. absolut stetig mit der Dichte

$$f_X(y) = \frac{1}{\pi} \cdot \frac{1}{1 + y^2}, \quad y \in \mathbb{R}.$$

Dann gilt für die momenterzeugende Funktion:

$$m_X(t) = \int_{\mathbb{R}} e^{ty} \cdot \frac{1}{\pi} \cdot \frac{1}{1 + y^2} dy = \begin{cases} +\infty, & t \neq 0, \\ 1, & t = 0. \end{cases}$$

Man kann auch andere Verteilungen konstruieren, die die gleiche momenterzeugende Funktion 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 momenterzeugende Funktion die Verteilung einer Zufallsvariable nicht eindeutig bestimmt.

Im weiteren werden wir voraussetzen, dass die momenterzeugende Funktion endlich in einem Intervall  $(-\varepsilon, \varepsilon)$  ist. Im nächsten Satz berechnen wir die Momente von  $X$  mithilfe der momenterzeugenden Funktion.

**Satz 10.4.5.** 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 10.4.6.** 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:

$$\text{Var } X = m_X''(0) - (m_X'(0))^2.$$

Der obige Satz erklärt auch die Bezeichnung ‘‘momenterzeugende Funktion’’ für  $m_X$ . Die Momente können nämlich aus der erzeugenden Funktion als Ableitungen an der Stelle 0 extrahiert werden.

**Idee des Beweises von Satz 10.4.5.** 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 10.4.5.** 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 dominierten Konvergenz 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 10.4.7.** 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 momenterzeugende Funktion 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:

$$\mathbb{E}X = m'_X(0) = \left( e^{\lambda(e^t-1)} \right)' \Big|_{t=0} = \lambda e^t e^{\lambda(e^t-1)} \Big|_{t=0} = \lambda.$$

Eine wichtige Eigenschaft der momenterzeugenden Funktionen besteht darin, dass einer Faltung der Verteilungen ein Produkt der momenterzeugenden Funktionen entspricht.

**Satz 10.4.8.** Seien  $X, Y$  unabhängige Zufallsvariablen. Dann gilt:

$$m_{X+Y}(t) = m_X(t) \cdot m_Y(t).$$

**Beweis.** Mit der Definition der momenterzeugenden Funktion 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.  $\square$

### 10.5. Charakteristische Funktion (Fourier-Transformierte)

Ein Nachteil der momenterzeugenden Funktion 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 momenterzeugende Funktion festgelegt. Wir werden deshalb eine andere Transformation betrachten, die sich von der momenterzeugenden Funktion 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 10.5.1.** 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 10.5.2.** 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  $\varphi_X(t)$  ist also (im Unterschied zur momenterzeugenden Funktion) für jede Zufallsvariable  $X$  wohldefiniert, denn  $|\cos(tX)| \leq 1$  und  $|\sin(tX)| \leq 1$ .

**Beispiel 10.5.3.** 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 10.5.4.** 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

$$\varphi_X(t) = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it} = \cos(t).$$

Im nächsten Satz listen wir Eigenschaften auf, die jede charakteristische Funktion hat.

**Satz 10.5.5.** Sei  $X$  eine Zufallsvariable. Dann hat die charakteristische Funktion  $\varphi_X$  die folgenden Eigenschaften:

- (a)  $\varphi_X(0) = 1$  und  $|\varphi_X(t)| \leq 1$  für alle  $t \in \mathbb{R}$ .
- (b)  $\varphi_X(-t) = \overline{\varphi_X(t)}$  für alle  $t \in \mathbb{R}$ .
- (c) Die Funktion  $\varphi_X$  ist gleichmäßig stetig auf  $\mathbb{R}$ .
- (d) Die Funktion  $\varphi_X$  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 ist  $(\varphi_X(t_k - t_j))_{k,j=1}^n$  eine positiv semidefinite Matrix.

**Beweis von (a).** Für den Wert an der Stelle  $t = 0$  gilt

$$\varphi_X(0) = \mathbb{E}[e^{i \cdot 0 \cdot X}] = \mathbb{E}1 = 1.$$

Für den Wert an einer beliebigen Stelle  $t \in \mathbb{R}$  gilt mit der Dreiecksungleichung

$$|\varphi_X(t)| = |\mathbb{E}e^{itX}| \leq \mathbb{E}|e^{itX}| = \mathbb{E}1 = 1.$$

**Beweis von (b).** 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 (c).** Seien  $t, h \in \mathbb{R}$ . Dann gilt

$$|\varphi_X(t+h) - \varphi_X(t)| = |\mathbb{E}[e^{i(t+h)X} - e^{itX}]| \leq \mathbb{E}|e^{itX}(e^{ihX} - 1)| = \mathbb{E}|e^{ihX} - 1| \stackrel{\text{def}}{=} g(h).$$

Die rechte eite hängt von  $t$  nicht ab, also folgt

$$\sup_{t \in \mathbb{R}} |\varphi_X(t+h) - \varphi_X(t)| \leq g(h).$$

Nun müssen wir zeigen, dass  $\lim_{h \rightarrow 0} g(h) = 0$ . Zunächst bemerken wir, dass  $\omega \in \Omega$ :  $\lim_{h \rightarrow 0} |e^{ihX(\omega)} - 1| = 0$  für jeden Ausgang gilt. Es folgt, dass

$$|e^{ihX} - 1| \xrightarrow[h \rightarrow 0]{f.s.} 0.$$

Um darauf den Satz von der dominierten Konvergenz anwenden zu dürfen, brauchen wir eine integrierbare Majorante. Eine solche lässt sich leicht konstruieren, denn es gilt 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  $\varphi_X$  gleichmäßig stetig.

**Beweis von (d).** 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 vorangegangenen Satzes, die wir hier nicht beweisen werden.

**Satz 10.5.6** (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  $\varphi_X = \varphi$ .

Im folgenden Lemma berechnen wir, wie sich eine lineare Transformation einer Zufallsvariable auf deren charakteristische Funktion auswirkt.

**Lemma 10.5.7.** Seien  $a, b \in \mathbb{R}$  und  $X$  eine Zufallsvariable. Dann gilt:

$$\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at).$$

**Beweis.** Mit der Definition der charakteristischen Funktion ergibt sich

$$\varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = \mathbb{E}[e^{itb} \cdot e^{i(at)X}] = e^{itb} \cdot \mathbb{E}[e^{i(at)X}] = e^{itb} \cdot \varphi_X(at).$$

□

**Satz 10.5.8.** Sei  $X \sim N(\mu, \sigma^2)$ . Dann gilt

$$\varphi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

**Beispiel 10.5.9.** Für eine standardnormalverteilte Zufallsvariable  $X \sim N(0, 1)$  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 momenterzeugende Funktion:

$$\begin{aligned} m_X(t) &= \int_{-\infty}^{\infty} e^{ty} f_X(y) dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} \cdot e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2} dy \\ &= e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \\ &= e^{\frac{t^2}{2}}, \end{aligned}$$

wobei wir die neue Variable  $z = y - t$  eingeführt haben.

SCHRITT 2. Aus der momenterzeugenden Funktion 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  $\mu, \sigma^2$ . Dann haben wir die Darstellung  $X = \sigma Y + \mu$ , wobei  $Y \sim N(0, 1)$ . Mit Lemma 10.5.7 gilt dann:

$$\varphi_X(t) = e^{it\mu} \cdot \varphi_Y(\sigma t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

□

**Satz 10.5.10.** Seien  $X, Y$  unabhängige Zufallsvariablen. Dann gilt

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t).$$

**Beweis.** Übung. □

In den folgenden Sätzen werden wir zeigen, dass die charakteristische Funktion die komplette Information über die Verteilung der Zufallsvariable enthält. Zuerst werden wir zeigen, wie man aus der charakteristischen Funktion die Momente der Verteilung extrahiert. Nun berechnen wir Momente mithilfe der charakteristischen Funktion.

**Satz 10.5.11.** Sei  $X$  eine Zufallsvariable mit  $\mathbb{E}|X|^n < \infty$  für ein  $n \in \mathbb{N}$ . Dann ist die charakteristische Funktion  $\varphi_X$   $n$ -mal stetig differenzierbar und es gilt:

$$\mathbb{E}[X^n] = i^{-n} \varphi_X^{(n)}(0).$$

Zum Beweis benötigen wir ein Lemma.

**Lemma 10.5.12.** 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 10.5.11.** SCHRITT 1. Wir betrachten den Fall  $n = 1$ . Sei also  $\mathbb{E}|X| < \infty$ . Wir werden zeigen, dass  $\varphi_X$  stetig differenzierbar ist und dass

$$\varphi_X'(t) = \mathbb{E}[iXe^{itX}]. \quad (10.5.1)$$

Wir stellen den Differenzenquotienten der Funktion  $\varphi_X$  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

$$\lim_{h \rightarrow 0} \left( e^{itX} \frac{e^{ihX} - 1}{h} \right) = iXe^{itX}.$$

Außerdem erhalten wir mit Lemma 10.5.12 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:

$$\lim_{h \rightarrow 0} \frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathbb{E} [iXe^{itX}].$$

Somit ist  $\varphi_X$  differenzierbar und die Formel (10.5.1) gilt. Außerdem ist die Funktion  $\mathbb{E}[iXe^{itX}]$  stetig. Der Beweis dafür ist identisch mit dem Beweis, dass  $\varphi_X$  stetig ist.

Nun setzen wir  $t = 0$  in (10.5.1) 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 10.5.13.** Sei  $X \sim N(0, 1)$  standardnormalverteilt. Die Taylor-Entwicklung der charakteristischen Funktion lautet:

$$\varphi_X(t) = e^{-t^2/2} = 1 - \frac{1}{2}t^2 + o(t^2).$$

Wir berechnen die ersten beiden Momente von  $X$  indem wir die charakteristische Funktion ableiten:

$$\begin{aligned} \varphi'_X(0) &= 0 & \text{also } \mathbb{E}X &= 0, \\ \varphi''_X(0) &= -1 & \text{also } \mathbb{E}X^2 &= i^{-2} \cdot (-1) = 1. \end{aligned}$$

Also gilt für die Varianz von  $X$ , dass  $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = 1$ .

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 10.5.14** (Umkehrformel). Sei  $Z$  eine Zufallsvariable mit charakteristischer Funktion  $\varphi_Z$  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 10.5.15.** 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  $\mathbb{P}[a < z < b]$ .  $\square$

**Satz 10.5.16** (Eindeutigkeitssatz). Die charakteristische Funktion bestimmt die Verteilungsfunktion eindeutig. Sind nämlich  $X$  und  $Y$  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 10.5.17** (Umkehrformel für die Dichte). Sei  $Z$  eine Zufallsvariable, deren charakteristische Funktion  $\varphi_Z$  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 10.5.16), dass  $g$  die Dichte von  $Z$  ist.  $\square$

**Beispiel 10.5.18.** 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 den folgenden Satz bewiesen:

**Satz 10.5.19.** Für die charakteristische Funktion einer Cauchy-verteilten Zufallsvariable  $Z$  gilt

$$\varphi_Z(t) = e^{-|t|}.$$

**Aufgabe 10.5.20.** Seien  $X_1, X_2, \dots$  unabhängige Cauchy-verteilte Zufallsvariablen. Zeigen Sie, dass

$$\frac{X_1 + \dots + X_n}{n}$$

ebenfalls Cauchy-verteilt ist.

**Beispiel 10.5.21.** 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

### 11.1. Einführung

Es seien  $X_1, X_2, \dots$  unabhängige und identisch verteilte Zufallsvariablen. Wir können uns z.B. vorstellen, dass mehrere unabhängige Runden eines Glücksspiels gespielt werden und dass  $X_i$  den Gewinn in der  $i$ -ten Runde modelliert. Der Gesamtgewinn in den ersten  $n$  Runden ist dann  $S_n := X_1 + \dots + X_n$ . Das Gesetz der großen Zahlen (in seinen verschiedenen Formulierungen) behauptet im Wesentlichen, dass  $S_n$  durch  $n\mathbb{E}X_1$  approximiert werden kann, wenn  $n \rightarrow \infty$ . Aber wie groß ist der Fehler dieser Approximation? Diese Frage wird durch den zentralen Grenzwertsatz beantwortet. Um diesen zu formulieren, nehmen wir an, dass  $\mathbb{E}[X_1^2] < \infty$  und bezeichnen mit  $\mu := \mathbb{E}X_1$  den Erwartungswert und mit  $\sigma^2 := \text{Var } X_1$  die Varianz von  $X_1$ . Zuerst stellen wir fest, dass

$$\mathbb{E}S_n = n\mu, \quad \text{Var } S_n = n\sigma^2.$$

Vereinfacht gesprochen behauptet der zentrale Grenzwertsatz, dass die Verteilung von  $S_n$  durch eine Normalverteilung mit Erwartungswert  $n\mu$  und Varianz  $n\sigma^2$  approximiert werden kann, wenn  $n$  groß ist:

$$S_n \approx N(n\mu, n\sigma^2) \quad [\text{in Verteilung, wenn } n \rightarrow \infty].$$

Bezeichnen wir mit  $\xi \sim N(0, 1)$  eine standardnormalverteilte Zufallsvariable, so können wir die Approximation auch wie folgt formulieren:

$$S_n \approx n\mu + \sigma\sqrt{n}\xi \quad [\text{in Verteilung, wenn } n \rightarrow \infty].$$

Approximieren wir  $S_n$  durch  $n\mu$ , so entsteht dabei ein zufälliger Fehler von ungefähr  $\sigma\sqrt{n}\xi$ . Es sei bemerkt, dass die Größe des Fehlers bei steigendem  $n$  wie  $\sqrt{n}$  wächst. Wir schauen uns nun einige Spezialfälle an.

**Beispiel 11.1.1.** Wir werfen einen fairen Würfel  $n = 100$  Mal. Die Augensumme kann als  $S_n = X_1 + \dots + X_n$  dargestellt werden, wobei die einzelnen Augenzahlen  $X_1, \dots, X_n$  unabhängig und uniform verteilt auf  $\{1, \dots, 6\}$  sind. Die Zähldichte von  $S_n$  wird auf dieser Abbildung dargestellt:

Es fällt auf, dass die Zähldichte die Form einer Glocke hat, deren Mittelpunkt in der Nähe von 350 liegt, was genau dem Erwartungswert  $\mathbb{E}S_n = 3.5n = 350$  entspricht.

**Beispiel 11.1.2.** Ein ähnliches Bild ergibt sich, wenn man Summen von u.i.v. Zufallsvariablen mit absolut stetigen Verteilungen betrachtet. Seien z.B.  $X_1, X_2, \dots$  unabhängig und mit Parameter  $\lambda = 1$  exponentialverteilt. Die Dichten von  $S_n = X_1 + \dots + X_n$  mit  $n = 100, 300, 500$  sehen folgendermaßen aus:

Mit steigendem  $n$  "verschiebt" sich die Glocke nach rechts, wobei ihr Mittelpunkt in der Nähe von  $n$  liegt. Dies ist nicht überraschend, denn für den Erwartungswert von  $S_n$  gilt  $\mathbb{E}S_n = n\mathbb{E}X_1 = n$ . Außerdem wird die Glocke mit steigendem  $n$  breiter. Dies liegt daran,

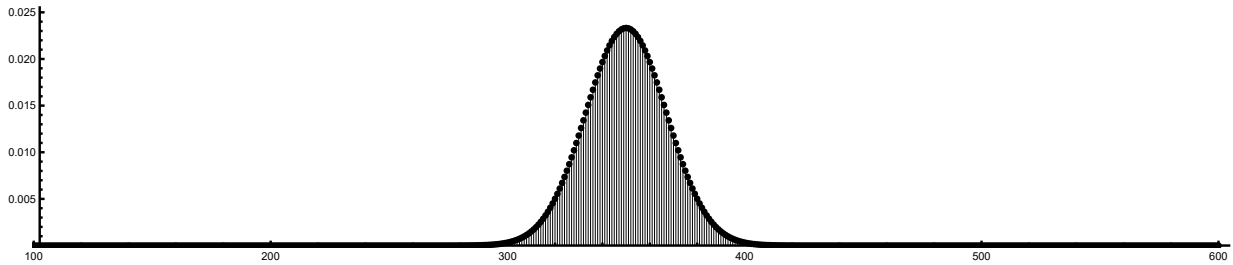


ABBILDUNG 1. Zähldichte der Augensumme der  $n = 100$  Würfe mit einem fairen Würfel.

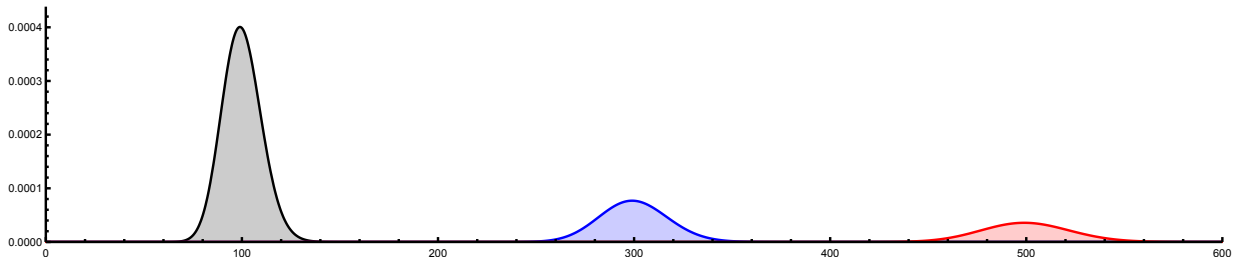


ABBILDUNG 2. Dichte der Summe  $S_n = X_1 + \dots + X_n$  mit  $X_k \sim \text{Exp}(1)$  und  $n = 100, 300, 500$ .

dass die Varianz von  $S_n$  steigt. In der Tat, wegen der Unabhängigkeit der  $X_i$ 's gilt  $\text{Var } S_n = n \text{Var } X_1 = n$ , siehe Aufgabe 7.1.14. Wegen der steigenden Breite der Glocke und wegen der Bedingung, dass das Integral einer Dichte 1 ist, muss die Glocke mit steigendem  $n$  außerdem flacher werden, was wir tatsächlich beobachten.

Es fällt auf, dass in allen Beispielen die Zähldichten (bzw. die Dichten) die Form einer Glocke besitzen, deren Mittelpunkt und Breite allerdings von Fall zu Fall variieren. Um dieser Beobachtung gerecht zu werden, können wir versuchen, alle Glocken zu standardisieren. Die beiden Parameter, die die Position des Mittelpunktes und die Breite der Glocke beschreiben, sind der Erwartungswert  $\mathbb{E}S_n$  und die Standardabweichung  $\sqrt{\text{Var } S_n}$ .

Um die obige Behauptung als eine Grenzwertaussage zu formulieren, betrachten wir anstelle von  $S_n$  deren zentrierte und normierte Version

$$T_n := \frac{S_n - \mathbb{E}S_n}{\sqrt{\text{Var } S_n}}.$$

Es gilt dann offenbar  $\mathbb{E}T_n = 0$  und  $\text{Var } T_n = 1$ . Der zentrale Grenzwertsatz behauptet, dass die Verteilung von  $T_n$  für  $n \rightarrow \infty$  in einem gewissen Sinne gegen die Standardnormalverteilung konvergiert.

**Satz 11.1.3.** Es seien  $X_1, X_2, \dots$  unabhängige und identisch verteilte Zufallsvariablen mit  $\mathbb{E}[X_1^2] < \infty$  und  $\text{Var } X_1 \neq 0$ . Dann gilt für alle  $x \in \mathbb{R}$ , dass

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \frac{S_n - \mathbb{E}S_n}{\sqrt{\text{Var } S_n}} \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Die rechte Seite ist die Verteilungsfunktion der Standardnormalverteilung. Der zentrale Grenzwertsatz wird als eine Aussage über die punktweise Konvergenz der Verteilungsfunktionen formuliert, denn diese existieren für beliebige Zufallsvariablen. Dichten oder Zähldichten, die wir in den obigen Beispielen betrachtet haben, existieren dagegen nicht immer.

## 11.2. Konvergenz in Verteilung

In diesem Kapitel werden wir einen weiteren Konvergenzbegriff für Zufallsvariablen einführen und seine Eigenschaften formulieren.

**Definition 11.2.1.** Eine Folge von Zufallsvariablen  $X_1, X_2, \dots$  konvergiert *in Verteilung* gegen eine Zufallsvariable  $X$ , falls

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t) \quad \text{für alle } t \in S(X),$$

wobei  $S(X)$  die Menge der Stetigkeitspunkte der Verteilungsfunktion von  $X$  bezeichnet, d.h.

$$S(X) = \{t \in \mathbb{R} : F_X \text{ ist stetig an der Stelle } t\}.$$

Wir schreiben dann  $X_n \xrightarrow[n \rightarrow \infty]{d} X$  oder  $F_{X_n} \xrightarrow[n \rightarrow \infty]{d} F_X$ . Dabei steht das “ $d$ ” für “distribution” (Verteilung).

**Bemerkung 11.2.2.** 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 11.2.3.** Das Komplement von  $S(X)$  stimmt mit der Menge der Atome von  $X$  überein und ist somit höchstens abzählbar. In der Definition der Verteilungskonvergenz mag die Einschränkung auf die Stetigkeitspunkte von  $F_X$  am Anfang unnatürlich erscheinen. In diesem Beispiel zeigen wir, dass sie dennoch sinnvoll ist. Seien  $c_1 > c_2 > \dots > 0$  Konstanten mit  $\lim_{n \rightarrow \infty} c_n = 0$ . Als Beispiel kann man etwa  $c_n = \frac{1}{n}$  betrachten. Definiere nun Zufallsvariablen  $X_n := c_n$  und  $X := 0$ . Es gilt:

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \\ 0, & t = 0 \end{cases} = F_X(t), \quad \text{für alle } t \neq 0.$$

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 11.2.4.** Seien  $X_1, X_2, \dots$  und  $X$  Zufallsvariablen mit  $X_n \xrightarrow[n \rightarrow \infty]{P} X$ . Dann gilt  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ .

**Bemerkung 11.2.5.** 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 11.2.4.** Sei  $X_n \xrightarrow[n \rightarrow \infty]{P} X$ . Für alle  $\varepsilon > 0$  gilt also:

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0.$$

Zu zeigen ist, dass

$$F_X(t) = \lim_{n \rightarrow \infty} F_{X_n}(t) \quad \text{für alle } t \in S(X).$$

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  $X_n \xrightarrow[n \rightarrow \infty]{P} X$  folgt, dass

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t + \varepsilon).$$

Das gilt für jedes  $\varepsilon > 0$ . Wir können also  $\varepsilon \downarrow 0$  gehen lassen und die Rechtsstetigkeit von  $F_X$  benutzen:

$$\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t).$$

Dies ist die Abschätzung nach oben.

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  $X_n \xrightarrow[n \rightarrow \infty]{P} X$  folgt nun, dass

$$F_X(t - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

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:

$$F_X(t) \leq \liminf_{n \rightarrow \infty} F_{X_n}(t).$$

Dies beweist die untere Abschätzung.

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 11.2.6.** 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  $X_n \xrightarrow[n \rightarrow \infty]{L^2} 0$ , 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 11.2.7.** Dieses Beispiel soll zeigen, dass die Umkehrung von Satz 11.2.4 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 11.2.4 richtig ist. Man muss nämlich voraussetzen, dass die Grenzwertzufallsvariable konstant ist.

**Satz 11.2.8.** 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 11.2.9.** 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. Wir zeigen, dass

$$X_n \xrightarrow[n \rightarrow \infty]{d} X,$$

wobei  $X$  gleichverteilt auf  $[0, 1]$  ist. Es sei bemerkt, dass  $X_n$  diskret, wohingegen  $X$  absolut stetig ist.

**Beweis.** Für  $t < 0$  gilt  $F_{X_n}(t) = F_X(t) = 0$ . Für  $t > 1$  gilt  $F_{X_n}(t) = F_X(t) = 1$ . Schließlich gilt im nichttrivialen Fall  $t \in [0, 1]$ , dass

$$F_{X_n}(t) = \frac{[tn]}{n} \xrightarrow[n \rightarrow \infty]{} t = F_X(t),$$

wobei  $[tn]$  die Gauß-Klammer von  $tn$  bezeichnet. Insgesamt 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 11.2.10** (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:  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ . □

### 11.3. Eine Charakterisierung der Konvergenz in Verteilung

**Definition 11.3.1.** Die Menge der stetigen und beschränkten Funktionen bezeichnen wir mit

$$C_b := \left\{ f : \mathbb{R} \rightarrow \mathbb{R} \mid f \text{ ist stetig und } \sup_{t \in \mathbb{R}} |f(t)| \leq \infty \right\}.$$

Der nächste Satz gibt eine äquivalente Charakterisierung der Verteilungskonvergenz an.

**Satz 11.3.2.** Seien  $X, X_1, X_2, \dots$  Zufallsvariablen. Die folgenden zwei Aussagen sind äquivalent:

- (a)  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ .
- (b) Für alle  $f \in C_b$  gilt:  $\lim_{n \rightarrow \infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ .

**Beweis von (a)  $\Rightarrow$  (b).** 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

$$\mathbb{P}[|X| > c] < \frac{\varepsilon}{b}.$$

Die Mengen der Atome von  $X_n$  und  $X$  sind höchstens abzählbar, die Vereinigung all dieser Mengen ebenso. Indem man  $c$  vergrößert, kann man also erreichen, dass  $c, -c \in S(X)$  und  $c, -c \in S(X_n)$  für alle  $n \in \mathbb{N}$  gilt.

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)  $\sup_{|t| \leq c} |f(t) - g(t)| \leq \varepsilon$ .
- (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  $\lim_{n \rightarrow \infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ , was den Beweis der Implikation (a)  $\Rightarrow$  (b) abschließt.

**Beweis von (b)  $\Rightarrow$  (a).** 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:

$$F_{X_n}(t) = \mathbb{E}\mathbb{1}_{X_n \leq t} \leq \mathbb{E}f(X_n).$$

(1'') Außerdem gilt  $\mathbb{1}_{X \leq t} \leq f(X) \leq \mathbb{1}_{X \leq t + \varepsilon}$ . Es folgt:

$$\mathbb{E}f(X) \leq \mathbb{E}\mathbb{1}_{X \leq t + \varepsilon} = F_X(t + \varepsilon).$$

Fasst man nun (1') und (1'') 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. Fasst man nun Schritte 1 und 2 zusammen, so folgt für alle  $t \in S(X)$ :

$$F_X(t) = \lim_{n \rightarrow \infty} F_{X_n}(t).$$

Somit haben wir gezeigt, dass  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ . □

**Aufgabe 11.3.3.** Zeigen Sie unter Anwendung von Satz 11.3.2, dass die diskrete uniforme Verteilung auf  $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$  gegen die stetige Uniformverteilung auf  $[0, 1]$  konvergiert.

#### 11.4. Straffheit und der Satz von Helly

Der Satz von Bolzano–Weierstraß besagt, dass man aus jeder beschränkten Folge  $a_1, a_2, \dots$  von reellen Zahlen eine konvergente Teilfolge  $a_{n_1}, a_{n_2}, \dots$  extrahieren kann. In diesem Kapitel beweisen wir eine Reihe ähnlicher Sätze für Folgen von Wahrscheinlichkeitsmaßen oder Verteilungsfunktionen.

**Satz 11.4.1** (Helly). Seien  $F_1, F_2, \dots$  beliebige Verteilungsfunktionen. Dann gibt es eine Teilfolge  $F_{n_1}, F_{n_2}, \dots$  sowie eine rechtsstetige monoton nichtfallende Funktion  $G : \mathbb{R} \rightarrow [0, 1]$  mit

$$\lim_{k \rightarrow \infty} F_{n_k}(t) = G(t) \quad \text{für alle } t \in S(G),$$

wobei  $S(G)$  die Menge der Stetigkeitspunkte der Funktion  $G$  bezeichnet.

**Beispiel 11.4.2.** Die Grenzwertfunktion  $G$  aus dem Satz von Helly ist zwar monoton nichtfallend und rechtsstetig, 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 tatsächlich verletzt werden können, betrachten wir die deterministische Folge  $X_n = n$ ,  $n \in \mathbb{N}$ . 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  $F_{X_{n_1}}, F_{X_{n_2}}, \dots$  identisch gleich 0. Die Funktion  $G(t) = 0$  ist aber keine Verteilungsfunktion. Ein anderes Beispiel kann man erhalten, indem man  $X_n = -n$  wählt. Dann gilt  $G(t) = 1$ .

**Beweis von Satz 11.4.1.** Die Menge der rationalen Zahlen ist abzählbar, also gibt es eine Abzählung  $\mathbb{Q} = \{r_1, r_2, \dots\}$ . Wir werden im Folgenden immer dünner werdende Teilfolgen von  $F_1, F_2, \dots$  konstruieren, die die Eigenschaft besitzen, dass die erste Teilfolge an der Stelle  $r_1$ , die zweite an den Stellen  $r_1, r_2$ , die dritte an den Stellen  $r_1, r_2, r_3$ , usw. konvergieren. Danach werden wir aus diesen Teilfolgen eine noch dünnere Diagonaltteilfolge extrahieren, die an allen rationalen Stellen konvergiert. Daraus werden wir schließlich die Konvergenz an allen reellen Stellen herleiten. Die Indizes der oben erwähnten Teilfolgen bilden eine geschachtelte Familie von unendlichen Mengen  $\mathbb{N} = K_0 \supset K_1 \supset \dots$ , die folgendermaßen konstruiert wird.

SCHRITT 1. Sei  $K_0 = \mathbb{N}$ . Es gilt offenbar, dass  $\{F_n(r_1)\}_{n \in \mathbb{N}} \subset [0, 1]$ . Mit dem Satz von Bolzano–Weierstraß können wir eine unendliche Teilmenge  $K_1 \subset K_0$  konstruieren, sodass

$$\lim_{n \rightarrow \infty, n \in K_1} F_n(r_1) = g_1.$$

Somit haben wir eine an der Stelle  $r_1$  konvergente Teilfolge konstruiert.

SCHRITT 2. Nun schauen wir uns die Werte dieser Teilfolge an der Stelle  $r_2$  an. Es gilt offenbar  $\{F_n(r_2)\}_{n \in K_1} \subset [0, 1]$ . Und damit folgt wiederum mit dem Satz von Bolzano–Weierstraß, dass

eine unendliche Teilmenge  $K_2 \subset K_1$  existiert mit

$$\lim_{n \rightarrow \infty, n \in K_2} F_n(r_2) = g_2.$$

Somit haben wir eine an den Stellen  $r_1$  und  $r_2$  konvergente Teilfolge konstruiert.

SCHRITT 3. Diese Argumentation kann induktiv fortgeführt werden. Es existieren also ineinander geschachtelte, unendliche Mengen  $\mathbb{N} = K_0 \supset K_1 \supset K_2 \supset \dots$  und

$$\lim_{n \rightarrow \infty, n \in K_j} F_n(r_j) = g_j \text{ für alle } j \in \mathbb{N}.$$

Ohne Einschränkung der Allgemeinheit können wir annehmen, dass das erste Element jeder Teilfolge in der nächsten Teilfolge nicht mehr präsent ist.

SCHRITT 4. Nun wollen wir aber eine Teilfolge extrahieren, die an allen rationalen Stellen konvergiert. Dazu benutzen wir einen eleganten Trick, die sogenannte Cantor–Diagonalisierung. Wir erklären die Methode zuerst an einem Beispiel. Als Ergebnis der ersten drei Beweisschritte könnte z.B. das folgende System von immer dünner werdenden Teilfolgen entstanden sein:

$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$	$F_9$	$F_{10}$	$F_{11}$	$F_{12}$	$\dots$	konv. an der Stelle $r_1$
	$F_3$	$F_4$	$F_5$	$F_6$			$F_8$	$F_9$	$F_{10}$		$F_{12}$	$\dots$	konv. an den Stellen $r_1, r_2$
				$F_6$			$F_9$			$F_{12}$	$\dots$		konv. an den Stellen $r_1, r_2, r_3$
$\dots$	$\dots$	$\dots$	$\dots$										

Wir behaupten nun, dass die ersten Elemente der jeweiligen Reihen, also in unserem Beispiel  $F_1, F_3, F_6, \dots$  eine an *allen* rationalen Stellen konvergente Teilfolge bilden.

Wir möchten nun die Notation für das oben beschriebene Verfahren einführen. In jeder Menge  $K_j$  wählen wir das kleinste Element und bezeichnen es mit  $\min K_j$ . Ohne Die Vereinigung dieser Elemente nennen wir  $K$ , so dass  $K = \{\min K_1, \min K_2, \dots\}$ . Nach Konstruktion ist die Menge  $K$  unendlich. Außerdem gilt für alle  $j \in \mathbb{N}$ , dass  $K$  bis auf endlich viele Elemente eine Teilmenge von  $K_j$  ist. Da sich der Limes beim Übergang zur Teilfolge und durch das Hinzufügen von endlich vielen Termen nicht ändert, ergibt sich, dass

$$\lim_{n \rightarrow \infty, n \in K} F_n(r_j) = g_j \text{ für alle } j \in \mathbb{N}.$$

Wir haben somit gezeigt, dass für alle  $j \in \mathbb{N}$  der folgende Grenzwert existiert:

$$H(r_j) := \lim_{n \rightarrow \infty, n \in K} F_n(r_j).$$

Somit haben wir eine Teilfolge von  $F_1, F_2, \dots$  konstruiert, die an allen rationalen Stellen punktweise konvergiert.

SCHRITT 5. Die Funktion  $H$  ist nur auf der Menge der rationalen Zahlen definiert. Wir können sie aber auf ganz  $\mathbb{R}$  wie folgt erweitern:

$$G(t) := \inf\{H(r) : r \in \mathbb{Q}, r > t\}, \quad t \in \mathbb{R}.$$

Der Leser möge als Übung zeigen, dass  $G$  monoton nichtfallend und rechtsstetig ist. Würde man in der Definition von  $G$  das Infimum über  $r \geq t$  bilden, wäre  $G$  linksstetig.

Nun wollen wir zeigen, dass

$$\lim_{n \rightarrow \infty, n \in K} F_n(t) = G(t) \quad \text{für alle } t \in S(G).$$

Sei also  $t \in S(G)$  und sei  $\varepsilon > 0$ . Wegen Stetigkeit von  $G$  an der Stelle  $t$  existiert ein  $y < t$  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:

$$G(t) - \varepsilon \leq G(y) \leq G(r) \leq G(s) \leq G(t) + \varepsilon.$$

Aus Schritt 4 folgt, dass

$$G(y) \leq H(r) \xrightarrow[n \rightarrow \infty]{n \in K} F_n(r) \leq F_n(t) \leq F_n(s) \xrightarrow[n \rightarrow \infty]{n \in K} H(s) \leq G(s).$$

Somit gilt

$$G(t) - \varepsilon \leq \liminf_{n \rightarrow \infty, n \in K} F_n(t) \leq \limsup_{n \rightarrow \infty, n \in K} F_n(t) \leq G(t) + \varepsilon.$$

Durch den Grenzwertübergang  $\varepsilon \downarrow 0$  ergibt sich schließlich  $G(t) = \lim_{n \rightarrow \infty, n \in K} F_n(t)$ .  $\square$

Der Satz von Helly kann als eine Kompaktheitsaussage über einen passenden Raum der Wahrscheinlichkeitsmaße interpretiert werden.

**Definition 11.4.3.** Ein metrischer Raum  $(M, \rho)$  heißt **kompakt**, wenn aus jeder Folge  $a_1, a_2, \dots \in M$  eine konvergente Teilfolge extrahiert werden kann.

**Beispiel 11.4.4.** Der Satz von Bolzano–Weierstraß besagt, dass jedes abgeschlossene Intervall  $[a, b]$  (mit der Metrik  $\rho(x, y) = |y - x|$ ) kompakt ist.

Sei nun  $\mathcal{M}$  die Menge aller Wahrscheinlichkeitsmaße auf  $\mathbb{R}$ . Da jedes Wahrscheinlichkeitsmaß eindeutig durch seine Verteilungsfunktion bestimmt ist, können wir auch

$$\mathcal{M} := \left\{ F : \mathbb{R} \rightarrow [0, 1] : F \text{ ist monoton nichtfallend, rechtsstetig mit} \right. \\ \left. \lim_{t \rightarrow -\infty} F(t) = 0 \text{ und } \lim_{t \rightarrow +\infty} F(t) = 1 \right\}$$

schreiben. Weiterhin können wir  $\mathcal{M}$  mit der Lévy–Metrik versehen (Definition???) , so dass Konvergenz in dieser Metrik mit der Konvergenz in Verteilung übereinstimmt. Ist nun der Raum  $\mathcal{M}$  kompakt? Die Antwort ist negativ, wie bereits in Beispiel 11.4.2 gezeigt wurde. Bezeichnen wir mit  $\delta_n$  die Dirac–Einheitsmasse (???) an der Stelle  $n$ , so hat die Folge  $\delta_1, \delta_2, \dots$  keine in  $\mathcal{M}$  konvergente Teilfolge. Dabei konvergiert die Folge der entsprechenden Verteilungsfunktionen punktweise gegen 0, was allerdings keine Verteilungsfunktion ist. Der Grund für die Nichtkompaktheit von  $\mathcal{M}$  ist, vereinfacht gesprochen, die Möglichkeit des Verschwindens der Masse in  $\pm\infty$ . Wir können auch etwas allgemeinere Beispiele dieser Art wie folgt konstruieren.

**Beispiel 11.4.5.** Seien  $p_+, p_-$  und  $q$  positive Zahlen mit  $p_+ + p_- + q = 1$ . Wir betrachten die folgenden Wahrscheinlichkeitsmaße auf  $\mathbb{R}$ :

$$\mu_n := p_- \delta_{-n} + p_+ \delta_n + (1 - p_+ - p_-) \delta_0.$$

Lassen wir  $n \rightarrow \infty$ , so “verschwinden” die Massen  $p_-$  und  $p_+$  in  $-\infty$  bzw.  $+\infty$ , während die Masse  $q$  an der Stelle 0 verbleibt. Man kann nachrechnen, dass die Verteilungsfunktionen von  $\mu_n$  punktweise gegen die folgende Funktion konvergieren:

$$G(t) = \begin{cases} p_-, & \text{falls } t < 0, \\ 1 - p_+, & \text{falls } t \geq 0. \end{cases}$$

Dabei ist  $G$  keine Verteilungsfunktion, so dass weder die Folge  $\mu_n$  noch deren Teilfolgen im Raum  $\mathcal{M}$  konvergieren.

Um der Tatsache Rechnung zu tragen, dass die Masse in  $\pm\infty$  verschwinden kann, werden wir den Raum  $\mathcal{M}$  wie folgt erweitern. Sei  $\overline{\mathcal{M}}$  die Menge aller Wahrscheinlichkeitsmaße auf  $\mathbb{R} \cup \{+\infty, -\infty\}$ . Solche Wahrscheinlichkeitsmaße dürfen Atome an den Stellen  $+\infty$  und  $-\infty$  haben. Dabei können wir jedes Element aus  $\overline{\mathcal{M}}$  als die Verteilung einer erweiterten Zufallsvariable  $X : \Omega \rightarrow \mathbb{R} \cup \{+\infty, -\infty\}$  auffassen, die auch die beiden Werte  $\pm\infty$  annehmen darf. Die “Verteilungsfunktion” (im erweiterten Sinne) von  $X$  ist dann definiert als

$$F_X(t) := \mathbb{P}[X \leq t] = \mu([-\infty, t]), \quad t \in \mathbb{R}.$$

Dabei ist  $F_X$  monoton nichtfallend und rechtsstetig, allerdings keine Verteilungsfunktion im üblichen Sinne, denn für ihre Grenzwerte gilt lediglich

$$\lim_{t \rightarrow -\infty} F_X(t) = \mathbb{P}[X = -\infty] \geq 0, \quad \lim_{t \rightarrow +\infty} F_X(t) = 1 - \mathbb{P}[X = +\infty] \leq 1.$$

Wir können die Elemente von  $\overline{\mathcal{M}}$  mit solchen Funktionen identifizieren:

$$\overline{\mathcal{M}} := \left\{ F : \mathbb{R} \rightarrow [0, 1] : F \text{ ist monoton nichtfallend und rechtsstetig} \right\}.$$

Die Konvergenz der Funktionen aus  $\overline{\mathcal{M}}$  wird genauso wie für übliche Verteilungsfunktionen definiert. Eine leichte Modifizierung des Beweises des Satzes von Helly ergibt die folgende Aussage:

**Satz 11.4.6.** Jede Folge  $F_1, F_2, \dots \in \overline{\mathcal{M}}$  besitzt eine gegen ein Element aus  $\overline{\mathcal{M}}$  konvergente Teilfolge.

Somit ist der Raum  $\overline{\mathcal{M}}$  kompakt. Dieses Resultat ist überraschend, denn auf den ersten Blick scheint der Raum  $\overline{\mathcal{M}}$  “sehr groß” zu sein. Eine mindestens ebenso überraschende (und eng verwandte!) Aussage über die Kompaktheit ist der folgende Satz von Tychonoff: Ein Produkt von beliebig vielen kompakten topologischen Räumen ist kompakt. So ist z.B. das Produkt von Kontinuum-vielen Kopien des Intervalls  $[0, 1]$  (das mit der Menge aller Funktionen  $f :$

$\mathbb{R} \rightarrow [0, 1]$  identifiziert werden kann), versehen mit der Produkttopologie, kompakt! Für mehr Einzelheiten verweisen wir auf das Buch von K. Jänich, "Topologie", Kapitel 10.

Wir kehren nun zum nichtkompakten Raum  $\mathcal{M}$  aller Verteilungsfunktionen zurück. Wir werden einen Begriff einführen, der im gewissen Sinne das Verschwinden der Masse in  $\pm\infty$  ausschließt.

**Definition 11.4.7.** Eine Familie  $\mathcal{X} = \{X_i : i \in I\}$  von Zufallsvariablen heißt *straff*, wenn für jedes  $\varepsilon > 0$  ein  $c > 0$  existiert mit

$$\sup_{X \in \mathcal{X}} \mathbb{P}[|X| > c] < \varepsilon.$$

Wir werden diese Definition meistens in der Situation anwenden, wenn  $\mathcal{X} = \{X_1, X_2, \dots\}$  eine Folge von Zufallsvariablen ist.

**Beispiel 11.4.8.** Sei  $A$  eine positive Zahl und betrachte die Familie  $\mathcal{X}$  aller Zufallsvariablen, die Werte im Intervall  $[-A, A]$  annehmen. Dann ist diese Familie straff, denn wir können in der obigen Definition  $c := A$  wählen.

**Beispiel 11.4.9.** Sei  $X$  eine beliebige Zufallsvariable. Die Familie  $\{X\}$  ist straff (Übung). Ebenso ist jede endliche Familie  $\mathcal{X} = \{X_1, \dots, X_k\}$  straff.

**Beispiel 11.4.10.** Die Folge  $X_n := n$ ,  $n \in \mathbb{N}$ , ist nicht straff.

**Aufgabe 11.4.11.** Es sei  $X_1, X_2, \dots$  eine Folge von Zufallsvariablen mit  $\sup_{n \in \mathbb{N}} \mathbb{E}|X_n|^p < \infty$  für ein  $p > 0$ . Zeigen Sie, dass diese Folge straff ist.

**Aufgabe 11.4.12.** Es seien  $\{X_1, X_2, \dots\}$  und  $\{Y_1, Y_2, \dots\}$  zwei straffe Familien von Zufallsvariablen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{P})$ . Zeigen Sie, dass auch die Familie  $\{X_n + Y_n : n \in \mathbb{N}\}$  straff ist.

**Aufgabe 11.4.13.** Es seien  $\{X_1, X_2, \dots\}$  eine straffe Familie von Zufallsvariablen, und  $Y_1, Y_2, \dots$  eine Folge von Zufallsvariablen, die gegen 0 in Wahrscheinlichkeit konvergiert. Zeigen Sie, dass  $X_n Y_n \rightarrow 0$  in Wahrscheinlichkeit.

**Aufgabe 11.4.14.** Seien  $X_1, X_2, \dots$  Poisson-verteilte Zufallsvariablen mit  $X_n \sim \text{Poi}(\lambda_n)$ . Zeigen Sie: ist die Folge  $\lambda_1, \lambda_2, \dots$  beschränkt, so ist die Familie  $\{X_n : n \in \mathbb{N}\}$  straff.

**Aufgabe 11.4.15.** Seien  $X_1, X_2, \dots$  normalverteilte Zufallsvariablen mit  $X_n \sim N(\mu_n, \sigma_n^2)$ . Zeigen Sie: sind die Folgen  $\mu_1, \mu_2, \dots$  und  $\sigma_1^2, \sigma_2^2, \dots$  beschränkt, so ist die Familie  $\{X_n : n \in \mathbb{N}\}$  straff.

Wir können den Begriff der Straffheit in die Sprache der entsprechenden Verteilungsfunktionen  $F_X(t) := \mathbb{P}[X \leq t]$  übersetzen.



**Definition 11.4.16.** Eine Familie  $\mathcal{F} = \{F_i : i \in I\}$  von Verteilungsfunktionen heißt *straff*, wenn für jedes  $\varepsilon > 0$  ein  $c > 0$  existiert, so dass

$$\sup_{F \in \mathcal{F}} (F(-c) + 1 - F(c)) < \varepsilon.$$

Der nächste Satz behauptet, dass man aus einer straffen Folge von Verteilungsfunktionen stets eine in Verteilung konvergente Teilfolge extrahieren kann. Er ist analog zum Satz von Bolzano–Weierstraß, der behauptet, dass man aus einer beschränkten Folge von reellen Zahlen stets eine konvergente Teilfolge extrahieren kann. Straffheit von Folgen von Verteilungsfunktionen ist somit analog zur Beschränktheit von Zahlenfolgen.

**Satz 11.4.17** (Helly–Prochorow). Sei  $F_1, F_2, \dots$  eine **straffe** Folge von Verteilungsfunktionen. Dann existieren eine **Verteilungsfunktion**  $G$  sowie eine Teilfolge  $F_{n_1}, F_{n_2}, \dots$ , so dass

$$\lim_{k \rightarrow \infty} F_{n_k}(t) = G(t) \quad \text{für alle } t \in S(G).$$

**Beweis.** Aus Satz 11.4.1 folgt, dass es eine Teilfolge  $F_{n_1}, F_{n_2}, \dots$  und eine nichtfallende, rechtsstetige Funktion  $G$  mit  $\lim_{k \rightarrow \infty} F_{n_k}(t) = G(t)$  für alle  $t \in S(G)$  gibt. Es bleibt noch zu zeigen, dass  $G$  eine Verteilungsfunktion ist (was der Satz von Helly ja nicht behauptet). Wir berechnen dazu die Grenzwerte von  $G(t)$  für  $t \rightarrow +\infty$  und  $t \rightarrow -\infty$ .

Sei  $\varepsilon > 0$ . Wegen Straffheit existiert ein  $c > 0$  mit

$$F_{n_k}(-c) < \varepsilon \text{ und } 1 - F_{n_k}(c) < \varepsilon \text{ für alle } k \in \mathbb{N}.$$

Indem wir  $c$  vergrößern, können wir zusätzlich annehmen, dass  $c$  und  $-c$  Stetigkeitspunkte von  $G$  sind. Nun lassen wir  $k \rightarrow \infty$  und folgern, dass

$$G(-c) \leq \varepsilon \text{ und } 1 - G(c) \leq \varepsilon.$$

Dies gilt für alle  $\varepsilon > 0$ , woraus sich  $\lim_{c \rightarrow +\infty} G(-c) = 0$  und  $\lim_{c \rightarrow +\infty} G(c) = 1$  ergibt. Also ist  $G$  eine Verteilungsfunktion.  $\square$

Der obige Satz kann als eine Aussage über die relative Kompaktheit im Raum  $\mathcal{M}$  der Verteilungsfunktionen interpretiert werden.

**Definition 11.4.18.** Sei  $(M, \rho)$  ein metrischer Raum.

- Eine Menge  $A \subset M$  heißt **kompakt**, wenn jede Folge  $a_1, a_2, \dots \in A$  eine gegen ein Element aus  $A$  konvergente Teilfolge besitzt.
- Eine Menge  $A \subset M$  heißt **relativ kompakt**, wenn jede Folge  $a_1, a_2, \dots \in A$  eine gegen ein Element aus  $M$  konvergente Teilfolge besitzt. Dabei muss der Grenzwert dieser Teilfolge nicht unbedingt in  $A$  liegen.

**Beispiel 11.4.19.** Eine Menge  $A \subset \mathbb{R}^d$  ist genau dann kompakt, wenn sie beschränkt und abgeschlossen ist. Eine Menge  $A \subset \mathbb{R}^d$  ist genau dann relativ kompakt, wenn sie beschränkt ist.

**Beispiel 11.4.20.** Die Menge  $\{1, \frac{1}{2}, \frac{1}{3}, \dots\} \subset \mathbb{R}$  ist relativ kompakt aber nicht kompakt. Erweitern wir diese Menge um den Punkt 0, so wird sie kompakt.

Der Satz von Helly–Prochorow behauptet, dass eine straffe Familie von Verteilungsfunktionen relativ kompakt ist. Es gilt auch die Umkehrung.

**Satz 11.4.21** (Satz von Prochorow). Eine Familie  $\mathcal{F} \subset \mathcal{M}$  von Verteilungsfunktionen ist genau dann relativ kompakt, wenn sie straff ist.

Wir werden diesen Satz hier nicht benötigen. Für mehr Einzelheiten verweisen wir auf das Buch von P. Billingsley, “Weak convergence of probability measures”.

### 11.5. Stetigkeitssatz von Lévy

Der nachfolgende Satz wird im Beweis des zentralen Grenzwertsatzes eine wichtige Rolle spielen. Es besagt, dass Konvergenz in Verteilung äquivalent zur punktweisen Konvergenz der entsprechenden charakteristischen Funktionen ist.

**Satz 11.5.1** (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 die folgenden zwei Aussagen äquivalent:

- (a)  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ .
- (b)  $\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t)$  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 (a)  $\Rightarrow$  (b).** Sei  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ . Dann gilt für alle  $t \in \mathbb{R}$  unter Verwendung von Satz 11.3.2, dass

$$\varphi_n(t) = \mathbb{E} \cos(tX_n) + i\mathbb{E} \sin(tX_n) \xrightarrow[n \rightarrow \infty]{} \mathbb{E} \cos(tX) + i\mathbb{E} \sin(tX) = \varphi(t).$$

Dabei haben wir benutzt, dass  $x \mapsto \cos(tx)$  und  $x \mapsto \sin(tx)$  stetige und global beschränkte Funktionen sind.  $\square$

**Beweis von (b)  $\Rightarrow$  (a).** 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 der Verteilungsfunktionen  $F_1, F_2, \dots$  straff ist. Wir definieren eine Funktion  $B_n(u)$ ,  $u > 0$ , folgendermaßen:

$$B_n(u) := \frac{1}{u} \int_{-u}^u (1 - \varphi_n(t)) dt = \frac{1}{u} \int_{-u}^u \mathbb{E}[1 - e^{itX_n}] dt = \frac{1}{u} \mathbb{E} \left[ \int_{-u}^u (1 - e^{itX_n}) dt \right],$$

wobei wir im letzten Schritt das Integral und den Erwartungswert (der eigentlich auch ein Integral ist) unter Berufung auf den Satz von Fubini vertauscht haben. Um die Anwendung des Satzes von Fubini zu rechtfertigen, bemerken wir, dass  $|1 - e^{itX_n}| \leq 2$  global beschränkt ist. Da nun  $\sin x$  eine gerade Funktion ist und der Imaginärteil somit nach der Integration verschwindet, erhalten wir, dass

$$B_n(u) = 2\mathbb{E} \left[ 1 - \frac{\sin(uX_n)}{uX_n} \right].$$

Um diesen Erwartungswert abzuschätzen, bemerken wir zuerst, dass stets

$$1 - \frac{\sin(uX_n)}{uX_n} \geq 0.$$

Außerdem gilt auf dem Ereignis  $\{|X_n| \geq \frac{2}{u}\}$  die Abschätzung

$$1 - \frac{\sin(uX_n)}{uX_n} \geq 1 - \frac{1}{u|X_n|} \geq \frac{1}{2}.$$

Insgesamt folgt, dass

$$B_n(u) = 2\mathbb{E} \left[ 1 - \frac{\sin(uX_n)}{uX_n} \right] \geq \mathbb{P} \left[ |X_n| \geq \frac{2}{u} \right].$$

Sei  $\varepsilon > 0$  vorgegeben. Als charakteristische Funktion ist  $\varphi$  stetig und erfüllt außerdem  $\varphi(0) = 1$ . Somit existiert ein hinreichend kleines  $u > 0$  mit

$$|B(u)| := \frac{1}{u} \cdot \left| \int_{-u}^u (1 - \varphi(t)) dt \right| < \varepsilon.$$

Da  $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$  für alle  $t \in \mathbb{R}$  gilt, folgt mit der dominierten Konvergenz, dass

$$B_n(u) = \frac{1}{u} \int_{-u}^u (1 - \varphi_n(t)) dt \xrightarrow[n \rightarrow \infty]{} \frac{1}{u} \int_{-u}^u (1 - \varphi(t)) dt = B(u).$$

Um die Verwendung der dominierten Konvergenz zu rechtfertigen, sei bemerkt, dass  $|1 - \varphi_n(t)| \leq 2$  für alle  $n \in \mathbb{N}$  und  $t \in \mathbb{R}$ . Aus den vorangegangenen Behauptungen folgt die Existenz eines  $n_0 \in \mathbb{N}$  mit  $B_n(u) < 2\varepsilon$  für alle  $n \geq n_0$ . Dann gilt 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  $t_0 \in \mathbb{R}$  mit

$$\lim_{n \rightarrow \infty} F_n(t_0) \neq F(t_0) \text{ und } \mathbb{P}[X = t_0] = 0. \quad (11.5.1)$$

Daraus folgt, dass es ein  $\varepsilon > 0$  und eine Teilfolge  $F_{n_1}, F_{n_2}, \dots$  gibt, sodass

$$|F_{n_k}(t_0) - F(t_0)| > \varepsilon \text{ für alle } k \in \mathbb{N}. \quad (11.5.2)$$

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 11.4.17 die Existenz einer Verteilungsfunktion  $G$  und einer Teilteilfolge  $F_{n_{k_1}}, F_{n_{k_2}}, \dots$  mit

$$F_{n_{k_j}} \xrightarrow[j \rightarrow \infty]{d} G. \quad (11.5.3)$$

Wir betrachten zwei Fälle:

FALL 1.  $G$  ist nicht stetig an der Stelle  $t_0$ . Dann gilt  $G \neq F$ , denn  $F$  ist stetig an der Stelle  $t_0$  wegen (11.5.1).

FALL 2.  $G$  ist stetig an der Stelle  $t_0$ . Dann folgt aus (11.5.3), dass  $\lim_{j \rightarrow \infty} F_{n_{k_j}}(t_0) = G(t_0)$  und somit  $G(t_0) \neq F(t_0)$  wegen (11.5.2).

In beiden Fällen gilt  $G \neq F$ . Aus (11.5.3) und wegen der bereits bewiesenen Richtung (a) $\Rightarrow$ (b) 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  $\varphi_G$  die charakteristische Funktion von  $G$ . Nach Voraussetzung (b) gilt aber auch, 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.  $\square$

**Aufgabe 11.5.2.** Zeigen Sie unter Verwendung des Stetigkeitssatzes von Lévy, dass für jedes  $\lambda > 0$ ,

$$\text{Bin}(n, \lambda/n) \xrightarrow[n \rightarrow \infty]{d} \text{Poi}(\lambda).$$

## 11.6. 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 11.6.1** (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:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N,$$

wobei  $N \sim N(0, 1)$  eine standardnormalverteilte Zufallsvariable ist.

**Bemerkung 11.6.2.** 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 11.6.3** (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 11.6.4.** 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 11.6.5.** 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

$$\mathbb{E}S = np = 100 \cdot \frac{1}{2} = 50 \quad \text{und} \quad \text{Var } S = np(1 - p) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25.$$

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  $\mathbb{P}[S \leq 55] = 0.86437$ .  $\square$

**Beispiel 11.6.6.** Seien  $X_1, X_2, \dots \sim N(\mu, \sigma^2)$  unabhängige normalverteilte Zufallsvariablen. Dann gilt nicht nur approximativ, sondern sogar exakt, dass

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0, 1).$$

**Satz von Poincaré-Borel, Maxwell-Verteilung, Galton-Brett, warum sich das Gas nicht in einer Hälfte des Raumes sammelt...**

**11.6.1. Satz von Poincaré-Borel.** Wir betrachten die Einheitssphäre im  $n$ -dimensionalen Euklidischen Raum  $\mathbb{R}^n$ :

$$\mathbb{S}^{n-1} := \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1^2 + \dots + x_n^2 = 1\}.$$

Nun wählen wir einen zufälligen Punkt, der auf der Sphäre uniform verteilt ist. Diesen Punkt bezeichnen wir mit  $\mathbf{X}_n = (X_{1,n}, X_{2,n}, \dots, X_{n,n})$ . Die Koordinaten dieses Punktes sind Zufallsvariablen. Wie ist z.B. die erste Koordinate verteilt, wenn die Dimension  $n \rightarrow \infty$ ? Zunächst einmal gilt  $\sum_{k=1}^n X_{k,n}^2 = 1$ , so dass man erwarten kann, dass jede einzelne Koordinate ungefähr  $1/\sqrt{n}$  ist. Es erscheint deshalb natürlich, die Koordinaten mit  $\sqrt{n}$  zu multiplizieren. Der nächste Satz behauptet, dass die einzelnen normierten Koordinaten  $\sqrt{n} X_{k,n}$  approximativ normalverteilt sind, wenn  $n \rightarrow \infty$ .

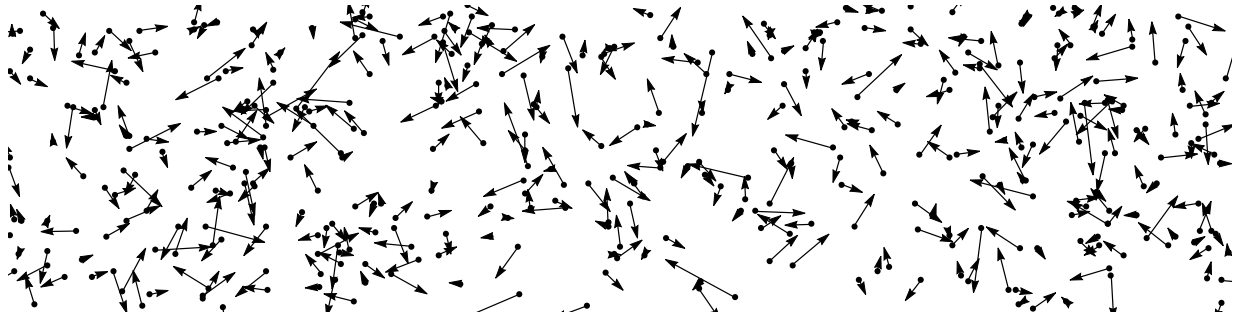


ABBILDUNG 3. Visualisierung der Positionen der Moleküle und ihrer Geschwindigkeiten.

**Satz 11.6.7** (Poincaré-Borel). Sei  $\mathbf{X}_n = (X_{1,n}, X_{2,n}, \dots, X_{n,n})$  ein auf der Sphäre  $\mathbb{S}^{n-1}$  uniform verteilter Zufallsvektor. Dann gilt für jedes  $k \in \mathbb{N}$ , dass

$$\sqrt{n} X_{k,n} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

**Beweis.** Zuerst müssen wir uns die Frage stellen, wie die Uniformverteilung auf der Sphäre definiert ist. Die Sphäre ist eine Fläche der Dimension  $n - 1$  in  $\mathbb{R}^n$  (weshalb wir sie mit  $\mathbb{S}^{n-1}$  bezeichnet haben) und hat als solche Lebesgue-Maß 0. Somit ist die übliche Definition der Uniformverteilung im Fall der Sphäre nicht anwendbar.

Wir werden hier die folgende Definition benutzen. Seien zuerst  $\xi_1, \dots, \xi_n$  unabhängige, standardnormalverteilte Zufallsvariablen. Der  $n$ -dimensionale Zufallsvektor  $\xi := (\xi_1, \dots, \xi_n)$  hat die folgende gemeinsame Dichte auf  $\mathbb{R}^n$ :

$$f(t_1, \dots, t_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(t_1^2 + \dots + t_n^2)} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|t\|^2},$$

wobei  $\|t\| := \sqrt{t_1^2 + \dots + t_n^2}$  die Länge des Vektors  $t = (t_1, \dots, t_n) \in \mathbb{R}^n$  bezeichnet. Es sei bemerkt, dass die gemeinsame Dichte nur von  $\|t\|$  abhängt. Das bedeutet, dass die Verteilung von  $\xi$  isotrop ist, d.h.  $O\xi$  hat die gleiche Verteilung wie  $\xi$  für jede orthogonale Transformation  $O : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Nun können wir  $\xi$  normieren, indem wir diesen Vektor durch die Norm  $\|\xi\|$  teilen.

**Definition 11.6.8.** Der Zufallsvektor

$$\frac{\xi}{\|\xi\|} = \left( \frac{\xi_1}{\sqrt{\xi_1^2 + \dots + \xi_n^2}}, \dots, \frac{\xi_n}{\sqrt{\xi_1^2 + \dots + \xi_n^2}} \right)$$

ist uniformverteilt auf der Einheitssphäre in  $\mathbb{R}^n$ .

Muss noch ergänzt werden...

□

### 11.6.2. Geschwindigkeitsverteilung der Moleküle im Gas.

## 11.7. Beweis des zentralen Grenzwertsatzes

**Beweis von Satz 11.6.1.** 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 11.5.1 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  $\varphi$  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  $1^\infty$  zu tun. Um die aufzulösen, werden wir die ersten zwei Terme der Taylor-Entwicklung von  $\varphi$  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 11.5.1, dass  $V_n$  in Verteilung gegen eine Standardnormalverteilte Zufallsvariable konvergiert. Das beweist die Behauptung.  $\square$



## 11.8. 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:

$$\begin{array}{c}
 X_{11} \\
 X_{21}, \quad X_{22} \\
 X_{31}, \quad X_{32}, \quad X_{33} \\
 \dots \\
 X_{n1}, \quad X_{n2}, \quad X_{n3}, \quad \dots, \quad X_{nn} \\
 \dots
 \end{array}$$

Wir werden annehmen, dass jede Zeile dieses Dreiecks aus unabhängigen Zufallsvariablen besteht. Abhängigkeiten zwischen den Variablen aus verschiedenen Zeilen sind jedoch erlaubt.

**Satz 11.8.1** (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  $\varepsilon > 0$

$$L_n(\varepsilon) := \sum_{k=1}^n \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}| > \varepsilon}] \xrightarrow{n \rightarrow \infty} 0. \quad (11.8.1)$$

Dann gilt:

$$X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

**Bemerkung 11.8.2.** Die Lindeberg-Bedingung besagt, dass alle Werte der Zufallsvariablen, die größer als  $\varepsilon$  sind, einen asymptotisch verschwindenden Beitrag zur Varianz der Summe der Zufallsvariablen leisten.

**Bemerkung 11.8.3.** Wir beweisen, dass der zentrale Grenzwertsatz ein Spezialfall von Satz 11.8.1 ist. Seien  $X_1, X_2, \dots$  unabhängige, identisch verteilte Zufallsvariablen mit  $\mathbb{E}[X_k] = \mu$  und  $\text{Var } X_k = \sigma^2 \in (0, \infty)$ . 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 11.8.1 erfüllt sind, müssen wir noch die Lindeberg-Bedingung (11.8.1) 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 dominierten Konvergenz anwenden. Somit gilt die Lindeberg-Bedingung (11.8.1).

Nun können wir Satz 11.8.1 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 11.8.1.

**Bemerkung 11.8.4.** 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 11.8.5.** 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

$$X_{n1} + \dots + X_{nn} = X \approx N(0, 1).$$

In diesem Fall konvergieren die Summen also nicht gegen die Normalverteilung. Die Ursache dafür ist, dass die Lindeberg-Bedingung (11.8.1) nicht erfüllt ist.

**Bemerkung 11.8.6.** Wenn die Lindeberg-Bedingung (11.8.1) 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 11.8.1 brauchen wir zunächst einige Hilfsaussagen.

**Lemma 11.8.7.** 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 11.8.8.** 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 11.8.9.** 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 11.8.8.**

SCHRITT 1. Zuerst beweisen wir eine Abschätzung für den Restterm in der Taylor-Entwicklung der Funktion  $e^{it}$ . Wir werden zeigen, dass für alle  $t \in \mathbb{R}$

$$\left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| \leq \min \left\{ \frac{|t|^{n+1}}{(n+1)!}, 2 \frac{|t|^n}{n!} \right\}. \quad (11.8.2)$$

Wir führen den Beweis mit Induktion.

INDUKTIONSANFANG. Für  $n = 0$  stimmt die zu beweisende Ungleichung (11.8.2), denn

$$|e^{it} - 1| \leq |t| \text{ und } |e^{it} - 1| \leq 2.$$

INDUKTIONSSCHRITT. Wir nehmen an, dass die Ungleichung (11.8.2) für ein  $n \in \mathbb{N}_0$  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 (11.8.2) 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 11.8.1.** Es ist zu zeigen, dass

$$X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

SCHRITT 1. Nach dem Stetigkeitssatz 11.5.1 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 11.8.8 erhalten wir

$$c_{nk} = \left| \varphi_{nk}(t) - \left( 1 - \frac{1}{2}\sigma_{nk}^2 t^2 \right) \right| \leq \mathbb{E} [\min \{ |tX_{nk}|^3, |tX_{nk}|^2 \}].$$

Und nun verwenden wir für  $X_{nk} \leq \varepsilon$  die Abschätzung  $|tX_{nk}|^3$  und für  $X_{nk} > \varepsilon$  die Abschätzung  $|tX_{nk}|^2$ :

$$c_{nk} \leq \mathbb{E} [ |tX_{nk}|^3 \mathbb{1}_{|X_{nk}| \leq \varepsilon} ] + \mathbb{E} [ |tX_{nk}|^2 \mathbb{1}_{|X_{nk}| > \varepsilon} ].$$

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 Lindenbergs-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 11.8.6

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n d_{nk} = 0.$$

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 11.5.1 folgt daraus, dass  $X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$ .  $\square$

Nun beweisen wir einen Satz, der eine vereinfachte Form des Satzes von Lindeberg ist.

**Satz 11.8.10** (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

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}|X_{nk}|^{2+\delta} = 0. \quad (11.8.3)$$

Dann gilt:

$$X_{n1} + \dots + X_{nn} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1).$$

**Beweis.** Wir zeigen, dass die Ljapunow-Bedingung (11.8.3) die Lindeberg-Bedingung (11.8.1) impliziert. Sei also (11.8.3) 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 (11.8.1).  $\square$

**Beispiel 11.8.11.** 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 für die Anzahl der Erfolge  $S_n \sim \text{Bin}(n, p_n)$ :

$$\frac{S_n - np_n}{\sqrt{np_n(1 - p_n)}} \xrightarrow[n \rightarrow \infty]{d} N \sim N(0, 1). \quad (11.8.4)$$

**Bemerkung 11.8.12.** 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 (11.8.4).** 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 11.8.13.** 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 12

### Irrfahrt

#### 12.1. 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 12.1.1.** 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

$$p_i = p \cdot p_{i+1} + q \cdot p_{i-1}, \quad i = 1, \dots, M - 1. \quad (12.1.1)$$

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

$$p_i = p \cdot p_i + q \cdot p_i, \quad i = 1, \dots, M - 1, \quad (12.1.2)$$

denn  $p + q = 1$ . Zieht man nun die Gleichungen (12.1.1) und (12.1.2) 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

$$d_i = \frac{q}{p} d_{i-1} = \left(\frac{q}{p}\right)^2 d_{i-2} = \dots = \left(\frac{q}{p}\right)^i d_0, \quad i = 1, \dots, M-1. \quad (12.1.3)$$

SCHRITT 2. Nun berechnen wir  $d_0$ . Dazu setzen wir die Randbedingungen  $p_0 = 1$  und  $p_M = 0$  ein:

$$1 = p_0 - p_M = \sum_{i=0}^{M-1} d_i = d_0 \sum_{i=0}^{M-1} \left(\frac{q}{p}\right)^i. \quad (12.1.4)$$

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 (12.1.4), dass  $1 = d_0 \cdot M$ , also  $d_0 = 1/M$ . Mit (12.1.3) 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 (12.1.4), dass  $1 = d_0 \cdot \frac{1-r^M}{1-r}$ . Daraus folgt, dass  $d_0 = \frac{1-r}{1-r^M}$ . Mit (12.1.3) 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

$$p_k = \begin{cases} \frac{M-k}{M}, & \text{falls } p = q = \frac{1}{2}, \\ \frac{r^k - r^M}{1 - r^M}, & \text{falls } r := \frac{q}{p} \neq 1, \end{cases} \quad k = 0, \dots, M. \quad (12.1.5)$$

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.  $\square$

**Beispiel 12.1.2.** 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 (12.1.5):

$$p_k = \begin{cases} 1, & \text{falls } p = q = 1/2, \\ 1, & \text{falls } p < q, \\ r^k, & \text{falls } p > q. \end{cases}$$

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.

## 12.2. Rückkehr der Irrfahrt zum Ursprung

**Definition 12.2.1.** 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 12.2.2.** 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 12.2.3.** Für die Zähldichte des ersten Rückkehrzeitpunktes gilt

$$f_{2n} = \frac{1}{2n-1} \binom{2n}{n} p^n q^n, \quad f_{2n+1} = 0. \quad (12.2.1)$$

**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)

$$u_{2n} = \mathbb{P}[S_{2n} = 0] = \binom{2n}{n} p^n q^n, \quad u_{2n+1} = 0. \quad (12.2.2)$$

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:

$$u_n = \sum_{k=1}^n f_k \cdot u_{n-k}. \quad (12.2.3)$$

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 (12.2.3) 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 12.2.4.** 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 12.2.5.** 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 12.2.6.** 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 12.2.7.** Zwei Folgen  $a_1, a_2, \dots$  und  $b_1, b_2, \dots$  heißen *asymptotisch äquivalent*, wenn

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1.$$

**Notation 12.2.8.**  $a_n \sim b_n$  für  $n \rightarrow \infty$ .

**Satz 12.2.9.** 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  $f_{2n} = \frac{u_{2n}}{2n-1} \sim \frac{1}{2n} \cdot \frac{1}{\sqrt{\pi n}} = \frac{1}{2\sqrt{\pi n^3}}$ . □

**Bemerkung 12.2.10.** Mit dem obigen Satz kann man Satz 12.2.6 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 12.2.11.** 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 12.2.9. 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.

### 12.3. 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 12.3.1.** 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

$$\begin{aligned} \mathbb{P}[M_n \geq m] &= \mathbb{P}[M_n \geq m, S_n < m] + \mathbb{P}[M_n \geq m, S_n = m] + \mathbb{P}[M_n \geq m, S_n > m] \quad (12.3.1) \\ &= \mathbb{P}[S_n > m] + \mathbb{P}[S_n = m] + \mathbb{P}[S_n > m] \\ &= 2\mathbb{P}[S_n > m] + \mathbb{P}[S_n = m] \\ &= 2\mathbb{P}[S_n \geq m] - \mathbb{P}[S_n = m]. \end{aligned}$$

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  $\Phi$

$$\lim_{n \rightarrow \infty} \mathbb{P}[S_n = m_n] = 0.$$

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. □

## 12.4. 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 12.4.1.** 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 (12.3.1) 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

$$p_{2k, 2n} = \frac{1}{2} \sum_{r=1}^k f_{2r} p_{2k-2r, 2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} p_{2k, 2n-2r}. \quad (12.4.1)$$

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

$$p_{2k, 2m} = u_{2k} \cdot u_{2m-2k} \quad \text{für alle } m = 1, \dots, n-1 \text{ und } k = 1, \dots, m-1. \quad (12.4.2)$$

Nun ist zu zeigen, dass diese Formel auch für  $m = n$  und beliebiges  $k = 1, \dots, n-1$  gilt. Mit (12.4.1) und (12.4.2) 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 (12.2.3), benutzt. Somit gilt  $p_{2k, 2n} = u_{2k} \cdot u_{2n-2k}$  und die Induktion ist beendet.  $\square$

**Satz 12.4.2** (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 12.4.3.** 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 12.4.2.** Hier soll nur eine Beweisidee vorgestellt werden. Wir haben in Satz 12.4.1 und Satz 12.2.9 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 12.4.2 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 12.4.4** (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 12.4.5** (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

$$\frac{R_{2n}}{2n} \xrightarrow[n \rightarrow \infty]{d} T.$$

## 12.5. 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 12.3.1 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 12.5.1** (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 12.5.2.** Aus dem Gesetz vom iterierten Logarithmus folgt, dass für alle  $\varepsilon > 0$  mit Wahrscheinlichkeit 1 gilt:

$$\lim_{n \rightarrow \infty} \frac{S_n}{n^{\frac{1}{2} + \varepsilon}} = 0 \text{ jedoch } \limsup_{n \rightarrow \infty} \frac{S_n}{n^{\frac{1}{2} - \varepsilon}} = +\infty.$$

Die Folge  $S_n$  wächst also langsamer als  $n^{\frac{1}{2} + \varepsilon}$ , aber nicht langsamer als  $n^{\frac{1}{2} - \varepsilon}$ .