

# Stochastik

Dominik Bullach

Hannes F. Funk

Stand: 19. Juli 2015, 14:57

Dies ist eine Mitschrift der Vorlesung von Prof. Dr. Peter Pickl im Sommersemester 2015. Sie stellt den Versuch dar, den Vorlesungsstoff übersichtlich und verständlich aufzubereiten. Sie ist sicher nicht fehlerfrei, für eine nette Mail über Verbesserungsvorschläge sind wir daher dankbar.

## Inhaltsverzeichnis

<b>I</b>	<b>Endliche Ergebnisräume</b>	<b>2</b>
1.1	Das Wahrscheinlichkeitsmaß . . . . .	2
1.2	Die Laplace-Annahme . . . . .	4
1.3	Kombinatorik . . . . .	5
1.4	Bedingte Wahrscheinlichkeit . . . . .	7
1.5	Unabhängigkeit von Ereignissen . . . . .	9
<b>II</b>	<b>Überabzählbare Ergebnisräume</b>	<b>11</b>
2.1	Das Wahrscheinlichkeitsmaß für überabzählbare Ergebnisräume . . . . .	11
2.2	$\sigma$ -Algebren . . . . .	12
2.3	Zufallsgrößen . . . . .	17
2.4	Verallgemeinerung des Unabhängigkeitsbegriffs . . . . .	22
2.5	Die Verteilungsfunktion . . . . .	25
2.6	Spezielle Verteilungsfunktionen . . . . .	30
<b>III</b>	<b>Invarianten von Zufallsgrößen</b>	<b>35</b>
3.1	Der Erwartungswert . . . . .	35
3.2	Die Varianz . . . . .	38
3.3	Kovarianz und Korrelationskoeffizient . . . . .	39
3.4	Die Ungleichungen von Markov und Tschebyscheff . . . . .	43
3.5	Schwaches und Starkes Gesetz der großen Zahlen . . . . .	45
3.6	Konvergenzbegriffe der Stochastik . . . . .	52
3.7	Der Zentrale Grenzwertsatz . . . . .	57
<b>IV</b>	<b>Statistik</b>	<b>61</b>
4.1	Testen von Hypothesen . . . . .	61
4.2	Schätzmethoden . . . . .	65
4.3	Das Neyman-Pearson-Lemma . . . . .	70
4.4	Der $\chi^2$ -Verteilungstest . . . . .	71
	<b>Index</b>	<b>75</b>

# I Endliche Ergebnisräume

Wir führen zunächst einige Grundbegriffe ein und entwickeln die grundlegende Theorie zur Beschreibung von endlichen Ergebnisräumen in enger Anlehnung an die Stochastik der Schule, um daran die Notwendigkeit der später folgenden Verallgemeinerung aufzuzeigen.

## 1.1 Das Wahrscheinlichkeitsmaß

Ziel der Stochastik ist es, mithilfe mathematischer Modelle Experimente zu beschreiben, deren Ausgang nicht festgelegt ist, sondern einem gewissen Zufall unterliegt.

**Beispiel 1.1.** Beim Wurf eines Würfels sind die möglichen Ausgänge des Experimentes die Augenzahlen  $\square$ ,  $\square$ , ... Man kann nun folgende Identifizierung vornehmen:

$$\begin{array}{ccc} \square & \longleftrightarrow & 1 \\ \square & \longleftrightarrow & 2 \\ \square & \longleftrightarrow & 3 \\ \vdots & & \vdots \end{array}$$

Das mathematische Modell verwendet also die Menge  $\{1, 2, \dots, 6\}$ .

**Definition 1.2.** Die Menge  $\Omega$ , deren Elemente für die möglichen Ausgänge eines Zufallsexperiments stehen, heißt *Grundmenge* oder *Ergebnisraum*. Dementsprechend werden die einzelnen Elemente von  $\Omega$  *Ergebnisse* genannt.

**Beispiele 1.3.** (a) Beim Würfelwurf ist  $\Omega = \{1, 2, \dots, 6\}$  (vgl. Bsp. 1.1). Die Grundmenge ist hier endlich.

(b) Beim Drehen eines Glücksrads kann jedes Ergebnis als Winkel interpretiert werden, d.h.  $\Omega = [0; 2\pi[$ . Diese Menge ist ein Beispiel für eine überabzählbar unendliche Grundmenge.

**Definition 1.4.** Falls  $\Omega$  endlich ist, nennen wir jede Teilmenge von  $\Omega$  ein *Ereignis*. Die Menge aller Ereignisse  $\mathfrak{B}(\Omega)$  nennen wir *Ereignisraum*.

Wir möchten nun den Ereignissen Wahrscheinlichkeiten zuordnen. Dies erreichen wir mithilfe eines Wahrscheinlichkeitsmaßes.

**Definition 1.5.** Sei  $\Omega$  eine endliche Menge und seien  $A, B \in \mathfrak{P}(\Omega)$  beliebig. Eine Abbildung  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  heißt *Wahrscheinlichkeitsmaß*, falls gilt:

- (a)  $P(A) \geq 0$ ,
- (b)  $P(\Omega) = 1$ ,
- (c) Sind  $A$  und  $B$  disjunkt, so gilt  $P(A \cup B) = P(A) + P(B)$ .

Die Bedingungen (a) bis (c) heißen Axiome von Kolmogoroff<sup>1</sup>. Diese sind Axiome im klassischen Sinne, d.h. es handelt sich um unmittelbar einleuchtende Prinzipien, die keines Beweises bedürfen. Im Hilbertschen Sinne sind Axiome dagegen lediglich widerspruchsfreie, unabgeleitete Aussagen, wobei es keine Rolle spielt, ob diese auch sinnvoll sind.

**Satz 1.6.** Sei  $\Omega$  eine endliche Menge,  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  ein Wahrscheinlichkeitsmaß und  $A, B \in \mathfrak{P}(\Omega)$ . Dann gilt:

- (a)  $P(A) \leq 1$
- (b) (*Satz vom Gegenereignis*)  $P(A^c) = 1 - P(A)$
- (c)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

*Beweis.* (b) Da  $A$  und  $A^c$  per Definition disjunkt sind, gilt nach Axiom 1.5 (c)  $P(A \cup A^c) = P(A) + P(A^c)$ . Wegen  $A \cup A^c = \Omega$  also:

$$1 \stackrel{1.5(b)}{=} P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

Auflösen liefert wie gewünscht  $P(A^c) = 1 - P(A)$ .

- (a) Wäre  $P(A) > 1$ , so wäre nach der eben bewiesenen Aussage (b)

$$P(A^c) = 1 - P(A) < 1 - 1 = 0$$

im Widerspruch zu Axiom 1.5 (a).

- (c) Wir zerlegen  $A \cup B$  in disjunkte Teilmengen und erhalten  $A \cup B = (A \setminus B) \cup B$  (vgl. dazu auch Abb. 1.1). Mit dem Axiom 1.5 (c) erhalten wir

$$P(A \cup B) = P(A \setminus B) + P(B)$$

Andererseits ist auch  $A = (A \setminus B) \cup (A \cap B)$  eine disjunkte Vereinigung, weshalb wiederum nach 1.5 (c) folgt:  $P(A) = P(A \setminus B) + P(A \cap B)$ . Auflösen nach  $P(A \setminus B)$  und Einsetzen in die erste Formel liefert nun

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

□

<sup>1</sup>nach Andrej Nikolajewitsch Kolmogoroff (1903 - 1987)

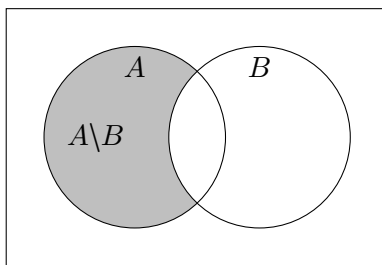


Abbildung 1.1: Mengendiagramm zur Illustration des Beweises von 1.5 (c)

## 1.2 Die Laplace-Annahme

Es ist offensichtlich, dass es verschiedene Möglichkeiten gibt, eine Abbildung zu definieren, die die Axiome von Kolmogoroff erfüllt. Was ist also das *richtige* Wahrscheinlichkeitsmaß, das ein gegebenes Experiment modelliert?

Im Allgemeinen ist das schwer zu beantworten (siehe später). Falls das Experiment eine gewisse Symmetrie in den Ergebnissen aufweist, lässt sich jedoch eine vereinfachende Annahme treffen:

**1.7 (Laplace-Annahme).** Gibt es keinen Grund zur Annahme, dass die verschiedenen Ausgänge eines Experimentes sich unterscheiden, nehmen wir an, dass diese gleichberechtigt sind, also gleiche Wahrscheinlichkeit haben.

Betrachte den Fall, dass  $\Omega$  endlich ist und treffe die Laplace-Annahme (LA) Dann lässt sich schreiben:

$$1 = P(\Omega) = P\left(\bigcup_{\omega \in \Omega} \{\omega\}\right) \stackrel{1.5(c)}{=} \sum_{\omega \in \Omega} P(\{\omega\}) \stackrel{LA}{=} |\Omega| \cdot P(\{\omega\})$$

Durch Auflösen folgt sofort  $P(\{\omega\}) = \frac{1}{|\Omega|}$  für alle  $\omega \in \Omega$ . Induktiv lässt sich dies auch auf Ereignisse verallgemeinern:

**Folgerung 1.8.** Sei  $\Omega$  endlich, dann gilt unter der Laplace-Annahme für jedes Ereignis  $A \subseteq \Omega$

$$P(A) = \frac{|A|}{|\Omega|}.$$

Wichtig ist, dass die Laplace-Annahme nicht für jedes Experiment sinnvoll ist.

**Beispiel 1.9.** Betrachte einen Wurf von zwei Würfeln. Man interessiere sich lediglich für die Augenzahl der Würfe, also  $\Omega = \{2, \dots, 12\}$ . Das Resultat “2” kommt nur zustande, wenn beide Würfel eine 1 zeigen. Für “7” gibt es hingegen mehrere Möglichkeiten (z.B.  $\begin{smallmatrix} 1 & 6 \\ 6 & 1 \end{smallmatrix}$ ,  $\begin{smallmatrix} 2 & 5 \\ 5 & 2 \end{smallmatrix}$ , ...). Aufgrund dieses Symmetriebruchs kann die Laplace-Annahme (zumindest ohne weitere Betrachtung) hier nicht verwendet werden.

**Beispiel 1.10 (Ziegenproblem).** Folgendes Problem erlangte Berühmtheit, als es als Leserbrief in einer amerikanischen Zeitschrift 1990 veröffentlicht wurde. Die deutsche Übersetzung der Problemstellung lautet:

Nehmen Sie an, Sie wären in einer Gameshow und hätten die Wahl zwischen drei Toren. Hinter einem der Tore ist ein Auto, hinter den anderen sind Ziegen. Sie wählen ein Tor, sagen wir, Tor Nummer 1, und der Showmaster, der weiß, was hinter den Toren ist, öffnet ein anderes Tor, sagen wir, Nummer 3, hinter dem eine Ziege steht. Er fragt Sie nun: "Möchten Sie zu Tor Nummer 2 wechseln?" Ist es von Vorteil, die Wahl des Tores zu ändern?

Die Antwort auf diese Frage gestaltet sich schwieriger als man zunächst meinen möchte. Die kurze Variante wäre: Das hängt von der Arbeitsweise des Moderators ab. Wir versuchen nun eine ausführlichere Antwort zu geben.

Dazu legen wir bei der Wahl des ersten Tores die Laplace-Annahme zu Grunde, da die Tore im Hinblick darauf, ob dahinter ein Gewinn zu erwarten ist, nicht zu unterscheiden sind. Die Wahrscheinlichkeit, dass sich hinter einem beliebig gewählten Tor ein Gewinn befindet, beträgt folglich  $\frac{1}{3}$ .

Im zweiten Schritt unterscheiden wir nun die Beweggründe des Moderators.

- (a) Nehmen wir an, es war von Anfang an klar, dass der Moderator eine der Türen öffnen wird. Es treten zwei Fälle auf:
  - (i) Im ersten Schritt wurde das Tor mit dem Gewinn gewählt. Durch einen Wechsel tauschen wir also, ohne es zu wissen, den Gewinn in eine Niete um.
  - (ii) Im ersten Schritt wurde eine Ziege gewählt. Da der Moderator uns die andere Ziege zeigt, führt ein Wechsel zum Gewinn.
 Da die Wahrscheinlichkeit für Fall (i) wie oben ausgeführt bei  $\frac{1}{3}$  liegt, für Fall (ii) jedoch bei  $\frac{2}{3}$ , ist ein Wechsel sinnvoll.
- (b) Falls es die Motivation des Moderators ist, uns vom Gewinn abzuhalten, könnte es sein, dass er ein weiteres Tor nur öffnet, falls im ersten Schritt das Tor mit dem Gewinn gewählt haben. Hier wäre ein Wechsel nicht vorteilhaft.

Man sieht hier deutlich, dass die Schwierigkeit in der Behandlung stochastischer Problemstellungen meist nicht auf der Seite der Mathematik zu suchen ist, sondern in der Wahl der Modellierung. Hier haben zusätzliche Informationen über die Arbeitsweise des Moderators zu verschiedenen Modellierungen und daher unterschiedlichen Ergebnissen geführt.

## 1.3 Kombinatorik

In der Kombinatorik geht es meist darum, die Anzahl der Elemente bestimmter Mengen abzuschätzen. Diese ist besonders bei der Behandlung von Laplace-Experimenten hilfreich, denn nach 1.8 gilt bei diesen für beliebige Ereignisse  $A \subseteq \Omega$  eines endlichen Ergebnisraums  $\Omega$  :

$$P(A) = \frac{|A|}{|\Omega|}.$$

Es ist daher von besonderem Interesse,  $|A|$  und  $|\Omega|$  abzuschätzen.

**Bemerkung 1.11** (Kartesisches Produkt). Seien  $B, C$  endliche Mengen und  $A = B \times C$ . Dann gilt  $|A| = |B| \cdot |C|$ .

**Beispiele 1.12.** (a) Betrachte ein zweistufiges Zufallsexperiment, bestehend aus einem Münzwurf (als Ergebnis lassen wir Kopf  $(\mathbb{K})$  oder Zahl  $(\mathbb{Z})$  zu) und anschließendem Würfelwurf. In der Notation von oben bedeutet das für den Ergebnisraum:

$$B = \{(\mathbb{K}), (\mathbb{Z})\}, \quad C = \{\square, \square, \dots, \boxplus\}$$

$$A = B \times C = \{((\mathbb{K}), \square), ((\mathbb{K}), \square), \dots, ((\mathbb{K}), \boxplus), \\ (\mathbb{Z}, \square), (\mathbb{Z}, \square), \dots, (\mathbb{Z}, \boxplus)\}$$

Allgemeiner gilt für ein kartesisches Produkt  $M = M_1 \times M_2 \times \dots \times M_k$  endlicher Mengen  $M_1, \dots, M_k$  mit  $k \in \mathbb{N}$  dann  $|M| = \prod_{i=1}^k |M_i|$ . Im Sonderfall  $A = B^k, |B| = n$ , also  $|A| = n^k$ , spricht man bei den Elementen von  $A$  von “ $k$ -Tupeln aus einer  $n$ -Menge, mit Zurücklegen”.

(b) Gegeben sei eine Urne mit sechs unterschiedlichen Kugeln, aus der drei mal ohne Zurücklegen gezogen wird. Hier beträgt die Anzahl der möglichen Kombinationen  $6 \cdot 5 \cdot 4 = 120$ . Allgemein lässt sich das folgendermaßen formulieren:

$$|\Omega| = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - k + 1) = \frac{n!}{(n - k)!}$$

Hier spricht man von einem “ $k$ -Tupel aus einer  $n$ -Menge, ohne Zurücklegen”.

(c) Im Gegensatz zum Beispiel oben spielt beim Zahlenlotto die Reihenfolge der gezogenen Zahlen keine Rolle. Mathematisch lässt sich das modellieren, indem man statt Tupeln Mengen betrachtet. Beispielsweise können so durch die Mengenschreibweise folgende Permutationen des Tupels  $(1, 7, 9)$  zusammengefasst werden:

$$\{1, 7, 9\} \leftrightarrow (1, 7, 9), (1, 9, 7), (7, 1, 9), (7, 9, 1), (9, 1, 7), (9, 7, 1)$$

Sei nun allgemein eine  $n$ -elementige Menge  $A$  gegeben. Wir interessieren uns dafür, wie viele Kombinationen mit  $k$  Elementen gebildet werden können, also für die Mächtigkeit der Menge

$$\Omega = \{\omega \in \mathfrak{P}(A) \mid |\omega| = k\}.$$

Dazu berechnet man zunächst die Anzahl der Kombinationen bei Beachtung der Reihenfolge wie in (b) und dividiert anschließend durch die Anzahl der Permutationen. Letztere beträgt für ein  $k$ -Tupel gerade  $k!$ . Insgesamt ist also

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

Man spricht hierbei von “Ziehen ohne Zurücklegen, Reihenfolge egal”.

(d) *Multinomialkoeffizient.* Beim “MISSISSIPPI-Problem” ist die Anzahl der Möglichkeiten gesucht, die Buchstaben des Wortes “Mississippi” anzuordnen. Die Schwierigkeit dabei ist, dass gleiche Buchstaben auftreten. Wir nehmen daher zunächst eine künstliche Unterscheidung vor, d.h. wir tun so, als wären alle Buchstaben unterschiedlich:

$$MI_1S_1S_2I_2S_3S_4I_3P_1P_2I_4$$

Es gibt nach (b)  $11!$  Möglichkeiten, diese elf Buchstaben anzuordnen. Da es jedoch hier nicht gewünscht ist, gleiche Buchstaben zu unterscheiden, dividieren wir durch

die Anzahl der Permutationen. Diese ist z.B.  $4!$  für die vier "I" und  $1!$  für das "M". Damit ergibt sich für die Anzahl der Möglichkeiten, Anagramme des Wortes "Mississippi" zu bilden:

$$\frac{11!}{2! \cdot 4! \cdot 4! \cdot 1!} = 34650$$

**Bemerkung 1.13.** Den Term  $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$  kennt man als *Binomialkoeffizient*. Er tritt im binomischen Lehrsatz, der Verallgemeinerung der binomischen Formel, auf:

$$(x + y)^n = x^n + nx^{n-1}y + \dots + \binom{n}{k}x^{n-k}y^k + \dots + y^n = \sum_{k=0}^n \binom{n}{k}x^{n-k}y^k$$

Diese Formel kann kombinatorisch interpretiert werden, denn die Anzahl der Möglichkeiten, aus einer Menge mit  $n$  Elementen der Form  $(x + y)$  das  $x$  genau  $k$ -mal auszuwählen beträgt nach (b) gerade  $\binom{n}{k}$ . Der Koeffizient vor dem Summanden  $x^{n-k}y^k$  ist daher der Binomialkoeffizient  $\binom{n}{k}$ .

## 1.4 Bedingte Wahrscheinlichkeit

**Beispiel 1.14.** Eine Fabrik stellt Glühbirnen her. Dabei ergibt eine stichprobenartige Überprüfung des Inhalts zweier Kartons zu je 10 000 Birnen Folgendes:

	1. Karton	2. Karton
defekt	20	250
o.k.	9980	9750

Wählt man nun zufällig eine Birne aus einem Karton, so hängt die Wahrscheinlichkeit für einen Defekt also davon ab, welchen Karton wir wählen. Die Wahrscheinlichkeit, eine defekte zu erwischen, falls man aus dem ersten Karton wählt, ist also  $\frac{20}{10^4}$ .

**Definition 1.15.** Sei  $\Omega$  eine endliche Menge,  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  ein Wahrscheinlichkeitsmaß und  $B \subseteq \Omega$  mit  $P(B) \neq 0$ . Dann nennt man für alle  $A \subseteq \Omega$  den Wert

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

*bedingte Wahrscheinlichkeit* von  $A$  unter der Bedingung  $B$ .

**Satz 1.16.** Sei  $\Omega$  eine endliche Menge,  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  ein Wahrscheinlichkeitsmaß und  $B \subseteq \Omega$  mit  $P(B) \neq 0$ . Dann ist die Abbildung

$$P_B : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}, A \mapsto P_B(A)$$

ein Wahrscheinlichkeitsmaß.

*Beweis.* Wir rechnen die drei Axiome von Kolmogorow nach, indem wir verwenden, dass es sich bei  $P$  um ein Wahrscheinlichkeitsmaß handelt.

(i) zu zeigen:  $P_B(\Omega) = 1$

Es ist gemäß der Definition

$$P_B(\Omega) = \frac{P(B \cap \Omega)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(ii) zu zeigen:  $P(A) \geq 0$  für alle  $A \in \mathfrak{P}(\Omega)$

Hier gilt

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \geq 0$$

denn sowohl Zähler als auch Nenner sind größer gleich Null, der Nenner ist aufgrund der Definition sogar strikt positiv.

(iii) zu zeigen:  $P_B(A \cup C) = P_B(A) + P_B(C)$  falls  $A \cap C = \emptyset$

$$\begin{aligned} P_B(A \cup C) &= \frac{P((A \cup C) \cap B)}{P(B)} \stackrel{(*)}{=} \frac{P((A \cap B) \cup (C \cap B))}{P(B)} = \\ &\stackrel{(**)}{=} \frac{P(A \cap B) + P(C \cap B)}{P(B)} = P_B(A) + P_B(C) \end{aligned}$$

Dabei haben wir an der Stelle  $(*)$  das Distributivgesetz der Mengenlehre verwendet und bei  $(**)$ , dass  $A \cap B \subseteq A$  und  $C \cap B \subseteq C$  disjunkt sind, da  $A$  und  $C$  nach Voraussetzung disjunkt sind.

□

**Satz 1.17** (Satz von Bayes). Sei  $\Omega$  eine Menge,  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  ein Wahrscheinlichkeitsmaß und seien  $A, B \subseteq \Omega$  Ereignisse mit  $P(A), P(B) \neq 0$ . Es gilt

$$P_B(A) = \frac{P_A(B)P(A)}{P(B)}$$

*Beweis.* Aus der Definition der bedingten Wahrscheinlichkeit erhält man direkt  $P(A \cap B) = P_A(B)P(A)$ . Erneutes Einsetzen, diesmal in die bedingte Wahrscheinlichkeit mit Bedingung  $B$ , ergibt die gewünschte Gleichung. □

**Definition 1.18.** Sei  $\Omega$  eine Menge. Eine Familie  $\{B_i\}_{i \in I}$  paarweise disjunkter Ereignissen  $B_i \in \mathfrak{P}(\Omega)$  mit einer Indexmenge  $I$  heißt *Zerlegung* von  $\Omega$ , falls  $\bigcup_{i \in I} B_i = \Omega$  gilt.

**Satz 1.19** (Satz von der totalen Wahrscheinlichkeit). Sei  $\Omega$  eine endliche Menge,  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  ein Wahrscheinlichkeitsmaß und  $\{B_i\}_{i \in I}$  eine Zerlegung von  $\Omega$  mit einer endlichen Indexmenge  $I$  und  $P(B_i) \neq 0$  für alle  $i \in I$ . Dann gilt

$$P(A) = \sum_{i \in I} P_{B_i}(A) \cdot P(B_i).$$



*Beweis.* Betrachte

$$\begin{aligned} \sum_{i \in I} P_{B_i}(A)P(B_i) &= \sum_{i \in I} \frac{P(A \cap B_i)}{P(B_i)}P(B_i) \\ &= \sum_{i \in I} P(A \cap B_i) \stackrel{(*)}{=} P\left(A \cap \left(\bigcup_{i \in I} B_i\right)\right) = P(A \cap \Omega) = P(A) \end{aligned}$$

An Stelle (\*) ging ein, dass die Mengen  $B_i$  laut Voraussetzung disjunkt sind. Die Wahrscheinlichkeit der Vereinigungsmenge ergibt dich damit laut 1.5 (ii) als Summe der einzelnen Wahrscheinlichkeiten.  $\square$

## 1.5 Unabhängigkeit von Ereignissen

In Beispiel 1.14 im letzten Kapitel wurde das Ziehen von Glühbirnen aus zwei verschiedenen Kisten betrachtet. Offensichtlich war hier die Wahrscheinlichkeit, eine defekte Birne zu ziehen, *abhängig* davon, aus welcher Kiste gezogen wurde. Ausgehend von dieser intuitiven Vorstellung werden wir uns in diesem Kapitel mit der Abhängigkeit zweier Ereignisse beschäftigen und definieren dazu zunächst die stochastische Unabhängigkeit.

**Definition 1.20.** Sei  $\Omega$  eine Menge und  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  ein Wahrscheinlichkeitsmaß. Zwei Ereignisse  $A, B \in \mathfrak{P}(\Omega)$  mit  $P(B) \neq 0$  heißen *stochastisch unabhängig* voneinander, falls

$$P(A) = P_B(A)$$

gilt.

**Satz 1.21.** Sei  $\Omega$  eine endliche Menge,  $P : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$  ein Wahrscheinlichkeitsmaß und  $A, B \in \mathfrak{P}(\Omega)$  zwei stochastisch unabhängige Ereignisse mit  $P(A) \neq 0 \neq P(B)$ . Dann gilt auch

- (i)  $B$  ist unabhängig von  $A$ ,
- (ii)  $A$  ist unabhängig von  $B^c$ , falls  $P(B) \neq 1$ ,
- (iii)  $P(A \cap B) = P(A)P(B)$ .

*Beweis.* zu (ii): zu zeigen ist  $P_{B^c}(A) = P(A)$

Es gilt die Mengengleichung

$$A = (A \cap B) \cup (A \cap B^c)$$

wobei die Mengen  $A \cap B$  und  $A \cap B^c$  offensichtlich disjunkt sind. Mit Kolmogoroff 1.5 (ii) erhalten wir nun

$$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c) = \\ &= P_B(A)P(B) + P_{B^c}(A)P(B^c) \end{aligned}$$

Subtraktion von  $P_B(A)P(B)$  auf beiden Seiten ergibt

$$\begin{aligned} P_{B^c}(A)P(B^c) &= P(A) - P_B(A)P(B) = P(A) - P(A)P(B) = \\ &= P(A)(1 - P(B)) = P(A)P(B^c) \end{aligned}$$

Nach Kürzen von  $P(B^c)$  erhält man die Behauptung:

$$P_{B^c}(A) = P(A)$$

zu (iii): Die Behauptung erhält man direkt, indem man die Voraussetzung wie folgt umformt

$$P(A) = P_B(A) \Leftrightarrow P(A) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A)P(B) = P(A \cap B)$$

zu (i): Da Aussage (ii) durch eine Äquivalenz bewiesen wurde, gilt auch die umgekehrte Richtung. Ist also die Gleichung  $P(A \cap B) = P(A)P(B)$  erfüllt, so ist  $A$  unabhängig von  $B$ . Ist also  $A$  von  $B$  unabhängig, dann folgt aufgrund der Symmetrie dieser Formel, dass  $B$  von  $A$  unabhängig ist.  $\square$

**Beispiele 1.22.** Wir führen zwei Beispiele an, um aufzuzeigen, dass eine solche Unabhängigkeit nicht immer einfach zu interpretieren ist.

- (a) In Österreich hat man durch Zählungen herausgefunden, dass die Anzahl der Störche von der Geburtenrate abhängig ist (d.h. dass diese beiden Ereignisse korreliert sind). Je höher also die Anzahl der Störche, desto mehr Neugeborene wurden im entsprechenden Gebiet gezählt. Woran liegt das?  
Der wahre Grund ist natürlich nicht, dass Kinder vom Storch gebracht werden, sondern eine gemeinsame Ursache aus der Vergangenheit (sowohl die Zahl der Kinder als auch die der Storchennester nehmen seit Jahren ab).
- (b) Die Zahl der in der USA gefällten Todesurteile ist mit der Hautfarbe der vermeintlichen Täter korreliert. Schwarze werden häufiger zum Tode verurteilt. Eine Erklärung könnte sein, dass die mildere Strafe durch die Hautfarbe der Täter beeinflusst wird. Da jedoch auch die Hautfarbe von Täter und Opfer stark korreliert sind, könnte auch angenommen werden, dass Gerichte die Morde, in denen Schwarze die Opfer sind, strenger verfolgen.

## II Überabzählbare Ergebnisräume

Da der Fall abzählbar unendlicher Ergebnisräume wenig Neues bietet, wenden wir uns gleich überabzählbaren Ergebnisräumen zu und verallgemeinern die bisher vorgestellten Konzepte.

### 2.1 Das Wahrscheinlichkeitsmaß für überabzählbare Ergebnisräume

**Definition 2.1** (Kolmogoroff-Axiome für überabzählbare Ergebnisräume). Sei  $\Omega$  eine (überabzählbar unendliche) Menge. Wir betrachten eine Menge  $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ , ein Ereignis  $A \in \mathcal{A}$  sowie eine Folge paarweise disjunkter Ereignisse  $\{A_i\}_{i \in I} \subseteq \mathcal{A}$ . Eine Abbildung  $P: \mathcal{A} \rightarrow \mathbb{R}$  heißt *Wahrscheinlichkeitsmaß*, falls gilt:

- (a)  $P(A) \geq 0$ ,
- (b)  $P(\Omega) = 1$ ,
- (c)  $P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$ .

An dieser Stelle kommt die Frage auf, wie die Menge  $\mathcal{A}$  aussieht und warum  $P$  nicht wieder auf der Potenzmenge von  $\Omega$  definiert wird. Tatsächlich würde dies die Möglichkeiten, ein  $P$  zu finden, das die eben gestellten Bedingungen erfüllt, stark einschränken. Wir verdeutlichen dies anhand des folgenden Beispiels.

**Beispiel 2.2** (Glücksrad revisited). Wir kommen auf Beispiel 1.3 zurück. Hier war  $\Omega = [0; 2\pi[$ . Wir zeigen, dass für diese Situation unter einer Laplace-Annahme kein Wahrscheinlichkeitsmaß auf ganz  $\mathfrak{P}(\Omega)$  konstruiert werden kann. Es erscheint sinnvoll, eine Art Laplace-Annahme in der Form zu treffen, dass jedem gleich großen Winkelsegment die gleiche Wahrscheinlichkeit zugeordnet werden soll. Wir fordern daher in Analogie zu 1.8, dass jedem Winkelsegment  $A \subseteq [0; 2\pi[$  die Wahrscheinlichkeit

$$P(A) = \frac{\lambda(A)}{\lambda(\Omega)}$$

zugeordnet werden soll, wobei  $\lambda(\cdot)$  das Lebesgue-Maß von  $A$  bzw.  $\Omega$  bezeichnet. Es gibt jedoch Teilmengen von  $[0; 2\pi[$ , für die das Lebesgue-Maß nicht definiert ist. Auf diesen Mengen kann daher auch  $P$  nicht definiert werden, weshalb der Definitionsbereich nicht ganz  $\mathfrak{P}(\Omega)$  sein kann.

**Lemma 2.3.** Es gibt eine nicht Lebesgue-messbare Teilmenge von  $[0; 2\pi[$ .<sup>1</sup>

*Beweis.* Wir konstruieren als nicht Lebesgue-messbare Menge die sogenannte *Vitali-Menge*. Wir sehen zunächst, dass für  $a, b \in \mathbb{R}$  durch

$$a \sim b \Leftrightarrow a - b \in \mathbb{Q}$$

eine Äquivalenzrelation definiert ist.

Wir benutzen nun das Auswahlaxiom, um aus jeder Äquivalenzklasse ein Element auszuwählen und bezeichnen die Menge dieser Repräsentanten mit  $M \subseteq [0; 2\pi[$ . Im Folgenden zeigen wir, dass dieser Menge keine sinnvolles Maß zugeordnet werden kann.

Sei  $0 \neq q \in \mathbb{Q}$  und  $M + q = \{m + q \mid m \in M\}$ .  $M + q$  und  $M$  sind disjunkt, denn gäbe es ein  $m \in M$ , das auch in  $M + q$  liegt, so wäre  $m = m' + q$  für ein  $m' \in M$  mit  $m \neq m'$ . Wegen  $m - m' = q \in \mathbb{Q}$  liegen  $m$  und  $m'$  in derselben Äquivalenzklasse. Da  $M$  aus jeder Äquivalenzklasse nur genau ein Element enthält, ist das unmöglich.

Es gilt zudem

$$[0; 2\pi[ \subseteq \bigcup_{q \in [-2\pi; 2\pi[ \cap \mathbb{Q}} (M + q) \subseteq [-2\pi; 4\pi[$$

Die zweite Inklusion ist klar, da  $M$  um höchstens  $\pm 2\pi$  verschoben wird. Um die die erste Inklusion einzusehen, sei  $x \in [0, 2\pi[$ , dann gibt es ein  $m \in M$  mit  $x \sim m$ . Das bedeutet gerade

$$x - m = q \in \mathbb{Q} \quad \Leftrightarrow \quad x = m + q \in (M + q)$$

mit einem  $x \in [0, 2\pi[$ .

Wäre  $M$  nun messbar, so träfe dies auch auf  $M + q$  zu, da diese Menge nur durch eine Translation verschoben wurde. Insbesondere wäre  $\lambda(M) = \lambda(M + q)$ . Wäre weiterhin  $\lambda(M) = 0$ , so würde aus der Additivität des Lebesgue-Maßes für abzählbare, disjunkte Vereinigungen auch

$$2\pi = \lambda([0, 2\pi[) \leq \lambda\left(\bigcup_{q \in [-2\pi; 2\pi[ \cap \mathbb{Q}} (M + q)\right) = \sum_{q \in [-2\pi; 2\pi[ \cap \mathbb{Q}} \lambda(M + q) = 0$$

folgen, was offensichtlich nicht der Fall ist. Wäre jedoch  $\lambda(M) = \varepsilon \in \mathbb{R}$ , dann wäre

$$6\pi = \lambda([-2\pi, 4\pi[) \geq \lambda\left(\bigcup_{q \in [-2\pi; 2\pi[ \cap \mathbb{Q}} (M + q)\right) = \sum_{q \in [-2\pi; 2\pi[ \cap \mathbb{Q}} \lambda(M + q) = \sum_{q \in [-2\pi; 2\pi[ \cap \mathbb{Q}} \varepsilon = \infty$$

was ebenfalls unmöglich sein kann. □

## 2.2 $\sigma$ -Algebren

Das Beispiel im vorigen Abschnitt zeigt, dass wir den Definitionsbereich des Wahrscheinlichkeitsmaßes einschränken müssen, um eine sinnvolle Modellierung zu ermöglichen. Dazu orientieren wir uns an den Axiomen von Kolmogoroff und formulieren eine allgemeine Definition, die diesen entgegen kommt.

<sup>1</sup>Den Beweis dieser Aussage haben wir im Vergleich zur Vorlesung leicht geändert.

**Definition 2.4.** Sei  $\Omega$  ein Ereignisraum.  $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$  heißt  $\sigma$ -Algebra über  $\Omega$ , falls

- (a)  $\Omega \in \mathcal{A}$ ,
- (b) für eine Folge  $\{A_n\}_{n \in \mathbb{N}}$  in  $\mathcal{A}$  auch  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$  erfüllt ist,
- (c) für jedes  $A \in \mathcal{A}$  auch  $A^c \in \mathcal{A}$  gilt.

**Beispiele 2.5.** (a) Sei  $\Omega = \{1, 2, 3\}$ , dann ist  $\mathcal{A} = \{\Omega, \emptyset, \{1\}, \{2, 3\}\}$  eine  $\sigma$ -Algebra.

Man prüft unmittelbar nach, dass die angegebene Menge unter Vereinigung und Komplementbildung abgeschlossen ist.

- (b) Für jedes  $\Omega$  ist  $\mathfrak{P}(\Omega)$  eine  $\sigma$ -Algebra.
- (c) Für jedes  $\Omega$  ist  $\{\Omega, \emptyset\}$  die kleinste und damit langweiligste  $\sigma$ -Algebra.

**Satz 2.6.** Für jede  $\sigma$ -Algebra  $\mathcal{A}$  über  $\Omega$  gilt:

- (a)  $\emptyset \in \mathcal{A}$ .
- (b) Falls  $\{A_n\}_{n \in \mathbb{N}}$  eine Folge in  $\mathcal{A}$  ist, so gilt auch  $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$ .
- (c) Endliche Vereinigungen und Schnitte von Ereignissen aus  $\mathcal{A}$  sind wieder Ereignisse aus  $\mathcal{A}$ .
- (d) Falls  $A, B \in \mathcal{A}$ , so folgt  $A \setminus B \in \mathcal{A}$ .

*Beweis.* (a) Laut Axiom (a) gilt  $\Omega \in \mathcal{A}$ , also ist nach (c) auch  $\Omega^c = \emptyset \in \mathcal{A}$ .

(b) Nach den DeMorgan'schen Regeln ist

$$\bigcap_{n \in \mathbb{N}} A_n = \left( \bigcup_{n \in \mathbb{N}} A_n^c \right)^c$$

Da auf der rechten Seite nur Operationen auf eine Folge in  $\mathcal{A}$  angewendet wurden, unter denen  $\mathcal{A}$  abgeschlossen ist, liegt auch die linke Seite in  $\mathcal{A}$ .

(c)  $\bigcup_{n=1}^k A_n = \bigcup_{n=1}^{\infty} A_n$  mit  $A_n = \emptyset$  für alle  $n > k$ . Mit a) ist  $\emptyset \in \mathcal{A}$ , daher ist die Vereinigungsmenge gemäß 2.4 b) in  $\mathcal{A}$  enthalten. Die Aussage für endliche Schnitte beweist man genauso.

(d) An der Darstellung  $A \setminus B = A \cap B^c$  sieht man, dass auch  $A \setminus B$  aus  $A$  und  $B$  durch elementare Operationen gebildet werden kann, unter denen  $\mathcal{A}$  abgeschlossen ist.  $\square$

**Definition 2.7.** Sei  $\mathcal{A}$  eine  $\sigma$ -Algebra und  $\{A_n\}_{n \in \mathbb{N}}$  eine Folge in  $\mathcal{A}$ , dann setzen wir:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} \left( \bigcup_{k \geq n} A_k \right)$$

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n \in \mathbb{N}} \left( \bigcap_{k \geq n} A_k \right)$$

**Satz 2.8.** Sei  $\mathcal{A}$  eine  $\sigma$ -Algebra und  $\{A_n\}_{n \in \mathbb{N}}$  eine Folge in  $\mathcal{A}$ , dann gilt auch

$$\limsup_{n \rightarrow \infty} A_n \in \mathcal{A} \quad \text{und} \quad \liminf_{n \rightarrow \infty} A_n \in \mathcal{A}.$$

*Beweis.* Wegen Axiom (b) der Definition ist  $\bigcup_{k \geq n} A_k \in \mathcal{A}$  für alle  $n \in \mathbb{N}$ . Durch Anwendung von Satz 2.6 (c) folgt  $\lim_{n \rightarrow \infty} \sup A_n \in \mathcal{A}$  wie gewünscht. Die Aussage für  $\lim_{n \rightarrow \infty} \inf A_n \in \mathcal{A}$  folgt genauso.  $\square$

**Satz 2.9.** Sei  $I$  eine (nicht notwendigerweise abzählbare) Indexmenge und  $\{\mathcal{A}_i\}_{i \in I}$  eine Folge von  $\sigma$ -Algebren über der selben Grundmenge  $\Omega$ . Dann ist auch  $\mathcal{B} = \bigcap_{i \in I} \mathcal{A}_i$  eine  $\sigma$ -Algebra.

*Beweis.* Wir prüfen die Axiome aus Definition 2.4 nach.

- (a) Da alle  $\mathcal{A}_i$   $\sigma$ -Algebren sind, ist  $\Omega \in \mathcal{A}_i$  für alle  $i \in I$  und liegt daher auch im Schnitt der  $\mathcal{A}_i$ .
- (b) Sei  $\{C_n\}_{n \in \mathbb{N}}$  eine Folge in  $\mathcal{B}$ . Da  $\mathcal{B}$  der Schnitt aller  $\mathcal{A}_i$  ist, gilt  $C_n \in \mathcal{A}_i$  für alle  $n \in \mathbb{N}$  und  $i \in I$ . Also gilt für alle  $i \in I$ , dass  $\bigcup_{n \in \mathbb{N}} C_n \in \mathcal{A}_i$ , da die  $\mathcal{A}_i$  jeweils  $\sigma$ -Algebren sind. Damit ist auch  $\bigcup_{n \in \mathbb{N}} C_n \in \mathcal{B}$ .
- (c) Sei  $C \in \mathcal{B}$ , dann ist  $C \in \mathcal{A}_i$  für alle  $i \in I$ . Also ist auch  $C^c \in \mathcal{A}_i$  für alle  $i \in I$  und es folgt  $C^c \in \mathcal{B}$ .

$\square$

**Bemerkung 2.10.** Die Vereinigung von  $\sigma$ -Algebren ist im Allgemeinen keine  $\sigma$ -Algebra. Als Gegenbeispiel dazu betrachten wir  $\Omega = \{1, 2, 3\}$ . Sei nun

$$\mathcal{A}_1 = \{\Omega, \emptyset, \{1\}, \{2, 3\}\} \quad \text{und} \quad \mathcal{A}_2 = \{\Omega, \emptyset, \{2\}, \{1, 3\}\}.$$

Wir haben bereits in Bsp 2.5 gesehen, dass  $\mathcal{A}_1$  und  $\mathcal{A}_2$   $\sigma$ -Algebren sind. Nun ist jedoch

$$\mathcal{B} = \mathcal{A}_1 \cup \mathcal{A}_2 = \{\Omega, \emptyset, \{1\}, \{2\}, \{2, 3\}, \{1, 3\}\}$$

keine  $\sigma$ -Algebra, denn  $\{1\}, \{2\} \in \mathcal{B}$ , aber  $\{1, 2\} \notin \mathcal{B}$ . Damit ist 2.6 (c) verletzt.

**Definition 2.11.** Sei  $\Omega$  eine Grundmenge,  $M \subseteq \mathfrak{P}(\Omega)$  und  $\mathcal{M} \subseteq \mathfrak{P}(\Omega)$  die kleinste  $\sigma$ -Algebra, die  $M$  enthält. Dann nennt man  $\mathcal{M}$  die von  $M$  erzeugte  $\sigma$ -Algebra. Dabei ist

$$\mathcal{M} = \bigcap \{ \mathcal{A} \subseteq \mathfrak{P}(\Omega) \mid \mathcal{A} \text{ ist } \sigma\text{-Algebra mit } M \subseteq \mathcal{A} \}$$

**Beispiel 2.12.** Sei  $\Omega = \{1, 2, 3\}$  und  $M = \{\emptyset, \{2, 3\}\}$ . Die von  $M$  erzeugte  $\sigma$ -Algebra ist dann  $\mathcal{A} = \{\Omega, \emptyset, \{2, 3\}, \{1\}\}$ .

**Definition 2.13.** Sei  $(\Omega, d)$  ein metrischer Raum. Die  $\sigma$ -Algebra, die von den offenen Teilmengen von  $\Omega$  erzeugt wird, heißt *Borelsche  $\sigma$ -Algebra* (oder Borelmenge) über  $\Omega$ .

**Bemerkung 2.14.** In Definition 2.13 kann statt einem metrischen Raum genauso ein topologischer Raum gefordert werden. Tatsächlich benutzt man auf einem metrischen Raum gerade die von der Metrik induzierte Topologie, d.h. eine Teilmenge  $U \subseteq \Omega$  ist genau dann offen, wenn für jeden Punkt  $x$  ein  $\varepsilon > 0$  existiert, sodass die Menge  $B_\varepsilon = \{y \in \Omega \mid d(x, y) < \varepsilon\}$  eine Teilmenge von  $U$  ist. Die offenen zusammenhängenden Teilmengen von  $\mathbb{R}$  bezüglich der Standardmetrik  $d(x, y) = |x - y|$  sind genau die offenen Intervalle.

**Definition 2.15.** Die Borelmenge über  $\mathbb{R}$  bzw. über einer Teilmenge von  $\mathbb{R}$  ist die  $\sigma$ -Algebra, die von allen offenen Intervallen  $]a, b[$  mit  $a, b \in \mathbb{R} \cup \{\pm\infty\}$  und  $a < b$  erzeugt wird.

**Satz 2.16.** In der Borelmenge über  $\mathbb{R}$  sind auch alle abgeschlossenen und alle halboffenen Intervalle enthalten.

*Beweis.* Sei  $\mathcal{B}$  die Borelmenge über  $\mathbb{R}$ . Wir zeigen, dass  $]a, b[ \in \mathcal{B}$  für beliebige  $a, b \in \mathbb{R}$  mit  $a < b$ , die anderen Fälle beweist man analog.

Dazu behaupten wir

$$\bigcap_{n \in \mathbb{N}} ]a - \frac{1}{n}, b[ = ]a, b[$$

Aufgrund von Satz 2.6 (b) ist die linke Menge als Schnitt von Elementen der Borelmenge selbst ein Element der Borelmenge, also genügt es, diese Gleichheit zu zeigen.

“ $\supseteq$ ” ist klar, da  $a \in ]a - \frac{1}{n}, b[$  für alle  $n \in \mathbb{N}$  gilt.

“ $\subseteq$ ” Angenommen, es gibt eine Zahl  $c$  in der linken Menge, die kleiner ist als  $a$  (also mit  $c = a - \varepsilon$ ). Wir können dann  $n$  so groß wählen, dass  $\frac{1}{n} < \varepsilon$  und also  $c \notin ]a - \frac{1}{n}, b[$  gilt. So ein  $c$  kann es daher nicht geben.  $\square$

Dieses Erzeugendensystem kann noch weiter reduziert werden.

**Proposition 2.17.** Auch die Intervalle der Form  $] - \infty, a[$  mit  $a \in \mathbb{Q}$  bilden ein Erzeugendensystem der Borelmenge über  $\mathbb{R}$ .

*Beweis.* Sei  $\mathcal{B}$  diejenige  $\sigma$ -Algebra, die von derartigen Intervallen erzeugt wird. Wir beweisen, dass  $\mathcal{B}$  auch alle offenen Teilmengen von  $\mathbb{R}$  enthält. Da die Borelmenge die kleinste  $\sigma$ -Algebra mit dieser Eigenschaft ist, folgt daraus, dass auch die ganze Borelmenge in  $\mathcal{B}$  enthalten ist.

1. *Schritt:* Wir zeigen, dass auch Intervalle der Form  $] - \infty, b[$  für  $b \in \mathbb{R}$  in  $\mathcal{B}$  enthalten sind. Dazu betrachten wir die Menge

$$\bigcup_{\substack{q \in \mathbb{Q} \\ q < b}} ] - \infty, q[ = ] - \infty, b[$$

Mit der selben Argumentation wie in Satz 2.16 zeigt man, dass diese Menge in  $\mathcal{B}$  enthalten ist. Zum Nachweis von “ $\supseteq$ ” merken wir an, dass für jedes Element  $c \in \mathbb{R}$  mit  $c < b$  eine rationale Zahl  $q \in \mathbb{Q}$  mit  $c < q < b$  gewählt werden kann, da  $\mathbb{Q}$  dicht in  $\mathbb{R}$  liegt. Es ist dann  $c \in ] - \infty, q[$  und liegt also in der Vereinigung. Die Zahl  $b$  selbst hingegen ist kein Element der Menge, da es in keinem der offenen Intervalle enthalten ist.

2. *Schritt:* Wir zeigen, dass  $]b, c[ \in \mathcal{B}$  für  $b, c \in \mathbb{R}$ .

Analog zum Beweis von 2.16 folgt, dass

$$] - \infty, b[ = \bigcap_{n \in \mathbb{N}} ] - \infty, b + \frac{1}{n}[ \in \mathcal{B}$$

Als Komplement dieser Menge ist damit auch  $]b, +\infty[$  in  $\mathcal{B}$  enthalten. Da  $\mathcal{B}$  unter Bildung von Schnittmengen abgeschlossen ist, folgt, dass die offenen Intervalle

$]b, c[ = ] - \infty, c[\cap]b, \infty[$  in  $\mathcal{B}$  enthalten sind.

Es bleibt zu zeigen, dass die angegebenen Intervalle keine größere Menge als die Borelmenge erzeugen. Dies ist jedoch klar, da  $\mathcal{B}$  laut der Definition von offenen Intervallen erzeugt wird, die Borelmenge aber von *allen* offenen Intervallen. Damit ist  $\mathcal{B}$  in der Borelmenge enthalten, die beiden Mengen sind also insgesamt identisch.  $\square$

**Bemerkung 2.18.** Als Erzeugendensystem können auch die abgeschlossenen, halboffenen Intervalle oder Intervalle der Form  $] - \infty, a]$  gewählt werden.

**Lemma 2.19.** Alle Elemente der Borelmenge über  $\mathbb{R}$  sind Lebesgue-messbar.

*Beweis.* Da offene Intervalle stets messbar sind, gilt dies zumindest für alle Elemente des Erzeugendensystems. Die Menge der Lebesgue-messbaren Mengen  $\mathcal{L}$  ist selbst eine  $\sigma$ -Algebra (sowohl Vereinigung als auch Komplemente messbarer Mengen sowie  $\mathbb{R}$  selbst sind messbar). Damit ist  $\mathcal{L}$  eine  $\sigma$ -Algebra, die alle offenen Mengen enthält, die Borelmenge die kleinste  $\sigma$ -Algebra mit dieser Eigenschaft. Es folgt, dass die Borelmenge eine Teilmenge von  $\mathcal{L}$  ist, d.h. alle Elemente der Borelmenge sind messbar.  $\square$

**Proposition 2.20.** Auf jedem Intervall  $I \subseteq \mathbb{R}$  kann ein Wahrscheinlichkeitsmaß auf der Borelmenge über  $I$  definiert werden, indem man

$$P(A) = \frac{\lambda(A)}{\lambda(I)}$$

setzt. Dabei bezeichnet  $\lambda : \mathcal{L} \rightarrow \mathbb{R}$  die Lebesgue'sche Maßfunktion.

*Beweis.* Die Messbarkeit der Mengen folgt direkt aus dem Lemma 2.19. Die Axiome des Wahrscheinlichkeitsmaßes rechnet der geübte Leser unmittelbar nach.  $\square$

Für das Glücksrad erhält man auf diese Art ein translationsinvariantes Wahrscheinlichkeitsmaß, d.h. das Drehen einer Menge um einen beliebigen Winkel  $\varphi$  ändert die Wahrscheinlichkeit nicht, da sich das Maß der Menge durch die Translation nicht ändert.

**Satz 2.21** (ohne Beweis). Sei  $\mathcal{B}$  die Borelsche  $\sigma$ -Algebra über  $\mathbb{R}$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  eine integrierbare Funktion mit der Eigenschaft

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Dann ist durch

$$P(A) = \int_A f(x) dx$$

ein Wahrscheinlichkeitsmaß definiert. Die Funktion  $f$  nennt man *Wahrscheinlichkeitsdichte*.

**Bemerkung 2.22.** Das Konzept der Wahrscheinlichkeitsdichte besitzt u.a. in der Physik eine große Bedeutung. Beipielsweise definiert das Betragsquadrat der Wellenfunktion in der Quantenmechanik eine Wahrscheinlichkeitsdichte.



## 2.3 Zufallsgrößen

Wir betrachten zunächst wieder endliche Wahrscheinlichkeitsräume und behandeln den Fall für allgemeine  $\sigma$ -Algebren in einem späteren Teil.

**Definition 2.23.** Seien  $\Omega, \vartheta$  endliche Ergebnisräume. Eine Abbildung

$$X : \Omega \rightarrow \vartheta$$

heißt *Zufallsvariable* oder *Zufallsgröße*. Gilt  $\vartheta \subseteq \mathbb{R}$ , dann spricht man von einer reellen Zufallsvariablen.

**Beispiel 2.24.** Wir beschäftigen uns mit Roulette und betrachten den vereinfachten Ereignisraum

$$\Omega = \{\text{rot, grün, schwarz}\}.$$

Dies definiert auf natürliche Weise eine Zufallsvariable

$$X : \Omega \rightarrow \vartheta \text{ mit } X(\text{rot}) = 1, X(\text{schwarz}) = -1, X(\text{grün}) = -1$$

Also ist hier  $\vartheta = \{\pm 1\}$ .

Falls wir ein Wahrscheinlichkeitsmaß auf der Potenzmenge  $\mathfrak{P}(\Omega)$  kennen, können wir ein entsprechendes Wahrscheinlichkeitsmaß  $\tilde{P}$  auf  $\mathfrak{P}(\vartheta)$  definieren, indem wir für jedes  $A \subseteq \mathfrak{P}(\vartheta)$

$$\tilde{P}(A) = P(X^{-1}(A))$$

setzen. Zum Beispiel ist  $\tilde{P}(\{-1\}) = P(\{\text{schwarz, grün}\})$  im Beispiel oben. Zu beachten ist, dass eine Umkehrabbildung

$$X^{-1} : \vartheta \rightarrow \Omega$$

im Allgemeinen nicht definiert ist. Jedoch lässt sich in jedem Fall die Menge

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$$

angeben.

**Bemerkung 2.25.** Im Endlichen ist auf diese Weise immer ein Wahrscheinlichkeitsmaß auf  $\vartheta$  gegeben, da das Wahrscheinlichkeitsmaß auf ganz  $\mathfrak{P}(\Omega)$  definiert ist. Ist  $|\Omega|$  hingegen überabzählbar unendlich, ist dies im Allgemeinen nicht mehr möglich. Falls es nämlich ein Ereignis  $A \subseteq \vartheta$  gibt, sodass  $X^{-1}(A)$  kein Ereignis in  $\Omega$  ist, ist  $P(X^{-1}(A))$  nicht definiert. Die Frage ist also, welche  $\sigma$ -Algebra man auf  $\vartheta$  verwendet.

**Definition 2.26.** Seien  $\Omega, \vartheta$  Mengen,  $\mathcal{A}$  eine  $\sigma$ -Algebra über  $\Omega$ ,  $\mathcal{B}$  eine  $\sigma$ -Algebra über  $\vartheta$  und  $X : \Omega \rightarrow \vartheta$  eine Abbildung.  $X$  heißt *messbar* in Bezug auf  $(\mathcal{A}, \mathcal{B})$ , falls

$$X^{-1}(B) \in \mathcal{A}$$

für alle  $B \in \mathcal{B}$  erfüllt ist. Man sagt auch  $X$  ist  $(\mathcal{A}, \mathcal{B})$ -messbar.

- Definition 2.27.** (a) Ein Paar  $(\Omega, \mathcal{A})$  bestehend aus einer Menge  $\Omega$  und einer  $\sigma$ -Algebra  $\mathcal{A}$  über  $\Omega$  nennt man *Messraum*.  
 (b) Ein Tripel  $(\Omega, \mathcal{A}, P)$  bestehend aus einer Menge  $\Omega$ , einer  $\sigma$ -Algebra  $\mathcal{A}$  über  $\Omega$  und einem Wahrscheinlichkeitsmaß  $P : \mathcal{A} \rightarrow \mathbb{R}$  nennt man *Wahrscheinlichkeitsraum*.

**Definition 2.28.** Eine  $(\mathcal{A}, \mathcal{B})$ -messbare Abbildung  $X : \Omega \rightarrow \vartheta$  für einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  und einen Messraum  $(\vartheta, \mathcal{B})$  nennt man *Zufallsgröße* oder *Zufallsvariable*.

**Bemerkung 2.29.** Mit dem Begriff der messbaren Abbildung lässt sich das Problem aus 2.25 beheben, denn mittels

$$\tilde{P}(A) = P(X^{-1}(A))$$

ist auf dem Messraum  $(\vartheta, \mathcal{B})$  durch die Zufallsvariable  $X : \Omega \rightarrow \vartheta$  auf natürliche Art ein Wahrscheinlichkeitsmaß definiert.

*Beweis.* Wir überprüfen dazu die Axiome von Kolmogoroff:

- (i) Da für alle  $\omega \in \Omega$  gilt, dass  $X(\omega) \in \vartheta$ , bedeutet dies nach Definition  $X^{-1}(\vartheta) = \Omega$ .  
 Also:

$$\tilde{P}(\vartheta) = P(X^{-1}(\vartheta)) = P(\Omega) = 1.$$

- (ii) Es ist  $\tilde{P}(B) = P(X^{-1}(B)) \geq 0$ , da es sich bei  $P$  um ein Wahrscheinlichkeitsmaß handelt.  
 (iii) Sei  $\{A_n\}_{n \in \mathbb{N}}$  eine Folge disjunkter Ereignisse in  $\mathcal{B}$ , dann ist

$$\tilde{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P\left(X^{-1}\left(\bigcup_{n \in \mathbb{N}} A_n\right)\right) = P\left(\bigcup_{n \in \mathbb{N}} X^{-1}(A_n)\right) = \sum_{n \in \mathbb{N}} P(X^{-1}(A_n)) = \sum_{n \in \mathbb{N}} \tilde{P}(A_n).$$

Dabei sind die  $X^{-1}(A_n)$  paarweise disjunkt, da jedes Element aus  $\Omega$  nur ein Bildelement haben kann. Gäbe es nämlich ein  $\omega \in X^{-1}(A_n) \cap X^{-1}(A_m)$  für gewisse  $n \neq m$ , so läge nach Definition des Urbilds  $X(\omega)$  sowohl in  $A_n$  als auch  $A_m$ . Diese sind nach Voraussetzung jedoch disjunkt.

- (iv) Auch der Satz vom Gegenereignis gilt wie gewohnt:

$$\tilde{P}(B^C) = P([X^{-1}(B^C)]) = P([X^{-1}(B)]^C) = 1 - P(X^{-1}(B^C))$$

□

In diesem Beweis kam unter (iii) und (iv) bereits folgendes Lemma zur Verwendung.

**Lemma 2.30** (Rechenregeln für Urbildmengen). Seien  $\Omega, \vartheta$  Mengen und  $X : \Omega \rightarrow \vartheta$  eine Abbildung. Sind  $A, B \subseteq \vartheta$  Teilmengen und  $\{A_n\}_{n \in \mathbb{N}}$  eine Folge von Teilmengen von  $\vartheta$ , so gilt

- (a)  $X^{-1}(A^C) = (X^{-1}(A))^C$   
 (b)  $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$   
 (c)  $X^{-1}(\bigcup_{n \in \mathbb{N}} A_n) = \bigcup_{n \in \mathbb{N}} X^{-1}(A_n)$

*Beweis.* (a) Es ist  $\omega \in X^{-1}(A^C)$  nach Definition genau dann, wenn  $X(\omega) \in A^C = \Omega \setminus A$ , also insbesondere  $X(\omega) \notin A$ . Daraus folgt gerade  $\omega \notin X^{-1}(A)$  bzw.  $\omega \in (X^{-1}(A))^C$ .

(b) folgt sofort aus (c).

(c) Sei  $\omega \in X^{-1}(\bigcup_{i \in \mathbb{N}} A_i)$ . Es ist dann  $X(\omega) \in \bigcup_{i \in \mathbb{N}} A_i$ , d.h. es gibt mindestens ein  $n$ , sodass  $X(\omega) \in A_n$ . Also ist  $\omega \in X^{-1}(A_n) \subseteq \bigcup_{i \in \mathbb{N}} X^{-1}(A_i)$ . □

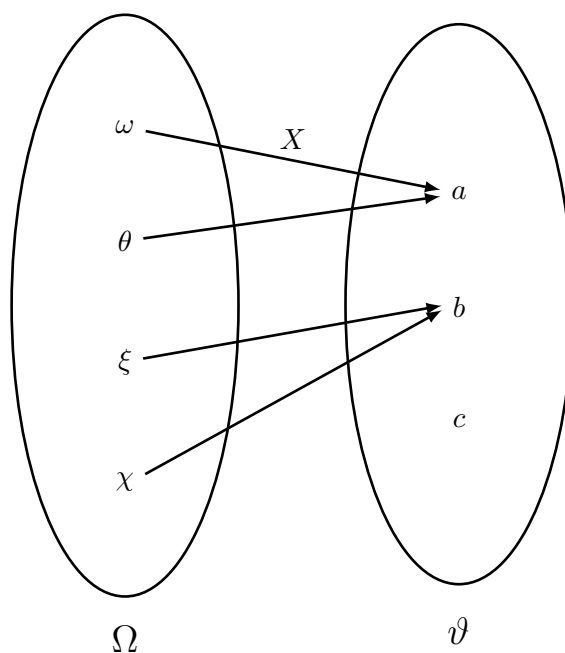


Abbildung 2.1: zur Veranschaulichung des Beweises von 2.30 ist hier eine Abbildung  $X$  zwischen zwei Mengen  $\Omega$  und  $\vartheta$  schematisch dargestellt, die weder injektiv noch surjektiv ist

**Definition 2.31.** Sei  $\mathcal{B}$  die Borel'sche  $\sigma$ -Algebra über  $\mathbb{R}$ . Eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt *reelle Zufallsgröße*.

**Satz 2.32.** Seien  $\Omega, \vartheta$  Mengen,  $\mathcal{A}$  eine  $\sigma$ -Algebra über  $\Omega$ ,  $\mathcal{B}$  eine  $\sigma$ -Algebra über  $\vartheta$  und  $E \subseteq \mathfrak{P}(\vartheta)$  ein Erzeugendensystem von  $\mathcal{B}$ . Eine Abbildung  $X : \Omega \rightarrow \vartheta$  ist genau dann  $(\mathcal{A}, \mathcal{B})$ -messbar, falls

$$X^{-1}(E) \subseteq \mathcal{A}$$

gilt.

In Proposition 2.17 haben wir festgestellt, dass die Intervalle der Form  $] -\infty, a[$  mit  $a \in \mathbb{R}$  ein Erzeugendensystem der Borelmenge über  $\mathbb{R}$  bilden. Wir erhalten daher aus Satz 2.32 die wichtige Folgerung:

**Folgerung 2.33.** Sei  $\Omega$  eine Menge,  $\mathcal{A}$  eine  $\sigma$ -Algebra über  $\Omega$  und  $\mathcal{B}$  die Borel'sche  $\sigma$ -Algebra über  $\mathbb{R}$ . Eine Abbildung  $X : \Omega \rightarrow \mathcal{B}$  ist genau dann  $(\mathcal{A}, \mathcal{B})$ -messbar, falls

$$X^{-1}(] - \infty, a]) \in \mathcal{A}$$

für alle  $a \in \mathbb{R}$  erfüllt ist.

Der Beweis von Satz 2.32 lässt sich auf folgenden Hilfssatz zurückführen:

**Lemma 2.34.** Seien  $\Omega, \vartheta$  Mengen,  $\mathcal{A}$  eine  $\sigma$ -Algebra über  $\Omega$ ,  $\mathcal{B}$  eine  $\sigma$ -Algebra über  $\vartheta$ . Dann sind folgende Mengen ebenfalls  $\sigma$ -Algebren:

- (i)  $X^{-1}(\mathcal{B}) = \{X^{-1}(B) \mid B \in \mathcal{B}\}$  über  $\Omega$ ,
- (ii)  $X_*(\mathcal{A}) = \{B \subseteq \vartheta \mid X^{-1}(B) \in \mathcal{A}\}$  über  $\vartheta$ .

*Beweis.* Wir beschränken uns auf den Beweis, dass (i) eine  $\sigma$ -Algebra ist. Der Nachweis für (ii) läuft analog. Wir überprüfen die definierenden Eigenschaften:

- (a)  $\Omega \in X^{-1}(\mathcal{B})$ : Wähle  $B = \vartheta$ , dann ist  $X^{-1}(\vartheta) = \Omega$ .
- (b) Sei  $A \in X^{-1}(\mathcal{B})$ , dann ist zu zeigen, dass  $A^C \in X^{-1}(\mathcal{B})$  gilt. Es gibt nun ein  $B \in \mathcal{B}$  mit  $A = X^{-1}(B)$ . Also ist auch  $X^{-1}(B^C) \in X^{-1}(\mathcal{B})$ . Aus der Rechenregeln 2.30 (a) folgt dann  $A^C = (X^{-1}(B))^C \in X^{-1}(\mathcal{B})$  wie gewünscht.
- (c) Sei  $\{A_i\}_{i \in I}$  eine Folge in  $X^{-1}(\mathcal{B})$ . Nach Definition gibt es es dann für jedes Folgenglied  $A_i$  ein  $B_i$  mit  $X^{-1}(B_i) = A_i$ . Auf diese Weise erhalten wir eine Folge  $\{B_i\}_{i \in I}$  in  $\mathcal{B}$ . Da  $\mathcal{B}$  eine  $\sigma$ -Algebra ist, liegt auch  $\bigcup_{i \in I} B_i \in \mathcal{B}$  und es folgt  $X^{-1}(\bigcup_{i \in \mathbb{N}} B_i) \in X^{-1}(\mathcal{B})$ . Rechenregel 2.30 (c) liefert dann  $\bigcup_{i \in \mathbb{N}} X^{-1}(B_i) \in X^{-1}(\mathcal{B})$ . Dies bedeutet gerade  $\bigcup_{i \in \mathbb{N}} A_i \in X^{-1}(\mathcal{B})$ . □

**Bemerkung 2.35.** Im Allgemeinen ist

$$X(\mathcal{A}) = \{X(A) \mid A \in \mathcal{A}\} \subseteq \mathfrak{P}(\vartheta)$$

keine  $\sigma$ -Algebra. Beispielsweise kann bei einer nicht-surjektiven Abbildung  $X$  der Fall  $\vartheta \notin X(\mathcal{B})$  auftreten, was laut Definition einer  $\sigma$ -Algebra unmöglich ist.

*Beweis von 2.32.* Sei  $E \subseteq \mathfrak{P}(\vartheta)$  ein Erzeugendensystem von  $\mathcal{B}$  mit  $X^{-1}(E) \in \mathcal{A}$ . Zu zeigen ist nun, dass  $X^{-1}(\mathcal{B}) \subseteq \mathcal{A}$ . Betrachte  $X_*(\mathcal{A})$ . Dies ist nach Lemma 2.34 eine  $\sigma$ -Algebra. Wir zeigen nun  $\mathcal{B} \subseteq X_*(\mathcal{A})$ .

Nach Voraussetzung liegen alle Elemente von  $E$  auch in  $X_*(\mathcal{A})$ .  $X_*(\mathcal{A})$  ist also eine  $\sigma$ -Algebra, die alle Elemente des Erzeugendensystems  $E$  von  $\mathcal{B}$  enthält.  $\mathcal{B}$  ist die kleinste  $\sigma$ -Algebra, die alle Elemente aus  $E$  enthält. Daher ist  $\mathcal{B} \subseteq X_*(\mathcal{A})$ . Es folgt  $X^{-1}(\mathcal{B}) \subseteq \mathcal{A}$ , d.h.  $X$  ist  $(\mathcal{A}, \mathcal{B})$ -messbar. Die Umkehrung ist klar. □

**Bemerkung 2.36.** Wegen Satz 2.32 genügt es also, die Eigenschaft der Messbarkeit für ein beliebiges Erzeugendensystem zu überprüfen, d.h. eine Abbildung  $X : \Omega \rightarrow \vartheta$  ist bereits eine  $(\mathcal{A}, \mathcal{B})$ -Zufallsgröße, wenn  $X^{-1}(B) \in \mathcal{A}$  für alle  $B \in E$  erfüllt ist. Dabei sei  $E$  ein Erzeugendensystem von  $\mathcal{B}$ .

Meist wird  $\vartheta = \mathbb{R}$  und  $\mathcal{B}$  die Borelmenge über  $\mathbb{R}$  sein. Für diese kennen wir verschiedene Erzeugendensysteme und dürfen uns je nach Situation eines aussuchen.

**Satz 2.37.** Sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  stetig und  $\mathcal{B}$  die Borelmengen über  $\mathbb{R}$ . Dann ist  $f$  messbar bezüglich  $(\mathcal{B}, \mathcal{B})$ .

*Beweis.* Wir verwenden das Erzeugendensystem  $E = \{A \subseteq \mathbb{R} \mid A \text{ offen}\}$ . Zu zeigen ist, dass  $f^{-1}(B) \in \mathcal{B}$  für alle  $B \in E$ .

Da  $f$  stetig ist, ist, wie wir aus der Analysis I wissen, das Urbild  $f^{-1}(B)$  in jedem Fall offen, und damit laut Definition in  $\mathcal{B}$  enthalten.  $\square$

**Satz 2.38.** Seien  $\Omega, \vartheta, \psi$  Mengen und  $\mathcal{A}, \mathcal{B}, \mathcal{C}$   $\sigma$ -Algebren über  $\Omega, \vartheta$  bzw.  $\psi$ . Sei weiter  $X : \Omega \rightarrow \vartheta$  messbar bezüglich  $(\mathcal{A}, \mathcal{B})$  und  $Y : \vartheta \rightarrow \psi$  messbar bezüglich  $(\mathcal{B}, \mathcal{C})$ . Dann ist  $Z : \Omega \rightarrow \psi$ , gegeben durch  $Z(\omega) = Y(X(\omega))$ , messbar bezüglich  $(\mathcal{A}, \mathcal{C})$ .

*Beweis.* Zu zeigen ist  $Z^{-1}(C) \in \mathcal{A}$  für alle  $C \in \mathcal{C}$ .

Da  $Y$   $(\mathcal{B}, \mathcal{C})$ -messbar ist, ist für alle  $C \in \mathcal{C}$  auch  $Y^{-1}(C) \in \mathcal{B}$  erfüllt.

Ebenso ist  $X^{-1}(B) \in \mathcal{A}$  für alle  $B \in \mathcal{B}$ , da  $X$  messbar ist. Wenden wir dies auf  $B = Y^{-1}(C)$  an, so erhalten wir insgesamt

$$Z^{-1}(C) = X^{-1}(Y^{-1}(C)) \in \mathcal{A}$$

für alle  $C \in \mathcal{C}$ .  $\square$

**Satz 2.39.** Sei  $\Omega$  eine Grundmenge,  $\mathcal{A}$  eine  $\sigma$ -Algebra über  $\Omega$  und  $X, Y : \Omega \rightarrow \mathbb{R}$  reelle Zufallsgrößen. Dann gilt

- (a)  $X + Y$  ist eine reelle Zufallsgröße,
- (b)  $f(X)$  ist eine reelle Zufallsgröße für jede stetige Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,
- (c)  $\alpha X + \beta Y$  ist eine reelle Zufallsgröße mit  $\alpha, \beta \in \mathbb{R}$

*Beweis.* (a) Sei  $Z = X + Y$ . Wählen wir als Erzeugendensystem  $E$  die Menge aller offenen Intervalle  $] - \infty; a[$  für  $a \in \mathbb{R}$ , so ist zu zeigen:  $Z^{-1}(B) \in \mathcal{A}$  für alle  $B \in E$ . Sei daher  $B \in E$  beliebig vorgegeben. Es gilt

$$Z^{-1}(B) = \{\omega \in \Omega \mid X(\omega) + Y(\omega) \in B\}.$$

Es ist  $B = ] - \infty; a[$  für ein  $a \in \mathbb{R}$ , also lässt sich diese Urbildmenge auch als

$$Z^{-1}(B) = \{\omega \in \Omega \mid X(\omega) + Y(\omega) < a\}$$

schreiben. Wir überprüfen nun folgende Mengengleichung

$$Z^{-1}(B) = \bigcup_{b \in \mathbb{Q}} \{\omega \in \Omega \mid X(\omega) < b, Y(\omega) < a - b\}$$

“ $\supseteq$ ”: Sei  $\omega$  aus der Menge auf der rechten Seite. Es gibt dann ein  $b \in \mathbb{Q}$ , sodass  $X(\omega) < b$  und  $Y(\omega) < a - b$  erfüllt ist. Addition dieser Ungleichungen liefert

$$X(\omega) + Y(\omega) < a - b + b = a,$$

also  $\omega \in Z^{-1}(B)$ .

“ $\subseteq$ ”: Sei  $\omega \in Z^{-1}(B)$ , also  $X(\omega) + Y(\omega) < a$ . Wir definieren  $\delta := a - (X(\omega) + Y(\omega))$  und wählen ein  $b \in \mathbb{Q}$  mit

$$X(\omega) < b < X(\omega) + \delta,$$

sodass also  $b < X(\omega) + \delta = X(\omega) + a - (X(\omega) + Y(\omega)) = a - Y(\omega)$ . Durch Umstellen der Ungleichung erhält man

$$Y(\omega) < a - b,$$

d.h.  $\omega$  liegt in der Menge auf der rechten Seite.

An dieser Darstellung sieht man nun leicht, dass  $Z^{-1}(B)$  tatsächlich in  $\mathcal{A}$  liegt, denn es ist

$$\begin{aligned} Z^{-1}(B) &= \bigcup_{b \in \mathbb{Q}} (\{\omega \in \Omega \mid X(\omega) < b\} \cap \{\omega \in \Omega \mid Y(\omega) < a - b\}) \\ &= \bigcup_{b \in \mathbb{Q}} (X^{-1}(] - \infty; b[) \cap Y^{-1}(] - \infty; a - b[)). \end{aligned}$$

Alle auftretenden Mengen liegen in  $\mathcal{A}$ , damit ist  $Z^{-1}(B)$  als abzählbare Vereinigung von Elementen aus  $\mathcal{A}$  wieder in  $\mathcal{A}$  enthalten.

(b) Diese Aussage folgt direkt aus den letzten beiden Sätzen 2.38 und 2.39 (setze dazu  $\vartheta = \psi = \mathbb{R}$  und  $\mathcal{C} = \mathcal{B}$  als Borelmenge sowie  $Y = f$ ).

(c) Die Multiplikation einer Zufallsgröße mit einer reellen Zahlen entspricht der Verkettung mit einer stetigen Funktion. Daher folgt die Aussage aus (a) und (b).  $\square$

## 2.4 Verallgemeinerung des Unabhängigkeitsbegriffs

Im Falle eines endlichen Ergebnisraums  $\Omega$  haben wir in Definition 1.20 festgelegt, dass zwei Ereignisse  $A, B \in \Omega$  genau dann unabhängig bzgl. eines Wahrscheinlichkeitsmaßes  $P$  voneinander sind, falls

$$P_B(A) = P(A)$$

gilt. Diese Definition nimmt keinen expliziten Bezug auf die Endlichkeit von  $\Omega$ , sodass wir sie wortwörtlich auf allgemeine (möglicherweise überabzählbar unendliche) Ereignisräume übertragen können. Im Beweis von Satz 1.21 haben wir gesehen, dass die Unabhängigkeit von  $A$  und  $B$  äquivalent zu

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

ist. Auch hier haben wir keinen Gebrauch von der Endlichkeit von  $\Omega$  gemacht, sodass diese Aussage weiterhin unverändert gilt. Die ursprüngliche Definition ist zwar anschaulicher, die zweite kann jedoch auch im Fall  $P(B) = 0$  verwendet werden, sodass wir im Allgemeinen Fall an diese anknüpfen werden.

**Definition 2.40.** Sei  $\Omega$  eine Menge,  $\mathcal{A}$  eine  $\sigma$ -Algebra und  $\{A_i\}_{i \in I}$  eine Folge von Ereignissen in  $\mathcal{A}$ .

- (a) Die  $\{A_i\}_{i \in I}$  heißen *paarweise unabhängig* genau dann, wenn für jedes  $j, k \in I$  mit  $k \neq j$  das Ereignis  $A_j$  unabhängig vom Ereignis  $A_k$  ist, d.h.

$$P(A_j \cap A_k) = P(A_k)P(A_j).$$

- (b) Die  $\{A_i\}_{i \in I}$  heißen *unabhängig*, wenn für jede endliche Teilmenge  $J \subseteq I$  gilt:

$$P\left(\bigcap_{k \in J} A_k\right) = \prod_{k \in J} P(A_k).$$

**Beispiel 2.41.** Sei  $\Omega = \{(1, 1), (-1, -1), (1, -1), (-1, 1)\}$ . Dies könnte beispielsweise einen zweifachen Wurf einer Münze darstellen, deren Seiten mit 1 bzw. -1 beschriftet sind. Weiter nehmen wir an, dass die Laplace-Annahme gilt und betrachten folgende Ereignisse:

$$\begin{array}{ll} A: \text{“erste Münze zeigt eine 1”}, & \text{also } A = \{(1, 1), (1, -1)\} \\ B: \text{“zweite Münze zeigt 1”}, & \text{also } B = \{(1, 1), (-1, 1)\} \\ C: \text{“Das Produkt der Würfe ist gleich 1”}, & \text{also } C = \{(1, 1), (-1, -1)\} \end{array}$$

$A, B, C$  sind *nicht* unabhängig. Es gilt nämlich

$$P(A) \cdot P(B) \cdot P(C) = \left(\frac{1}{2}\right)^3 = \frac{1}{8} \neq P(A \cap B \cap C) = P(\{(1, 1)\}) = \frac{1}{4}$$

Allerdings sind beispielsweise die beiden Ereignisse  $A$  und  $B$  unabhängig, denn es gilt

$$\frac{1}{4} = P(A) \cdot P(C) = P(A \cap C) = P(\{(1, 1)\}).$$

In analoger Weise kann man auch die anderen Kombinationen nachrechnen, sodass  $A, B, C$  zumindest *paarweise* unabhängig sind.

**Definition 2.42.** Seien  $X, Y$  zwei reelle Zufallsvariablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ .  $X$  und  $Y$  heißen *unabhängig* voneinander genau dann, wenn die Mengen

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$$

$$Y^{-1}(B) = \{\omega \in \Omega \mid Y(\omega) \in B\}$$

für jedes Paar  $A, B \in \mathcal{B}$  unabhängig sind.

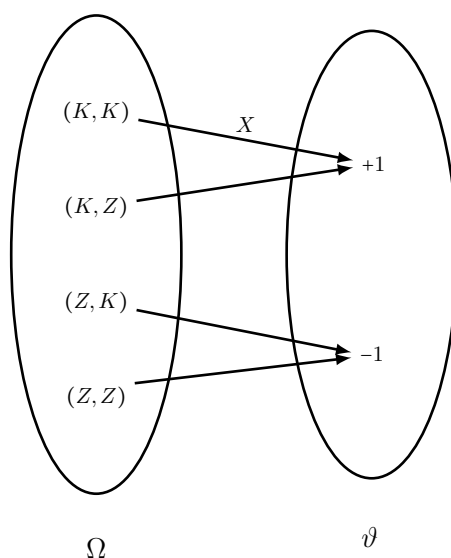


Abbildung 2.2: Illustration der Zufallsvariable  $X$ .

**Beispiel 2.43.** Wir betrachten den zweimaligen Wurf einer Laplace-Münze, d.h.

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

Wir definieren nun reelle Zufallsgrößen  $X, Y, Z$  mittels

$$X(\omega) = \begin{cases} 1 & \text{falls erster Wurf } K \\ -1 & \text{sonst.} \end{cases} \quad Y(\omega) = \begin{cases} 1 & \text{falls zweiter Wurf } K \\ -1 & \text{sonst.} \end{cases}$$

$$\text{und } Z(\omega) = X(\omega) \cdot Y(\omega).$$

Zunächst zeigen wir, dass  $X$  und  $Y$  unabhängig sind. Dazu untersuchen wir, welche Möglichkeiten sich für  $X^{-1}(A)$  ergeben können. Da die Zufallsgröße nur die Werte  $\pm 1$  annimmt, reicht es, vier verschiedene Fälle für  $A$  zu unterscheiden:

- (i) Weder  $+1$  noch  $-1$  in  $A$ : Dann ist  $X^{-1}(A) = X^{-1}(\emptyset) = \emptyset$
- (ii) nur  $+1$  in  $A$ : Dann ist  $X^{-1}(A) = \{(K, K), (K, Z)\}$  (vgl. Abb. 2.2).
- (iii) nur  $-1$  in  $A$ : Dann ist  $X^{-1}(A) = \{(Z, K), (Z, Z)\}$  (vgl. Abb. 2.2).
- (iv) beide in  $A$ : Dann ist  $X^{-1}(A) = X^{-1}(\mathbb{R}) = \Omega$

Durch ähnliche Überlegungen ergibt sich für  $Y^{-1}(B)$  dann

$$Y^{-1}(B) \in \{\Omega, \emptyset, \{(K, K), (Z, K)\}, \{(K, Z), (Z, Z)\}\}.$$

Nun muss für beliebige Kombinationen die Gleichung aus Definition 2.40 überprüft werden. Einige Fälle lassen sich dabei schnell abhandeln:

- (a) Ist eine der Mengen  $X^{-1}(A)$  oder  $Y^{-1}(B)$  die leere Menge, so steht auf beiden Seiten der Gleichung Null, wie man schnell nachrechnet. Die Ereignisse sind damit unabhängig.



(b) Ist eine der Mengen ganz  $\Omega$  (wir nehmen o.B.d.A an, dass  $X^{-1}(A) = \Omega$ ), so ist

$$\begin{aligned} P(X^{-1}(A) \cap Y^{-1}(B)) &= P(X^{-1}(A)) \cdot P(Y^{-1}(B)) \\ \Leftrightarrow P(\Omega \cap Y^{-1}(B)) &= P(\Omega)P(Y^{-1}(B)) \Leftrightarrow P(Y^{-1}(B)) = 1 \cdot P(Y^{-1}(B)) \end{aligned}$$

Da die letzte Gleichung offensichtlich erfüllt ist, muss auch die erste gelten. Also sind auch für diesen Fall die Ereignisse immer unabhängig.

Die restlichen Kombinationen müssen "zu Fuß" abgearbeitet werden, was wir an einem Beispiel verdeutlichen:

$$P(\{(K, K), (K, Z)\} \cap \{(K, K), (Z, K)\}) = P(\{(K, K)\}) = \frac{1}{4}$$

$$P(\{(K, K), (K, Z)\}) \cdot P(\{(K, K), (Z, K)\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

**Definition 2.44.** Seien  $\{X_i\}_{i \in I}$  Zufallsgrößen. Diese nennt man

- (a) *paarweise unabhängig*, falls  $X_i$  und  $X_j$  für  $i \neq j$  unabhängig sind.
- (b) *unabhängig*, falls die Ereignisse  $A_i = X_i^{-1}(B_i)$  für jede mögliche Wahl von  $\{B_i\}_{i \in I} \subseteq \mathcal{B}$  unabhängig sind.

**Beispiel 2.45.** Für das obige Beispiel 2.43 gilt die paarweise Unabhängigkeit für  $X, Y, Z$ , jedoch nicht die Unabhängigkeit. Um die Abhängigkeit zu zeigen, wähle

$$B_1 = B_2 = B_3 = \{1\} \quad \text{sodass} \quad X^{-1}(B_1) = \{(K, K), (K, Z)\}$$

$$Y^{-1}(B_2) = \{(K, K), (Z, K)\}$$

$$Z^{-1}(B_3) = \{(K, K), (Z, Z)\}$$

Diese drei Ereignisse sind abhängig, da das Produkt die Wahrscheinlichkeit  $\frac{1}{8}$  ergibt, es jedoch im Zufallsexperiment gar kein Ereignis mit dieser Wahrscheinlichkeit gibt.

## 2.5 Die Verteilungsfunktion

**Definition 2.46.** Sei  $X$  eine reelle Zufallsvariable auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ . Die Funktion

$$V_X : \mathbb{R} \rightarrow [0, 1], \quad V_X(t) = P(X^{-1}([-\infty, t]))$$

nennt man *Verteilungsfunktion* von  $X$ .

Die Wahrscheinlichkeit  $P(X^{-1}([-\infty, t]))$  wird oft mit  $P(X \leq t)$  abgekürzt. Allgemein meint man mit  $P(\text{Aussage})$  die Wahrscheinlichkeit für die Menge aller  $\omega \in \Omega$ , für die die Aussage erfüllt ist:

$$P(X + Y = 3) = P(\{\omega \in \Omega \mid X(\omega) + Y(\omega) = 3\})$$

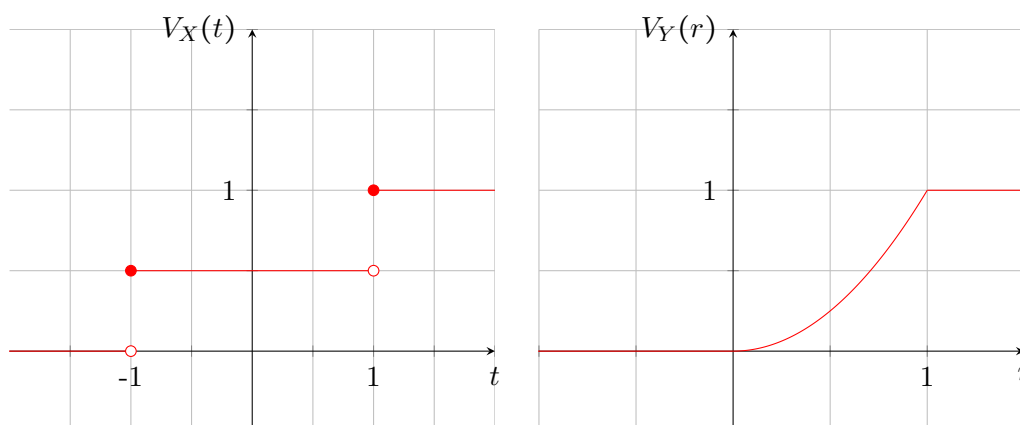


Abbildung 2.3: Graphen der Verteilungsfunktionen aus Beispiel 2.48

**Bemerkung 2.47.** Für eine gegebene Verteilungsfunktion ist das Wahrscheinlichkeitsmaß auf  $\mathbb{R}$  eindeutig bestimmt. Ist nämlich  $V_X$  gegeben, so lässt sich daraus  $P(X^{-1}(B))$  für alle  $B \in \mathcal{B}$  bestimmen. Der Beweis dieser Aussage bedarf umfangreicher Vorarbeit, weshalb wir darauf verzichten.

**Beispiele 2.48.** (a) Wir betrachten abermals den Wurf einer Laplace-Münze mit der Zufallsvariable  $X$ , gegeben durch

$$X(\{K\}) = 1 \quad \text{und} \quad X(\{Z\}) = -1.$$

Die zugehörige Verteilungsfunktion hat dann zwei Unstetigkeitsstellen bei 1 und  $-1$  (vgl. Abb. 2.3).

(b) (*Dartwurf*) Wähle zufällig einen Punkt  $\omega$  auf der Einheitskreisscheibe  $B_1(0, 0)$ , wobei die Laplace-Annahme bzgl. der Fläche gelte, d.h. die Wahrscheinlichkeit, Segmente gleicher Fläche auszuwählen, sei gleich groß. Definiere nun eine Zufallsgröße  $Y$  durch den Abstand des Punktes vom Ursprung, d.h.

$$Y(\omega) = \|\omega\|.$$

Die zugehörige Verteilungsfunktion  $V_Y$  kann man sich anschaulich so vorstellen, dass diese die Wahrscheinlichkeit angibt, mit der gewählte Punkt  $\omega$  Abstand  $\leq r$  vom Kreismittelpunkt hat. Es handelt sich dabei also um die Wahrscheinlichkeit,  $\omega$  bereits in einer kleineren Kreisscheibe  $B_r(0, 0)$  mit Radius  $r$  zu finden. Wählen wir  $r \geq 1$ , so ist es also ein sicheres Ereignis, den Punkt  $\omega$  zu erwischen. Die Wahrscheinlichkeit den Nullpunkt zu treffen, ist dagegen 0. Allgemein gilt für  $r \in ]0; 1[$

$$P(Y \leq r) = P(\{\omega \in B_r(0, 0)\}) = \frac{\text{Fläche von } B_r(0, 0)}{\text{Fläche von } B_1(0, 0)} = \frac{r^2\pi}{1^2\pi} = r^2$$

Wir fassen nun einige grundlegende Eigenschaften von Verteilungsfunktionen zusammen und beweisen als Vorbereitung des nächsten Satzes den folgenden Hilfssatz.

**Lemma 2.49** (Stetigkeit des Wahrscheinlichkeitsmaßes). Seien  $\{A_n\}_{n \in \mathbb{N}}$  und  $\{B_n\}_{n \in \mathbb{N}}$  Folgen von Ereignissen einer  $\sigma$ -Algebra  $\mathcal{A}$  über einer Grundmenge  $\Omega$ . Dabei gelte für alle  $n \in \mathbb{N}$

$$A_n \subseteq A_{n+1} \quad \text{und} \quad B_n \supseteq B_{n+1}$$

Weiterhin seien  $A = \bigcup_{n \in \mathbb{N}} A_n$  und  $B = \bigcap_{n \in \mathbb{N}} B_n$ , dann gilt

$$\lim_{n \rightarrow \infty} P(B_n) = P(B) \quad \text{und} \quad \lim_{n \rightarrow \infty} P(A_n) = P(A)$$

für jedes Wahrscheinlichkeitsmaß  $P$  auf  $\mathcal{A}$ .

*Beweis.* Wir zeigen zunächst die Aussage für die Vereinigung. Dazu definieren wir eine neue Folge  $\{C_n\}_{n \in \mathbb{N}}$  durch

$$C_1 = A_1, \quad C_n = A_n \setminus A_{n-1} \quad \text{für } n \geq 2.$$

Es ist dann  $A_n = \bigcup_{j=1}^n C_j$ , wie man unmittelbar überprüft. Außerdem besteht diese Folge aus paarweise disjunkte Mengen. Deshalb können wir das dritte Axiom von Kolmogorow (für endliche Vereinigungen) anwenden und erhalten

$$P(A_n) = P\left(\bigcup_{j=1}^n C_j\right) \stackrel{2.1(c)}{=} \sum_{j=1}^n P(C_j)$$

sowie mit dem Axiom von Kolmogorow für abzählbar unendliche Vereinigungen:

$$P(A) = P\left(\bigcup_{j=1}^{\infty} C_j\right) \stackrel{2.1(c)}{=} \sum_{j=1}^{\infty} P(C_j)$$

Den Grenzwert einer unendlichen Summe hatten wir in der Analysis I als Grenzwert der Folge der Partialsummen definiert, also ist

$$P(A) = \sum_{j=1}^{\infty} P(C_j) = \lim_{n \rightarrow \infty} \sum_{j=1}^n P(C_j) = \lim_{n \rightarrow \infty} P(A_n)$$

Die zweite Aussage beweisen wir mithilfe der Regeln von DeMorgan. Da  $B_n \supseteq B_{n+1}$  ist  $B_n^C \subseteq B_{n+1}^C$ . Aufgrund der eben bewiesenen Aussage erhalten wir

$$\lim_{n \rightarrow \infty} P(B_n^C) = P\left(\bigcup_{n \in \mathbb{N}} B_n^C\right) = P\left(\left(\bigcap_{n \in \mathbb{N}} B_n\right)^C\right) = P(B^C)$$

Der Satz vom Gegenereignis liefert dann

$$\lim_{n \rightarrow \infty} 1 - P(B_n) = 1 - P(B) \Rightarrow \lim_{n \rightarrow \infty} P(B_n) = P(B).$$

□

**Satz 2.50.** Für jede Verteilungsfunktion  $V_X$  einer reellen Zufallsgröße  $X$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  gilt:

- (a)  $\lim_{t \rightarrow -\infty} V_X(t) = 0$  und  $\lim_{t \rightarrow +\infty} V_X(t) = 1$ .
- (b)  $V_X(t)$  ist monoton steigend.

- (c)  $V_X(t)$  ist rechtsseitig stetig, d.h.  $\lim_{n \rightarrow \infty} V_X(t + h_n) = V_X(t)$  für jede Folge  $(h_n)_{n \in \mathbb{N}}$  mit  $h_n \geq 0$  für alle  $n \in \mathbb{N}$  und  $\lim_{n \rightarrow \infty} h_n = 0$ .
- (d)  $V_X(t)$  hat höchstens abzählbar viele Sprungstellen.

*Beweis.* (a) Es gilt für  $n \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} V_X(n) = \lim_{n \rightarrow \infty} P(X \leq n)$$

Sei nun  $A_n = \{\omega \in \Omega \mid X(\omega) \leq n\}$ . Aufgrund der Definition gilt  $A_n \subseteq A_{n+1}$ , sodass man also unter Verwendung von Lemma 2.49 erhält

$$\lim_{n \rightarrow \infty} V_X(n) = P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$$

Für den anderen Grenzwert betrachten wir

$$\lim_{t \rightarrow -\infty} V_X(t) = \lim_{n \rightarrow \infty} V_X(-n) = \lim_{n \rightarrow \infty} P(X \leq -n)$$

Sei  $B_n = \{\omega \in \Omega \mid X(\omega) \leq -n\}$ , wobei hier gilt  $B_n \supseteq B_{n+1}$ . Wieder erhalten wir mit 2.49

$$\lim_{t \rightarrow -\infty} V_X(t) = P\left(\bigcap_{n \in \mathbb{N}} B_n\right) = P(\emptyset) = 0$$

- (b) Zu zeigen ist  $V_X(t) \geq V_X(s)$  für  $t, s \in \mathbb{R}$  mit  $t > s$ .

$$\begin{aligned} V_X(t) - V_X(s) &= P(X \leq t) - P(X \leq s) = \\ &= P(\{\omega \in \Omega \mid X(\omega) \leq s\} \cup \{\omega \in \Omega \mid s < X(\omega) \leq t\}) - P(X \leq s) = \\ &\stackrel{*}{=} P(\{\omega \in \Omega \mid X(\omega) \leq s\}) + P(\{\omega \in \Omega \mid s < X(\omega) \leq t\}) - P(X \leq s) = \\ &= P(X \leq s) + P(\{\omega \in \Omega \mid s < X(\omega) \leq t\}) - P(X \leq s) = \\ &= P(\{\omega \in \Omega \mid s < X(\omega) \leq t\}) \geq 0 \end{aligned}$$

An der Stelle (\*) ging Komogoroff 2.1 (c) ein, da die Mengen disjunkt sind.

- (c) Sei  $(h_n)_{n \in \mathbb{N}}$  eine solche Folge, o.B.d.A. nehmen wir an, dass diese Folge monoton fallend ist (ansonsten erhält man aus einer beliebigen Folge eine monoton fallende durch Umsortierung, denn dadurch ändert sich Grenzwert nicht). Es ist dann

$$\lim_{n \rightarrow \infty} V_X(t + h_n) = \lim_{n \rightarrow \infty} P(X \leq t + h_n)$$

Sei nun  $B_n = \{\omega \in \Omega \mid X(\omega) \leq t + h_n\}$ . Da die  $h_n$  aufgrund unserer Annahme monoton fallen, ist  $B_n \supseteq B_{n+1}$ . Wir verwenden wieder Lemma 2.49 und erhalten so

$$\lim_{n \rightarrow \infty} V_X(t + h_n) = \lim_{n \rightarrow \infty} P(B_n) = P\left(\bigcap_{n \in \mathbb{N}} B_n\right) = P(X \leq t) = V_X(t)$$

- (d) Sei  $M_n$  für  $n \in \mathbb{N}$  die Menge aller Sprungstellen mit Höhe größer gleich  $\frac{1}{n}$ . Die Menge aller Sprungstellen ist  $\bigcup_{n \in \mathbb{N}} M_n$ . Aufgrund der Monotonie und Teil

(a) dieses Satzes gibt es höchstens  $n$  Sprungstellen in  $M_n$  (sonst wäre die Gesamtsprunghöhe größer als  $\frac{1}{n} \cdot n = 1$  im Widerspruch zum Grenzwert aus a)). Damit ist die Vereinigung aller  $M_n$  als abzählbare Vereinigung endlicher Mengen abzählbar. □

**Beispiel 2.51.** Wir kombinieren einen Münz- und einen Dartwurf zu einem Zufallsexperiment. Letzterer werde wie in Beispiel 2.48) durch das zufällige Auswählen eines Punktes  $\omega$  auf der Einheitskreisscheibe  $B_1(0, 0)$  modelliert. Betrachtet wird nun die Zufallsgröße

$$X : \{K, Z\} \times B_1(0, 0) \rightarrow [0, 1], (M, \omega) \mapsto \begin{cases} \frac{1}{2} & \text{falls } M = K, \text{ d.h. es wird K geworfen} \\ |\omega| & \text{falls } M = Z, \text{ d.h. es wird Z geworfen} \end{cases}$$

Unser Ziel ist es, die Wahrscheinlichkeitsverteilung  $V_X(t) = P(X \leq t)$  anzugeben. Wir betrachten dazu einige Einzelfälle:

- (a) Da  $|\omega|$  keine negativen Werte annimmt und auch  $\frac{1}{2} > 0$  ist, ist  $X$  nicht-negativ und damit  $V_X(t) = 0$  für  $t < 0$ . Ebenso nimmt die Zufallsgröße nur Werte unter 1 an, also ist  $V_X(t) = 1$  für  $t \geq 1$ .
- (b) Betrachten wir ein  $t \in [0; \frac{1}{2}[$ . Dann muss im Wurf  $Z$  eingetreten sein (sonst wäre  $X = \frac{1}{2}$ ). Wir müssen also das Ereignis  $A$  untersuchen, wobei  $A$  die Ergebnisse beschreibt, bei denen zunächst Zahl geworfen wurde und dann ein Radius unter  $\frac{1}{2}$  getroffen wurde. Aufgrund der Unabhängigkeit dieser beiden Elemente (und mit der Formel aus 2.48 erhält man

$$P(A) = \frac{1}{2} \cdot t^2$$

- (c) Bei Annäherung an die  $t = \frac{1}{2}$  von links erhält man mit eben dieser Formel die Wahrscheinlichkeit  $P(X < \frac{1}{2}) = \frac{1}{2} \cdot (\frac{1}{2})^2$ .
- (d) Weiter ist  $P(X \leq \frac{1}{2}) = P(X < \frac{1}{2}) + P(X = \frac{1}{2}) = \frac{1}{8} + \frac{1}{2} = \frac{5}{8}$
- (e) Für den Fall  $t > \frac{1}{2}$  betrachten wir die disjunkte Zerlegung

$$\{(M, \omega) \mid X(M, \omega) \leq t\} = \{(M, \omega) \mid X(M, \omega) \leq \frac{1}{2}\} \cup \{(M, \omega) \mid \frac{1}{2} < X(M, \omega) \leq t\}$$

Den Wert der Verteilungsfunktion für die erste Menge auf der rechten Seite haben wir bereits bestimmt. Für die zweite Menge brauchen wir wieder nur Ergebnisse zu berücksichtigen, bei denen  $Z$  geworfen wurde. Daher ist

$$P(\{(M, \omega) \mid \frac{1}{2} < X(\omega) \leq t\}) = P(\{(M, \omega) \mid |\omega| \leq t\}) - P(\{(M, \omega) \mid |\omega| \leq \frac{1}{2}\}) = \frac{1}{2} \cdot t^2 - \frac{1}{2} \cdot (\frac{1}{2})^2$$

Also Insgesamt:

$$V_X(X \leq t) = \frac{5}{8} + \frac{1}{2}t^2 - \frac{1}{8} = \frac{1}{2}t^2 + \frac{1}{2}$$

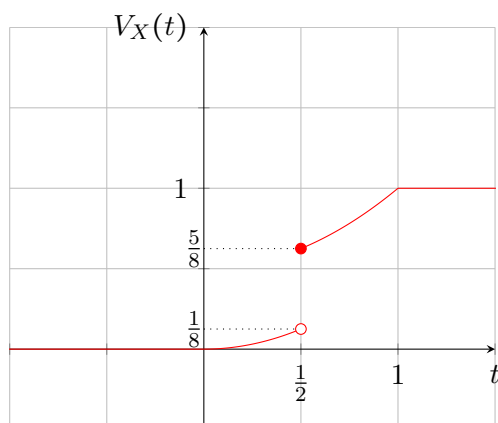


Abbildung 2.4: Graph der Verteilungsfunktion aus Beispiel 2.51

## 2.6 Spezielle Verteilungsfunktionen

Im Folgenden betrachten wir einige klassische Beispiele für stochastische Modelle, die häufig zur mathematischen Beschreibung von Zufallsexperimenten herangezogen werden. Diese werden zwar üblicherweise als Verteilungsfunktionen bezeichnet, geben jedoch jeweils die Wahrscheinlichkeit eines Elementarereignisses an, sodass es sich nicht um Verteilungsfunktionen im Sinne des letzten Abschnitts handelt.

### Binomialverteilung

**Definition 2.52** (Binomialverteilung). Für jedes  $p \in [0, 1]$  und  $n \in \mathbb{N}$  heißt die Abbildung

$$B_{n,p} : \{0, 1, \dots, n\} \rightarrow \mathbb{R}, \quad B_{n,p}(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

*Binomialverteilung* für  $n$  Stichproben mit Erfolgswahrscheinlichkeit  $p$ .

**Definition 2.53.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Eine Zufallsgröße  $X : \Omega \rightarrow \{0, 1, \dots, n\}$  heißt *binomialverteilt*, falls es ein  $p \in [0, 1]$  gibt, sodass

$$P(X = k) = B_{n,p}(k)$$

erfüllt ist.

Um diese Formel plausibel zu machen, betrachten wir einen  $n$ -fachen Münzwurf. Die Zufallsgröße  $X$  gebe an, wie oft “Kopf” gefallen ist. Wir interessieren uns nun für die Wahrscheinlichkeit, dass  $k$  mal “Kopf” auftritt, d.h.  $P(X = k)$ .

Im konkreten Fall, dass  $n = 6$  und  $k = 3$  ist, wäre ein günstiger Ausgang des Zufallsexperiments dann beispielsweise

$$(Z, K, K, Z, Z, K)$$

Ist  $p = \frac{1}{2}$ , so berechnet sich die Wahrscheinlichkeit dafür, genau dieses Ergebnis zu erhalten, zu

$$P(\{(Z, K, K, Z, Z, K)\}) = (1-p) \cdot p \cdot p \cdot (1-p) \cdot (1-p) \cdot p = p^k \cdot (1-p)^{n-k}$$

Es gibt jedoch noch andere Möglichkeiten, genau 3 Treffer zu landen, z.B.  $(Z, Z, Z, K, K, K)$ . Da die Wahrscheinlichkeit der einzelnen Ausgänge identisch ist, müssen wir nun die Anzahl aller günstigen Möglichkeiten bestimmen. Es handelt sich dabei um eine ähnliche Situation wie beim MISSISSIPPI-Problem, vgl. dazu 1.13 und 1.12 (d). Somit ergibt sich für die Anzahl der Permutationen  $\binom{n}{k} = \binom{6}{3}$  und als Wahrscheinlichkeit für  $k = 3$  Treffer

$$\binom{6}{3} \cdot p^3 \cdot (1-p)^3 = \binom{n}{k} \cdot p^k (1-p)^{n-k} = B_{n,p}(k)$$

Also ist die Anzahl von "Kopf" beim Münzwurf binomialverteilt.

Wir haben bereits in 2.29 gesehen, dass eine Zufallsgröße  $X$  auf der Bildmenge ein Wahrscheinlichkeitsmaß definiert. Insbesondere gilt dies natürlich für binomialverteilte  $X$ , so dass es nahe liegt, dass ein Wahrscheinlichkeitsmaß mithilfe der Binomialverteilung definiert werden kann.

**Lemma 2.54.** Für jedes  $p \in [0, 1]$  und  $n \in \mathbb{N}$  ist durch

$$P: \mathfrak{P}(\{0, 1, \dots, n\}) \rightarrow \mathbb{R}, \quad P(A) = \sum_{k \in A} B_{n,p}(k)$$

ein Wahrscheinlichkeitsmaß auf  $\{0, 1, \dots, n\}$  gegeben.

*Beweis.* Wir überprüfen die Axiome von Kolmogoroff für endliche Ergebnismengen.

(i) Es ist

$$P(\{0, 1, \dots, n\}) = \sum_{k \in \{0, 1, \dots, n\}} B_{n,p}(k) = \sum_{k=0}^n B_{n,p}(k).$$

Aus dem Binomischen Lehrsatz 1.13 folgt dann direkt mit  $x = 1, y = 1 - p$ :

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

(ii) Die Nicht-Negativität folgt aus  $p \in [0, 1]$ .

(iii) Seien  $A, B \subseteq \{0, 1, \dots, n\}$  disjunkt, dann liefern die Rechenregeln für Summen

$$P(A \cup B) = \sum_{k \in A \cup B} B_{n,p}(k) = \sum_{k \in A} B_{n,p}(k) + \sum_{k \in B} B_{n,p}(k) = P(A) + P(B)$$

□

## Normalverteilung

Bereits in Satz 2.21 wurde der Begriff der Wahrscheinlichkeitsdichte eingeführt. Es handelt sich dabei um eine Funktion  $\rho: \mathbb{R} \rightarrow \mathbb{R}_0^+$  mit  $\int_{-\infty}^{\infty} \rho(x) dx = 1$ . Aus dieser lässt sich ein Wahrscheinlichkeitsmaß auf der Borelmenge über  $\mathbb{R}$  konstruieren, in dem man  $P(A) = \int_A \rho(x) dx$  für ein Ereignis  $A$  setzt. Wir betrachten nun einen besonderen Fall für eine solche Dichtefunktion.

**Definition 2.55.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $X : \Omega \rightarrow \mathbb{R}$  eine reelle Zufallsgröße und  $\rho : \mathbb{R} \rightarrow \mathbb{R}_0^+$  eine Wahrscheinlichkeitsdichte. Dann besitzt  $X$  Dichte  $\rho$ , falls

$$P(X \in B) = \int_B \rho(x) \, dx$$

für alle Teilmengen  $B$  der Borelmenge über  $\mathbb{R}$  gilt.

**Definition 2.56** (Normalverteilung). Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Eine Zufallsgröße  $X : \Omega \rightarrow \mathbb{R}$  heißt *normalverteilt*, falls  $X$  eine Dichte der Form

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

mit gewissen  $\mu \in \mathbb{R}$  und  $\sigma \in \mathbb{R}^+$  besitzt.

Wir werden später zeigen, dass sich binomial verteilte Zufallsgrößen in einem gewissen Sinne einer Normalverteilung annähern. Auch die Bedeutung der Parameter  $\mu$  und  $\sigma$  wird sich später klären.

**Lemma 2.57.** Für jedes  $\mu \in \mathbb{R}$  und  $\sigma \in \mathbb{R}^+$  ist die in 2.56 definierte Funktion

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

eine Dichtefunktion.

*Beweis.* Zu zeigen ist

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = 1$$

Dazu bemerken wir zunächst, dass dieses Integral durch die Substitution  $y = \sigma\sqrt{2}x + \mu$  übergeht in

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot \sigma\sqrt{2}e^{-y^2} \, dy = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-y^2} \, dy$$

Zur Berechnung dieses Integrals betrachtet man zunächst das Quadrat des gesuchten Integrals<sup>2</sup>

$$\left( \int_{-\infty}^{+\infty} e^{-x^2} \, dx \right)^2 = \left( \int_{-\infty}^{+\infty} e^{-x^2} \, dx \right) \cdot \left( \int_{-\infty}^{+\infty} e^{-y^2} \, dy \right) = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} \, dx \right) \, dy$$

Hierbei wurde in der letzten Gleichung der Satz von Fubini angewendet. Durch den Wechsel auf Polarkoordinaten erhalten wir dann

$$\int_0^{+\infty} \left( \int_0^{2\pi} e^{-r^2} r \, d\phi \right) \, dr = \int_0^{+\infty} e^{-r^2} \cdot 2\pi r \, dr = \left[ -\pi e^{-r^2} \right]_0^{+\infty} = \pi$$

Durch Wurzelziehen erhält man das oben benötigte Integral, sodass wir insgesamt

$$\frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1$$



als Ergebnis der Integration erhalten. □

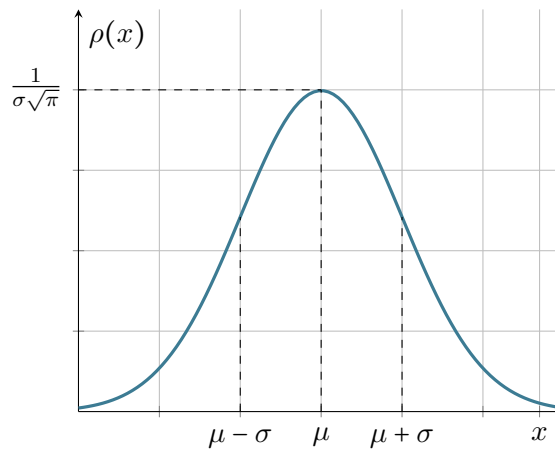


Abbildung 2.5: Graph der Dichtefunktion  $\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

### Poisson-Verteilung

Diese Verteilung wurde nach dem französischen Wort “poisson” (zu deutsch “Fisch”) benannt, da sie das große Fischsterben im sibirischen Irkutsk 1873 verblüffend genau mathematisch beschreibt.

Allgemein tritt diese Verteilung vor allem bei Ereignissen auf, die einzeln gesehen sehr selten statt finden. Eine sehr große Stichprobe führt jedoch dazu, dass sie trotzdem beobachtet werden können. In einer klassischen Studie konnte beispielsweise Bortkiewicz im 19. Jahrhundert die Anzahl der von einem Pferd zu Tode getretenen preußischen Soldaten mithilfe dieser Verteilung erstaunlich genau beschreiben. Weitere Beispiele sind der radioaktive Zerfall, die Anzahl der geschossenen Tore in einem Fußballspiel oder der Blitzschläge in einem bestimmten Gebiet. Eine besondere Bedeutung besitzt dieses Modell daher in der Versicherungsmathematik.

**Definition 2.58** (Poisson-Verteilung). Für jedes  $\lambda \in \mathbb{R}^+$  heißt die Abbildung

$$F_\lambda : \mathbb{N}_0 \rightarrow \mathbb{R}, \quad F_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

*Poisson-Verteilung* zum Parameter  $\lambda$ .

Genau wie im Falle der Binomialverteilung erhält man durch

$$P(A) = \sum_{k \in A} F_\lambda(k)$$

<sup>2</sup>Ein Ergebnis der Differentialalgebra besagt, dass die Stammfunktion von  $e^{-y^2}$  nicht durch Elementarfunktionen (z.B. Polynome,  $e$ -Funktion etc.) ausgedrückt werden kann. Der Wert des Integrals lässt sich also nur indirekt bestimmen.

ein Wahrscheinlichkeitsmaß auf  $\mathbb{N}_0$ . Wir überprüfen nur die Normierungsbedingung. Hier erhält man unter Verwendung der Reihendarstellung der Exponentialfunktion

$$\sum_{k \in \mathbb{N}_0} F_\lambda(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^\lambda = 1$$

**Definition 2.59.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Eine Zufallsgröße  $X : \Omega \rightarrow \mathbb{N}_0$  heißt *Poisson-verteilt* falls es ein  $\lambda \in \mathbb{R}^+$  gibt, sodass

$$P(X = k) = F_\lambda(k)$$

erfüllt ist.

Nun wollen wir den Zusammenhang mit der Binomialverteilung genauer untersuchen.

**Proposition 2.60.** Sei  $n \gg 1$  und  $p = \frac{\lambda}{n}$ , dann nähert sich die Binomialverteilung der Poisson-Verteilung an, d.h. es gilt

$$B_{n, \frac{\lambda}{n}}(k) \approx \frac{1}{k!} \lambda^k e^{-\lambda}$$

Der Beweis dieser Aussage verwendet, dass  $(1 - \frac{\lambda}{n})^n \xrightarrow{n \rightarrow \infty} e^{-\lambda}$ . Um dies zu sehen, leiten wir den Term nach  $\lambda$  ab und erhalten für große  $n$ :

$$\frac{d}{d\lambda} \left(1 - \frac{\lambda}{n}\right)^n = \left(1 - \frac{\lambda}{n}\right)^{n-1} \cdot n \cdot \left(-\frac{1}{n}\right) \approx (-1) \cdot \left(1 - \frac{\lambda}{n}\right)^n$$

Da die Exponentialfunktion  $\lambda \mapsto e^{-\lambda}$  ebenfalls eine Lösung dieser ‘‘Differentialgleichung’’ ist, erscheint die obige Behauptung plausibel. Tatsächlich kann dieser Argumentationsansatz auch zu einem formalen Beweis ausgebaut werden.

*Beweisskizze von 2.60.* Einsetzen in die Formel der Binomialverteilung ergibt zunächst

$$\begin{aligned} B_{n, \frac{\lambda}{n}}(k) &= \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{(n-k)! \cdot k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &\approx \frac{n \cdot (n-1) \dots (n-k+1)}{n^k \cdot k!} \cdot \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

Verwenden wir nun die Vorbemerkung  $(1 - \frac{\lambda}{n})^n \approx e^{-\lambda}$ , ergibt dies eingesetzt:

$$\frac{n \cdot (n-1) \dots (n-k+1)}{n^k \cdot k!} \cdot \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \approx \frac{1}{k!} \lambda^k e^{-\lambda}$$

Beachte dabei, dass unter der Voraussetzung, dass  $n$  deutlich größer als  $k$  ist, der Bruch  $\frac{n \cdot (n-1) \dots (n-k+1)}{n^k}$  näherungsweise 1 ist.  $\square$

# III Invarianten von Zufallsgrößen

Jeder reellen Zufallsgröße  $X$  können Invarianten zugeordnet werden, zu denen Erwartungswert und Varianz zählen. Diese werden wir später verwenden, um Konvergenzaussagen wie das Schwache bzw. Starke Gesetz der Großen Zahlen zu formulieren.

## 3.1 Der Erwartungswert

Der Erwartungswert bestimmt das Mittel der Werte einer Zufallsgröße, wobei die möglichen Ergebnisse mit ihrer jeweiligen Wahrscheinlichkeit gewichtet werden.

**Definition 3.1.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine reelle Zufallsgröße.

(a) Ist  $X$  eine diskrete Zufallsgröße, so ist der *Erwartungswert* von  $X$  gegeben durch

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \cdot P(X = x)$$

(b) Besitzt  $X$  eine stetige Dichtefunktion  $\rho$ , so ist der *Erwartungswert* von  $X$  gegeben durch

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \rho(x) dx,$$

falls das Integral existiert. Trifft dies nicht zu, so sagt man,  $X$  hat keinen Erwartungswert.

**Bemerkung 3.2.** Eine Verallgemeinerung kann in folgender Weise getroffen werden:

$$\mathbb{E}(X) = \int_{\mathbb{R}} X(\omega) dP(\omega)$$

Hierbei handelt es sich um ein Lebesgue-Integral, wobei als Maß das jeweilige Wahrscheinlichkeitsmaß verwendet wird.

Wir werden uns im Folgenden auf diskrete und stetig verteilte Fälle konzentrieren und einige Sätze über die Verallgemeinerung in den Übungen behandeln.

**Beispiele 3.3.** Wir berechnen die Erwartungswerte der Wahrscheinlichkeitsverteilungen von oben.

- (a) Sei  $X$  binomialverteilt mit den Parametern  $n, p$ . Hier gilt mit der Notation  $q = 1 - p$ :

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k \cdot \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^n k \cdot \frac{n!}{k!(n-k)!} p^k q^{n-k} = \\ &= 0 + \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} = n \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^k q^{n-k} = \\ &= n \sum_{k=1}^n \binom{n-1}{k-1} \cdot p \cdot p^{k-1} q^{(n-1)-(k-1)} = n \cdot p \left( \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} \right) \stackrel{*}{=} n \cdot p \end{aligned}$$

Bei (\*) wurde die Gleichung aus Lemma 2.54 verwendet, die besagt, dass die Summe 1 ergibt.

- (b) Sei  $X$  normalverteilt mit den Parametern  $\mu, \sigma$ . Hier gilt

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\ &= \mu \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \stackrel{*}{=} \\ &= \mu + \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \mu \end{aligned}$$

In der Gleichung (\*) wurde die Gleichung aus Lemma 2.57 und eine Substitution verwendet. Das Integral in der letzten Zeile verschwindet aufgrund der Punktsymmetrie des Integranden.

- (c) Sei  $X$  nun poisson-verteilt mit Parameter  $\lambda$ . Man erhält

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \\ &= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \end{aligned}$$

Auch hier wurde im letzten Schritt die Normierungseigenschaft der Poisson-Verteilung verwendet.

**Definition 3.4.** Man sagt, ein Ereignis  $A$  tritt *fast sicher* ein, wenn  $P(A) = 1$ .

Ein Beispiel für ein solches Ereignis ist: Bei einem Dartwurf wird der Mittelpunkt einer Dartscheibe fast sicher *nicht* getroffen.

**Proposition 3.5.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine diskrete Zufallsgröße.

- (a) Sei  $\mu \in \mathbb{R}$  und das Ereignis  $X = \mu$  fast sicher. Dann ist  $\mathbb{E}(X) = \mu$ .  
 (b) Ist  $X \geq 0$  fast sicher, so ist  $\mathbb{E}(X) \geq 0$

*Beweis.* (a) Wegen  $P(X = \mu) = 1$  muss  $P(X \neq \mu) = 0$  sein. Deshalb ist

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} xP(X = x) = \mu P(X = \mu) + \sum_{x \neq \mu} P(X = x) = \mu + 0 = \mu$$

(b) Durch die gleiche Überlegung wie oben ist hier  $P(X < 0) = 0$  und somit

$$\mathbb{E}(X) = \sum_{x \geq 0} xP(X = x) + \sum_{x < 0} xP(X = x) = \sum_{x \geq 0} xP(X = x) \geq 0$$

□

**Satz 3.6** (Rechenregeln für Erwartungswerte). Seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $X, Y : \Omega \rightarrow \mathbb{R}$  diskrete Zufallsgrößen und  $a \in \mathbb{R}$ . Dann gilt:

(a)  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

(b)  $\mathbb{E}(aX) = a\mathbb{E}(X)$

*Beweis.* (a) Es ist klar, dass  $P(X = x) = \sum_{y \in Y(\Omega)} P(X = x \text{ und } Y = y)$ . Daraus folgt

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} xP(X = x) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xP(X = x \text{ und } Y = y)$$

Man erhält für die Zufallsgröße  $X + Y$ :

$$\begin{aligned} \mathbb{E}(X) + \mathbb{E}(Y) &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} (x + y)P(X = x \text{ und } Y = y) = \\ &= \sum_{z \in (X+Y)(\Omega)} \sum_{\substack{x,y \\ x+y=z}} zP(X = x, Y = y) = \sum_{z \in (X+Y)(\Omega)} zP(X + Y = z) = \mathbb{E}(X + Y) \end{aligned}$$

(b) Setze  $Y = aX$  und berechne

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{y \in Y(\Omega)} yP(Y = y) = \sum_{x \in X(\Omega)} axP(aX = ax) = \\ &= \sum_{x \in X(\Omega)} axP(X = x) = a\mathbb{E}(X) \end{aligned}$$

□

**Bemerkung 3.7.** Die Aussage  $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$  ist im Allgemeinen falsch. Als Gegenbeispiel betrachten wir den einmaligen Wurf einer Laplace-Münze mit  $\Omega = \{K, Z\}$  sowie die Zufallsgröße  $X$ , gegeben durch  $X(K) = 1, X(Z) = -1$ . Es gilt

$$\mathbb{E}(X) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$$

Gehen wir nun zum Quadrat der Zufallsgröße, also  $X^2$ , über, so erhalten wir

$$X^2(K) = 1^2 = 1 \quad \text{und} \quad X^2(Z) = (-1)^2 = 1$$

Das wiederum bedeutet

$$\mathbb{E}(X^2) = 1 \neq \mathbb{E}(X) \cdot \mathbb{E}(X) = 0$$

### 3.2 Die Varianz

**Definition 3.8.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X : \Omega \rightarrow \mathbb{R}$  eine reelle Zufallsgröße. Die *Varianz* von  $X$  ist dann gegeben durch

$$\text{Var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2).$$

**Bemerkung 3.9.** Die Varianz ist nicht-negativ. Falls  $X = \mu$  ein fast sicheres Ereignis ist, so ist  $\mathbb{E}(X) = \mu$  und wegen  $\mathbb{E}(0) = 0$  folgt  $\text{Var}(X) = 0$ .

**Lemma 3.10.** Für die Varianz einer diskreten reellen Zufallsgröße  $X$  gilt folgende Formel:

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

*Beweis.* Sei  $\mathbb{E}(X) = \mu$ . Es gilt

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}([X - \mathbb{E}(X)]^2) = \mathbb{E}([x - \mu]^2) = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \\ &= \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mathbb{E}(\mu^2) = \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \end{aligned}$$

□

**Beispiele 3.11.** Wie oben für den Erwartungswert betrachten wir nun die Varianz der eingeführten Verteilungen.

(a) Sei  $X$  binomialverteilt mit Parametern  $n, p$ . Wir berechnen zunächst  $\mathbb{E}(X^2)$ :

$$\mathbb{E}(X^2) = \sum_{a_i \in X(\Omega)} a_i P(X^2 = a_i) = \sum_{k=0}^n k^2 P(X = k)$$

Wobei wir im letzten Schritt ausgenutzt haben, dass  $X^2$  nur Quadratzahlen als Werte annehmen kann. Einsetzen der Binomialverteilung für  $P(X = k)$  liefert nun

$$\begin{aligned} \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} + \sum_{k=0}^n k(k-1) \binom{n}{k} p^k q^{n-k} = \\ &= np + \underbrace{\sum_{k=0}^n k(k-1) \binom{n}{k} p^k q^{n-k}}_{=\mathbb{E}(X)} = \\ &= np + n(n-1) \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{(n-2)-(k-2)} p^2 = \\ &= np + n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-2-k)!} p^k q^{n-2-k} = np + n(n-1)p^2 \end{aligned}$$

Einsetzen in die eben bewiesene Formel aus 3.10 ergibt

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = np + n(n-1)p^2 - (np)^2 = \\ &= np + n^2p^2 - np^2 - n^2p^2 = np(1-p) = npq \end{aligned}$$

- (b) Sei nun  $X$  eine normalverteilte Zufallsgröße mit Parametern  $\sigma$  und  $\mu$ . Wir haben bereits  $\mathbb{E}(X) = \mu$  nachgerechnet. Es ist also dann

$$\text{Var}(X) = \mathbb{E}((X - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Durch Substitution und anschließende partielle Integration (P.I.) geht dieses Integral über in

$$\begin{aligned} \int_{-\infty}^{+\infty} u^2 \cdot \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{u^2}{2\sigma^2}} du &= \sigma^2 \cdot \frac{1}{\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} u \cdot \left( \frac{u}{\sigma^2} e^{-\frac{u^2}{2\sigma^2}} \right) du = \\ &\stackrel{\text{P.I.}}{=} \sigma^2 \frac{1}{\sigma\sqrt{\pi}} \left[ u \cdot \left( -e^{-\frac{u^2}{2\sigma^2}} \right) \right]_{-\infty}^{+\infty} + \sigma^2 \cdot \frac{1}{\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} 1 \cdot e^{-\frac{u^2}{2\sigma^2}} du \end{aligned}$$

Da die Exponentialfunktion an den Rändern viel stärker abfällt als jedes Polynom, gibt der erste Term 0, sodass wir

$$\sigma^2 \cdot \frac{1}{\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2\sigma^2}} du = \sigma^2$$

als Ergebnis erhalten. Im letzten Schritt wurde dabei verwendet, dass die Normalverteilung normiert ist (vgl. dazu den Beweis von 2.57).

- (c) Die Varianz der Poisson-Verteilung wurde in den Übungen berechnet. Man erhält für eine Poisson-verteilte Zufallsgröße  $X$  mit Parameter  $\lambda$

$$\text{Var}(X) = \lambda.$$

**Bemerkung 3.12.** Die Wurzel der Varianz nennt man auch *Standardabweichung*.

### 3.3 Kovarianz und Korrelationskoeffizient

Wir untersuchen nun, wie wir die umgangssprachliche *Korrelation* am Besten in eine mathematische Größe übersetzen können.

**Definition 3.13.** Seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und seien  $X, Y : \Omega \rightarrow \mathbb{R}$  reelle Zufallsgrößen. Die *Kovarianz* von  $X$  und  $Y$  ist definiert durch

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

**Lemma 3.14** (Rechenregeln für die Kovarianz). Seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $X, Y, Z : \Omega \rightarrow \mathbb{R}$  diskrete Zufallsgrößen und  $a \in \mathbb{R}$ . Dann gilt:

- (a)  $\text{Cov}(X, X) = \text{Var}(X)$
- (b)  $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- (c)  $\text{Cov}(X, aY + Z) = a\text{Cov}(X, Y) + \text{Cov}(X, Z)$

*Beweis.* (a) Ausformulieren der Voraussetzung ergibt direkt die Varianz:

$$\text{Cov}(X, X) = \mathbb{E}([X - \mathbb{E}(X)] \cdot [X - \mathbb{E}(X)]) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \text{Var}(X)$$

(b) Mit den Abkürzungen  $\mathbb{E}(X) = \mu$  und  $\mathbb{E}(Y) = \nu$  erhalten wir:

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mu)(Y - \nu)) = \mathbb{E}(XY - \mu Y - \nu X + \mu\nu)$$

Nach den Rechenregeln für Erwartungswerte 3.6 ist dies gerade

$$\mathbb{E}(XY) - \mathbb{E}(\mu Y) - \mathbb{E}(\nu X) + \mu\nu = \mathbb{E}(XY) - \mu\nu - \nu\mu + \mu\nu = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

(c) Wir verwenden (b) und die Rechenregeln für Erwartungswerte und erhalten

$$\begin{aligned} \text{Cov}(X, aY + Z) &\stackrel{(b)}{=} \mathbb{E}(X \cdot (aY + Z)) - \mathbb{E}(X)(aY + Z) = \\ &= \mathbb{E}(aXY + XZ) - \mathbb{E}(X) \cdot \mathbb{E}(aY + Z) = \\ &\stackrel{3.6}{=} a\mathbb{E}(XY) + \mathbb{E}(XZ) - a\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Z) = \\ &= a\mathbb{E}(XY) - a\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z) \stackrel{(b)}{=} a\text{Cov}(X, Y) + \text{Cov}(X, Z) \end{aligned}$$

□

**Satz 3.15.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X, Y : \Omega \rightarrow \mathbb{R}$  diskrete Zufallsgrößen. Falls  $X$  unabhängig von  $Y$  ist, ist  $\text{Cov}(X, Y) = 0$ .

*Beweis.* Aus dem eben bewiesenen Lemma 3.14 folgt, dass es genügt,  $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$  nachzurechnen. Es ist

$$\begin{aligned} \mathbb{E}(X) \cdot \mathbb{E}(Y) &= \left( \sum_{x \in X(\Omega)} x \cdot P(X = x) \right) \cdot \left( \sum_{y \in Y(\Omega)} y \cdot P(Y = y) \right) = \\ &= \sum_{(x, y) \in X(\Omega) \times Y(\Omega)} xy \cdot P(X = x)P(Y = y) \end{aligned}$$

Wir verwenden nun die Unabhängigkeit von  $X$  und  $Y$  erhalten:

$$\sum_{(x, y) \in X(\Omega) \times Y(\Omega)} xy \cdot P(X = x \text{ und } Y = y)$$

Definieren wir nun eine neue Zufallsgröße  $Z : \Omega \times \Omega \rightarrow \mathbb{R}$ ,  $Z(\omega, \theta) = X(\omega) \cdot Y(\theta)$ , so wird dies zu

$$\begin{aligned} &\sum_{z \in Z(\Omega \times \Omega)} \sum_{\substack{(x, y) \in X(\Omega) \times Y(\Omega) \\ x \cdot y = z}} z \cdot P(X = x \text{ und } Y = y) = \\ &= \sum_{z \in Z(\Omega \times \Omega)} z \cdot P(Z = z) = \mathbb{E}(Z) = \mathbb{E}(X \cdot Y) \end{aligned}$$

□



Zu bemerken ist, dass die Aussage des letzten Satzes keine Äquivalenzaussage ist: Verschwindet die Kovarianz, müssen die zugehörigen Zufallsgrößen nicht notwendigerweise unabhängig sein.

Außerdem besitzt die Kovarianz die problematische Eigenschaft, dass die Vervielfachung einer Zufallsgröße zu einer Vervielfachung der Kovarianz führt. Dies widerspricht allerdings der intuitiven Vorstellung der Korrelation: Nur weil beispielsweise beim Münzwurf Kopf bzw. Zahl doppelt zählen, sollte die entsprechende Zufallsgröße nicht stärker mit einer anderen Zufallsgröße korreliert sein, als wenn Kopf und Zahl einfach gezählt würden. Wir führen deshalb eine neue Größe ein, die diese Anforderungen erfüllt.

**Definition 3.16.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X, Y$  diskrete reelle Zufallsgrößen. Dann nennt man

$$\kappa(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$$

den *Korrelationskoeffizienten* von  $X$  und  $Y$ , falls dieser definiert ist.

**Proposition 3.17.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X, Y$  reelle Zufallsgrößen. Für alle  $a \in \mathbb{R}$  gilt dann

$$|\kappa(X, Y)| = |\kappa(aX, Y)|$$

*Beweis.* Wir verwenden

$$\text{Var}(aX) = \mathbb{E}((aX)^2) - \mathbb{E}(aX)^2 = \mathbb{E}(a^2 X^2) - a^2 \mathbb{E}(X)^2 = a^2 \mathbb{E}(X^2) - a^2 \mathbb{E}(X)^2 = a^2 \text{Var}(X)$$

Damit erhält man

$$\frac{\text{Cov}(aX, Y)}{\sqrt{\text{Var}(aX) \cdot \text{Var}(Y)}} = \frac{a \cdot \text{Cov}(X, Y)}{|a| \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \text{sgn}(a) \cdot \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$$

Daraus folgt gerade die Behauptung.  $\square$

**Beispiel 3.18.** Sei  $X$  eine Zufallsgröße, wegen  $\text{Cov}(X, X) = \text{Var}(X)$  gilt

$$\kappa(X, X) = \frac{\text{Var}(X)}{\sqrt{\text{Var}(X) \cdot \text{Var}(X)}} = 1$$

Der Beweis von 3.17 zeigt zudem, dass  $\kappa(X, -X) = -1$  gilt.

Der Beweis des nächsten Satzes benutzt die *Cauchy-Schwarz'sche-Ungleichung (CSU)*: Für alle Vektoren  $v, w$  eines Hilbertraums mit Skalarprodukt  $\langle \cdot, \cdot \rangle$  gilt

$$\langle v, w \rangle^2 \leq \langle v, v \rangle \cdot \langle w, w \rangle$$

Um diese anwenden zu können, überprüfen wir zunächst, dass die Menge aller reellen Zufallsgrößen auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  eine Vektorraum-Struktur besitzt.

Wir haben bereits gesehen, dass die Summe zweier Zufallsgrößen sowie Vielfache von Zufallsgrößen wieder Zufallsgrößen sind. Als Nullvektor identifizieren wir diejenige Zufallsgröße, die jedem Ergebnis die Zahl 0 zuordnet. Die restlichen Rechenregeln und Gruppen-Axiome folgen direkt aus den Gruppen-Eigenschaften von  $\mathbb{R}$ .

Auf diesem Vektorraum definieren wir nun ein Abbildung vermöge

$$\langle X, Y \rangle = \mathbb{E}(X \cdot Y)$$

Diese Abbildung ist ein Skalarprodukt: Rechenlemma 3.6 stellt sicher, dass sie bilinear ist. Die Symmetrie folgt aus  $\mathbb{E}(XY) = \mathbb{E}(YX)$ , die positive Definitheit aus  $\mathbb{E}(X^2) > 0$  (falls  $X \neq 0$ ).

Insgesamt ist damit die Menge der Zufallsgrößen ein Hilbertraum, d.h. ein Vektorraum mit einem Skalarprodukt, und wir können die CSU anwenden.

**Satz 3.19.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum. Für beliebige diskrete reelle Zufallsgrößen  $X$  und  $Y$  gilt

$$-1 \leq \kappa(X, Y) \leq 1$$

*Beweis.* Sei o.B.d.A.  $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ . Ist nämlich  $\mathbb{E}(X) = \mu \neq 0$ , so definiere eine neue Zufallsgröße  $A = X - \mu$ . Diese erfüllt nun  $\mathbb{E}(A) = 0$  und  $X$  und  $A$  haben sowohl die gleiche Varianz als auch Kovarianz:

$$\begin{aligned} \text{Var}(A) &= \mathbb{E}(A^2) - \underbrace{\mathbb{E}(A)^2}_{=0} = \mathbb{E}((X - \mu)^2) = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \\ &= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \text{Var}(X) \end{aligned}$$

Genauso gilt:

$$\begin{aligned} \text{Cov}(A, Y) &\stackrel{3.14}{=} \mathbb{E}(AY) - \underbrace{\mathbb{E}(A) \mathbb{E}(Y)}_{=0} = \\ &= \mathbb{E}([X - \mu]Y) = \mathbb{E}(XY) - \mu \mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y) = \text{Cov}(X, Y). \end{aligned}$$

Gilt also obige Annahme  $\mathbb{E}(X) = \mathbb{E}(Y) = 0$  nicht, so können wir  $X$  und  $Y$  wie beschrieben durch andere Zufallsgrößen ersetzen, ohne dass sich der Korrelationskoeffizient ändert.

Wir zeigen nun  $\kappa(X, Y)^2 \leq 1$ . Da die Erwartungswerte verschwinden, ist

$$\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y) = \mathbb{E}(X \cdot Y)$$

sowie

$$\text{Var}(X) = \mathbb{E}(X^2) \quad \text{und} \quad \text{Var}(Y) = \mathbb{E}(Y^2)$$

Wie in der Vorbemerkung angekündigt, verwenden wir nun die Cauchy-Schwarz'sche

Ungleichung (CSU) und erhalten

$$\begin{aligned} \text{Cov}(X, Y)^2 &= \mathbb{E}(X \cdot Y)^2 \stackrel{CSU}{\leq} \mathbb{E}(X^2) \cdot \mathbb{E}(Y^2) = \text{Var}(X) \cdot \text{Var}(Y) \\ \Leftrightarrow \left( \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \right)^2 &\leq 1 \quad \Leftrightarrow \quad \kappa(X, Y)^2 \leq 1 \end{aligned}$$

□

### 3.4 Die Ungleichungen von Markov und Tschebyscheff

Das Ziel dieses Abschnitts ist es, aus möglichst wenig Kenntnissen über eine Zufallsgröße  $X$  eine Abschätzung dafür zu finden, dass der Wert von  $X$  in bestimmtem Maße vom Erwartungswert  $\mathbb{E}(X) = \mu$  abweicht, also für die Wahrscheinlichkeit  $P(|X - \mu| \geq a)$ . Intuitiv wird schnell klar, dass die Varianz hier eine ausschlaggebende Rolle spielen wird, da sie die Streuung einer Zufallsvariable um den Erwartungswert anzeigt.

Wir werden uns zunächst mit einem allgemeineren Fall beschäftigen, aus dem eine Abschätzung für diese Situation als Spezialfall folgt.

**Satz 3.20** (Markov'sche Ungleichung). Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $X : \Omega \rightarrow \mathbb{R}$  eine diskrete Zufallsgröße. Sei weiter  $g : D \rightarrow \mathbb{R}_0^+$  eine stetige und monoton steigende Funktion mit  $X(\Omega) \subseteq D$  und  $a \in D$  mit  $g(a) \neq 0$ . Dann gilt

$$P(X \geq a) \leq \frac{\mathbb{E}(g(X))}{g(a)}$$

*Beweis.* Nach 2.39 ist auch  $g(X)$  eine Zufallsgröße und gemäß Aufgabe 2(b), Blatt 6 ist dann

$$E(g(X)) = \sum_{x \in X(\Omega)} g(x)P(X = x)$$

Nun teilen wir die Summe geeignet auf und erhalten

$$\mathbb{E}(g(X)) = \sum_{\substack{x \in X(\Omega) \\ x \geq a}} g(x)P(X = x) + \sum_{\substack{x \in X(\Omega) \\ x < a}} g(x)P(X = x)$$

Laut Voraussetzung gilt  $g(x) \geq 0$  für alle  $x \in \mathbb{R}$ , weshalb der zweite Summand nicht negativ ist und die Summe kleiner wird, wenn wir ihn weglassen. Da  $g$  zudem monoton steigend ist, gilt  $g(x) \geq g(a)$  für alle  $x \in X(\Omega)$  mit  $x \geq a$ . Man erhält also als Abschätzung

$$\begin{aligned} \mathbb{E}(g(X)) &\geq \sum_{\substack{x \in X(\Omega) \\ x \geq a}} g(x)P(X = x) \geq \sum_{\substack{x \in X(\Omega) \\ x \geq a}} g(a)P(X = x) \\ &= g(a) \sum_{\substack{x \in X(\Omega) \\ x \geq a}} P(X = x) = g(a)P(X \geq a) \end{aligned}$$

Sodann lösen wir nach  $P(X \geq a)$  auf und erhalten

$$P(X > a) \leq \frac{\mathbb{E}(g(X))}{g(a)}$$

□

Wir sehen uns nun an, wie die Funktion  $g$  in geeigneter Weise gewählt werden kann, um sinnvolle Abschätzungen zu erhalten.

**Beispiel 3.21.** Sei  $X$  eine reelle diskrete Zufallsgröße mit  $\mu = \mathbb{E}(X) = 0$ . Betrachte die Zufallsgröße

$$Y : \Omega \rightarrow \mathbb{R}, Y(\omega) = |X(\omega)|$$

Es ist dann  $Y(\Omega) = \mathbb{R}_0^+$  und die Abbildung  $g(x) = x^2$  ist auf  $Y(\Omega)$  monoton wachsend und stetig. Die Ungleichung von Markov liefert dann also für beliebige  $a \in \mathbb{R}^+$

$$P(|X| \geq a) \leq \frac{\mathbb{E}(|X|^2)}{a^2} = \frac{\mathbb{E}[(X - \mu)^2]}{a^2} = \frac{\text{Var}(X)}{a^2}.$$

Nun ist  $P(|X| \geq a) = P(X \geq a) + P(X < -a)$  und daher

$$P(X \geq a) \leq P(|X| \geq a) \leq \frac{\text{Var}(x)}{a^2}$$

**Folgerung 3.22** (Ungleichung von Tschebyscheff). Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $X : \Omega \rightarrow \mathbb{R}$  eine diskrete Zufallsgröße mit  $\mathbb{E}(X) = \mu$ . Dann gilt für beliebige  $a > 0$  die Abschätzung

$$P(|X - \mu| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

*Beweis.* Die Vorgehensweise ist analog zum Beispiel eben. Definiere eine neue Zufallsgröße  $Y = |X - \mu|$  und betrachte die Abbildung  $g(x) = x^2$ . Dann vergewissert man sich leicht, dass auch hier die Voraussetzungen der Markovschen Ungleichung (3.20) erfüllt sind, und erhält

$$P(|X - \mu| \geq a) \leq \frac{\mathbb{E}(|X - \mu|^2)}{a^2} = \frac{\text{Var}(X)}{a^2}$$

□

Zu bemerken ist, dass die Ungleichung von Tschebyscheff einerseits sehr praktisch, ist, weil die Kenntnis der Varianz einer Zufallsgröße ausreicht, um mit ihrer Hilfe eine Abschätzung zu bekommen. Andererseits stellt sich schnell heraus, dass diese Abschätzung im Allgemeinen relativ ungenau ist und mit mehr Informationen über die Verteilung der Zufallsgröße deutlich verbessert werden kann.

**Beispiel 3.23.** Sei  $X$  eine binomialverteilte Größe mit den Parametern  $n = 10\,000$  und  $p = \frac{1}{2}$ , beispielsweise könnte die Zufallsgröße  $X$  die Anzahl der Treffer bei einer Treffwahrscheinlichkeit  $\frac{1}{2}$  angeben (Kopf beim Münzwurf, hat Hannes ein Karohemd an, etc.). Mit der Ungleichung von Tschebyscheff können wir folgende Abschätzung treffen

$$P(X \geq 6000) \leq P(|X - 5000| \geq 1000) \leq \frac{np(1-p)}{1\,000\,000} = \frac{1}{400}$$

### 3.5 Schwaches und Starkes Gesetz der großen Zahlen

Wir wollen in diesem Abschnitt das sogenannte schwache Gesetz der Großen Zahlen beweisen. Die Aussage ist dem Leser vermutlich bereits aus der Schule vertraut: führt man viele unabhängige Wiederholungen eines Zufallsexperiments aus, so nähert sich die relative Häufigkeit eines Treffers seiner Wahrscheinlichkeit an. Wir führen zunächst eine für den Beweis notwendige Hilfsaussage an, formulieren dann den Satz mathematisch präzise und beschäftigen uns später mit dessen Bedeutung.

**Lemma 3.24.** Es seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $X, Y : \Omega \rightarrow \mathbb{R}$  zwei diskrete und unabhängige Zufallsgrößen. Dann gilt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) = \text{Var}(X) + \text{Var}(Y)$$

*Beweis.* Als Übungsaufgabe (Blatt 8, Aufgabe 1). □

**Satz 3.25** (Schwaches Gesetz der großen Zahlen). Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(X_i)_{i \in \mathbb{N}}$  eine Folge von diskreten und unabhängigen Zufallsgrößen, die alle den gleichen Erwartungswert  $\mu$  und die gleiche Varianz haben. Ist  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$  jeweils der Mittelwert der ersten  $n$  Folgenglieder, so gilt für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|M_n - \mu| \geq \varepsilon) = 0.$$

*Beweis.* Wir schreiben  $\text{Var}(X_1) = \sigma^2$  und erhalten für die obigen Zufallsgrößen wegen der Unabhängigkeit und aufgrund von Lemma 3.24:

$$\text{Var}(M_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n} \text{Var}(X_1) = \frac{1}{n} \sigma^2$$

Dabei haben wir verwendet, dass die  $X_n$  alle gleiche Varianz haben. Es folgt mit Tschebyscheff

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad \text{für alle } \varepsilon > 0$$

□

Falls eine Zufallsgröße  $X$  nur die Werte  $\{0, 1\}$  annimmt, so zählt sie gewissermaßen die Treffer bzw. Nicht-Treffer bei einem Zufallsexperiment. Auf diese Weise können wir eine Folge  $(X_n)_{n \in \mathbb{N}}$  definieren, wobei jedes  $X_n$  angibt, ob bei der  $n$ -ten Wiederholung des Experimentes ein Treffer oder Nicht-Treffer aufgetreten ist.

Ist  $p$  die Trefferwahrscheinlichkeit, so gilt für jedes  $n \in \mathbb{N}$

$$\mathbb{E}(X_n) = \sum_{x \in \text{im } X_n} X_n(x) P(X = x) = 1 \cdot p + 0 \cdot (1 - p) = p$$

Der Erwartungswert entspricht damit der Wahrscheinlichkeit eines Treffers. Das schwache Gesetz der Großen Zahlen besagt nun, dass für jedes  $\varepsilon \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P(|M_n - p| \geq \varepsilon) = 0$$

gilt. Der Mittelwert von Treffer/Nicht-Treffer, d.h. die relative Häufigkeit der Treffer, nähert sich also in einem gewissen Sinne der Trefferwahrscheinlichkeit an. Genauer gesagt wird die Wahrscheinlichkeit, dass die relative Häufigkeit von der Trefferwahrscheinlichkeit abweicht, beliebig klein. Das Schwache Gesetz der großen Zahlen setzt damit eine empirische Mess-Größe mit einer Modellgröße in Verbindung.

Interessanter als die Kovergenzaussage ist im Hinblick auf ein konkretes Experiment die Fehlerabschätzung  $\frac{\sigma^2}{n\varepsilon^2}$ , die wir im Beweis gesehen haben.

Im Folgenden wenden wir uns dem Starken Gesetz der Großen Zahlen zu. Ausgang ist die Frage, ob sich die Bildung des Grenzwerts und die Wahrscheinlichkeitsberechnung in 3.25 vertauschen lassen. Wie sich heraus stellt, ist dies nur unter zusätzlichen Voraussetzungen möglich.

**Satz 3.26** (Starkes Gesetz der Großen Zahlen). Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(X_k)_{k \in \mathbb{N}}$  eine Folge von diskreten und unabhängigen Zufallsgrößen. Weiterhin sei  $\mathbb{E}(X_k^4)$  für jedes  $k \in \mathbb{N}$  endlich und die  $X_k$  gleich verteilt (d.h. haben gleichen Erwartungswert  $\mu$  und Varianz). Dann gilt für den Mittelwert der ersten  $n$  Folgenglieder  $M_n = \frac{1}{n} \sum_{k=1}^n X_k$

$$P\left(\lim_{n \rightarrow \infty} M_n = \mu\right) = 1$$

**Bemerkung 3.27.** Das Schwache Gesetz gab uns eine Möglichkeit, Wahrscheinlichkeit zu interpretieren: ein Treffer hat eine bestimmte Wahrscheinlichkeit  $p$ , wenn auf lange Frist die Wahrscheinlichkeit, dass die relative Häufigkeit von  $p$  abweicht, gegen Null geht. Diese Erklärung ist dahingehend unbefriedigend, dass sie auf den Begriff Wahrscheinlichkeit selbst angewiesen ist. Das Starke Gesetz sagt nun aus, dass die Wahrscheinlichkeit, dass sich der Mittelwert einer Folge von Zufallsvariablen auf lange Sicht genau der Erwartungswert ist, 1 ist. Damit tritt dieses Ereignis (fast) sicher ein.

Der eigentliche Beweis des Satzes bedarf einiger Vorarbeit, was unter anderem daran liegt, dass wir, um die Wahrscheinlichkeit des Grenzwertes berechnen zu können, auf einen überabzählbaren Ereignisraum zurückgreifen müssen.

Betrachtet man als Zufallsexperiment beispielsweise das  $n$ -fache Werfen einer Münze, so ist die Kovergenzaussage nur interessant, falls wir tatsächlich einen Übergang  $n \rightarrow \infty$  durchführen können. Der Ergebnisraum besteht in diesem Fall also aus unendlich langen Binärfolgen der Form 0101111... und ist folglich überabzählbar unendlich. Daher müssen wir zunächst sicherstellen, dass die Wahrscheinlichkeit überhaupt wohldefiniert ist. Wir behandeln dies in einigen Lemmata.

**Lemma 3.28.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(X_i)_{i \in \mathbb{N}}$  eine Folge von diskreten und unabhängigen reellen Zufallsgrößen. Außerdem seien die  $X_i$  gleich verteilt (d.h. sie haben gleichen Erwartungswert  $\mu$ ) und der Mittelwert der ersten  $n$  Folgenglieder werde für jedes  $\omega \in \Omega$  mit  $M_n(\omega) = \frac{1}{n} \sum_{i=1}^n X_i(\omega)$  bezeichnet. Dann gilt

$$E = \left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} M_n(\omega) = \mu \right\} = \bigcap_{l \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} \left\{ \omega \in \Omega \mid |M_k(\omega) - \mu| < \frac{1}{l} \right\}$$

*Beweis.* Zum Beweis führen wir folgende Bezeichnung ein:

$$A_k^l = \{\omega \in \Omega \mid |M_k(\omega) - \mu| < \frac{1}{l}\} \quad B_n^l = \bigcap_{k \geq n} A_k^l \quad C^l = \bigcup_{n \in \mathbb{N}} B_n^l$$

Zu zeigen ist also  $\bigcap_{l \in \mathbb{N}} C^l = E$ .

Nun gilt laut Definition von  $A_k^l$ ,  $B_n^l$  und  $C^l$  als Schnitt- bzw. Vereinigungsmenge

$$(i) \omega \in B_n^l \Leftrightarrow \omega \in A_k^l \text{ für alle } k \geq n \quad (ii) \omega \in C^l \Leftrightarrow \exists n \in \mathbb{N} : \omega \in B_n^l$$

Beides zusammen ergibt

$$\omega \in C^l \Leftrightarrow \exists n \in \mathbb{N} : \forall k \geq n : \omega \in A_k^l$$

Da  $\omega \in A_k^l$  gleichbedeutend mit  $|M_k(\omega) - \mu| < \frac{1}{l}$  ist, folgt

$$\omega \in \bigcap_{l \in \mathbb{N}} C^l \Leftrightarrow \forall l \in \mathbb{N} : \omega \in C^l \Leftrightarrow \forall l \in \mathbb{N} : \exists n \in \mathbb{N} : \forall k \geq n : |M_k(\omega) - \mu| < \frac{1}{l}$$

Damit ist dem Nachweis der Konvergenz Genüge getragen. Also ist

$$\omega \in \bigcap_{l \in \mathbb{N}} C^l \Leftrightarrow \lim_{n \rightarrow \infty} M_n(\omega) = \mu \Leftrightarrow \omega \in E$$

□

**Folgerung 3.29.** Unter den Voraussetzungen und mit den Bezeichnungen aus 3.28 gilt

$$E = \left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} M_n(\omega) = \mu \right\} \in \mathcal{A}.$$

*Beweis.* Es gilt für beliebige  $k, l \in \mathbb{N}$

$$|M_k(\omega) - \mu| < \frac{1}{l} \Leftrightarrow M_k(\omega) \in ]\frac{1}{l} - \mu, \frac{1}{l} + \mu[$$

Es ist also  $A_k^l = M_k^{-1}(] \frac{1}{l} - \mu, \frac{1}{l} + \mu [)$ . Da Summen und skalare Vielfache von Zufallsgrößen wieder Zufallsgrößen sind, ist  $M_k$  eine Zufallsgröße und als Urbild eines offenen Intervalls (also einem Element der Borelmenge) unter einer Zufallsgröße liegt  $A_k^l$  in  $\mathcal{A}$ . Nach dem eben bewiesenen Lemma 3.28 entsteht  $E$  durch Schnitte bzw. Vereinigung über  $A_k^l$ , liegt also ebenfalls in  $\mathcal{A}$ . □

Wir erinnern an dieser Stelle an die sogenannte “Stetigkeit des Wahrscheinlichkeitsmaßes”, d.h. für eine beliebige Folge von Ereignissen  $(A_n)_{n \in \mathbb{N}}$  mit  $A_{n+1} \subseteq A_n$  gilt  $\lim_{n \rightarrow \infty} P(A_n) = P(\bigcap_{n \in \mathbb{N}} A_n)$  (vgl. Lemma 2.49). Diese Aussage wird im nächsten Beweis beitragen.

**Lemma 3.30** (Borel-Cantelli). Seien die Mengen  $E, A_k^l, B_n^l, C^l$  definiert wie oben. Falls gilt, dass

$$\sum_{k=1}^{\infty} P[(A_k^l)^C] < \infty \quad \text{für alle } l \in \mathbb{N},$$

so ist  $P(E) = 1$ .

*Beweis.* Wir benutzen die Stetigkeit von  $P$  bzgl. der  $C^l$ . Es gilt  $A_k^{l+1} \subseteq A_k^l$ , und daher auch  $C^{l+1} \subseteq C^l$ , also sind die Bedingungen von Lemma 2.49 erfüllt. Wir erhalten

$$P(E) = \lim_{l \rightarrow \infty} P(C^l)$$

Wir zeigen nun  $P(C^l) = 1$  für alle  $l \in \mathbb{N}$ . Dazu rechnen wir nach, dass  $P((C^l)^C) = 0$ . Hier gilt aufgrund der Regeln von De Morgan

$$P((C^l)^C) = P\left[\bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k^l\right]^C = P\left[\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} (A_k^l)^C\right]$$

Für alle  $n \in \mathbb{N}$  gilt

$$P((C^l)^C) \leq P\left(\bigcup_{k \geq n} (A_k^l)^C\right) \leq \sum_{k=n}^{\infty} P((A_k^l)^C)$$

Da die Reihe  $\sum_{k=1}^{\infty} P[(A_k^l)^C]$  laut Voraussetzung konvergiert, bilden die Partialsummen eine Cauchy-Folge und für jedes  $\varepsilon > 0$  gibt es ein  $N \in \mathbb{N}$ , sodass für  $m \geq n > N$  gilt

$$\sum_{k=n}^m P[(A_k^l)^C] = \sum_{k=0}^m P[(A_k^l)^C] - \sum_{k=0}^n P[(A_k^l)^C] < \varepsilon$$

Insbesondere bedeutet dies für alle  $n > N$ , dass

$$\sum_{k=n}^{\infty} P[(A_k^l)^C] < \varepsilon$$

Kombinieren wir dies mit der Abschätzung oben, erhalten wir  $P((C^l)^C) = 0$  und damit  $P(E) = 1$ .  $\square$

Um nun den Beweis des Starken Gesetzes zu führen, genügt es also, die Voraussetzung von Lemma 3.30 nachzuweisen. Wir müssen also zeigen, dass

$$\sum_{k=1}^{\infty} P(A_k^l)^C = \sum_{k=1}^{\infty} P(|M_k - \mu| > \frac{1}{l}) < \infty$$

ist. Beim Beweis des schwachen Gesetzes haben wir bereits folgende Abschätzung gefunden:

$$P(|M_k - \mu| > \frac{1}{l}) < \frac{\sigma^2}{n(\frac{1}{l})^2}$$

Nun gilt jedoch

$$\sum_{k=1}^{\infty} \frac{\sigma^2}{k(\frac{1}{l})^2} = \frac{\sigma^2}{(\frac{1}{l})^2} \sum_{k=1}^{\infty} \frac{1}{k} \rightarrow \infty$$

Da diese Reihe also divergiert, müssen wir die Abschätzung verbessern, um Lemma 3.30 anwenden zu können. Dazu verwenden wir die Ungleichung von Markov.



*Beweis von 3.26.* Wir setzen o.B.d.A.  $\mu = 0$  voraus, denn andernfalls können wir stattdessen die Folge  $(Y_i)_{i \in \mathbb{N}}$  mit  $Y_i = X_i - \mu$  betrachten, die dann Erwartungswert 0 hat und ebenfalls unabhängig ist. Außerdem haben auch die  $Y_i$  alle die gleiche Varianz und  $\mathbb{E}(Y_i)$  existiert jeweils. Ist  $\overline{M}_k$  der Mittelwert der ersten  $k$  Folgenglieder dieser Folge, so ist

$$\overline{M}_k = \frac{1}{k} \sum_{i=1}^k Y_k = \frac{1}{k} \sum_{i=1}^k (X_i - \mu) = \left( \frac{1}{k} \sum_{i=1}^k X_i \right) - \frac{1}{k} \cdot k\mu = M_k - \mu$$

Das bedeutet

$$\lim_{k \rightarrow \infty} \overline{M}_k = 0 \quad \Leftrightarrow \quad \lim_{k \rightarrow \infty} M_k = \mu,$$

sodass also aus der Gültigkeit des starken Gesetzes für die  $Y_i$  auch die Gültigkeit für die  $X_i$  folgt. Wir wenden nun die Markov'sche Ungleichung auf  $|M_k|$  und  $g(x) = x^4$  an und erhalten:

$$P(|M_k| > \frac{1}{l}) < \frac{\mathbb{E}(M_k^4)}{(\frac{1}{l})^4} \quad (*)$$

Wir berechnen nun Schritt für Schritt den Ausdruck auf der rechten Seite:

$$M_k^4 = \left( \frac{1}{k} \sum_{i=1}^k X_i \right)^4 = \frac{1}{k^4} \sum_{(a,b,c,d) \in \{1, \dots, k\}^4} X_a X_b X_c X_d$$

Die Summanden der letzten Summe sortieren wir als nächstes im Hinblick darauf, wie viele gleiche Faktoren auftreten:

$$\begin{aligned} & \sum_{a=1}^k X_a^4 + \sum_{\substack{(a,b) \in \{1, \dots, k\}^2 \\ a \neq b}} \binom{4}{3} X_a^3 X_b + \sum_{\substack{(a,b) \in \{1, \dots, k\}^2 \\ a \neq b}} \binom{4}{2} X_a^2 X_b^2 + \sum_{\substack{(a,b,c) \in \{1, \dots, k\}^3 \\ a \neq b, b \neq c, a \neq c}} \frac{4!}{2!} X_a^2 X_b X_c \\ & + \sum_{\substack{(a,b,c,d) \in \{1, \dots, k\}^4 \\ \text{paarw. verschieden}}} X_a X_b X_c X_d \end{aligned}$$

Bei der Berechnung des Erwartungswertes dieser Summe nutzen wir aus, dass sich Summation und Erwartungswertbildung vertauschen lassen, sowie dass aus der Unabhängigkeit der  $X_i$  auch  $\mathbb{E}(X_i \cdot X_j) = \mathbb{E}(X_i) \cdot \mathbb{E}(X_j)$  für  $i \neq j$  folgt (vgl. dazu den Beweis von 3.15). Da wir vorausgesetzt hatten, dass die Erwartungswerte verschwinden, verschwinden also alle Summanden, in denen eine einfache Potenz  $X_i$  auftaucht. Dadurch erhalten wir

$$\mathbb{E}(M_k^4) = \frac{1}{k} \left[ \sum_{a=1}^k \mathbb{E}(X_a^4) + 6 \cdot \sum_{\substack{(a,b) \in \{1, \dots, k\}^2 \\ a \neq b}} \mathbb{E}(X_a^2) \mathbb{E}(X_b^2) \right]$$

Nach Voraussetzung sind die  $X_i$  gleich verteilt, sodass auch die  $\mathbb{E}(X_a^4)$  bzw.  $\mathbb{E}(X_b^2)$  alle gleich sind. Es genügt daher, die Anzahl der Summanden zu zählen:

$$\mathbb{E}(M_k^4) = \frac{1}{k^4} [k \cdot \mathbb{E}(X_1^4) + 6 \cdot k(k-1) \cdot \mathbb{E}(X_1^2)^2] \quad (**)$$

Nach Annahme ist  $\mathbb{E}(X_1^4)$  endlich. Damit existiert auch  $\mathbb{E}(X_1^2)^2$ , denn es kann  $0 \geq \text{Var}(X) = \mathbb{E}(X^4) - \mathbb{E}(X^2)^2$  nur erfüllt sein, falls  $\mathbb{E}(X^2)^2 \leq \mathbb{E}(X^4) < \infty$  gilt. Diese Abschätzung auf  $(**)$  angewendet, liefert

$$\mathbb{E}(M_k^4) \leq \frac{1}{k^4} [k \cdot \mathbb{E}(X_1^4) + 6 \cdot k(k-1) \cdot \mathbb{E}(X_1^4)] \leq \frac{6}{k^4} k^2 \mathbb{E}(X_1^4)$$

Wir können dies nun in  $(*)$  einsetzen und erhalten

$$P(M_k > \frac{1}{l}) < \frac{6}{k^2} \mathbb{E}(X_1^4) l^4$$

und damit

$$\sum_{k=1}^{\infty} P(M_k > \frac{1}{l}) < 6 \mathbb{E}(X_1^4) l^4 \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

Die Voraussetzung von Lemma 3.30 ist also erfüllt und wir erhalten die Aussage des Starken Gesetzes der Großen Zahlen.  $\square$

**Bemerkung 3.31.** Falls  $X$  eine beschränkte Zufallsgröße ist (das ist insbesondere dann der Fall, wenn  $X$  eine diskrete Zufallsgröße ist), so existiert  $\mathbb{E}(X^4)$  immer. Eine noch bessere Abschätzung als im Beweis erhält man durch Verwendung von  $g(x) = e^x$  anstatt  $g(x) = x^4$ . Die Abschätzung wird dann zu

$$P(|M_k - \mu| > \frac{1}{l}) < C e^{-k}$$

mit einer gewissen Konstante  $C$ . Man benötigt allerdings, dass  $\mathbb{E}(e^{|X|})$  existiert.

Der Name des Starken Gesetzes der großen Zahlen rührt daher, dass es als mathematische Aussage stärker ist, als das Schwache Gesetz, d.h. aus der Gültigkeit des Starken Gesetzes folgt die des Schwachen. Bevor wir diese Aussage jedoch beweisen, sehen wir uns zunächst an einem Beispiel an, dass die Umkehrung im Allgemeinen falsch ist. Dazu konstruieren wir eine Folge  $(Y_n)_{n \in \mathbb{N}}$ , die

$$(i) \lim_{n \rightarrow \infty} P(|Y_n| > \varepsilon) = 0 \quad \text{und} \quad (ii) P(\lim_{n \rightarrow \infty} Y_n = 0) \neq 1$$

erfüllt.

**Beispiel 3.32.** Wir betrachten alle Winkel  $\omega$ , die beim Drehen eines Glücksrades auftreten können, d.h. unsere Grundmenge ist  $\Omega = [0; 2\pi[$ . Wir definieren nun zunächst induktiv eine Folge von Intervallen  $I_n = [a_n; b_n]$  mittels

$$I_1 = [0; \frac{1}{n}[ \quad \text{und} \quad I_n = [a_{n-1} + \frac{1}{n}; b_{n-1} + \frac{1}{n}[ = [\sum_{k=1}^{n-1} \frac{1}{k}; \sum_{k=1}^n \frac{1}{k}[$$

und anschließend eine Folge von Zufallsgrößen

$$Y_n(\omega) = \begin{cases} 1 & \text{falls } \omega \in I_n + 2\pi\mathbb{Z} \\ 0 & \text{sonst} \end{cases}$$

Dabei bedeutet die Addition von  $2\pi\mathbb{Z}$ , dass für ein  $\omega$  also auch  $Y_n(\omega) = 1$  sein soll, falls  $\omega$  durch die Addition eines ganzzahligen Vielfachen von  $2\pi$  in  $I_n$  landet.

Es gelte weiterhin die Laplace-Annahme in der Form, dass

$$P(\omega \in I_n) = \frac{b_n - a_n}{2\pi}$$

sein soll.

zu (i): Zunächst zeigen wir, dass  $\lim_{n \rightarrow \infty} P(|Y_n| > \varepsilon) = 0$ .

Falls  $\varepsilon \geq 1$ , so ist  $|Y_n| > \varepsilon$  unmöglich und  $P(|Y_n| > \varepsilon) = 0$ .

Falls  $\varepsilon < 1$  ist, betrachten wir das Ereignis  $|Y_n| > \varepsilon$ , d.h. die Menge

$$\{\omega \in \Omega \mid Y_n(\omega) > \varepsilon\} = \{\omega \in \Omega \mid Y_n(\omega) = 1\}$$

Aufgrund unserer Definition ist diese Menge ein Intervall der Länge  $\frac{1}{n}$ . Damit ist  $P(|Y_n| > \varepsilon) = \frac{1}{2\pi n}$  und wegen  $\lim_{n \rightarrow \infty} \frac{1}{2\pi n} = 0$  ist die behauptete Gleichung wahr.

zu (ii): Wir zeigen nun, dass  $\lim_{n \rightarrow \infty} Y_n = 0$  ein unmögliches Ereignis ist. Dazu zeigen wir, dass

$$\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} Y_n = 0\} = \emptyset,$$

indem wir nachweisen, dass für alle  $\omega \in \Omega$  der Grenzwert  $\lim_{n \rightarrow \infty} Y_n(\omega)$  nicht existiert.

Sei dazu  $\omega \in \Omega$  beliebig. Da man für jedes Intervall  $I_n$  von oben auch als

$$I_n = \left[ \sum_{k=1}^{n-1} \frac{1}{k}; \sum_{k=1}^n \frac{1}{k} \right[$$

schreiben kann, sind die Intervallgrenzen durch die harmonische Reihe gegeben. Diese ist divergent, wächst also unbeschränkt und für jedes  $l \in \mathbb{N}$  gibt es ein  $n \in \mathbb{N}$ , sodass

$$\omega + l \cdot 2\pi \in I_n$$

erfüllt ist. Die Menge aller  $n \in \mathbb{N}$ , für die  $Y_n(\omega) = 1$  ist, ist also unendlich. Ebenso ist die Menge aller  $m \in \mathbb{N}$ , für die  $Y_n(\omega) = 0$  ist, unendlich. Daher kann die Folge  $Y_n$  nicht konvergieren.

**Bemerkung 3.33** ( $\sqrt{n}$ -Gesetz). Das schwache Gesetz sagt aus, dass der Mittelwert  $M_n$  der ersten  $n$  Folgenglieder einer Folge identisch verteilter Zufallsgrößen  $(X_n)_{n \in \mathbb{N}}$  für genügend großes  $n$  mit hoher Wahrscheinlichkeit gerade zum Erwartungswert  $\mu$  wird. Um detailliertere Aussagen zu erhalten, erinnern wir daran, dass wir im Beweis zu 3.25 gesehen hatten, dass

$$\text{Var}(M_n) = \frac{1}{n} \text{Var}(X_1)$$

gilt. Bezeichnen wir nun mit  $\sigma$  die jeweilige Standardabweichung, so kann diese Gleichheit auch als

$$\sigma_{M_n} = \frac{\sigma_X}{\sqrt{n}}$$

interpretiert werden. Intuitiv ist die Standardabweichung ein Maß für die Streuung einer Zufallsgröße um ihren Mittelwert. Nehmen wir eine normalverteilte Zufallsgröße  $X$  an, so lässt sich diese Aussage weiter quantifizieren:

	$P( X - \mu  > \varepsilon)$
$1\sigma$	$\approx 30\%$
$2\sigma$	$\approx 5\%$
$3\sigma$	$\approx 0.3\%$
$4\sigma$	$\approx 0.01\%$
$5\sigma$	$\approx 0.000\ 05\%$

Als Faustregel kann man daher eine Abweichung von  $1\sigma$  als ungewöhnlich ansehen, in der Teilchenphysik dagegen gilt die Existenz ein neues Teilchens z.B. erst als gesichert, wenn eine Streuung von  $5\sigma$  ausgeschlossen werden kann. Im Kapitel über Hypothesentests werden wir später dazu sagen, dass ein Signifikanzniveau von  $0.000\ 05\%$  gewählt wird.

Aus den Überlegungen lassen sich Abschätzungen der folgenden Art ableiten: Angenommen, Peter hat eine Laplace-Münze, die er  $10\ 000$  Mal wirft. Dabei tritt  $5\ 200$  mal "Kopf" auf. Ist das bereits eine verdächtig große Abweichung vom Erwartungswert  $5000$  oder liegt die Schwankung noch im natürlichen Bereich?

Wir haben bereits gesehen, dass die Varianz einer binomialverteilten Zufallsgröße zu den Parametern  $n$  und  $p$  gerade  $npq$  beträgt, für große  $n$  also ungefähr  $n$  ist. Die Standardabweichung ist also  $\sqrt{n}$  und nach oben erwähnter Faustregel können wir eine größere Abweichung vom Mittelwert als  $\sqrt{n}$  als verdächtig ansehen. Diese Faustregel wird auch als  $\sqrt{n}$ -Gesetz bezeichnet.

In unserem Beispiel ist  $\sqrt{n} = 100$ , also ist obige Abweichung von  $200$  tatsächlich ungewöhnlich.

### 3.6 Konvergenzbegriffe der Stochastik

Wir betrachten in diesem Abschnitt Folgen von Zufallsgrößen  $(X_n)_{n \in \mathbb{N}}$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  sowie eine Zufallsgröße  $X$ , an die sich die Folge (in einem gewissen noch genauer zu bestimmenden Sinne) annähert:

$$X_n \xrightarrow{n \rightarrow \infty} X$$

In der Analysis hatten wir für eine derartige "Annäherung" verschiedene Konvergenzbegriffe zur mathematischen Beschreibung eingeführt. Beispielsweise sagt man, dass eine Folge von Funktionen  $(f_n)_{n \in \mathbb{N}}$  *punktweise* gegen  $f$  konvergiert, falls für alle  $x$  des Definitionsbereichs gilt

$$\lim_{n \rightarrow \infty} f_n(x) = f(x)$$

Dagegen konvergiert die Folge *gleichmäßig* gegen  $f$ , falls

$$\lim_{n \rightarrow \infty} \|f_n - f\|_\infty = \lim_x \sup |f_n(x) - f(x)| = 0$$

erfüllt ist. So wie in der Analysis ist auch in der Stochastik eine Unterscheidung verschiedener Konvergenzbegriffe nötig. Von diesen werden wir im Folgenden drei kennen lernen.

**Definition 3.34.** Eine Folge von reellen Zufallsgrößen  $(X_n)_{n \in \mathbb{N}}$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  konvergiert *fast sicher* gegen eine reelle Zufallsgröße  $X$ , falls

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Ausgeschrieben bedeutet diese Bedingung:

$$P(\{\omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$$

Die Menge in den Klammern ist genau die Menge der  $\omega$ , auf denen  $X_n$  punktweise gegen  $X$  im Sinne der Analysis konvergiert. Daher ist fast sichere Konvergenz gleichbedeutend damit, dass für fast alle  $\omega \in \Omega$  die Folge  $X_n$  punktweise gegen  $X$  konvergiert.

In Anlehnung an das Schwache Gesetz der Großen Zahlen definieren wir einen weiteren Konvergenzbegriff.

**Definition 3.35.** Eine Folge von reellen Zufallsgrößen  $(X_n)_{n \in \mathbb{N}}$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  konvergiert *stochastisch* (oder *in Wahrscheinlichkeit*) gegen eine reelle Zufallsgröße  $X$ , falls

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$$

für jedes  $\varepsilon \in \mathbb{R}^+$  erfüllt ist.

**Beispiele 3.36.** Die beiden Gesetze der Großen Zahlen liefern Beispiele für fast sichere sowie stochastische Konvergenz.

- Erfüllt eine Folge  $(X_n)_{n \in \mathbb{N}}$  die Voraussetzungen des Starken Gesetzes der Großen Zahlen, so konvergiert die Folge der zugehörigen Mittelwerte  $(M_n)_{n \in \mathbb{N}}$  fast sicher gegen den Erwartungswert  $\mu$  der Folge.
- Ist andererseits  $(X_n)_{n \in \mathbb{N}}$  eine Folge von Zufallsgrößen, für die das Schwache Gesetz der großen Zahlen gilt, so konvergiert die Folge der zugehörigen Mittelwerte  $(M_n)_{n \in \mathbb{N}}$  stochastisch gegen den Erwartungswert  $\mu$  der Folge.

**Definition 3.37.** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge von reellen Zufallsgrößen auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ . Sei weiter  $V_{X_n} : \mathbb{R} \rightarrow [0, 1]$  die Verteilungsfunktion von  $X_n$  und  $V_X : \mathbb{R} \rightarrow [0, 1]$  die Verteilungsfunktion von  $X$ . Man sagt,  $X_n$  konvergiert *in Verteilung* gegen  $X$ , falls für alle  $t \in \mathbb{R}$ , an denen  $V_X$  stetig ist, die Bedingung

$$\lim_{n \rightarrow \infty} V_{X_n}(t) = V_X(t)$$

erfüllt ist.

**Bemerkung 3.38.** An Unstetigkeitsstellen ist es schwierig, auf sinnvolle Weise eine Konvergenz zu definieren, da man beispielsweise zwischen Grenzwerten von links und von rechts unterscheiden muss. Deshalb schließen wir diese Stellen in der Definition aus. Noch einfacher wird es natürlich, falls wir von vornherein stetige Funktionen betrachten.

**Satz 3.39.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(X_n)_{n \in \mathbb{N}}$  eine Folge von reellen Zufallsgrößen, die fast sicher gegen die Zufallsgröße  $X$  konvergiert. Dann konvergiert  $(X_n)_{n \in \mathbb{N}}$  auch stochastisch gegen  $X$ .

*Beweis.* Der Beweis wurde in den Übungen besprochen (Blatt 9, Aufgabe 1). □

**Bemerkung 3.40.** Aus Satz 3.39 folgt, dass tatsächlich aus der Gültigkeit des Starken Gesetzes der Großen Zahlen für eine Folge von Zufallsgrößen auch die die des Schwachen Gesetzes folgt. Wie in den Beispielen 3.36 kann nämlich das Starke Gesetz in eine Aussage über fast sichere Konvergenz und das Schwache Gesetz über stochastische Konvergenz umformuliert werden. Dies bedeutet, dass das Starke Gesetz eine stärkere Aussage im mathematischen Sinne ist.

Die Umkehrung von Satz 3.39 gilt jedoch im Allgemeinen nicht. Wie wir nun zeigen, ist eine weitere Bedingung nötig, um aus stochastischer Konvergenz fast sichere Konvergenz zu folgern.

**Satz 3.41.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(X_n)_{n \in \mathbb{N}}$  eine Folge von reellen Zufallsgrößen, die stochastisch gegen die Zufallsgröße  $X$  konvergiert. Gilt zusätzlich für jedes  $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty,$$

so konvergiert  $X_n$  auch fast sicher gegen  $X$ .

*Beweis.* Wie oben angeführt, betrachten wir das Ereignis  $E = \{\omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$ . Man vergewissert sich leicht, dass diese Menge mit der Menge übereinstimmt, die wir in Lemma 3.28 eingeführt haben (statt der Zufallsgröße  $M_n$  verwenden wir nun  $X_n$ ). Die Voraussetzung von 3.30 ist durch die hier geforderte Voraussetzung sicher gestellt, damit erhalten wir

$$P(\{\omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$$

und somit konvergiert  $X_n$  fast sicher gegen  $X$ . □

**Satz 3.42.** Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $(X_n)_{n \in \mathbb{N}}$  eine Folge von reellen Zufallsgrößen, die stochastisch gegen die Zufallsgröße  $X$  konvergiert. Dann konvergiert  $X_n$  auch in Verteilung gegen  $X$ .

*Beweis.* Aufgrund der stochastischen Konvergenz gilt  $\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$  für alle  $\varepsilon > 0$ . Sei weiterhin  $t \in \mathbb{R}$  so, dass  $V_X$  stetig bei  $t$  ist. Zu zeigen ist nun  $\lim_{n \rightarrow \infty} V_{X_n}(t) = V_X(t)$ , d.h.

$$\forall \varepsilon > 0 : \exists n_0 \in \mathbb{N} : \forall n \geq n_0 : |V_{X_n} - V_X(t)| < \varepsilon$$

Da  $V_X(t)$  stetig ist, existiert ein  $\delta > 0$ , sodass

$$|V_X(t) - V_X(s)| < \frac{\varepsilon}{2} \quad \text{für alle } |t - s| \leq \delta$$

Wir zeigen nun in zwei Schritten, dass für hinreichend große  $n_0$  gilt:

- (1)  $V_{X_n}(t) - V_X(t) < \varepsilon$
- (2)  $V_X(t) - V_{X_n}(t) < \varepsilon$ .

Zu (1)

Laut Definition ist  $V_{X_n}(t) - V_X(t) = P(X_n \leq t) - P(X \leq t)$ . Um eine Abschätzung für den ersten Term zu bekommen, überlegen wir uns, welche Fälle bei  $X \leq t$  auftreten können. Hierbei kann prinzipiell  $X \leq t + \delta$  oder  $X > t + \delta$  sein. Damit aber immer noch  $X_n \leq t$  erfüllt sein kann, folgt im zweiten Fall, dass  $|X_n - X| > \delta$  sein muss. Wir bekommen aus diesem Grund

$$V_{X_n}(t) = P(X \leq t) \leq P(X \leq t + \delta \text{ oder } |X - X_n| > \delta)$$

Wir setzen dies ein und erhalten

$$\begin{aligned} V_{X_n}(t) - V_X(t) &\leq P(X \leq t + \delta \text{ oder } |X - X_n| > \delta) - P(X \leq t) \\ &\leq P(X \leq t + \delta) + P(|X - X_n| > \delta) - P(X \leq t) \end{aligned}$$

Unter der Verwendung der Definition der Verteilungsfunktion erhält man dafür

$$V_X(t + \delta) - V_X(t) + P(|X - X_n| > \delta)$$

Aufgrund der Stetigkeit von  $V_X$  ist wie oben bemerkt  $V_X(t + \delta) - V_X(t) < \frac{\varepsilon}{2}$ . Wegen  $\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$  gibt es außerdem ein  $n_0 \in \mathbb{N}$  sodass  $P(|X - X_n| > \delta) < \frac{\varepsilon}{2}$  für alle  $n \geq n_0$ . Es folgt daher insgesamt

$$V_{X_n}(t) - V_X(t) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

wie gewünscht.

Zu (2)

Wir gehen ähnlich wie oben vor und bemerken zunächst, dass laut Definition  $V_X(t - \delta) = P(X \leq t - \delta)$  gilt. Wiederum unterscheiden wir die beiden Fälle  $X_n \leq t$  und  $X_n > t$ , was im zweiten Fall genauso wie oben  $|X - X_n| > \delta$  bedeutet.

Wiederum erhalten wir folglich

$$\begin{aligned} P(X \leq t - \delta) &\leq P(X_n \leq t) + P(|X - X_n| > \delta) \\ \Leftrightarrow V_X(t - \delta) &\leq V_{X_n}(t) + P(|X - X_n| > \delta) \end{aligned}$$

Umstellen liefert

$$-V_{X_n} \leq -V_X(t - \delta) + P(|X - X_n| > \delta)$$

Nun verwenden wir, dass aufgrund der Stetigkeit  $V_X(t) - V_X(t - \delta) < \frac{\varepsilon}{2}$  ist und für  $n \geq n_0$  die Ungleichung  $P(|X_n - X| > \delta) < \frac{\varepsilon}{2}$  erfüllt ist. Damit erhält man nun

$$V_X(t) - V_{X_n}(t) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Aus (1) und (2) folgt insgesamt

$$|V_{X_n}(t) - V_X(t)| < \varepsilon \quad \text{für alle } n \geq n_0$$

Damit haben wir für alle  $t$ , bei denen  $V_X$  stetig ist,  $\lim_{n \rightarrow \infty} V_{X_n}(t) = V_X(t)$  gezeigt.  $\square$

**Bemerkung 3.43.** Auch hier ist die Umkehrung im Allgemeinen falsch: Konvergiert eine Folge  $X_n$  gegen  $X$  in Verteilung, so konvergiert sie nicht zwangsläufig auch stochastisch

gegen  $X$ .

*Beweis.* Dazu betrachten wir folgendes Gegenbeispiel. Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $A \subseteq \Omega$  eine Menge, sodass  $P(A) = \frac{1}{2}$ . Definiere nun eine Zufallsgröße  $X : \Omega \rightarrow \mathbb{R}$  mittels

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ -1 & \text{falls } \omega \notin A \end{cases}$$

Ein Beispiel hierfür wäre wiederum das Werfen einer Laplace-Münze (vgl. Beispiel 2.48 (a)). Weiterhin definieren wir nun eine Folge  $(X_n)_{n \in \mathbb{N}}$  von Zufallsgrößen durch

$$X_n = (-1)^n X$$

Die Zufallsgrößen  $X_n$  haben alle die gleiche Verteilungsfunktion, welche identisch mit der Verteilungsfunktion von  $X$  ist:

1. *Fall:* Für  $t < -1$  ist aufgrund der Definition klar, dass  $V_{X_n}(t) = P(X_n \leq t) = P(\emptyset) = 0$  und ebenso  $V_X(t) = P(X \leq t) = 0$  ist.
2. *Fall:* Für  $t \geq 1$  erhält man sofort  $V_{X_n}(t) = V_X(t) = 1$  aus der Definition.
3. *Fall:* Für  $t \in [-1; 1[$  betrachten wir

$$V_{X_n}(t) = P(X_n \leq t) = P(X_n = -1) = P(A^A) = \frac{1}{2} \text{ für gerade } n$$

$$V_{X_n}(t) = P(X_n \leq t) = P(-X_n = -1) = P(X = -1) = P(A^C) = \frac{1}{2} \text{ für ungerade } n$$

Zusammen mit  $V_X(t) = P(X \leq t) = P(X = -1) = \frac{1}{2}$  in diesem Fall erhalten wir insgesamt

$$V_{X_n}(t) = V_X(t) \quad \text{für alle } t \in \mathbb{R}, n \in \mathbb{N}$$

Damit konvergieren die  $V_{X_n}$  offensichtlich für jeden Punkt gegen  $V_X$ , und  $X_n$  damit in Verteilung gegen  $X$ .

Die  $X_n$  konvergieren jedoch nicht stochastisch gegen  $X$ . Laut unserer Konstruktion gilt nämlich

$$X_n = X \text{ falls } \omega \text{ gerade} \quad X_n = -X \text{ falls } n \text{ ungerade}$$

und damit

$$|X_n - X| = 0 \text{ falls } n \text{ gerade} \quad |X_n - X| = 2 \text{ falls } n \text{ ungerade}$$

Daraus folgt, dass beispielsweise

$$P(|X_n - X| < \frac{1}{4}) = \begin{cases} 1 & \text{falls } n \text{ gerade} \\ 0 & \text{falls } n \text{ ungerade} \end{cases}$$

Der Grenzwert dieser Wahrscheinlichkeit für  $n \rightarrow \infty$  existiert damit nicht. □



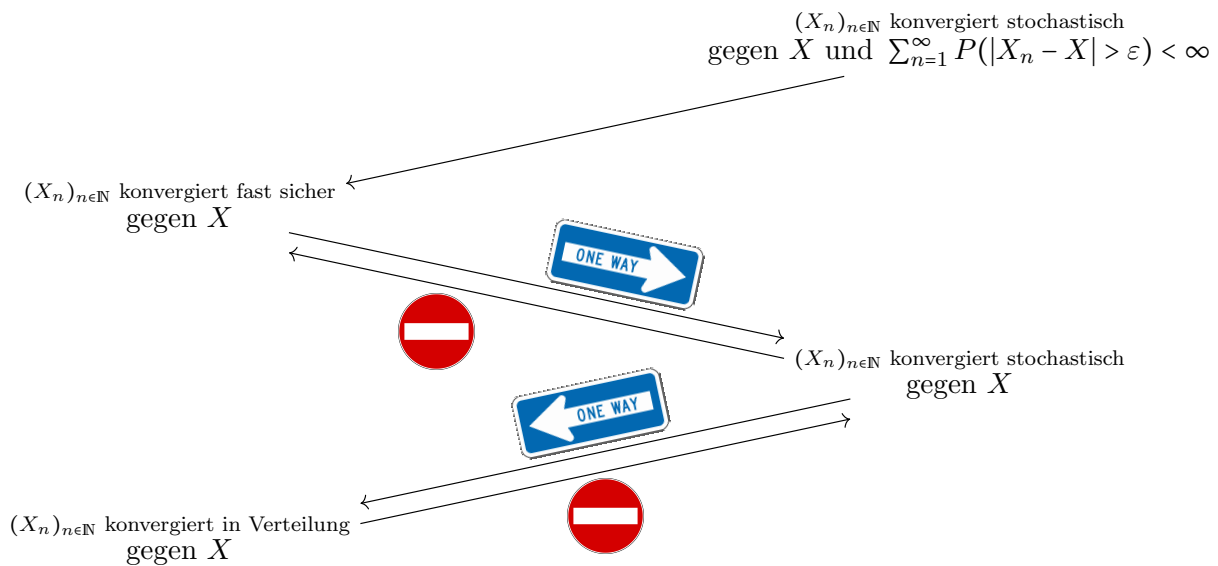


Abbildung 3.1: Graphische Darstellung der Ergebnisse aus dem letzten Abschnitt über den Zusammenhang der verschiedenen Konvergenzbegriffe

### 3.7 Der Zentrale Grenzwertsatz

In diesem Abschnitt beschäftigen wir uns mit einer wichtigen Konvergenzaussage für Zufallsgrößen. Der so genannte *Zentrale Grenzwertsatz* besagt, dass sich bei großer Stichprobenlänge ein Zusammenhang mit der Normalverteilung einstellt. Zur Motivation dieser Tatsache betrachten wir zunächst den Spezialfall Bernoulli-verteilter Zufallsgrößen.

Sei dazu  $(X_n)_{n \in \mathbb{N}}$  ein Folge unabhängiger Zufallsgrößen auf  $\Omega = \{0, 1\}$ . Dann ist die Summe der ersten  $n$  Folgenglieder  $S_j = \sum_{j=1}^n X_j$  binomialverteilt. Wir haben bereits gesehen, dass für die Mittelwerte  $M_n = \frac{1}{n}S_n$  sowohl das Schwache als auch das Starke Gesetz der Großen Zahlen gelten.

Wir hatten die Binomialverteilung zur Stichprobenzahl  $n$  und Trefferwahrscheinlichkeit  $p$  (und der Abkürzung  $q = 1 - p$ ) als  $B_{n,p}(k) = \binom{n}{k} p^k q^{1-k}$  definiert.

Damit gilt für die Differenz zweier aufeinanderfolgender Trefferzahlen

$$\begin{aligned} B_{n,p}(k) - B_{n,p}(k-1) &= \binom{n}{k} p^k q^{n-k} - \binom{n}{k-1} p^{k-1} q^{n-k+1} = \\ &= \binom{n}{k} p^k q^{n-k} - \frac{q}{p} \binom{n}{k-1} p^k q^{n-k} \end{aligned}$$

Wir berechnen nun zunächst

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

und schreiben damit den obigen Ausdruck um zu

$$\begin{aligned} \binom{n}{k} p^k q^{n-k} \left[ 1 - \frac{k}{n-k+1} \cdot \frac{q}{p} \right] &= B_{n,p}(k) \cdot \left[ \frac{p(n-k+1) - kq}{p(n-k+1)} \right] = \\ &= B_{n,p}(k) \cdot \left[ \frac{pn - pk + p - kq}{p(n-k+1)} \right] = B_{n,p}(k) \left[ \frac{pn - k + p}{pn - pk + p} \right] \end{aligned}$$

Die Idee ist nun,  $B_{n,p}(k) - B_{n,p}(k-1)$  im Sinne einer Ableitung als Tangentensteigung aufzufassen. Da der Erwartungswert der Binomialverteilung  $np$  ist, hat diese dort ein Maximum und die Tangentensteigung 0, falls  $k \approx pn$ . Dadurch kommen wir zu einer näherungsweisen Differentialgleichung für  $g(x) = B_{n,p}(pn+x)$ .

$$g'(x) \approx g(x) \cdot \frac{pn - pn - x + p}{pn - p^2n - px + p} = g(x) \frac{-x + p}{npq - px + p}$$

Interessant ist sehr großes  $n$  und  $x \approx \sqrt{n}$ . Also nähern wir in der folgenden Weise:

$$g'(x) \approx g(x) \cdot \frac{-x}{npq}$$

Diese Differentialgleichung lösen wir nun mittels Trennen der Variablen zur Anfangsbedingung  $g(x_0) = y_0$ :

$$\begin{aligned} \ln g(x) - \ln y_0 &= \int_{y_0}^{g(x)} \frac{1}{g} dg = \int_{x_0}^x \frac{-t}{npq} dt = \frac{-1}{npq} \cdot (x^2 - x_0) \\ \Leftrightarrow \ln g(x) &= \frac{-1}{npq} \cdot (x^2 - x_0) + \ln y_0 \end{aligned}$$

Also ist  $g(x) = e^{-x^2/(2npq)} \cdot C$ , wobei wir in der Konstante  $C$  die Abhängigkeit von der Anfangsbedingung zusammengefasst haben. Diese Konstante kann auch rückwärts anhand der Normierungsforderung ohne Kenntnis der Anfangsbedingung bestimmt werden. Also stellen wir fest, dass sich die Binomialverteilung der Normalverteilung zu den Parametern

$$\sigma = \sqrt{npq} \quad \text{und} \quad \mu = np$$

annähert. Unsere heuristische Überlegung hat uns damit zu einer Aussage geführt, die als Satz von *Moivre-Laplace* bekannt ist. Diesen Satz wollen wir jedoch nicht formulieren und beweisen, sondern gehen gleich zum allgemeineren Fall über, dem zentralen Grenzwertsatz.

**Satz 3.44** (Zentraler Grenzwertsatz, ohne Beweis). Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge unabhängiger, identisch verteilter reeller Zufallsgrößen auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , deren Erwartungswert  $\mu < \infty$  und Varianz  $\sigma^2 < \infty$  jeweils existieren. Dann konvergiert

$$\frac{1}{\sigma\sqrt{n}} \left( \sum_{j=1}^n X_j - n\mu \right)$$

in Verteilung gegen die Standardnormalverteilung, d.h. der Normalverteilung zu den Parametern  $\mu = 0$  und  $\sigma = 1$ .

Wir geben nun nochmal eine Übersicht über die auftretenden Zufallsgrößen. Sei dazu  $(X_n)_{n \in \mathbb{N}}$  eine Folge von Zufallsgrößen, die die Voraussetzungen des Zentralen Grenzwertsatzes erfüllt (mit Erwartungswert  $\mu$ , Standardabweichung  $\sigma$ ). Wir führen nun die Bezeichnungen

$$M_n = \frac{1}{n} \sum_{j=1}^n X_j \quad S_n = \sum_{j=1}^n X_j \quad Z_n = \frac{1}{\sigma\sqrt{n}}(S_n - n\mu)$$

für den Mittelwert, die Summe und die standardisierte Summe der ersten  $n$  Folgenglieder ein. Für die jeweiligen Erwartungswerte bzw. Varianzen rechnet man dann leicht nach:

	$\mathbb{E}$	$\text{Var}$
$M_n$	$\mu$	$\frac{1}{n}\sigma^2$
$S_n$	$n\mu$	$n\sigma^2$
$Z_n$	$0$	$1$

Die Zufallsgröße  $Z_n$  in der letzten Zeile konvergiert nach dem Zentralen Grenzwertsatz in Verteilung gegen die Normalverteilung zu den Parametern  $\mu = 0$  und  $\sigma = 1$  (Standardnormalverteilung). Die beiden anderen Zufallsgrößen dagegen gegen die Normalverteilung zu den jeweiligen anderen Parametern.

Speziell für Bernoulli-verteilte Zufallsgrößen ist  $\mathbb{E}(X) = p$  und  $\text{Var}(X) = pq$ , also erhalten wir als Konsequenz für großes  $n$ :

$$P(a \leq M_n \leq b) \approx V(b) - V(a)$$

wobei  $V$  die Verteilungsfunktion von  $M_n$  bezeichnet. Dabei haben wir  $P(M_n = a)$  vernachlässigt, da diese Wahrscheinlichkeit in der Regel sehr gering ist. Wir führen nun die Bezeichnung  $\Phi_{\sigma^2}^{\mu}$  für die Verteilungsfunktion der Normalverteilung zu den Parametern  $\sigma$  und  $\mu$  ein. Unter Verwendung des Zentralen Grenzwertsatzes können wir dann weiter in folgender Weise nähern:

$$P(a \leq M_n \leq b) \approx \Phi_{pq/n}^p(b) - \Phi_{pq/n}^p(a)$$

Verwendet man nun die Dichtefunktion der Normalverteilung (vergleiche Definition 2.56), dann erhält man

$$\Phi_{\sigma^2}^{\mu}(a) = P(X \leq a) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

Damit wird die Differenz von oben zu

$$\Phi_{pq/n}^p(b) - \Phi_{pq/n}^p(a) = \int_a^b \exp\left(-\frac{(x-p)^2}{2pq/n}\right) \cdot \sqrt{\frac{n}{2\pi pq}} dx$$

Die Werte  $\Phi_{\sigma^2}^{\mu}(a)$  sind in Tafelwerken erfasst und können also leicht nachgeschlagen werden, während es nicht möglich ist, eine Stammfunktion direkt anzugeben.

In ähnlicher Weise gilt

$$P(a \leq S_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(x-np)^2}{2npq}\right) dx$$

Wie oben erwähnt, haben wir dabei die Wahrscheinlichkeit dafür, die Randwerte zu erhalten, vernachlässigt. Eine Verbesserung der Genauigkeit bekommt man deshalb, wenn man den Integrationsbereich leicht vergrößert und stattdessen

$$P(a \leq S_n \leq b) = \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} \dots dx \text{ bzw. } P(a < S_n < b) = \int_{a+\frac{1}{2}}^{b-\frac{1}{2}} \dots dx$$

berechnet. Man spricht hierbei von der *Stetigkeitskorrektur*.

**Bemerkung 3.45.** Durch Substitution kann man zeigen, dass  $\Phi_{\sigma^2}^\mu(a) = \Phi_1^0\left(\frac{a-\mu}{\sigma}\right)$  ist. Es genügt also ein Tafelwerk, in dem die Werte für die Standardnormalverteilung tabelliert sind, um die entsprechenden Werte für beliebige Normalverteilungen bestimmen zu können.

## IV Statistik

Wir versuchen als nächstes, die bisher entwickelte Wahrscheinlichkeitstheorie in Verbindung zur Empirie zu bringen.

### 4.1 Testen von Hypothesen

Zunächst widmen wir uns Hypothesentests und illustrieren die Vorgehensweise dabei an einem Beispiel. Gegeben sei eine Münze, von der wir annehmen, dass es sich um eine Laplace-Münze handelt. Diese Annahme eines geeigneten Modells für die Münze nennt man *Nullhypothese*  $H_0$ . Ziel ist es nun, anhand eines Experiment festzustellen, ob diese Annahme gerechtfertigt ist.

Wir entscheiden uns dazu, die Münze  $N$ -mal zu werfen, mit einem großen  $N$ , z.B.  $N = 1000$ .  $N$  bezeichnet man dabei als Stichprobenzahl. Ausgehend von dem Konzept einer Laplace-Münze erwarten wir als Ergebnis, dass die Münze in etwa  $\mu = 500$  Mal "Kopf" zeigt. Zur Durchführung des Tests legen wir als nächstes einen *Annahmebereich* fest. Dies geschieht beispielsweise in folgender Weise:

Wähle ein  $k$  so, dass eine (theoretisch perfekte) Laplace-Münze, die der Nullhypothese genügt, mit einer Wahrscheinlichkeit von 95% bei 1000 Würfeln eine Kopffzahl im Annahmebereich  $[\mu - k, \mu + k]$  liefert. Die zugelassenen 5% Abweichung werden *Fehlerniveau* oder *Signifikanzniveau* genannt.

Es stellt sich nun die Frage, wie man dieses  $k$  bestimmt. Da sich die zugelassenen 5% Abweichung aufgrund der Symmetrie der Normalverteilung zu gleichen Teilen auf den Bereich links bzw. rechts des Annahmebereichs verteilen, erhalten wir die Gleichung

$$P(X \leq 500 - k) = \Phi_{250}^{500}(500 - k) = 2.5\%$$

Hierbei haben wir die Varianz der Binomialverteilung zu  $\sigma^2 = npq = 250$  berechnet. Mit der Substitution aus 3.45 erhält man

$$\Phi_1^0\left(\frac{-k}{\sqrt{250}}\right) = 2.5\%$$

Mit einem Tafelwerk erhält man  $x \approx 2$  und damit

$$\frac{-k}{\sqrt{250}} = 2 \quad \Leftrightarrow \quad k = -\sqrt{250} \cdot 2 \approx -30$$

Dies bedeutet, dass eine Laplace-Münze bei 1000 Würfeln mit einer Wahrscheinlichkeit von 95% eine Anzahl von "Kopf"-Ergebnissen liefert, die zwischen 470 und 530 liegt. Liegt also ein experimenteller Befund unserer Münze vor, vergleichen wir diesen mit dem Annahmebereich. Sollte der experimentelle Wert außerhalb liegen, verwerfen wir die Hypothese. Falls die Anzahl dagegen im Intervall liegt, behält man die Hypothese bei.

Anzumerken ist an dieser Stelle, dass das Verwerfen der Hypothese in entscheidender Weise von der Wahl des Signifikanzniveaus abhängt. Diese Wahl kann nun dazu führen, dass eine eigentlich zutreffende Hypothese verworfen oder eine falsche Hypothese angenommen wird. Diese Fehlentscheidungen tragen die Bezeichnung Fehler 1. bzw. 2. Art und die Wahrscheinlichkeit dafür, eine dieser Fehlentscheidungen zu treffen wird häufig als  $\alpha$  bzw.  $\beta$  notiert. In unserem Beispiel beträgt die Wahrscheinlichkeit für den Fehler 1. Art, d.h. die Wahrscheinlichkeit, dass die Nullhypothese zwar wahr ist, aber dennoch verworfen wird, nach Wahl gerade 5%. Die Wahrscheinlichkeit für den Fehler 1. Art lässt sich damit als die bedingte Wahrscheinlichkeit  $P_{H_0}(H_0 \text{ wird abgelehnt})$  schreiben.

Wir stellen die Fehler hier in einer Tabelle zusammen.

	$H_0$ stimmt	$H_0$ ist falsch
$H_0$ wird beibehalten	richtig	Fehler 2. Art $\beta$
$H_0$ wird abgelehnt	Fehler 1. Art $\alpha$ hier: $\alpha = 5\%$	richtig

- Bemerkung 4.1.** (a) Man sollte beachten, dass der Fehler 2. Art nicht berechnet werden kann. Münzen sind beispielsweise niemals perfekte Laplace-Münzen, da diese z.B. von der Prägung und der Wurftechnik beeinflusst werden. Man sollte daher bei der Bestätigung von Nullhypothesen vorsichtig sein. Die Entscheidung, eine Nullhypothese abzulehnen, kann mathematisch gerechtfertigt werden, während auch bei Ergebnissen im Annahmereich die Nullhypothese niemals mit vollständiger Sicherheit bestätigt werden kann. Auch wenn die Ergebnisse eines Tests in einem sehr eng gewählten Annahmereich liegen, wäre beispielsweise die Hypothese, es handle sich bei einer Münze um eine vollkommene Laplace-Münze, dennoch falsch.
- (b) Das Fehlerniveau entspricht der Wahrscheinlichkeit, dass ausgehend von der Wahrheit der Hypothese das Experiment einen gewissen Ausgang hat. Das bedeutet, die Berechnung geht vom theoretischen Modell aus. Für Anwendungen wünschenswert hingegen wäre eine Aussage darüber, wie plausibel eine gewissen Hypothese ausgehend von empirischen Fakten ist. Dies ist jedoch im Allgemeinen schwer bis unmöglich.

### Einseitige Hypothesentests

Wir führen einige Beispiele für Tests an, um die soeben angesprochene Problematik zu verdeutlichen.

**Beispiel 4.2.** Von einer Kiste voller Würfel wissen wir, dass sich darin nur Würfel befinden, die entweder (in sehr guter Näherung) Laplace-Würfel sind oder so gezinkt wurden, dass die Wahrscheinlichkeit für einen Sechser  $p = \frac{1}{4}$  ist. Wir wählen einen Würfel aus und möchten herausfinden, ob es sich hierbei um einen Laplace-Würfel oder einen gezinkten Würfel handelt. Dazu werfen wir ihn  $N = 1000$  mal und zählen die Anzahl der aufgetretenen Sechser. Wir stellen also folgende Hypothesen auf:

$$\begin{array}{ll} \text{Nullhypothese } H_0 : & \text{Würfel ist Laplace-Würfel, d.h. } p = \frac{1}{6} \\ \text{Alternativhypothese } H_1 : & \text{Würfel ist gezinkt, d.h. } p = \frac{1}{4} \end{array}$$

Vorgegeben sei weiter ein Fehlerniveau, d.h. die Wahrscheinlichkeit für einen Fehler 1. Art von 5%. Anders ausgedrückt:

$$P_{H_0}(H_0 \text{ abgelehnt}) = 5\%$$

Daraus bestimmen wir das sogenannte *Konfidenzintervall*, das heißt einen Annahmebereich, der so gewählt ist, dass die Wahrscheinlichkeit, Treffer außerhalb des Intervalls zu landen, gerade 5% beträgt. Wie man der Tabelle auf Seite 52 leicht entnimmt, ist bei der Normalverteilung  $[\mu - 2\sigma; \mu + 2\sigma]$  ein geeignetes Intervall. Nun ist

$$\mu = \frac{1}{6} \cdot 1000 \approx 166 \quad \text{und} \quad \sigma = \sqrt{1000 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 12$$

Hieraus könnte dann das Konfidenzintervall angegeben werden.

Nun ist aber die Wahrscheinlichkeit, dass man links vom Konfidenzintervall landet, bei der Alternativhypothese noch ungünstiger. Da es nur zwei Alternativen gibt und  $H_1$  eine höhere Wahrscheinlichkeit als  $H_0$  hat, würden Werte in diesem Bereich eher für  $H_0$  sprechen. Daher dehnen wir die untere Grenze des Konfidenzintervalls bis  $-\infty$  aus und führen einen einseitigen (genauer: rechtsseitigen) Hypothesentest durch.

Wir suchen dazu ein  $k$ , sodass die Wahrscheinlichkeit, dass  $X \geq \mu + k$  ist, kleiner als 5% ist. Dazu muss  $P(X \notin [\mu - k; \mu + k]) < 10\%$  sein. Die Tabelle liefert:  $k \approx 1.65\sigma$ . Damit erhalten wir als Obergrenze des Konfidenzintervalls  $\mu + k \approx 167 + 20 \approx 187$ . Falls  $H_1$  hingegen stimmt, so ist

$$\tilde{\mu} = 1000 \cdot \frac{1}{4} = 250 \quad \text{und} \quad \tilde{\sigma} = \sqrt{1000 \cdot \frac{1}{4} \cdot \frac{3}{4}} \approx 14$$

Wegen  $250 - 187 \approx 4\sigma$  liegen zwischen 250 und 187 mehr als 4 Standardabweichungen. Durch erneute Zuhilfenahme der Tabelle sieht man, dass damit die Wahrscheinlichkeit, bei wahrer Alternativhypothese im Konfidenzintervall zu landen, kleiner als 1% ist. Dies ist genau die Wahrscheinlichkeit für einen Fehler 2. Art.

Nun kann man anstelle des Fehlers 1. Art auch den Fehler 2. Art vorgeben oder versuchen, beide Fehler vorzugeben, z.B.  $\alpha = \beta$  oder  $\alpha = \beta = 5\%$ . Bei letzterer Variante wird i.d.R. die Stichprobenlänge  $N$  als Parameter betrachtet und dann entsprechend gewählt. Eine entsprechende Aufgabe werden wir in den Übungen behandeln.

Geben wir eine Schranke für die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, vor, so berechnen wir wie oben dargestellt die bedingte Wahrscheinlichkeit

$$P_{H_0}(H_0 \text{ abgelehnt}) = P_{H_0}(\text{Treffer rechtes Intervall}).$$

Wir hätten jedoch gerne einen Wert für  $P_{\text{Treffer bei } 240}(H_0 \text{ stimmt})$ . Dies ist allgemein nicht möglich. In der folgenden Erweiterung des Beispiels können wir die letztere Wahrscheinlichkeit jedoch berechnen.

**Beispiel 4.3.** Nehmen wir dazu an, in einer Kiste liegen 1 000 Würfel, von denen genau einer in der Weise gezinkt ist, dass die Wahrscheinlichkeit für eine  $\text{⊠}$  gerade  $\frac{1}{4}$  beträgt. Wir ziehen nun zufällig einen Würfel aus dieser Kiste und würfeln damit 1 000 Mal. Weiterhin nehmen wir an, dass dabei 240 Mal eine Sechs auftritt.

Unter Verwendung der Abkürzungen

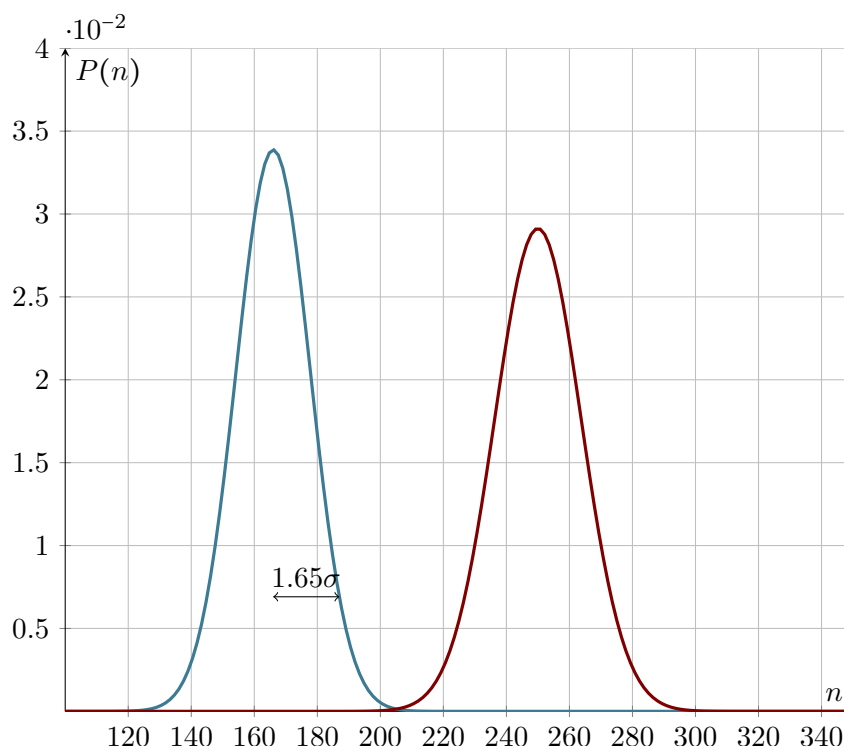


Abbildung 4.1: Im Diagramm ist eine Normalverteilung mit Mittelwert  $\mu = 166$  und Standardabweichung  $\sigma = 11.8$  in blau dargestellt, dies entspricht im Beispiel 4.2 der Verteilung der Sechsen mit einem nicht gezinkten Würfel. In rot ist eine Normalverteilung mit  $\mu = 250$  und  $\sigma = 13.7$  dargestellt, diese korrespondiert zu dem gezinkten Würfel.

	Würfel gezinkt	Würfel nicht gezinkt
240 Treffer	a	b

können wir nun folgende Berechnungen anstellen:

$$\begin{aligned}
 a &= P(240 \text{ Treffer und Würfel gezinkt}) = P_{\text{Würfel gezinkt}}(240) \cdot P(\text{Würfel gezinkt}) = \\
 &= \binom{1000}{240} \left(\frac{1}{4}\right)^{240} \left(\frac{3}{4}\right)^{760} \frac{1}{1000} \\
 b &= \binom{1000}{240} \cdot \left(\frac{1}{6}\right)^{240} \cdot \left(\frac{5}{6}\right)^{760} \cdot \frac{999}{1000}
 \end{aligned}$$

Damit ist

$$P_{240 \text{ Treffer}}(\text{Würfel gezinkt}) = \frac{P(240 \text{ Treffer und Würfel gesezinkt})}{P(240 \text{ Treffer})} = \frac{a}{a+b}$$

Dies entspricht der Wahrscheinlichkeit, dass der Würfel tatsächlich gezinkt ist, wenn 240 Treffer auftreten. Die Tatsache, dass nur ein gezinkter unter 1000 Würfeln vorhanden ist, verkleinert  $a$  und somit auch  $P_{240 \text{ Treffer}}(\text{Würfel gezinkt})$ . Tatsächlich wird es umso unwahrscheinlicher, dass ein Würfel gezinkt ist, je weniger gezinkte Würfeln vorhanden sind. Dies würde man auch intuitiv erwarten.



**Bemerkung 4.4.** Häufig ist die Nullhypothese auch von der Form “ $H_0 : p \leq a$ ”, z.B.  $H_0 : p \leq \frac{1}{2}$  beim Münzwurf. Man erhält auf diese Weise einen einseitigen Test. Beim  $n$ -fachen Münzwurf mit  $n = 100$  bestimmt man beispielsweise ein  $k$  so, dass

$$P_{p=\frac{1}{2}}(S_n > 50 + x) \leq \alpha$$

für einen vorgegebenen  $\alpha$ -Fehler erfüllt ist.

**Beispiel 4.5.** Wir behandeln nun ein Beispiel für einen Hypothesentest, der Gebrauch von der Poisson-Verteilung anstatt der Normalverteilung macht. Anwendung findet dieser Test v.a. in der Versicherungsmathematik, da dort seltene Ereignisse wie das Auftreten von Naturkatastrophen mathematisch beschrieben werden müssen.

Stellen wir uns also vor, wir wären Versicherungsunternehmer und wissen, dass die Anzahl der Brände in einer bestimmten Stadt pro Jahr Poisson-verteilt ist. Anders ausgedrückt

$$P(k \text{ Brände/ Jahr}) = \frac{\lambda^k}{k!} e^{-\lambda}$$

für ein gewisses  $\lambda \in \mathbb{R}^+$ . Die entscheidende Frage ist nun natürlich, welchen Wert dieses  $\lambda$  hat, weshalb wir die Nullhypothese

$$H_0 : \lambda = 1$$

aufstellen und diese Vermutung im Folgenden durch einen Hypothesentest mathematisch überprüfen wollen. Sei dazu ein  $\alpha$ -Fehler von 5% vorgegeben. Für welche  $k$  gilt nun  $\sum_{j=0}^k P_{\lambda=1}(j) \geq 95\%$ ? Dies rechnen wir zu Fuß nach:

$$\begin{aligned} \sum_{j=0}^k P_{\lambda=1}(j) &= \sum_{j=0}^k \frac{1}{j!} \cdot e^{-1} \stackrel{!}{\geq} 95\% \\ \Leftrightarrow \sum_{j=0}^k \frac{1}{j!} &\approx 2.584 \\ \Leftrightarrow 1 + 1 + \frac{1}{2} + \frac{1}{6} + \dots &= 2.\bar{6} + \dots \approx 2.584 \end{aligned}$$

Also ist der Ablehnungsbereich  $\{4, 5, 6, \dots\}$ . Für größere Werte von  $\lambda$  ist evtl. ein zweiseitiges Ablehnungsintervall zu bestimmen. Auch hier ist wieder eine Alternativhypothese (z.B.  $\lambda = 6$ ) denkbar, außerdem kann man den  $\beta$ -Fehler bestimmen. Letzterer berechnet sich gemäß:

$$\beta = P_{\lambda=6}(k \in \{0, 1, 2, 3\}) = 1 \cdot e^{-6} + \frac{6}{1} e^{-6} + \frac{6^2}{2} e^{-6} + \frac{6^3}{6} e^{-6} = 61e^{-6} \approx 0.151$$

## 4.2 Schätzmethoden

**Beispiel 4.6** (Taxi-Problem). Stellen wir uns vor, wir befinden uns in einer unbekanntem Stadt, in der nummerierte Taxis die Straßen bevölkern. Beispielsweise sehen wir Taxis mit den Nummern

$$17, 23, 7, 40, 36$$

vorbeifahren und stellen uns anschließend die wichtige Frage, wie viele Taxis es denn in dieser Stadt gibt. Sinnvollerweise, würde man nun die maximale auftretende Nummer tippen, oder auch den doppelten Mittelwert der beobachteten Nummern. Um dieses Problem mathematisch eleganter behandeln zu können, wandeln wir es etwas ab.

**Beispiel 4.7.** Stellen wir uns nun vor, wir befinden uns in einer unbekanntem Spielshow, in der ein unbekannter Moderator zufällig eine positive reelle Zahl  $a$  auswählt, die er uns nicht verrät. Anschließend bildet er 10 Zufallszahlen  $a_1, \dots, a_{10}$  im Intervall  $[0, a]$  (gleichverteilt), die er uns mitteilt. Die Aufgabe ist es nun, die Zahl  $a$  zu schätzen.

Diesem Problem liegt also die gleiche Fragestellung wie dem Taxi-Problem oben zu Grunde, mit dem einzigen Unterschied, dass nun kontinuierliche Zufallszahlen im Gegensatz zu den diskreten Taxi-Nummern zugelassen sind.

### Methode 1: Doppelter Mittelwert

Wir bilden den Mittelwert der 10 Zahlen und schätzen, dass  $a$  der doppelte Mittelwert ist, wir betrachten also die Zufallsgröße

$$A = 2M = 2 \cdot \frac{a_1 + \dots + a_{10}}{10} = \frac{2}{10} \sum_{i=1}^{10} a_i$$

Für den doppelten Mittelwert erhält man folgenden Erwartungswert:

$$\mathbb{E}(2M) = \frac{2}{10} \sum_{i=1}^{10} \mathbb{E}(a_i) = 2 \mathbb{E}(a_1)$$

Dabei haben wir verwendet, dass  $a_1, \dots, a_{10}$  gleichverteilt sind, also gleichen Erwartungswert besitzen. Die Dichtefunktion ist gegeben durch  $\rho(x) = \frac{1}{a}$ , somit erhält man

$$\mathbb{E}(a_1) = \int_0^a x \rho(x) dx = \frac{1}{a} \int_0^a x dx = \frac{1}{a} \cdot \frac{a^2}{2} = \frac{a}{2}$$

Also ist tatsächlich  $\mathbb{E}(2M) = a$ . Um die Güte dieser Schätzmethode beurteilen zu können, ist auch die Streuung von  $2M$  von großem Interesse. Wir berechnen nun deshalb die Varianz:

$$\begin{aligned} \text{Var}(2M) &= \frac{4 \cdot 10}{10^2} \cdot \text{Var}(a_1) = \frac{4}{10} \cdot \int_0^a (x - \frac{a}{2})^2 \cdot \rho(x) dx = \frac{4}{10} \cdot \frac{1}{a} \left( \int_0^a x^2 - ax + \frac{a^2}{4} dx \right) = \\ &= \frac{4}{10a} \left[ \frac{x^3}{3} - \frac{ax^2}{2} + \frac{a^2}{4}x \right]_0^a = \frac{a^2}{30} \end{aligned}$$

Wobei hier im ersten Schritt zu Anwendung kam, dass sich die Varianz einer Summe unabhängiger Zufallsgrößen als Summe der Einzelvarianzen ergibt.

### Methode 2: Maximaler Wert

Man könnte genauso gut als Schätzwert den größten auftretenden Wert verwenden. Wir betrachten deshalb die Zufallsgröße  $B = \max\{a_1, a_2, \dots, a_{10}\}$ . Es ist nun

$$V_B(x) = P(B \leq x) = P(\text{alle } a_i \leq x) = \left(\frac{x}{a}\right)^{10}$$

Wir bestimmen zunächst die zu  $B$  gehörige Dichtefunktion  $\rho_B$ . Definiere dazu für ein beliebiges  $\delta \in \mathbb{R}^+$  die beiden Mengen

$$\begin{aligned} M &= \{(a_1, \dots, a_{10}) \mid a_i \leq x \text{ für alle } i \in \{1, \dots, 10\}\} \\ N &= \{(a_1, \dots, a_{10}) \mid a_i \leq x + \delta \text{ für alle } i \in \{1, \dots, 10\}\} \end{aligned}$$

Es ist dann

$$N \setminus M = \{(a_1, \dots, a_{10}) \mid a_i \leq x + \delta \ \forall i \in \{1, \dots, 10\}, \text{ aber } \exists j \in \{1, \dots, 10\} : a_j > x\}$$

also die Menge, in der alle Zahlen kleiner als  $x + \delta$  sind, aber mindestens eine zugleich echt größer als  $x$  ist. Damit ist gerade  $(a_1, \dots, a_{10}) \in N \setminus M$ , wenn für den Maximalwert gilt  $B \in ]x, x + \delta]$ . Wegen  $P(B \in ]x, x + \delta]) \approx \rho_B(x) \cdot \delta$  für genügend kleines  $\delta$  erhalten wir deshalb

$$\rho_B(x) = \frac{1}{\delta} \cdot P(N \setminus M) = \frac{1}{\delta} \cdot [P(N) - P(M)] = \frac{V_B(x + \delta) - V_B(x)}{\delta}$$

Wir können nun einen Grenzübergang  $\delta \rightarrow 0$  durchführen und erhalten für die Wahrscheinlichkeitsdichte

$$\rho_B(x) = \lim_{\delta \rightarrow 0} \frac{V_B(x + \delta) - V_B(x)}{\delta} = V'_B(x) = \frac{d}{dx} \left( \frac{x}{a} \right)^{10} = 10 \cdot \frac{x^9}{a^{10}}.$$

Daraus ergibt sich als Erwartungswert

$$\mathbb{E}(B) = \int_0^a \rho_B(x) \cdot x \, dx = \int_0^a \frac{x^{10}}{a^{10}} \cdot 10 \, dx = \frac{10}{11} a$$

und entsprechend als Varianz:

$$\text{Var}(B) = \left( \int_0^a x^2 \cdot \rho_B(x) \, dx \right) - \left( \frac{10}{11} a \right)^2 = \frac{a^2}{12} \cdot 10 - \frac{10^2}{11^2} a^2 = \frac{10}{11^2 \cdot 12} a^2$$

Wir sehen also, dass der Erwartungswert zu klein ist, d.h. mit dieser Methode das  $a$  systematisch unterschätzt wird, jedoch ist die Varianz bemerkenswert klein.

### Methode 3: Hybrid-Methode

Wir kombinieren die Vorteile der beiden ersten Methoden. Dazu muss die Abweichung im Erwartungswert der zweiten Schätzmethode ausgeglichen werden. Deshalb definieren wir im vorliegenden Fall eine Zufallsgröße

$$C = \frac{11}{10} B = \frac{11}{10} \max\{a_1, a_2, \dots, a_{10}\}$$

Auch hier berechnen wir die Varianz sowie den Erwartungswert und sehen :

$$\mathbb{E}(C) = a \quad \text{und} \quad \text{Var}(C) = \text{Var}\left(\frac{11}{10} B\right) = \frac{11^2}{10^2} \cdot \text{Var}(B) = a^2 \frac{1}{120}$$

**Bemerkung 4.8.** Sind  $n$  Zahlen vorgegeben und die höchste soll geschätzt werden, so lässt sich leicht nachrechnen, dass durch

$$C = \frac{n+1}{n} B$$

eine Verallgemeinerung der Zufallsgröße aus der dritten Methode gegeben ist.

Unter den drei betrachteten Methoden ist die dritte die sinnvollste, da sie den richtigen Erwartungswert hat und gleichzeitig wenig streut. Die allgemein beste Schätzmethode jedoch kann nur schwer bestimmt werden.

### Maximum Likelihood

Zur Illustration dieser Methode beschäftigen wir uns zunächst mit einem konkreten Beispiel.

**Beispiel 4.9** (See mit Fischen). Wir stehen vor einem See und möchten wissen, wie viele Fische darin schwimmen. Dazu gehen wir wie folgt vor: wir fangen eine Zahl  $n$  von Fischen, die wir markieren und dann wieder frei lassen. Anschließend fangen wir erneut eine Anzahl von Fischen und zählen, wie viele darunter markiert sind. Bezeichnen wir die Anzahl markierter Fische in diesem Fang mit  $a$  und die Anzahl nicht-markierter Fische mit  $b$ , so lässt sich aus dem Verhältnis  $\frac{a}{b}$  die Gesamtzahl der Fische im See abschätzen.<sup>1</sup> Dies werden wir nun begründen.

Zunächst schätzen wir die Anzahl nicht-markierter Fische im See, bezeichnet als  $N$ , ab. Wir suchen dazu den Wert für  $N$ , für den die Wahrscheinlichkeit,  $a$  markierte und  $b$  unmarkierte Fische zu fangen, maximal ist. Diese Wahrscheinlichkeit ist gegeben durch

$$P_N = \frac{\binom{n}{a} \cdot \binom{N}{b}}{\binom{N+n}{a+b}}$$

Das Maximum von  $P_N$  schätzen wir ab, indem wir die Stelle suchen, an der  $P_N \approx P_{N-1}$  ist. Wir erhalten

$$\begin{aligned} P_N &= \frac{\binom{n}{a} \cdot \binom{N}{b}}{\binom{N+n}{a+b}} = \binom{n}{a} \frac{N!}{b! \cdot (N-b)!} \frac{(a+b)! \cdot (N+n-a-b)!}{(N+n)!} = \\ &= \binom{n}{a} \cdot \frac{(a+b)!}{b!} \cdot \frac{N! \cdot (N+n-a-b)!}{(N-b)! \cdot (N+n)!} \end{aligned}$$

Analog erhalten wir auch

$$P_{N-1} = \binom{n}{a} \cdot \frac{(a+b)!}{b!} \cdot \frac{(N-1)! \cdot (N+n-a-b-1)!}{(N-b-1)! \cdot (N+n-1)!}$$

Man sieht nun

$$P_N = P_{N-1} \cdot \frac{N(N+n-a-b)}{(N-b) \cdot (N+n)}$$

Um das  $N$  zu finden, für das  $P_N \approx P_{N-1}$  ist, bestimmen wir  $N$  so, dass  $\frac{P_N}{P_{N-1}} = 1$  ist. Dazu berechnen wir

$$\begin{aligned} \frac{N(N+n-a-b)}{(N-b)(N+n)} &\stackrel{!}{=} 1 \\ N(N+n-a-b) &= (N-b)(N+n) \\ N^2 + Nn - Na - Nb &= N^2 - Nb + Nn - nb \\ -Na &= -nb \\ N &= \frac{nb}{a} \end{aligned}$$

<sup>1</sup>Um die Anzahl der Ratten in New York zu bestimmen, wollte der Statistiker Jonathan Auerbach genauso vorgehen, jedoch beschreibt er in seinem *Fachartikel* dazu folgendes Problem: "Unfortunately, NYC's Department of Health and Mental Hygiene is unlikely to approve a large-scale rat-releasing experiment (I know, because I asked)."

Hat man also  $a$  markierte und  $b$  nicht-markierte Fische gefangen, nachdem zuvor  $n$  Fische markiert wurden, so ist die Zahl nicht-markierter Fische im ganzen See höchstwahrscheinlich gleich  $\frac{nb}{a}$ .

*Zusammenfassung:* Falls die Wahrscheinlichkeitsfunktion von einem (oder mehreren) Parametern abhängt, können wir die Werte berechnen, für die die Wahrscheinlichkeit der vorgegebenen Parameterwerte maximal wird. Dies nennt man *Maximum-Likelihood-Verfahren*.

**Beispiel 4.10** (Maximum-Likelihood bei Poisson-Verteilung). Man denke an die Versicherung gegen Brand von oben. Wir nehmen an, dass im letzten Jahr  $n$  Brände passiert sind und diese Poisson-verteilt sind. Für welches  $\lambda$  ist dieser Ausgang am wahrscheinlichsten?

Die Wahrscheinlichkeitsfunktion der Poisson-Verteilung ist gegeben durch

$$P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Wir suchen also das Maximum von  $P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda}$ , wobei wir  $\lambda$  als Variable betrachten und der Parameter  $n$  für die Zahl der Brände steht, die sich im letzten Jahr ereignet haben. Mit der Produktregel erhalten wir folgende Ableitung

$$\frac{d}{d\lambda} P_\lambda(n) = \frac{n\lambda^{n-1}}{n!} e^{-\lambda} + \frac{\lambda^n}{n!} (-1) e^{-\lambda} = \frac{e^{-\lambda}}{n!} \lambda^{n-1} (n - \lambda)$$

Die einzige (sinnvolle) Nullstelle der Ableitung ist bei  $\lambda = n$ . An dieser Stelle befindet sich also das Maximum  $P_\lambda(n)$  und damit ist  $\lambda = n$  derjenige Wert für  $\lambda$ , bei dem es am wahrscheinlichsten ist, dass  $n$  Brände pro Jahr auftreten. Bei der Poisson-Verteilung ist die Anzahl der Ereignisse in einem bestimmten Zeitraum also zugleich der beste Schätzwert für den Parameter  $\lambda$ .

Angenommen, es liegen mehrere Werte vor, z.B. 5 Brände im letzten Jahr, und 3 Brände im vorletzten Jahr. Man würde annehmen, dass die beste Wahl für  $\lambda$  dann 4 wäre. Wir überprüfen dies, in dem wir erneut das Maximum ausrechnen, diesmal für die Wahrscheinlichkeit  $P_\lambda(5 \text{ Brände, dann } 3 \text{ Brände})$ . Man erhält

$$P_\lambda(5 \text{ Brände, dann } 3 \text{ Brände}) = \frac{\lambda^5}{5!} e^{-\lambda} \cdot \frac{\lambda^3}{3!} e^{-\lambda} = \frac{\lambda^8}{3! \cdot 5!} e^{-2\lambda}$$

Wiederum betrachten wir die Ableitung nach  $\lambda$

$$\frac{d}{d\lambda} = \frac{1}{5! \cdot 3!} (8\lambda^7 e^{-2\lambda} - 2\lambda^8 e^{-2\lambda}) = \frac{\lambda^7 e^{-2\lambda}}{5! \cdot 3!} (8 - 2\lambda)$$

Und tatsächlich erhalten wir als (sinnvolle) Nullstelle der Ableitung  $\lambda = 4$ , wie erwartet.

**Beispiel 4.11** (Maximum-Likelihood bei zwei Parametern). Gegeben sei eine normalverteilte Zufallsgröße  $X$ , d.h. die zugehörige Dichte habe die Form

$$\rho(x) = C \cdot \frac{1}{\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

mit dem Normierungsfaktor  $C = \frac{1}{\sqrt{2\pi}}$  und den zwei Parametern  $\sigma$  und  $\mu$ . Wir führen ein Experiment durch und erhalten den Wert  $x_1$ . Lediglich anhand dieses einen Messwertes die beiden Parameter  $\mu$  und  $\sigma$  zu schätzen, macht wenig Sinn. Rein formal ist dann zwar das Maximum bei  $\sigma = 0$  und  $\mu = x_1$ , doch ist dies aus Modellierungssicht wenig hilfreich (es handelt sich dabei um die konstante Zufallsgröße  $\omega \mapsto x_1$ ).

Bei zwei Messwerten  $x_1$  und  $x_2$  bestimmen sich  $\sigma$  und  $\mu$  mittels Maximum Likelihood wie folgt:

$$\begin{aligned} P_{\mu,\sigma}(x_1, x_2) &= \left( C \cdot \frac{1}{\sigma} \cdot \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right) \right) \cdot \left( C \cdot \frac{1}{\sigma} \cdot \exp\left(-\frac{(x_2-\mu)^2}{2\sigma^2}\right) \right) = \\ &= C^2 \frac{1}{\sigma^2} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

Von dieser Funktion bestimmen wir nun das Maximum mittels Ableiten nach  $\mu$  bzw.  $\sigma$ . Um das Maximum bezüglich  $\mu$  zu erhalten, bestimmen wir die Nullstellen der partiellen Ableitung nach  $\mu$ :

$$\begin{aligned} \frac{\partial}{\partial \mu} P_{\mu,\sigma}(x_1, x_2) &= C^2 \cdot \frac{1}{\sigma^2} \cdot \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2}\right) \cdot \left[ -\frac{2(x_1-\mu)}{2\sigma^2} - \frac{2(x_2-\mu)}{2\sigma^2} \right] \stackrel{!}{=} 0 \\ &\Leftrightarrow -x_1 + \mu - (x_2 - \mu) = 0 \quad \Leftrightarrow 2\mu = x_1 + x_2 \\ &\Leftrightarrow \mu = \frac{x_1 + x_2}{2} \end{aligned}$$

Analoges Vorgehen für  $\sigma$  liefert:

$$\begin{aligned} &\frac{\partial}{\partial \sigma} P_{\mu,\sigma}(x_1, x_2) \stackrel{!}{=} 0 \\ \Leftrightarrow &\frac{-2C^2}{\sigma^3} \cdot \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2}\right) + \frac{C^2}{\sigma^2} \cdot \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2}\right) \cdot \left[ -\frac{(x_1-\mu)^2 + (x_2-\mu)^2}{2} \cdot \frac{-2}{\sigma^3} \right] = 0 \\ &\Leftrightarrow \frac{C^2}{\sigma^2} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2}\right) \left[ -2 + \frac{-2}{2\sigma^2} \left( -(x_1-\mu)^2 - (x_2-\mu)^2 \right) \right] = 0 \\ \Leftrightarrow &-2 + \frac{-2}{2\sigma^2} \left[ -(x_1-\mu)^2 - (x_2-\mu)^2 \right] = 0 \quad \Leftrightarrow 1 = \frac{1}{2\sigma^2} \left[ (x_1-\mu)^2 + (x_2-\mu)^2 \right] \\ &\Leftrightarrow \sigma^2 = \frac{1}{2} \left[ (x_1-\mu)^2 + (x_2-\mu)^2 \right] \end{aligned}$$

Man sieht also, dass das auf diese Weise geschätzte  $\mu$  dem Mittelwert von  $x_1$  und  $x_2$  entspricht und  $\sigma$  der gemittelten quadratischen Abweichung von  $\mu$ . Da der Erwartungswert dem langfristigen Mittelwert und die Varianz der quadratischen Abweichung entspricht, erscheint dieses Resultat plausibel.

Die Ergebnisse verallgemeinern sich für  $n$  vorliegende Messwerte in entsprechender Weise als Mittelwert bzw. gemittelte quadratische Abweichung.

### 4.3 Das Neyman-Pearson-Lemma

Wir kehren nochmals zu Hypothesentests zurück und betrachten für eine reelle Zufallsgröße  $X$  die spezielle Situation

$$H_0 : X \text{ hat Dichte } \rho_0$$

$$H_1 : X \text{ hat Dichte } \rho_1$$

und fragen uns nun, wie man hier den Hypothesentest optimieren kann. Anders ausgedrückt: für einen vorgegebenen  $\alpha$ -Fehler suchen wir unter allen sinnvollen Annahmebereichen  $A \subseteq \mathbb{R}$  denjenigen, für den der  $\beta$ -Fehler minimal wird. Es soll also gelten

$$P_{H_0}(X \in A) = 1 - \alpha \quad \text{und} \quad \beta = P_{H_1}(X \in A) \text{ ist minimal}$$

Dazu betrachten wir nun die Funktion  $\frac{\rho_0(x)}{\rho_1(x)}$ ,

wobei wir dabei formal  $\frac{\rho_0(x)}{\rho_1(x)} = +\infty$  setzen, falls  $\rho_1(x) = 0$  ist. Weiterhin wählen wir ein  $C \in \mathbb{R}^+$  (zunächst beliebig) und bilden die Menge

$$A_C = \left\{ x \in \Omega \mid \frac{\rho_0(x)}{\rho_1(x)} \geq C \right\}$$

Anschließend suchen wir dasjenige  $C$ , für das

$$P_{H_0}(X \in A_C) = 1 - \alpha$$

erfüllt ist und nennen das zugehörige  $A_C$  dann  $A$ . Dabei setzen wir stillschweigend voraus, dass es ein solches  $C$  gibt. Wir zeigen nun, dass diese Methode einen optimalen Annahmehereich  $A$  liefert.

**Lemma 4.12** (Neyman-Pearson). Mit den Bezeichnungen von oben gilt:

Ist  $B \subseteq \mathbb{R}$  eine Menge mit

$$P_{H_0}(X \in B) = 1 - \alpha$$

so gilt  $P_{H_1}(X \in B) \geq P_{H_1}(X \in A)$ . Dies bedeutet, dass der  $\beta$ -Fehler mit  $B$  als Annahmehereich größer ist, als mit  $A$  als Annahmehereich.

*Beweis.* Sei  $B$  eine Menge mit  $P_{H_0}(X \in B) = 1 - \alpha$  und definiere

$$I_1 = A \setminus B \quad \text{und} \quad I_2 = B \setminus A$$

Da nach Annahme  $P_{H_0}(X \in B) = P_{H_0}(X \in A)$  gilt, ist auch

Also ist insbesondere

$$\int_{I_1} \rho_0(x) \, dx = \int_{I_2} \rho_0(x) \, dx$$

Außerdem gilt  $\frac{\rho_0(x)}{\rho_1(x)} \leq c$  falls  $x \in I_2$  und  $\frac{\rho_0(x)}{\rho_1(x)} \geq c$ , falls  $x \in I_1$ . Die Änderung des  $\beta$ -Fehlers ist nun:

$$\begin{aligned} P_{H_1}(X \in A) - P_{H_1}(X \in B) &= \int_A \rho_1(x) \, dx - \int_B \rho_1(x) \, dx = \\ &= \int_{I_1} \rho_1(x) \, dx - \int_{I_2} \rho_1(x) \, dx + \int_{A \cap B} \rho_1(x) \, dx - \int_{A \cap B} \rho_1(x) \, dx \\ &\leq \frac{1}{c} \int_{I_1} \rho_0(x) \, dx - \frac{1}{c} \int_{I_2} \rho_0(x) \, dx = 0 \end{aligned}$$

Also wird der Fehler größer. □

## 4.4 Der $\chi^2$ -Verteilungstest

Wir betrachten einen beliebigen Würfel, wobei uns als Ergebnis wiederholten Werfens folgende Daten vorliegen:

$j$	1	2	3	4	5	6
$n_j$	11	7	8	9	13	12

Dabei bezeichnet  $j$  die Augenzahl und  $n_j$  die jeweilige Häufigkeit, mit der diese Augenzahl im Experiment aufgetreten ist. Wir stellen nun die Nullhypothese

$$H_0 : \text{Würfel ist Laplace}$$

und wollen diese mathematisch überprüfen. Wir könnten nun für jede einzelne Augenzahl  $j$  die Nullhypothese  $p_j = \frac{1}{6}$  für die Wahrscheinlichkeit  $p_j$ , die Augenzahl  $j$  zu erhalten, aufstellen und anschließend in bekannter Weise in einem Hypothesentest überprüfen. Da dies jedoch bereits bei einem solch einfachen Beispiel vergleichsweise aufwendig ist, überlegen wir uns stattdessen eine Methode, die es uns erlaubt, gleich eine ganze Verteilung zu testen.

Dazu betrachten wir die Größe

$$\sum_{j=1}^6 \frac{(np_j - n_j)^2}{np_j}$$

wobei  $n$  die Summe der  $n_j$  bezeichnet. Diese Summe beschreibt anschaulich eine gewichtete Abweichung vom Mittelwert. Die Entscheidungsregel lautet nun, dass die Nullhypothese zu verwerfen ist, falls der Wert dieser Summe einen gewissen Wert überschreitet. Wieso dies sinnvoll ist, führen wir im Folgenden aus.

Gegeben seien unabhängige reelle Zufallsvariablen  $X_1, \dots, X_f$ , die standard-normalverteilt sind. Die Frage ist, wie die Verteilung von  $X_1^2$ , von  $X_1^2 + X_2^2$ , von  $X_1^2 + X_2^2 + X_3^2 + \dots$  aussieht. Die entsprechende Verteilung nennt man dann  $\chi^2$ -Verteilung. Diese hängt von der Anzahl der betrachteten Zufallsgrößen  $f$  ab, welche häufig wird auch als *Anzahl der Freiheitsgrade* bezeichnet wird. Bevor wir die genaue Form der  $\chi^2$ -Verteilung berechnen, werden wir zunächst die Verbindung des Beispiels mit dieser Verteilung aufzeigen.

**Beispiel 4.13.** Sei zunächst  $j \in \{1, 2\}$ . Wir zeigen, dass

$$\sum_{j=1}^2 \frac{(nP_j - n_j)^2}{nP_j}$$

einer  $\chi^2$ -Verteilung mit Freiheitsgrad  $f = 1$  genügt.

Definiere dazu

$$Z = \frac{n_1 - nP_j}{\sqrt{np_j q_j}}$$

Diese Zufallsgröße ist Standard-Normalverteilt. Also ist  $Z^2$  per Definition  $\chi^2$ -verteilt mit Freiheitsgrad 1. Nun ist

$$Z^2 = \frac{(n_1 - np_1)^2 (p_1 + q_1)}{np_1 q_1} = \frac{(n_1 - np_1)^2 p_1}{np_1 q_1} + \frac{(n_1 - np_1)^2 q_1}{np_1 q_1} = \frac{(n_1 - np_1)^2}{nq_1} + \frac{(n_1 - np_1)^2}{np_1}$$

Da  $n_1 = n - n_2$  und  $np_1 = n - nq_1$  ist, ergibt sich

$$\frac{(n - n_2 - (n - nq_1))^2}{nq_1} + \frac{(n - n_2 - n + nq_1)^2}{n - nq_1} = \frac{(nq_1 + n_2)^2}{n}$$

**Bemerkung 4.14.** Ähnlich ließe sich zeigen, dass die Formel von oben für beliebiges  $j \in \{1, 2, \dots, k\}$  einer  $\chi^2$ -Verteilung mit Freiheitsgrad mit  $k - 1$  gehorcht.



Wir bestimmen nun die Dichtefunktion der  $\chi^2$ -Verteilung mit Freiheitsgrad  $f$ . Sei dazu  $X$  standard-normalverteilt und  $Z$  die Zufallsgröße gegeben durch  $Z = X^2$

1. Fall:  $f = 1$ . Sei  $V(t)$  die Verteilungsfunktion von  $Z$ , also

$$V(t) = P(Z \leq t) = P(-\sqrt{t} \leq Y \leq \sqrt{t}) = \Phi(\sqrt{t}) - \Phi(-\sqrt{t})$$

Die Dichte von  $Z$  ist gegeben durch die Ableitung der Verteilungsfunktion, also erhalten wir für  $t > 0$

$$\begin{aligned} \frac{d}{dt}V(t) &= \frac{d}{dt}\Phi(\sqrt{t}) - \frac{d}{dt}\Phi(-\sqrt{t}) = \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\sqrt{t}^2}{2}\right) \cdot \frac{2}{2\sqrt{t}} + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(-\sqrt{t})^2}{2}\right) \cdot \frac{1}{2\sqrt{t}} \end{aligned}$$

Also Insgesamt

$$\rho(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t}} e^{-\frac{t}{2}}$$

2. Fall:  $a = 2$ . Sei nun also  $Z = X_1^2 + X_2^2$  und  $\eta$  die Dichte von  $Z$ . Aus den Übungen wissen wir, dass  $\eta$  durch die Faltung der beiden Dichten gegeben ist, also

$$\begin{aligned} \eta(t) = (\rho * \rho)(t) &= \int_0^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s}{2}\right) \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t-s}} \exp\left(-\frac{(t-s)}{2}\right) ds = \\ &= \frac{1}{2\pi} e^{-\frac{t}{2}} \cdot \int_0^t \frac{1}{\sqrt{s(t-s)}} ds = \frac{e^{-\frac{t}{2}}}{2} \end{aligned}$$

Den Wert des bestimmten Integrals  $\int_0^t \frac{1}{\sqrt{s(t-s)}} ds$  bestimmt man durch eine geeignete Substitution oder durch das Tafelwerk als  $\pi$ .

**Bemerkung 4.15.** Man kann auch dies auf eine allgemeine Formel verallgemeinern, nämlich

$$\rho(t) = \frac{1}{2^{\frac{n}{2}} \left(\frac{n}{2} - 1\right)!} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}$$

Wobei man folgende Konvention für Fakultäten nicht-ganzzahliger Zahlen verwendet

$$k! = k(k-1)! \quad \text{und} \quad \left(\frac{1}{2}\right)! = \sqrt{\pi}$$

Nun wissen wir für den Würfelwurf: Stimmt  $H_0$ , so ist  $\sum_{j=1}^6 \frac{(np_j - n_j)^2}{np_j}$  nach unseren Überlegungen  $\chi^2$ -verteilt mit  $f = 5$ . Das Vorgehen ist nun ein "normaler" Hypothesentest: Ist  $\alpha = 5\%$  gegeben, so bestimmen wir den Annahmehereich gemäß des Neyman-Pearson-Lemmas.

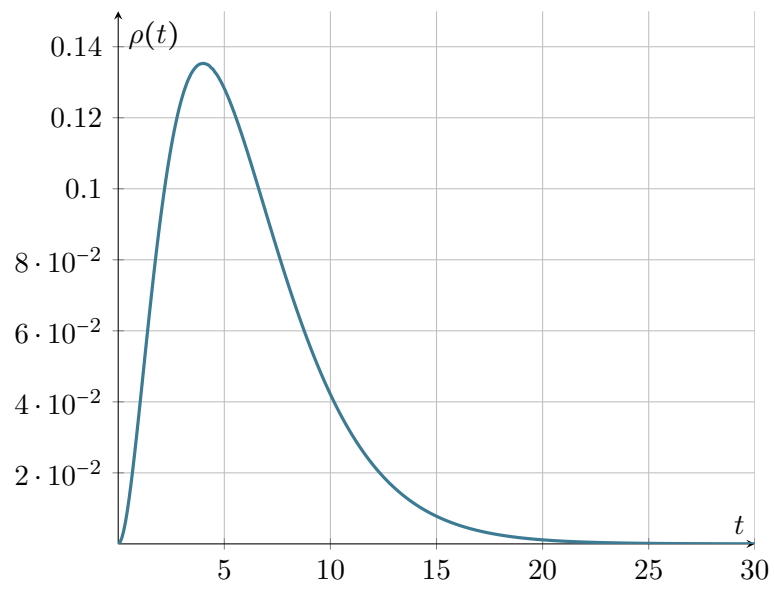


Abbildung 4.2: Dargestellt ist die  $\chi^2$ -Verteilung zu  $f = 6$ , d.h.  $\rho(t) = \frac{1}{16}t^2e^{-t/2}$ .

# Index

- $\chi^2$ -Verteilung, 72
- $\sqrt{n}$ -Gesetz, 51
- $\sigma$ -Algebra
  - Borelsche, 14
  - Definition, 13
  - Eigenschaften, 13, 14
  - erzeugte, 14
- Annahmehereich, 61
- bedingte Wahrscheinlichkeit, 7
- Binomialkoeffizient, 7
- Binomialverteilung, 30, 35, 38
- Borelmenge, 14
- Cauchy-Schwarz'sche Ungleichung, 41
- Dartwurf, 26, 29
- Dichtefunktion, 16, 32
- Ereignisraum, 2
- Ergebnisraum, 2
- Fehlerniveau, 61
- Freiheitsgrad, 72
- Gesetz der großen Zahlen,
  - Schwaches, 45
  - Starkes, 46
- Glücksrad, 2, 11, 50
- Grundmenge, 2
- Hypothesentest
  - einseitig, 62
  - zweiseitig, 61
- Kolmogoroff-Axiome
  - für überabzählbare  $\Omega$ , 11
  - für endliche  $\Omega$ , 3
- Konvergenz
  - fast sicher, 53
  - gleichmäßig, 52
  - in Wahrscheinlichkeit, 53
  - punktweise, 52
  - stochastisch, 53
- Korrelationskoeffizient, 41
- Kovarianz, 39
- Laplace-Annahme, 4
- Maximum Likelihood, 68
- Messbarkeit, 17, 19
- Mississippi-Problem, 6
- Neyman-Pearson-Lemma, 71
- Normalverteilung, 32, 36
- Nullhypothese, 61
- Poisson-Verteilung, 34, 36
- Satz
  - vom Gegenereignis, 3
  - von Bayes, 8
  - von der totalen Wahrscheinlichkeit, 8
  - von Moivre-Laplace, 58
- Signifikanzniveau, 61
- Stetigkeitskorrektur, 60
- Taxi-Problem, 65
- Unabhängigkeit
  - von Ereignissen, 9, 22
  - von Zufallsgrößen, 23, 25
- Ungleichung
  - von Markov, 43
  - von Tschebyscheff, 44
- Varianz, 38
- Verteilungsfunktion, 25
- Vitali-Menge, 12
- von einem Pferd zu Tode getretene Soldaten, 33
- Wahrscheinlichkeitsdichte, 16
- Wahrscheinlichkeitsmaß, 3, 11, 26
- Zentraler Grenzwertsatz, 58
- Ziegenproblem, 4
- Zufallsgröße, 17, 18
- Zufallsvariable, 17, 18