

# Elementare Wahrscheinlichkeitstheorie und Statistik

Kurzfassung: Definitionen und Sätze

Beweise und Beispiele werden in der Vorlesung behandelt

Vorlesung SS 2003 für Bio-Informatiker und Lehramtskandidaten

**Prof. Dr. Gerd Wittstock**

Universität des Saarlandes

FR 6.1 Mathematik

Version: 9. Juli 2003

## Inhaltsverzeichnis

<b>1</b>	<b>Endliche W-Räume</b>	<b>1</b>
1.1	W-Maß und W-Funktion . . . . .	1
1.2	Elementare Kombinatorik . . . . .	3
1.3	Rechnen mit Indikatorfunktionen . . . . .	6
1.4	Produkt von endl. W-Räumen . . . . .	8
1.5	Bernoulli- und Binomialverteilung . . . . .	11
1.6	Zufallsvariable und ihre Verteilung . . . . .	13
1.7	Hypergeometrische Verteilung . . . . .	15
1.8	Gemeinsame Verteilung von Z-Var. . . . .	17
1.9	Unabhängige Zufallsvariable . . . . .	19
1.10	Unabhängige Ereignisse . . . . .	20
1.11	Bedingte Wahrscheinlichkeit . . . . .	22
1.12	Zusammengesetzte W-Maße . . . . .	24
1.13	Mehrstufige Experimente . . . . .	26
1.14	Polya's Urnenmodell . . . . .	29
1.15	Erwartungswert . . . . .	30
1.16	Bedingte Erwartung . . . . .	33
1.17	Erwartungswert des Produktes . . . . .	35
1.18	Varianz und Kovarianz . . . . .	36
1.19	Bedingte Varianz . . . . .	39
1.20	Schwaches Gesetz der großen Zahl . . . . .	40
<b>2</b>	<b>Diskrete und stetige W-Räume</b>	<b>44</b>
2.1	Uniforme und geometrische Verteilung . . . . .	44
2.2	Diskrete W-Räume . . . . .	45
2.3	Konvergenz gegen geomtr. Vertlng. . . . .	49
2.4	Poisson-Verteilung . . . . .	50
2.5	Allgemeine W-Räume . . . . .	53
2.6	Zufallsvariable und ihre Verteilung . . . . .	55
2.7	Eindeutigkeit eines W-Maßes . . . . .	56
2.8	Verteilungsfunktion . . . . .	56
2.9	Dichtefunktion, stetige Verteilungen . . . . .	58
2.10	Exponentialverteilung . . . . .	59
2.11	Normalverteilung . . . . .	61
2.12	Grenzwertsatz von Moivre-Laplace . . . . .	62

## 1 Endliche W-Räume

In diesem Kapitel werden einige Grundbegriffe der Wahrscheinlichkeitstheorie (W-Theorie) für den besonders einfachen Fall eines endlichen Wahrscheinlichkeitsraumes (W-Raumes) eingeführt und einfache Folgerungen gezogen.

Die genaue Regelung der Bezeichnungen mag für den endlichen Fall etwas spitzfindig klingen, bietet aber die folgenden Vorteile:

- Die Definitionen, Bezeichnungen und Folgerungen gelten für allgemeinere W-Räume sinngemäß weiter. Es kommen nur noch einige mathematische Feinheiten hinzu, die im endlichen Fall keine Bedeutung haben.
- Die genaue Festlegung der Bezeichnungen erlaubt es, in Beispielen kurz und präzise zu sagen, wie das Modell aussehen soll und was zu berechnen ist.
- In den zugehörigen Übungen geht es häufig darum, umgangssprachlich formulierte Probleme, sogenannte *Textaufgaben*, in die streng geregelte Sprache der W-Räume zu übersetzen. Die anschließende *W-Rechnung* ist dann zumeist nur noch *Fleißarbeit*. Bei endlichen W-Räumen kommt man mit den Grundrechenarten „+ – · /“ und etwas Elementarmathematik aus.

### 1.1 W-Maß und W-Funktion

#### 1.1.1 Bez. (endlicher Grundraum)

**Math.:** Der Grundraum  $\Omega$  ist eine nichtleere endliche Menge. Die Elemente werden mit  $\omega \in \Omega$  bezeichnet. Die Anzahl der Elemente  $|\Omega| \in \mathbb{N}$  ist eine natürliche Zahl.

$2^\Omega$  bezeichnet die Potenzmenge von  $\Omega$ . Für die Teilmengen von  $\Omega$  verwenden wir die Buchstaben  $A, B \in 2^\Omega$ .

**Modell:** Ein Experiment mit endlich vielen möglichen Ausgängen.  $\Omega$  ist die **Ergebnismenge** oder der **Ergebnisraum** des Experimentes und enthält alle möglichen Ausgänge und eventuell auch einige unmögliche Ausgänge.

Es kann aus mathematischen Gründen praktisch sein, einige unmögliche Ausgänge hinzuzunehmen, wenn  $\Omega$  dadurch einfacher anzugeben ist ( $\rightarrow$  z. B. Bemerkung 1.5.5)

Die Elemente  $\omega$  nennt man die **Elementarereignisse** des Experiments. Die Teilmengen  $A \in 2^\Omega$  heißen **Ereignisse**

Je nach Beispiel sind noch weitere Bezeichnungen für  $\Omega$  üblich: *Stichprobenraum* ...

#### 1.1.2 Bsp. (Würfelexperiment)

**Ergebnisraum:**  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,

**Elementarereignis:**  $6 \in \Omega$ . Es ist eine 6 gewürfelt worden.

**Ereignis**  $A := \{2, 4, 6\} \in 2^\Omega$ . Es ist eine gerade Augenzahl gewürfelt worden.

#### 1.1.3 Bez. (Wahrscheinlichkeitsfunktion)

**Math.:** Eine W-Funktion  $p$  ist eine Funktion

$$p : \Omega \rightarrow [0, 1] \quad \text{mit} \quad \sum_{\omega \in \Omega} p_\omega = 1.$$

Für eine Teilmenge  $A \in 2^\Omega$  setzt man

$$P(A) := \sum_{\omega \in A} p_\omega$$

und nennt die Zahl  $P(A)$  die Wahrscheinlichkeit von  $A$ . Die Abbildung

$$P : 2^\Omega \rightarrow [0, 1]$$

heißt das zu  $p$  gehörende **Wahrscheinlichkeitsmaß**.

Man kann die W-Funktion aus dem W-Maß auf einfache Weise rekonstruieren. Offensichtlich gilt  $p_\omega = P(\{\omega\})$ .

**Modell:** Die Ergebnisse eines Experiments sind zufällig. Für jedes Elementarergebnis  $\omega$  setzt man die Wahrscheinlichkeit  $p_\omega$  fest, daß  $\omega$  eintritt. Die Wahrscheinlichkeiten  $p_\omega$  sind nichtnegative Zahlen und so *normiert*, daß die Summe der Wahrscheinlichkeiten 1 ist.

Ein Ereignis  $A$  tritt ein, wenn der Ausgang  $\omega$  des Experimentes in  $A$  liegt. Die Wahrscheinlichkeit, das das Ereignisses  $A$  eintritt, ist die Summe der Wahrscheinlichkeiten  $p_\omega$  mit  $\omega \in A$ .

Man interpretiert die Werte  $P(A) = 0$  als *unmögliches* Ereignis und  $P(A) = 1$  als *sicheres* Ereignis.

**Anmerkung.** Man findet in der Literatur ( $\rightarrow$  [7][Hinderer]) für die W-Funktion  $p$  auch die Bezeichnung *Zähldichte*. Die Begründung hierfür ist, daß auf endlichen W-Räumen die Funktion  $p$  die Dichte des W-Maßes  $P$  bezüglich des *Zählmaßes* ist. Das *Zählmaß* ordnet jeder Teilmenge die Anzahl ihrer Elemente zu.

Bei „*kontinuierlichen*“ W-Maßen tritt an die Stelle des *Zählmaßes* häufig das Lebesguesche Maß (Länge, Fläche bzw. Rauminhalt). In diesen Problemen haben die W-Maße eine Dichtefunktion  $p$  bezüglich des Lebesgueschen Maßes.

#### 1.1.4 Bsp. (fairer Würfel)

**Ergebnisraum:**  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,

**W-funktion:** Fairer Würfel bedeutet, daß alle sechs Elementarereignis gleichwahrscheinlich sind. Also ist:

$$p_\omega = \frac{1}{6} \quad \text{für } \omega = 1, 2, \dots, 6.$$

**W-Maß:** Das Ereignis, es ist eine gerade Augenzahl gewürfelt worden, hat die Wahrscheinlichkeit

$$P(\{2, 4, 6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

**Anmerkung.** Wir werden später auch Zufallsexperimente mit einer *kontinuierlichen* Zahl von Elementarereignissen untersuchen. Der Grundraum ist dann z. B. ein Intervall  $\Omega := [a, b] \subset \mathbb{R}$ .

Man denke an eine zufällige Drehung mit Drehwinkel  $\omega \in [0, 2\pi)$ . In diesem Modell ist die Wahrscheinlichkeit, daß der Drehwinkel exakt  $45^\circ$  ist, wohl 0. Die Wahrscheinlichkeit, daß der Drehwinkel im ersten Quadranten liegt wird  $1/4$  sein. Etwas allgemeiner:

$$P(\{\omega\}) = 0 \quad \text{und} \quad P([a, b]) = \frac{b-a}{2\pi} \quad \text{für } 0 \leq a \leq b < 2\pi.$$

Wie man sieht, ist nun die Funktion  $\omega \mapsto P(\{\omega\}) \equiv 0$  nicht sehr interessant! Die richtige Verallgemeinerung der W-Funktion ist die Dichtefunktion, die aber i.a. nicht so einfach zu erklären ist.

Im Hinblick auf allgemeinere Situationen stützen wir die axiomatische Beschreibung endlicher W-Räume auf den Begriff des W-Maßes und nicht auf den zunächst einfacher erscheinenden Begriff der W-Funktion. Die W-Funktion ist aber ein sehr nützliches Hilfsmittel bei der Berechnung des W-Maßes.

#### 1.1.5 Def. (endlicher W-Raum)

Ein *endlicher Wahrscheinlichkeitsraum* ist ein Paar  $(\Omega, P)$ , bestehend aus einer nichtleeren endlichen Menge  $\Omega$  und einer Abbildung  $P : 2^\Omega \rightarrow [0, \infty)$  mit den folgenden Eigenschaften:

- (i)  $P(\Omega) = 1$ ,  
(ii)  $P(A \cup B) = P(A) + P(B)$  für disjunkte  $A, B \in 2^\Omega$ .

$P$  heißt *W-Maß* auf  $\Omega$ .

Die Eigenschaft (i) ist eine sinnvolle Normierung. Die Eigenschaft (ii) heißt (endliche) Additivität.

**Anmerkung.** Im physikalischen Sprachgebrauch bezeichnet man W-Maße als *Statistiken*, z. Bsp. Boltzmann-Statistik in der kinetischen Gastheorie. Das kollidiert mit dem sonstigen Gebrauch des Wortes *Statistik*.

### 1.1.6 Festst. (Rechenregeln für W-Maße)

Es seien  $A, B, A_1, \dots, A_n \in 2^\Omega$ .

Wenn  $A_1, \dots, A_n$  paarweise disjunkt sind, gilt:

$$P\left(\bigcup_{\nu=1}^n A_\nu\right) = \sum_{\nu=1}^n P(A_\nu) \quad (1.1.1)$$

Für beliebige  $A_1, \dots, A_n$  gilt:

$$P\left(\bigcup_{\nu=1}^n A_\nu\right) \leq \sum_{\nu=1}^n P(A_\nu); \quad (1.1.2)$$

Speziell ist:

$$P(A^c) + P(A) = 1 \quad (1.1.3)$$

$$P(A \setminus B) + P(A \cap B) = P(A); \quad (1.1.4)$$

$$P(\emptyset) = 0; \quad (1.1.5)$$

$$P(A \cup B) + P(A \cap B) = P(A) + P(B), \quad (1.1.6)$$

Wenn  $A \subset B$  ist, dann gilt

$$P(A) \leq P(B). \quad (1.1.7)$$

### 1.1.7 Satz (W-Maß $\Leftrightarrow$ W-Funktion)

Es sei  $\Omega$  eine nichtleere, endliche Menge.

- (i) Ist  $P$  ein W-Maß auf  $\Omega$ , so ist

$$p : \Omega \rightarrow [0, 1] \quad \text{mit } p_\omega := P(\{\omega\})$$

eine W-funktion auf  $\Omega$ .

- (ii) Zu einer W-Funktion  $p : \Omega \rightarrow [0, 1]$  gibt es genau ein W-Maß  $P : 2^\Omega \rightarrow [0, 1]$  mit  $p_\omega = P(\{\omega\})$ .

In beiden Fällen gilt

$$P(A) = \sum_{\omega \in A} p_\omega$$

**Anmerkung.** 1. Der axiomatische Zugang zur W-Theorie (Kolmogorov 1933) erklärt den W-Raum durch die Regeln (Axiome), die ein W-Raum erfüllen muß. Im Falle eines endlichen W-Raumes sind dies die Regeln aus Definition 1.1.5.

2. Diese Regeln enthalten keine Angaben, wie in einem konkreten Beispiel die Wahrscheinlichkeiten festzulegen sind.

Dazu benötigt man weitere Informationen über das konkrete Modell. Folgerungen aus den Axiomen helfen dann bei der Festlegung der Wahrscheinlichkeiten, wie wir in dem sehr einfachen (!) Beispiel 1.1.4 gesehen haben.

### 1.1.8 Bem. (Häufigkeitsinterpretation)

Man wiederhole ein Zufallsexperiment sehr häufig und achte darauf, daß die verschiedenen Experimente sich nicht beeinflussen. Man sagt hierzu, die Experimente sind unabhängig.

Bei  $n$ -maliger Wiederholung erhält man die Ergebnisse<sup>1</sup>

$$x_1, x_2, x_3, \dots, x_n.$$

mit  $x_\nu \in \Omega$ . Für eine Teilmenge  $A \in 2^\Omega$  bezeichne

$$R_n(A) := \frac{1}{n} \sum_{\nu=1}^n \mathbb{1}_A(x_\nu)$$

die *relative Häufigkeit*, mit der das Ereignis  $A$  in den  $n$  Experimenten eingetreten ist. Dabei ist  $\mathbb{1}_A$  die charakteristische Funktion oder Indikatorfunktion von  $A$ :

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

Man wird naiv erwarten, daß für große  $n$  die relative Häufigkeit  $R_n(A)$  ungefähr gleich der Wahrscheinlichkeit  $P(A)$  ist.

### 1.1.9 Bez. (empirische Wahrscheinlichkeit)

Die bei der Wiederholung zufälliger Experimente unter gleichen Bedingungen beobachteten relativen Häufigkeiten verwendet man als sogenannte *empirische Wahrscheinlichkeiten*. Bei festem  $n$  ist

$$2^\Omega \ni A \mapsto R_n(A)$$

ein W-Maß auf  $\Omega$ .

**Anmerkung.** Hierauf beruht die *statistische Schätzung* der Wahrscheinlichkeit, z. Bsp. Sterbetafeln. Man möchte aufgrund von beobachteten relativen Häufigkeiten in einer Versuchsserie Vorhersagen für die relative Häufigkeit in einer zukünftigen Versuchsserie machen.

Dies setzt voraus, das wir die Ergebnisse des Experiments als *zufällig* betrachten können. D. h., es gibt keinen Anlaß an der Erfahrungstatsache zu zweifeln, daß bei Wiederholungen des Experiment mit mindestens derselben Anzahl von Versuchen, die relative Häufigkeiten nur sehr wenig schwanken werden.

**Problem:** Was *sehr wenig* ist, hängt von der Fragestellung ab. Wie groß soll man  $n$  wählen. Wie groß ist die Wahrscheinlichkeit, daß das beobachtete  $n$ -Tupel  $x_1, x_2, x_3, \dots, x_n$  ein *Ausreißer* ist: *n-mal Kopf beim Münzwurf!*

Die axiomatische W-Theorie leistet bei der Beantwortung dieser Frage Hilfestellung  $\rightarrow$  Schwaches Gesetz der großen Zahl.

### 1.1.10 Bez. (Laplace-Wahrscheinlichkeit)

**Math.:** Die Grundmenge hat  $|\Omega| := n$  Elemente. Alle Elementarereignisse sind gleichwahrscheinlich. Also gilt für  $\omega \in \Omega$  und  $A \in 2^\Omega$

$$p_\omega := \frac{1}{n} \quad \text{und} \quad P(A) := \frac{|A|}{|\Omega|}.$$

Man nennt diese W-Maß *Laplace-Wahrscheinlichkeit* oder *Gleichverteilung*<sup>2</sup>

<sup>1</sup>Zur Bezeichnung:  $\omega \in \Omega$  bezeichnet ein beliebiges Element des Grundraumes. Eine konkrete Beobachtung bezeichnen wir dagegen mit  $x$  ( $\rightarrow$  Zufallsvariable,  $\rightarrow$  Statistik)

<sup>2</sup>W-Maße nennt man auch W-Verteilungen

**Model:** Fairer Würfel, fairer Münzwurf, gut gemischtes Kartenspiel usw.

Hieraus konstruiert man dann weitere W-Räume wie:

- Augensumme beim zweimaligen Würfeln.
- Wann fällt zum ersten mal *Kopf* beim Münzwurf.

Bei Berechnung der Laplace-Wahrscheinlichkeit helfen Formeln der Kombinatorik ( $\rightarrow$  Abschnitt 1.2), um die Anzahl der Elemente von Mengen einer bestimmten Bauart zu bestimmen.

## 1.2 Elementare Kombinatorik

**Elementare Kombinatorik:** Lehre von der Abzählung endlicher strukturierter Mengen.

Die folgenden Tatsachen sind aus der Grundvorlesung bekannt:

### 1.2.1 Bem. (Permutation, Potenzmenge, Produkt)

(i)  $M$  ist eine endliche Menge mit  $|M| := n$  Elementen, wenn es eine bijektive Abbildung von  $\{1, 2, \dots, n\} \subset \mathbb{N}$  auf die Menge  $M$  gibt. Die Zahl  $n$  heißt die **Mächtigkeit** der Menge  $M$ .

Für die leere Menge sei  $|\emptyset| = 0$ .

(ii) Bijektive Abbildungen  $\sigma : \{1, 2, \dots, n\} \rightarrow M$  heißen **Permutationen** oder **Anordnungen** oder **Aufzählungen** der Menge  $M$ .

Für eine Menge mit  $n$  Elementen gibt es

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

Permutationen. Man setzt  $0! = 1$ .

(iii) Eine injektive Abbildung  $\sigma : \{1, \dots, k\} \rightarrow M$  gibt eine Aufzählung einer  $k$ -elementigen Teilmenge von  $M$  an. Es gibt genau

$$(n)_k := n \cdot (n-1) \cdot \dots \cdot (n-k+1)$$

injektive Abbildungen von  $\{1, \dots, k\}$  in  $M$ . Beachte:  $(n)_k = 0$  für  $k > n$ . Das leere Produkt ergibt  $(n)_0 = 1$ .

(iv) Eine Menge  $M$  mit  $n$  Elementen hat genau

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

$k$ -elementige Teilmengen. Es ist

$$\binom{n}{k} := \begin{cases} \frac{n!}{k!(n-k)!} & \text{für } 0 \leq k \leq n \\ 0 & \text{für } n < k. \end{cases}$$

Insbesondere ist  $\binom{n}{0} = 1$ . Der Ausdruck  $\binom{n}{k}$  heißt **Binomialkoeffizient**.

(v) Die Menge aller Teilmengen von  $M$  heißt die **Potenzmenge** von  $M$ . Jeder Teilmenge  $A \subset M$  entspricht eineindeutig ihre charakteristische Funktion oder Indikatorfunktion  $\mathbb{1}_A$ . Auf diese Weise erhält man eine Bijektion von  $\{0, 1\}^M$  mit der Potenzmenge.

Die Potenzmenge von  $M$  hat genau  $2^{|M|}$  Elemente. Daher rührt die Bezeichnung  $2^M$  für die Potenzmenge.

(vi) Es seien  $M_1, \dots, M_k$  endliche Mengen. Die **Produktmenge**  $M := \prod_{\varkappa=1}^k M_{\varkappa}$  hat

$$|M| = \prod_{\varkappa=1}^k |M_{\varkappa}|$$

Elemente. Nach Definition ist das Produkt leer, wenn zumindest einer der Faktoren  $M_{\varkappa} = \emptyset$  ist.

### 1.2.2 Festst. (Abzählprinzip)

Bemerkung 1.2.1(vi) nennt man das Abzählprinzip, daß man umgangssprachlich folgendermaßen formuliert:

Es sei  $\Omega$  eine Menge von  $n$ -Tupeln  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ . Der Raum  $\Omega$  sei der Ergebnisraum eines Experimentes, das aus  $n$  Telexperimenten besteht. Für jedes  $\nu \in \{1, 2, \dots, n\}$  sei  $k_{\nu}$  die Zahl der möglichen Ergebnisse des  $\nu$ -ten Experiments. Die Zahl  $k_{\nu}$  sei unabhängig davon, wie die anderen Telexperimente ausgehen. Dann ist

$$|\Omega| = k_1 \cdot k_2 \cdot \dots \cdot k_n.$$

**Anmerkung.** Oft lassen sich die Anzahlen auf eine von vier typischen Grundsituationen zurückführen ( $\rightarrow$  Stichproben 1.2.3 und Belegungen 1.2.5)

### 1.2.3 Bez. (Stichproben)

Aus einer  $n$ -elementigen Menge  $M$  kann man auf verschiedene Weise eine Stichprobe vom Umfang  $k$  entnehmen. Man nennt die Menge  $M$  auch eine **Urne**, aus der man die Stichprobe zieht. Bei der Stichprobe unterscheidet man, ob es zum Schluß auf die Reihenfolge ankommt oder nicht und ob man nach jedem Ziehen wieder zurücklegt oder nicht. Das ergibt vier Typen

#### geordnete Stichprobe ohne Wiederholung:

$k$ -maliges Ziehen aus der Urne ohne Zurücklegen unter Beachtung der Reihenfolge.

Bsp: Ziehung im Zahlenlotto 6 aus 49.

**Math.:** injektive Abbildung  $\{1, 2, \dots, k\} \rightarrow M$ , Aufzählung einer  $k$ -elementigen Teilmenge.

#### geordnete Stichprobe mit Wiederholung:

$k$ -maliges Ziehen aus der Urne mit Zurücklegen unter Beachtung der Reihenfolge.

Bsp.  $n = 6$ ,  $k$ -maliges Würfeln mit einem Würfel.

**Math.:**  $k$ -Tupel, Abbildung  $\{1, 2, \dots, k\} \rightarrow M$ .

#### ungeordnete Stichprobe ohne Wiederholung:

$k$ -maliges Ziehen aus der Urne ohne Zurücklegen und ohne Beachtung der Reihenfolge.

Bsp: Gewinnzahlen im Zahlenlotto 6 aus 49.

**Math.:**  $k$ -elementige Teilmenge von  $M$   
Abbildung  $f : M \rightarrow \{0, 1\}$  mit  $\sum_{\omega \in M} f(\omega) = k$ .

#### ungeordnete Stichprobe mit Wiederholung:

$k$ -maliges Ziehen aus der Urne mit Zurücklegen ohne Beachtung der Reihenfolge.

Bsp.  $n = 6$ ,  $k$  nicht unterscheidbare Würfel in einem Becher.

**Math.:** Abbildung  $f : M \rightarrow \mathbb{N}$  mit  $\sum_{\omega \in M} f(\omega) = k$ .

**Anmerkung.** Wir bestimmen nun die Anzahl der möglichen Stichproben in den obigen vier Fällen. Das besagt noch nichts über geeignete W-Maße auf dem Raum der Stichproben!

**1.2.4 Satz (Stichprobenanzahl)**

Für die Anzahl der möglichen Stichproben vom Umfang  $k$  aus einer Menge mit  $n$  Elementen gilt:

Stichprobe vom Umfang $k$ aus $M$ , $ M  = n$	ohne Wiederholung	mit Wiederholung
geordnete Stichprobe	$(n)_k$	$n^k$
ungeordnete Stichprobe	$\binom{n}{k}$	$\binom{n+k-1}{k}$

Dabei ist

$$(n)_k := \frac{n!}{(n-k)!} = n \cdot (n-1) \cdots (n-k+1)$$

**1.2.5 Bez. (Belegungen)**

Eine Stichprobe vom Umfang  $k$  aus  $M$  kann man auch anders interpretieren:

Die Menge  $M$  besteht aus Zellen. Man hat  $k$  Objekte, die auf die Zellen zu verteilen sind. Eine Stichprobe  $x_1, x_2, \dots, x_k$  gibt an, in welche Zelle die Objekte zu legen sind.

Bei geordneten Stichproben werden die Objekte entsprechend angeordnet und kommen der Reihe nach in die angegebene Zelle. Bei Stichproben mit Wiederholung kommen mehrere Objekte in dieselbe Zelle.

Sind die Objekte *ununterscheidbar*, so kommt es auf die Reihenfolge, in der die Objekte auf die Zellen verteilt werden, nicht an. Dieser Fall entspricht der *ungeordneten* Stichprobe. Die ungeordnete Stichprobe gibt nur an, in welche Zellen Objekte kommen.

Man hat also die folgenden vier Fälle:

**unterscheidbare Objekte, keine Mehrfachbelegung:** geordnete Stichprobe ohne Wiederholung

**Math.:** Injektive Abbildung von der Menge  $K$  der Objekte in die Menge  $M$  der Zellen.

**unterscheidbare Objekte, Mehrfachbelegung:** geordnete Stichprobe mit Wiederholung

**Math.:** Abbildung von der Menge  $K$  der Objekte in die Menge  $M$  der Zellen.

**nicht unterscheidbare Objekte, keine Mehrfachbelegung:** ungeordnete Stichprobe ohne Wiederholung

**Math.:**  $k$ -elementige Teilmenge von  $M$  oder Abbildung  $f : M \rightarrow \{0, 1\}$  mit  $\sum_{\omega \in M} f(\omega) = k$ .

**nicht unterscheidbare Objekte, mit Mehrfachbelegung:** ungeordnete Stichprobe mit Wiederholung

**Math.:** Abbildung  $f : M \rightarrow \mathbb{N}$  mit  $\sum_{\omega \in M} f(\omega) = k$ .

**Anmerkung.** Wir bestimmen nun die Anzahl der möglichen Belegungen in den obigen vier Fällen. Das besagt noch nichts über geeignete W-Maße auf dem Raum der Belegungen!

**1.2.6 Folg. (Anzahl der Belegungen)**

Für die Anzahl der möglichen Belegungen von  $n$  Zellen mit  $k$  Objekten gilt:

Belegung von $M$ mit $k$ Objekten, $ M  = n$	ohne Mehrfachbelegung	mit Mehrfachbelegung
unterscheidbare Objekte	$(n)_k := \frac{n!}{(n-k)!}$	$n^k$
Objekte nicht unterscheidbar	$\binom{n}{k}$	$\binom{n+k-1}{k}$

**Formeln für Binomialkoeffizienten.** Wir bringen einige Beispiele. Man kann die Formeln einerseits durch Nachrechnen zeigen. Dazu wird man zumeist einen Induktionsbeweis führen. Andere folgen leicht aus der binomischen Formel

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \quad \text{für } n \in \mathbb{N}_0.$$

Man setze für  $x, y$  spezielle Werte ein, differenziere die Formel ...

Man kann solche Formeln auch durch Bildung von passenden Mengen von ungeordneten Stichproben ohne Wiederholung aus  $\{1, \dots, n\}$ , d.h. passenden Teilmengen der Potenzmenge von  $\{1, \dots, n\}$ , **interpretieren:**

**1.2.7 Bsp. (Komplementärmengen)**

Für  $0 \leq k \leq n$  gilt

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

Die Komplementbildung ist eine bijektive Abbildung der Potenzmenge von  $\{1, \dots, n\}$  auf sich. Es gibt also genau soviele  $k$ -elementige Teilmengen wie  $(n-k)$ -elementige Teilmengen.

**1.2.8 Bsp. (Additionstheorem)**

Für  $1 \leq k \leq n$  gilt

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

**linke Seite:** Anzahl der ungeordneten Stichproben ohne Wiederholung vom Umfang  $k$  aus  $\{1, \dots, n\}$ .

Man zerlege diese in zwei disjunkte Typen. Deren Anzahlen ergeben die rechte Seite.

**$n$  nicht in der Stichprobe:** Dies sind alle ungeordneten Stichproben ohne Wiederholung vom Umfang  $k$  aus  $\{1, \dots, n-1\}$ . Anzahl ist  $\binom{n-1}{k}$ .

**$n$  ist in der Stichprobe:** Nimmt man das Element  $n$  davon weg, erhält man alle ungeordneten Stichproben ohne Wiederholung vom Umfang  $k-1$  aus  $\{1, \dots, n-1\}$ . Anzahl ist  $\binom{n-1}{k-1}$ .

Man kann die Methode von Beispiel 1.2.8 verallgemeinern:

**1.2.9 Bsp.** Für  $1 \leq k \leq n$  ist

$$\binom{n}{k} = \sum_{m=k-1}^{n-1} \binom{m}{k-1} = \binom{k-1}{k-1} + \binom{k}{k-1} + \dots + \binom{n-1}{k-1}.$$

**linke Seite:** Anzahl der  $k$ -elementigen Teilmengen von  $\{1, \dots, n\}$ .

Man zerlege die Menge der  $k$ -elementigen Teilmengen in folgende disjunkte Typen. Deren Anzahlen ergeben die rechte Seite.

**Teilmengen, die  $\{k+1, \dots, n\}$  nicht enthalten:**  
 Hiervon gibt es genau eine:  $\{1, \dots, k\}$ . Anzahl ist  $1 = \binom{k-1}{k-1}$

**Teilmengen, die  $\{k+2, \dots, n\}$  nicht enthalten und  $k+1$  enthalten:** Anzahl ist  $\binom{k}{k-1}$   
 .....

**Teilmengen, die  $\{n-1, n\}$  nicht enthalten und  $n-2$  enthalten:** Anzahl ist  $\binom{n-3}{k-1}$ .

**Teilmengen, die  $\{n\}$  nicht enthalten und  $n-1$  enthalten:** Anzahl ist  $\binom{n-2}{k-1}$ .

**Teilmengen, die  $\{n\}$  enthalten:** Anzahl ist  $\binom{n-1}{k-1}$ .

**1.2.10 Bsp.** Für  $1 \leq k \leq n$  ist

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

**linke Seite:** Für  $k = 1, 2, \dots, n$  bilde man die Anzahlen der  $k$ -elementigen Teilmengen von  $\{1, \dots, n\}$  und summiere diese auf. Dies ergibt die Anzahl aller Teilmengen von  $\{1, \dots, n\}$ .

**rechte Seite:** Mächtigkeit der Potenzmenge von  $\{1, \dots, n\}$ .

**1.2.11 Bsp.** Für  $n \in \mathbb{N}$  ist

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \dots$$

**linke Seite:** Anzahl der Teilmengen von  $\{1, \dots, n\}$  mit gerader Anzahl von Elementen.

**rechte Seite:** Anzahl der Teilmengen von  $\{1, \dots, n\}$  mit ungerader Anzahl von Elementen.

Es sei  $\Omega := \{1, \dots, n\}$ . Die folgende Abbildung  $\varphi : 2^\Omega \rightarrow 2^\Omega$  ist bijektiv und bildet Teilmengen  $A \in 2^\Omega$  mit gerader Elementanzahl auf solche mit ungerader Elementanzahl ab und umgekehrt.

$$\varphi : A \mapsto \begin{cases} A \cup \{n\} & \text{wenn } n \notin A. \\ A \setminus \{n\} & \text{wenn } n \in A. \end{cases}$$

Folglich gibt es genausoviel Teilmengen mit gerader Anzahl wie mit ungerader Anzahl.

**1.2.12 Bsp.** Für  $n \in \mathbb{N}_0$  ist

$$\sum_{k=1}^n k \binom{n}{k} = n2^{n-1}.$$

Es sei  $\Omega = \{1, \dots, n\}$ .

**linke Seite:** Mächtigkeit der Menge

$$L := \{(a, A) \mid A \in 2^\Omega, a \in A\}.$$

**rechte Seite:** Mächtigkeit der Menge

$$R := \{(b, B) \mid b \in \Omega, B \subset \Omega \setminus \{b\}\} = \bigcup_{b \in \Omega} \{(b, B) \mid B \subset \Omega \setminus \{b\}\}.$$

Man beachte, daß dies eine Vereinigung von  $n$  disjunkten Mengen ist, die alle die gleiche Mächtigkeit  $2^{n-1}$  haben.

Die folgende Abbildung  $\varphi : L \rightarrow R$  ist bijektiv:

$$L \ni (a, A) \mapsto (a, A \setminus \{a\}) \in R.$$

Also sind  $L$  und  $R$  gleichmächtig.

**1.2.13 Bsp.** Für  $n \in \mathbb{N}_0$  ist

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

Die erste Gleichung folgt aus der Komplementärformel 1.2.7.

Für die zweite Gleichung realisieren wir die linke und rechte Seite als Mächtigkeiten. Es sei  $\Omega := \{1, \dots, n\}$ .

**linke Seite:** Mächtigkeit der Menge

$$L := \{(A, B) \mid A, B \in 2^\Omega, |A| + |B| = n\}$$

**rechte Seite:** Mächtigkeit der Menge

$$R := \{C \mid C \subset \Omega \times \{0, 1\}, |C| = n\}.$$

Die folgende Abbildung  $\varphi : L \rightarrow R$  ist bijektiv:

$$L \ni (A, B) \mapsto A \times \{0\} \cup B \times \{1\} \in R.$$

Also sind  $L$  und  $R$  gleichmächtig.

**Anmerkung.** Der *Binomialkoeffizient*  $\binom{n}{k}$  gibt an, auf wieviele Arten man eine Menge mit  $n$  Elementen in zwei Gruppen aufteilen kann, daß die erste Gruppe  $k$  Elemente und die zweite  $n-k$  Elemente hat.

Die *Multinomialkoeffizienten* geben an, auf wie viele Arten man eine Menge mit  $n$  Elementen in  $m$  Gruppen aufteilen kann, so daß die *erste* Gruppe  $k_1$ , die *zweite*  $k_2$  und die  $m$ -te Gruppe  $k_m$  Elemente hat. Dabei kommt es auf die Reihenfolge der Gruppen an und es muß natürlich  $k_1 + k_2 + \dots + k_m = n$  gelten.

**1.2.14 Bez. (Multinomialkoeffizienten)**

Es seien  $m, n \in \mathbb{N}$ ,  $1 \leq m \leq n$  und  $(k_1, \dots, k_m) \in \mathbb{N}_0^m$  mit  $\sum_{\mu=1}^m k_\mu = n$ . Es gibt

$$\binom{n}{k_1} \binom{n-k_1}{k_2} \dots \binom{n-k_1-k_2-\dots-k_{m-1}}{k_m} = \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_m!}$$

Einteilungen einer  $n$ -elementigen Menge in  $m$  durchnummerierte Gruppen mit  $k_1, k_2, \dots, k_m$  Elementen.

Die so bestimmten Zahlen heißen *Multinomialkoeffizienten* und werden mit  $\binom{n}{k_1, \dots, k_m}$  bezeichnet. Man setzt

$$\binom{n}{k_1, \dots, k_m} = 0, \text{ wenn } \sum_{\mu=1}^m k_\mu \neq n \text{ ist.}$$

**Anmerkung. (Multinomialformel)** Analog zur Binomialformel gibt es die Multinomialformel:

$$(p_1 + \dots + p_m)^n = \sum_{\substack{(k_1, \dots, k_m) \in \mathbb{N}_0^n \\ k_1 + \dots + k_m = n}} \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_m!} p_1^{k_1} \cdot \dots \cdot p_m^{k_m}. \tag{1.2.1}$$

Ist speziell  $p_1 + \dots + p_m = 1$  so ist in Gleichung 1.2.1 die rechte Seite gleich 1.

### 1.3 Rechnen mit Indikatorfunktionen

#### 1.3.1 Bem. (Indikatorfunktion)

Für  $A \subset \Omega$  bezeichnet  $\mathbb{1}_A$  die charakteristische Funktion oder Indikatorfunktion von  $A$ :

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

Es gelten die folgenden Regeln:

- (i)  $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B$ , wenn  $A, B$  disjunkt sind;
- (ii)  $\mathbb{1}_{\bigcup_{\nu=1}^n A_\nu} = \sum_{\nu=1}^n \mathbb{1}_{A_\nu}$ , wenn  $A_1, \dots, A_n$  paarweise disjunkt sind;

Für beliebige Teilmengen von  $\Omega$  gilt:

- (iii)  $\mathbb{1}_{A \cap B} = \mathbb{1}_A \mathbb{1}_B = \min\{\mathbb{1}_A, \mathbb{1}_B\}$ ,
- (iv)  $\mathbb{1}_{\bigcap_{\nu=1}^n A_\nu} = \prod_{\nu=1}^n \mathbb{1}_{A_\nu}$ ,
- (v)  $\mathbb{1}_{A \setminus B} = \mathbb{1}_A - \mathbb{1}_{A \cap B}$ ,
- (vi)  $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$ ,
- (vii)  $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B} = \max\{\mathbb{1}_A, \mathbb{1}_B\}$ ,
- (viii)  $A \subset B \iff \mathbb{1}_A \leq \mathbb{1}_B$ .

Mit der folgenden Feststellung kann man die Rechenregeln für endliche W-Räume auf die entsprechenden Regeln für endliche Summen zurückführen.

#### 1.3.2 Festst.

Für eine Menge  $\Omega$  und eine endliche Teilmenge  $A \subset \Omega$  gilt

$$|A| = \sum_{\omega \in A} \mathbb{1}_A(\omega)$$

#### 1.3.3 Bem.

Mit der Morganschen Regel

$$\left(\bigcup_{\nu=1}^n A_\nu\right)^c = \bigcap_{\nu=1}^n A_\nu^c \tag{1.3.1}$$

folgt aus

$$\mathbb{1}_{\bigcap_{\nu=1}^n A_\nu} = \prod_{\nu=1}^n \mathbb{1}_{A_\nu} \tag{1.3.2}$$

und  $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$ , daß

$$\mathbb{1}_{\bigcup_{\nu=1}^n A_\nu} = 1 - \mathbb{1}_{\bigcap_{\nu=1}^n A_\nu^c} \tag{1.3.3}$$

$$= 1 - \prod_{\nu=1}^n (1 - \mathbb{1}_{A_\nu}) \tag{1.3.4}$$

$$= \sum_{i=1}^n (-1)^{i+1} s_i. \tag{1.3.5}$$

Dabei sind die  $s_i$  die Summe der  $i$ -fachen Produkte der Indikatorfunktionen:

$$s_i = \sum_{1 \leq \nu_1 < \dots < \nu_i \leq n} \prod_{j=1}^i \mathbb{1}_{A_{\nu_j}}.$$

Wendet man nochmal (1.3.2) an, so erhält man

$$s_i = \sum_{1 \leq \nu_1 < \dots < \nu_i \leq n} \mathbb{1}_{\bigcap_{j=1}^i A_{\nu_j}}. \tag{1.3.6}$$

#### 1.3.4 Folg. (Siebformel)

Es seien  $A_1, \dots, A_n \subset \Omega$ . Für die Mächtigkeit der Vereinigung gilt

$$\left|\bigcup_{\nu=1}^n A_\nu\right| = \sum_{i=1}^n (-1)^{i+1} m_i,$$

wobei die  $m_i$  die summierten Mächtigkeiten der  $i$ -fachen Schnittmengen sind:

$$m_i = \sum_{1 \leq \nu_1 < \dots < \nu_i \leq n} \left|\bigcap_{j=1}^i A_{\nu_j}\right|.$$

**Anmerkung.** Die folgende Formel ist ein Spezialfall des  $\rightarrow$  Erwartungswertes einer Zufallsvariablen.

#### 1.3.5 Festst.

Für einen W-Raum  $(\Omega, P)$  und  $A \in 2^\Omega$  gilt

$$p(A) = \sum_{\omega \in A} \mathbb{1}_A(\omega) P(\{\omega\}).$$

Manchmal ist die die Wahrscheinlichkeit von Schnittmengen leichter zu berechnen als die von Vereinigungen. Dann hilft die folgende Formel:

#### 1.3.6 Satz (Einschluß-Ausschluß-Formel)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $A_1, \dots, A_n \in 2^\Omega$ . Dann ist

$$P\left(\bigcup_{\nu=1}^n A_\nu\right) = \sum_{i=1}^n (-1)^{i+1} S_i,$$

wobei die  $S_i$  die summierten Wahrscheinlichkeiten der  $i$ -fachen Schnittmengen sind:

$$S_i = \sum_{1 \leq \nu_1 < \dots < \nu_i \leq n} P\left(\bigcap_{j=1}^i A_{\nu_j}\right).$$

**1.3.7 Bsp.**

Wie groß ist die Wahrscheinlichkeit, daß bei einer Permutation von  $M := \{1, \dots, n\}$  alle Elemente vertauscht werden, also keines an seinem Platz bleibt. Dabei seien alle Permutationen gleichwahrscheinlich.

Der Grundraum sind die Permutationen

$$\Omega := \{(x_1, \dots, x_n) \in \{1, \dots, n\}^n \mid x_i \neq x_j \text{ für } i \neq j\}$$

mit Laplace-Wahrscheinlichkeit. Es ist  $|\Omega| = n!$ .

Für  $\nu = 1, \dots, n$  sei

$$A_\nu = \{(x_1, \dots, x_n) \in \Omega \mid x_\nu = \nu\}.$$

Die Vereinigung  $\bigcup_{\nu=1}^n A_\nu$  ist die Menge der Permutationen, die zumindest eine Stelle festlassen, d.h.  $x_i = i$  für **irgendein**  $i \in M$ .

Für beliebige  $1 \leq \nu_1 < \dots < \nu_i \leq n$  hat die Schnittmenge

$$\bigcap_{j=1}^i A_{\nu_j} = \{(x_1, \dots, x_n) \in \Omega \mid x_{\nu_j} = \nu_j \text{ für } j = 1, \dots, i\}$$

genau  $(n-i)!$  Elemente. Die Laplace-Wahrscheinlichkeit des Durchschnitts hängt also nur von der Anzahl  $i$  der Schnittmengen ab.

$$P\left(\bigcap_{j=1}^i A_{\nu_j}\right) = \frac{(n-i)!}{n!}.$$

Es gibt  $\binom{n}{i}$  mögliche Auswahlen für die Indizes  $1 \leq \nu_1 < \dots < \nu_i \leq n$ .

Die Einschluß-Ausschluß-Formel ergibt nun

$$\begin{aligned} P\left(\bigcup_{\nu=1}^n A_\nu\right) &= \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} \frac{(n-i)!}{n!} \\ &= \sum_{i=1}^n (-1)^{i+1} \frac{1}{i!} \end{aligned} \quad (1.3.7)$$

**Anmerkung.** Man erinnere sich an die Exponentialreihe:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \text{ für } x \in \mathbb{R}.$$

Setzt man  $x = -1$ , so folgt

$$\begin{aligned} \lim_{k \rightarrow \infty} P\left(\bigcup_{\nu=1}^n A_\nu\right) &= \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k!} \\ &= 1 - \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \\ &= 1 - e^{-1} \approx 0,632. \end{aligned} \quad (1.3.8)$$

Da es sich um eine Leibniz-Reihe (alternierendes Vorzeichen der Summanden) handelt, ist

$$\left|P\left(\bigcup_{\nu=1}^n A_\nu\right) - (1 - e^{-1})\right| \leq \frac{1}{(n+1)!}.$$

Für  $n = 6$  ist  $\frac{1}{7!} \leq 0,0002$ . Die gesuchte Wahrscheinlichkeit ist fast konstant  $\approx 0,632$

### 1.4 Produkt von endl. W-Räumen

#### 1.4.1 Bez. (Produkt von zwei Laplace-Räumen)

**Math.:** Es seien  $(\Omega_1, P_1)$  und  $(\Omega_2, P_2)$  zwei W-Räume mit Laplace-Wahrscheinlichkeit. Der Grundraum  $\Omega := \Omega_1 \times \Omega_2$  mit Laplace-Wahrscheinlichkeit  $P$  heißt das Produkt der beiden Räume  $(\Omega_1, P_1)$  und  $(\Omega_2, P_2)$ .

Für die Anzahl gilt

$$|\Omega_1 \times \Omega_2| = |\Omega_1| |\Omega_2|.$$

Für  $\omega = (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$  gilt also:

$$P(\omega) = \frac{1}{|\Omega|} = \frac{1}{|\Omega_1|} \frac{1}{|\Omega_2|} = P_1(\omega_1) P_2(\omega_2)$$

Für Teilmengen der Form  $A \times B \subset \Omega_1 \times \Omega_2$  erhält man

$$\begin{aligned} P(A \times B) &= \sum_{\omega \in A \times B} P(\omega) \\ &= \sum_{\substack{\omega_1 \in A \\ \omega_2 \in B}} P(\omega_1) P(\omega_2) \\ &= \sum_{\omega_1 \in A} P(\omega_1) \sum_{\omega_2 \in B} P(\omega_2) \\ &= P_1(A) P_2(B). \end{aligned} \tag{1.4.1}$$

Diese Produktformeln sind der Anlaß für die Bezeichnung des Produktraumes:

$$\Omega := \Omega_1 \times \Omega_2 \quad \text{und} \quad P := P_1 \otimes P_2$$

**Modell:** Wir betrachten zwei Zufallsexperimente  $E_1$  und  $E_2$  die durch W-Räume  $(\Omega_1, P_1)$  und  $(\Omega_2, P_2)$  mit Laplace-Wahrscheinlichkeit beschrieben werden. Führt man beide Experimente  $E_1$  und  $E_2$  durch, so ist dies ein neues Experiment  $E$ .

Die Ergebnismenge von  $E$  ist das Kartesische Produkt

$$\Omega := \Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

Wenn  $E_1$  und  $E_2$  sich nicht gegenseitig beeinflussen, dann werden die Elementarereignisse in  $\Omega$  wieder gleichwahrscheinlich sein. D.h.  $E$  wird durch den W-Raum  $(\Omega, P)$  modelliert, wobei  $P$  die Laplace-Wahrscheinlichkeit ist.

Für Ereignisse  $A \in 2^{\Omega_1}$  und  $B \in 2^{\Omega_2}$  ist  $A \times B$  das Ereignis, daß im ersten Experiment  $A$  eintritt und im dem anderen  $B$ . Die Wahrscheinlichkeit von  $A \times B$  ist die Wahrscheinlichkeit, daß  $A$  eintritt, mal der Wahrscheinlichkeit, daß  $B$  eintritt.

#### 1.4.2 Bem. (Produkt von Laplace-Räumen)

Es seien  $(\Omega_\nu, P_\nu)$  Laplace-Räume. Für den Produktraum

$$\Omega := \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$$

mit Laplace-Wahrscheinlichkeit  $P$  gilt die Produktformel:

$$P(A_1 \times A_2 \times \dots \times A_n) = P_1(A_1) \cdot P_2(A_2) \cdot \dots \cdot P_n(A_n).$$

für  $A_\nu \in 2^{\Omega_\nu}$ .

**Anmerkung.** Man kann auch ganz andere W-Verteilungen auf dem Produktraum bilden.

**Bsp.** Es seien  $\Omega_1 = \Omega_2 = \{0, 1\}$  mit Laplaceverteilung. Man denke an den Wurf mit einer fairen Münze.

Auf  $\Omega := \Omega_1 \times \Omega_2$  betrachte man die folgende W-Funktion  $p$ :

$$p(i, j) := \begin{cases} \frac{1}{2} & \text{für } i = j \\ 0 & \text{für } i \neq j. \end{cases}$$

Dieses  $(\Omega, P)$  beschreibt den zweimaligen Wurf mit einer *Trickmünze mit Gedächtnis*. Beim ersten Wurf sind die Elementarereignisse Kopf und Zahl gleichverteilt. Das Ergebnis beim zweiten Wurf ist immer das gleiche wie beim ersten Wurf.

Man beachte, schaut man sich nur die Ergebnisse des zweiten Wurfs an, ohne den ersten zu kennen, so sind Kopf und Zahl wieder gleichverteilt:

$$P(\Omega_1 \times \{0\}) = P(\Omega_1 \times \{1\}) = \frac{1}{2}$$

(→ Bezeichnung 1.8.4 Marginalverteilungen).

#### 1.4.3 Ziel (Häufigkeiten bei unabh. Exper.)

Das folgende Überlegung beruht auf Plausibilität, sie ist so nicht mathematisch beweisbar, sondern dient als Motivation für die anschließende Definition → Satz 1.4.5

Wir betrachten zwei Zufallsexperimente  $E_1$  und  $E_2$  die durch W-Räume  $(\Omega_1, P_1)$  und  $(\Omega_2, P_2)$  beschrieben werden.  $P_1$  bzw.  $P_2$  können beliebig sein.

$E := (E_1, E_2)$  sei das Experiment, in dem man jeweils  $E_1$  und  $E_2$  durchführt. Dies soll so geschehen, daß sich die Ergebnisse von  $E_1$  und  $E_2$  nicht beeinflussen.  $E$  hat den Ergebnisraum  $\Omega := \Omega_1 \times \Omega_2$  mit einem noch zu bestimmenden W-Maß  $P$ .

Wir betrachten zwei Ereignisse  $A \in 2^{\Omega_1}$  und  $B \in 2^{\Omega_2}$  mit positiven Wahrscheinlichkeiten  $P_1(A) > 0$  und  $P_2(B) > 0$ . Wir nehmen an, daß in beiden Fällen für alle hinreichend großen  $n$  die relative Häufigkeiten  $R_n(A) \approx P_1(A)$  bzw.  $R_n(B) \approx P_2(B)$  seien.

Man führe  $E$   $N$ -mal durch, wobei  $N$  sehr groß sei. Wie groß, werden wir gleich festlegen. Wir erhalten die Ergebnisse

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N).$$

Hiermit bilden wir die relativen Häufigkeiten:

$$\begin{aligned} R_N(A \times B) &:= \frac{1}{N} \sum_{\nu=1}^N \mathbb{1}_{A \times B}(x_\nu, y_\nu) \\ R_N(A) &:= \frac{1}{N} \sum_{\nu=1}^N \mathbb{1}_A(x_\nu) \\ R_N(B) &:= \frac{1}{N} \sum_{\nu=1}^N \mathbb{1}_B(y_\nu) \end{aligned}$$

Wir zeigen, daß unter den getroffenen Annahmen

$$\boxed{R_N(A \times B) \approx R_N(A) R_N(B)} \tag{1.4.2}$$

ist.

Die Voraussetzung, daß sich die Experimente nicht beeinflussen, interpretieren wir dahingehend, daß  $R_N(A) \approx P_1(A)$  und  $R_N(B) \approx P_2(B)$  ist. D.h., kein Experiment beeinflusst den Ausgang des anderen.

Wir wählen unter den  $(x_\nu, y_\nu)$  diejenigen aus, für die die erste Komponente in  $A$  liegt:

$$(\tilde{x}_1, \tilde{y}_1), (\tilde{x}_2, \tilde{y}_2), \dots, (\tilde{x}_m, \tilde{y}_m).$$

Wenn  $N$  groß genug ist, wird nach Annahme

$$\frac{m}{N} = R_N(A) \approx P_1(A) > 0.$$

sein. Für hinreichend großes  $N$  ist also

$$m = NR_N(A) \approx NP_1(A) \geq n.$$

Die getroffene Auswahl der zweiten Komponenten  $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_m$  hängt zwar vom Ausgang des ersten Experimentes ab. Daß sich die Experimente nicht beeinflussen, interpretieren wir dahingehend, daß diese Auswahl keinen Einfluß auf die relative Häufigkeit im zweiten Experiment hat. Es sei

$$\tilde{R}_m(B) := \frac{1}{m} \sum_{\mu=1}^m \mathbb{1}_B(\tilde{y}_\mu)$$

die relative Häufigkeit mit der die  $\tilde{y}_\mu$  in  $B$  liegen.

Da  $m$  hinreichend groß ist, ist nach unserer Annahme

$$\tilde{R}_m(B) \approx P_2(B) \approx R_N(B).$$

Somit folgt

$$\begin{aligned} R_N(A \times B) &= \frac{1}{N} \sum_{\nu=1}^N \mathbb{1}_{A \times B}(x_\nu, y_\nu) \\ &= \frac{1}{N} \sum_{\nu=1}^N \mathbb{1}_A(x_\nu) \mathbb{1}_B(y_\nu) \\ &= \frac{1}{N} \sum_{\mu=1}^m \mathbb{1}_B(\tilde{y}_\mu) \\ &= \frac{m}{N} \tilde{R}_m(B) = R_N(A) \tilde{R}_m(B) \\ &\approx R_N(A) R_N(B). \end{aligned}$$

Die obige Formel 1.4.2 legt die folgende Festsetzung nahe: Für  $A \in 2^{\Omega_1}$ ,  $B \in 2^{\Omega_2}$  ist

$$P(A \times B) := P_1(A)P_2(B) \tag{1.4.3}$$

Der folgende Satz 1.4.5 besagt, daß es genau ein W-Maß  $P_{1,2}$  auf  $\Omega_1 \times \Omega_2$  gibt, für daß die Formel 1.4.3 gilt. Wir verwenden in der folgenden anschaulichen Überlegung bereits die übliche Bezeichnung  $P_1 \otimes P_2$  für  $P_{1,2}$ . Für diese Verknüpfung von W-Maßen ist assoziativ  $\rightarrow$  Feststellung 1.4.7. Wir geben zuvor eine anschauliche Interpretation des Assoziativgesetzes.

### 1.4.4 Ziel (Assoziativgesetz)

Man führe drei Zufallsexperimente  $E_1, E_2, E_3$  so durch, daß sie sich nicht beeinflussen. Die Experimente werden durch die W-Räume  $(\Omega_i, P_i)$  ( $i = 1, 2, 3$ ) modelliert. Das Gesamtexperiment  $(E_1, E_2, E_3)$  hat den Ergebnisraum  $\Omega_1 \times \Omega_2 \times \Omega_3$ . Wie üblich identifizieren wir

$$\Omega_1 \times \Omega_2 \times \Omega_3 = (\Omega_1 \times \Omega_2) \times \Omega_3 = \Omega_1 \times (\Omega_2 \times \Omega_3).$$

D.h., man darf bei der Bildung von Tupeln innere Klammern weglassen:  $(\omega_1, \omega_2, \omega_3) = (\omega_1, (\omega_2, \omega_3)) = ((\omega_1, \omega_2), \omega_3)$ .

Wir denken uns zwei Experimentatoren, von denen einer zwei Experimente durchführt und der andere das übrige Experiment. Dann führen die beiden ihre Beobachtungen zu einem Experiment zusammen.

- Der erste Experimentator führt die ersten beiden Experimente durch und der andere das dritte. Der erste wird sein Experiment  $E_{12} := (E_1, E_2)$ , wie in  $\rightarrow$  Ziel 1.4.3 diskutiert, durch den Produktraum  $(\Omega_{12}, P_{12}) := (\Omega_1 \times \Omega_2, P_1 \otimes P_2)$  modellieren. Da  $E_{12}$  und  $E_3$  sich nicht beeinflussen, werden die beiden Experimentatoren das Gesamtexperiment  $E_{(12)3} := (E_{12}, E_3)$  durch den Produktraum  $(\Omega_{1,2} \times \Omega_3, P_{12} \otimes P_3) = (\Omega_1 \times \Omega_2 \times \Omega_3, (P_1 \otimes P_2) \otimes P_3)$  beschreiben.
- Nun teilen die beiden Experimentatoren die Experimente anders auf: der erste führt  $E_1$  durch und der andere die übrigen beiden. Der zweite Experimentator wird sein Experiment  $E_{23} = (E_2, E_3)$  durch den W-Raum  $(\Omega_{23}, P_{23}) := (\Omega_2 \times \Omega_3, P_2 \otimes P_3)$  beschreiben. Die Zusammenführung zu einem Experiment  $E_{1(23)}$  ergibt nun den W-Raum  $(\Omega_1 \times \Omega_{23}, P_1 \otimes P_{23}) = (\Omega_1 \times \Omega_2 \times \Omega_3, P_1 \otimes (P_2 \otimes P_3))$

Die Aufteilung der Experimente auf die Experimentatoren soll natürlich das Ergebnis nicht verändern, d.h. die beobachteten Häufigkeiten für  $E_{(12)3}$  und  $E_{1(23)}$  sind ungefähr gleich. Für die mathematische Konstruktion des Produktes von W-Räumen wird das folgende Assoziativgesetz gelten:

$$(P_1 \otimes P_2) \otimes P_3 = P_1 \otimes (P_2 \otimes P_3).$$

Man kann also wie beim Produkt von Zahlen die Klammern weglassen.

Das Assoziativgesetz ist aber keine zusätzliche Forderung an das Produkt von W-Räumen, sondern folgt bereits aus der Eindeutigkeit der Konstruktion  $\rightarrow$  Satz 1.4.5

Eine analoge Betrachtung ergibt, daß bei einer Permutation  $\sigma$  der Experimente die Produkträume *maßgleich* sind:

$$\begin{aligned} &(\Omega_1 \times \Omega_2 \times \Omega_3, P_1 \otimes P_2 \otimes P_3) \\ &\cong (\Omega_{\sigma_1} \times \Omega_{\sigma_2} \times \Omega_{\sigma_3}, P_{\sigma_1} \otimes P_{\sigma_2} \otimes P_{\sigma_3}) \end{aligned}$$

### 1.4.5 Satz (Produkt von W-Räumen)

Es seien  $n \in \mathbb{N}$ ,  $(\Omega_\nu, P_\nu)$ ,  $(\nu = 1, \dots, n)$  endliche W-Räume. Dann gibt es auf dem Produktraum

$$\Omega := \prod_{\nu=1}^n \Omega_\nu$$

genau ein W-Maß  $P$  mit der folgenden Eigenschaft:

Für alle  $n$ -Tupel  $(A_1, A_2, \dots, A_n)$  mit  $A_\nu \in 2^{\Omega_\nu}$  ist

$$P(A_1 \times A_2 \times \dots \times A_n) = P_1(A_1)P_2(A_2) \dots P_n(A_n)$$

Man schreibt kurz  $P := P_1 \otimes P_2 \otimes \dots \otimes P_n$  und nennt  $P$  das Produkt der Maße  $P_1, P_2, \dots, P_n$ .

Die zugehörige W-Funktion  $p : \Omega \rightarrow [0, 1]$  hat die Form

$$p(\omega_1, \omega_2, \dots, \omega_n) = p_1 \omega_1 \cdot p_2 \omega_2 \cdots p_n \omega_n. \quad (1.4.4)$$

**1.4.6 Folg. (W-Funktion des Produktmaßes)**

Bei endlichen W-Räumen bestimmt die Formel 1.4.4 das Produktmaß bereits eindeutig, denn es gilt:

$$\begin{aligned} P(A_1 \times \dots \times A_n) &= \sum_{\substack{(\omega_1, \omega_2, \dots, \omega_n) \\ \in A_1 \times \dots \times A_n}} p(\omega_1, \omega_2, \dots, \omega_n) \\ &= \sum_{\omega_1 \in A_1} \cdots \sum_{\omega_n \in A_n} p(\omega_1, \omega_2, \dots, \omega_n) \\ &= \sum_{\omega_1 \in A_1} \cdots \sum_{\omega_n \in A_n} p_1 \omega_1 \cdot p_2 \omega_2 \cdots p_n \omega_n \\ &= \sum_{\omega_1 \in A_1} p_1 \omega_1 \cdots \sum_{\omega_n \in A_n} p_n \omega_n \\ &= P_1(A_1) \cdots P_n(A_n). \end{aligned}$$

**Anmerkung.** 1. Man spricht  $P_1 \otimes P_2$  als „ $P_1$  tensor  $P_2$ “ oder kurz „ $P_1$  mal  $P_2$ “. Das Wort *tensor* beugt einer Verwechslung mit dem Produkt von Zahlen vor.

2. **Zur Information:** Die Benutzung des Wortes *tensor* rührt daher, daß es sich beim Produkt von Maßen um einen Spezialfall des *Tensorproduktes*  $V_1 \otimes V_2$  von Vektorräumen  $V_1, V_2$  handelt.

Die Maße auf einer  $n$ -elementigen Menge  $\Omega$  bilden einen  $2^n$ -dimensionalen Vektorraum  $\mathcal{M}(\Omega)$ . Für endliche Mengen  $\Omega_1, \Omega_2$  ist der Vektorraum  $\mathcal{M}(\Omega_1 \times \Omega_2)$  isomorph zum Tensorprodukt der Räume  $\mathcal{M}(\Omega_1)$  und  $\mathcal{M}(\Omega_2)$ , oder kurz

$$\mathcal{M}(\Omega_1 \times \Omega_2) \cong \mathcal{M}(\Omega_1) \otimes \mathcal{M}(\Omega_2).$$

In diesem Sinne ist  $P_1 \otimes P_2 \in \mathcal{M}(\Omega_1) \otimes \mathcal{M}(\Omega_2)$ .

Eine verwandte Bildung ist das *Tensorprodukt reeller Funktionen*: Für  $f_i : \Omega_i \rightarrow \mathbb{R}$  ( $i = 1, 2$ ) definiert man

$$f_1 \otimes f_2 : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R} \quad \text{durch} \quad f_1 \otimes f_2 : (\omega_1, \omega_2) \mapsto f_1(\omega_1) f_2(\omega_2).$$

Die reellen Funktionen auf einer  $n$ -elementigen Menge  $\Omega$  bilden einen  $n$ -dimensionalen Vektorraum  $\mathcal{F}(\Omega)$ . Für endliche Mengen  $\Omega_1, \Omega_2$  ist

$$\mathcal{F}(\Omega_1 \times \Omega_2) \cong \mathcal{F}(\Omega_1) \otimes \mathcal{F}(\Omega_2).$$

Für zwei Funktionen  $f, g \in \mathcal{F}(\Omega)$  unterscheidet man

$$\begin{aligned} f \otimes g &\in \mathcal{F}(\Omega \times \Omega), & f \otimes g(\omega_1, \omega_2) &= f(\omega_1) g(\omega_2), \\ fg &\in \mathcal{F}(\Omega), & fg(\omega) &= f(\omega) g(\omega). \end{aligned}$$

Die Gleichung (1.4.4) für die W-Funktionen kann man etwas abstrakter auch so schreiben:

$$p = p_1 \otimes p_2 \otimes \cdots \otimes p_n.$$

3. Man findet für das Produktmaß auch die Bezeichnung ( $\rightarrow$  [1][Krengel])

$$P_1 \times P_2 := P_1 \otimes P_2$$

**Anmerkung.** Das kartesische Produkt ist assoziativ. Wie üblich identifiziert man

$$\begin{aligned} \Omega_1 \times \cdots \times \Omega_n &= (\Omega_1 \times \cdots \times \Omega_{n-1}) \times \Omega_n \\ &= \Omega_1 \times (\Omega_2 \times \cdots \times \Omega_n) \end{aligned}$$

und für  $1 < k < n - 1$

$$\Omega_1 \times \cdots \times \Omega_n = (\Omega_1 \times \cdots \times \Omega_k) \times (\Omega_{k+1} \times \cdots \times \Omega_n)$$

**1.4.7 Festst. (Assoziativgesetz)**

Es seien  $n \in \mathbb{N}$ ,  $(\Omega_\nu, P_\nu)$ ,  $(\nu = 1, \dots, n)$  endliche W-Räume. Die Bildung des Produktmaßes ist assoziativ:

$$\begin{aligned} P_1 \otimes \cdots \otimes P_n &= (P_1 \otimes \cdots \otimes P_{n-1}) \otimes P_n \\ &= P_1 \otimes (P_2 \otimes \cdots \otimes P_n) \end{aligned}$$

und für  $1 < k < n - 1$

$$P_1 \otimes \cdots \otimes P_n = (P_1 \otimes \cdots \otimes P_k) \otimes (P_{k+1} \otimes \cdots \otimes P_n)$$

Man darf also beliebig klammern:

Für  $1 < k_1 < k_n < \cdots < k_l = n$  gilt

$$\begin{aligned} (\Omega_1 \times \cdots \times \Omega_{k_1}) \times (\Omega_{k_1+1} \times \cdots \times \Omega_{k_2}) \times \cdots \\ \cdots \times (\Omega_{k_{l-1}+1} \times \cdots \times \Omega_{k_l}) = \Omega_1 \times \cdots \times \Omega_n \end{aligned}$$

und

$$\begin{aligned} (P_1 \otimes \cdots \otimes P_{k_1}) \otimes (P_{k_1+1} \otimes \cdots \otimes P_{k_2}) \otimes \cdots \\ \cdots \otimes (P_{k_{l-1}+1} \otimes \cdots \otimes P_{k_l}) = P_1 \otimes \cdots \otimes P_n \end{aligned}$$

**1.4.8 Bez.**

Bei der Bildung des Produktes von  $n$  gleichen Faktoren  $(\Omega', P')$ . schreiben wir kurz

$$\Omega := (\Omega')^n \quad \text{und} \quad P := (P')^{\otimes n}.$$

## 1.5 Bernoulli- und Binomialverteilung

### 1.5.1 Bez. ( $n$ -faches Bernoulli-Experiment)

1. Ein W-Maß  $P$  auf  $\{0, 1\}$  heißt *Bernoulli-Experiment*.  $P$  ist durch die Zahl  $p := P(\{1\} \in [0, 1])$  eindeutig festgelegt. Man nennt ein W-Maß auf  $\{0, 1\}$  kurz eine einfache *Bernoullische* Verteilung mit dem Parameter  $p$ .

2. Ein Zufallsexperiment mit nur zwei Ausgängen, die man mit 0 und 1 bezeichnet, heißt ein einfaches Bernoulli-Experiment mit Parameter  $p = P(\{1\})$ .

Man sagt auch Bernoullische Verteilung mit *Erfolgswahrscheinlichkeit*  $p$ . Das kann aber in konkreten Beispielen sehr unpassend klingen.

3. Wir betrachten nun das Experiment, das in der  $n$ -fachen unabhängigen Wiederholung eines einfachen Bernoulli-Experimentes mit Parameter  $p$  besteht. Wir nennen dies kurz ein  $n$ -faches *Bernoulli-Experiment* mit Parameter  $p$ .

Es sei also  $\Omega_\nu := \{0, 1\}$  und  $P_\nu$  die einfache Bernoulli-Verteilung mit Parameter  $p$ . Das  $n$ -fache Bernoulli-Experiment wird dann durch

$$\begin{aligned}\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n &= \{0, 1\}^n \\ P_1 \otimes P_2 \otimes \cdots \otimes P_n &= P^{\otimes n}\end{aligned}$$

beschrieben. Nach Formel 1.4.4 ist

$$p_{(\omega_1, \omega_2, \dots, \omega_n)} = p^k (1-p)^{n-k} \quad (1.5.1)$$

wenn

$$k := \omega_1 + \omega_2 + \cdots + \omega_n$$

die Zahl der Einsen in  $(\omega_1, \omega_2, \dots, \omega_n)$  ist. Das durch Gleichung 1.5.1 bestimmte W-Maß  $P^{\otimes n}$  auf  $\{0, 1\}^n$  heißt  $n$ -fache *Bernoulli-Verteilung*.

4. Das Ereignis, das insgesamt genau  $k$  Einsen auftreten wird durch die Menge

$$A_k := \left\{ \omega \in \{0, 1\}^n \mid \sum_{\nu=1}^n \omega_\nu = k \right\}$$

beschrieben. Alle Elemente  $\omega \in A_k$  haben die gleiche Wahrscheinlichkeit  $p_\omega = p^k (1-p)^{n-k}$ . Wir haben eine bijektive Abbildung von  $A_k$  auf die Menge der ungeordneten Stichproben ohne Wiederholung vom Umfang  $k$  aus  $\{1, 2, \dots, n\}$ . Die Stichprobe gibt an, auf welchen Plätzen die Einsen in  $\omega$  sitzen. Also ist die Anzahl der Elemente von  $A_k$  gleich der Anzahl der  $k$ -elementigen Stichproben, also gleich  $\binom{n}{k}$ . Die Wahrscheinlichkeit von  $A_k$  bezeichnet man mit

$$b_{n,p}(k) := \binom{n}{k} p^k (1-p)^{n-k}. \quad (1.5.2)$$

Da die  $A_k$  paarweise disjunkt sind, ist

$$\sum_{k=1}^n b_{n,p}(k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = 1. \quad (1.5.3)$$

Man beachte, daß die rechte Seite von Gleichung 1.5.3 sich auch aus der binomischen Formel ergibt:

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

### 1.5.2 Bez. (Binomialverteilung)

Das W-Maß auf  $\Omega := \{0, 1, 2, \dots, n\}$  mit der W-Funktion

$$p_k := b_{n,p}(k) := \binom{n}{k} p^k (1-p)^{n-k}$$

heißt die Binomialverteilung mit Parameter  $n$  und  $p$ . Es wird mit  $\mathcal{B}_{n,p}(A)$  für  $A \in 2^\Omega$  bezeichnet.

**Anmerkung.** 1. Bei einem  $n$ -fachen Bernoulli-Experiment mit Parameter  $p$  ist die Binomialverteilung auf  $\{0, 1, 2, \dots, n\}$  die Wahrscheinlichkeit für die Anzahl der Einsen (Erfolge)

$$X(\omega) := \sum_{\nu=1}^n \omega_\nu \quad \text{für } \omega = (\omega_1, \omega_2, \dots, \omega_n) \in \{0, 1\}^n.$$

Die Funktion  $X : \{0, 1\}^n \rightarrow \{0, 1, 2, \dots, n\}$  ist ein Beispiel einer  $\rightarrow$  Zufallsvariablen und die Binomialverteilung ist die  $\rightarrow$  Verteilung von  $X$ .

2. Es gibt nicht die Binomialverteilung sondern eine Familie solcher Verteilungen, die durch die Parameter  $n$  und  $p$  unterschieden werden.

Man kann die Binomialverteilung auf Experimente mit mehr als zwei Ausgängen verallgemeinern.

### 1.5.3 Bsp. (Multinomialverteilung)

Bei einem Zufallsexperiment treten die möglichen Elementarereignisse  $\omega_1, \omega_2, \dots, \omega_m$  mit den Wahrscheinlichkeiten  $p_1, p_2, \dots, p_m$  auf. Es ist  $\sum_{\mu=1}^m p_\mu = 1$ .

Wie groß ist die Wahrscheinlichkeit bei  $n$ -maliger unabhängiger Wiederholung des Experimentes  $k_1$  mal das Ergebnis  $\omega_1$ ,  $k_2$  mal das Ergebnis  $\omega_2$ ,  $\dots$  und  $k_m$  mal das Ergebnis  $\omega_m$  zu erhalten? Dabei sind die  $k_\mu \in \mathbb{N}_0$  und es ist  $\sum_{\mu=1}^m k_\mu = n$ .

Das Experiment wird durch den Grundraum  $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$  und dem W-Maß  $P$  mit  $P\{\omega_i\} = p_i$  beschrieben. Die  $n$ -malige unabhängige Wiederholung des Experimentes wird durch den Produktraum  $\Omega^n$  und das Produktmaß  $P^{\otimes n}$  modelliert.

Für  $(x_1, x_2, \dots, x_n) \in \Omega^n$  mit

$$k_\mu := \sum_{\nu=1}^n \mathbb{1}_{\{\omega_\mu\}}(x_\nu)$$

ist

$$\begin{aligned}P^{\otimes n}\{(x_1, x_2, \dots, x_n)\} &= \prod_{\nu=1}^n P\{x_\nu\} \\ &= p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m}.\end{aligned} \quad (1.5.4)$$

Gleichung 1.5.4 verallgemeinert die  $n$ -fache Bernoulli-Verteilung ( $\rightarrow$  Gleichung 1.5.1) auf Experimente mit  $m$  möglichen Ausgängen.

Es sei  $A$  das Ereignis, daß für  $\mu = 1, 2, \dots, m$  das Ergebnis  $\omega_\mu$  genau  $k_\mu$  mal eintritt:

$$\begin{aligned}A := \{(x_1, x_2, \dots, x_n) \in \Omega^n \mid \sum_{\nu=1}^n \mathbb{1}_{\{\omega_\mu\}}(x_\nu) = k_\mu \\ \text{für } \mu = 1, 2, \dots, m\}\end{aligned} \quad (1.5.5)$$

Die Anzahl der Elemente von  $A$  ist nach Bezeichnung 1.2.14

$$|A| = \frac{n!}{k_1! \cdot k_2! \cdots k_m!}.$$

Da nach Gleichung 1.5.4 alle Elemente von  $A$  die gleiche Wahrscheinlichkeit haben ist

$$P^{\otimes n}(A) = n! \prod_{\mu=1}^m \frac{p_{\mu}^{k_{\mu}}}{k_{\mu}!}. \quad (1.5.6)$$

#### 1.5.4 Bez. (Multinomialverteilung)

Es seien  $m \in \mathbb{N}$ ,  $m \geq 2$ ,  $p_{\mu} \in [0, 1]$  mit  $\sum_{\mu=1}^m p_{\mu} = 1$ . Man erhält auf dem Raum

$$\Omega := \left\{ (k_1, k_2, \dots, k_m) \in \mathbb{N}_0^m \mid \sum_{\mu=1}^m k_{\mu} = n \right\}$$

ein W-Maß  $P$  durch die Vorschrift

$$P\{(k_1, k_2, \dots, k_m)\} := n! \prod_{\mu=1}^m \frac{p_{\mu}^{k_{\mu}}}{k_{\mu}!}. \quad (1.5.7)$$

Dieses W-Maß heißt *Multinomialverteilung* mit den Parametern  $n$  und  $(p_1, p_2, \dots, p_m)$  und wird mit  $\mathcal{M}_{n;p_1, \dots, p_m}$  bezeichnet. Die durch Gleichung 1.5.7 erklärte W-Funktion bezeichnet man mit  $m_{n;p_1, \dots, p_m}(k_1, \dots, k_m)$ .

#### 1.5.5 Bem.

Mitunter wird die *Multinomialverteilung* auch auf dem größeren Raum

$$\tilde{\Omega} = \{0, 1, \dots, n\}^m$$

erklärt. Man setzt dann

$$\mathcal{M}_{n;p_1, \dots, p_m}(k_1, \dots, k_m) = 0 \quad \text{für} \quad \sum_{\mu=1}^m k_{\mu} > n.$$

### 1.6 Zufallsvariable und ihre Verteilung

In einem Zufallsexperiment  $(\Omega, P)$  interessiert häufig nicht das Elementarereignis  $\omega \in \Omega$  sondern der Wert  $f(\omega)$ , den eine Abbildung  $f : \Omega \rightarrow M$  an der Stelle  $\omega$  annimmt. Z. Bsp. bei zwei Würfeln die *Augensumme* oder *Pasch* oder *nicht-Pasch*. In diesem Zusammenhang nennt man die Abbildung  $f$  eine *Zufallsvariable*.

In der Stochastik wählt man gewöhnlich große Buchstaben  $X, Y, N, S \dots$  zur Bezeichnung von Zufallsvariablen. Wenn  $M$  oder seine Elemente besondere Namen haben, verwendet man diese auch bei der Benennung von  $X$ .

- $X : \Omega \rightarrow \mathbb{R}$  *reelle Zufallsvariable*.
- $X : \Omega \rightarrow \mathbb{R}^d$  *Zufallsvektor*.
- $N : \Omega \rightarrow \mathbb{N}$  *zufällige Zahl*.

Wenn die Elemente von  $M$  üblicherweise mit  $x, y, z$  bezeichnet werden, so verwendet man entsprechend  $X, Y, Z$  für  $M$ -wertige Zufallsvariable. Führt man das Zufallsexperiment  $n$ -mal durch, so beobachtet man für die Zufallsvariable  $X$  die Werte  $x_1, x_2, \dots, x_n$ .

#### 1.6.1 Bez. (Zufallsvariable)

Es sei  $(\Omega, P)$  ein W-Raum,  $M$  eine Menge. Eine Abbildung  $X : \Omega \rightarrow M$  heißt eine  $M$ -wertige Zufallsvariable.

#### Anmerkung. (Verwendung des Wortes Zufallsvariable)

1. Neben der in der Mengenlehre üblichen Bezeichnung  $f : \Omega \rightarrow M$  für Abbildungen gibt es die hergebrachte Schreibweise  $y = f(x)$  oder ganz kurz  $y = y(x)$ . Man drückt aus, daß ein Meßwert  $y$  mit einem Parameter  $x$  variiert. In den beiden letzteren Schreibweisen heißt  $y$  die abhängige Variable und  $x$  die unabhängige Variable.

In diesem Sinn ist eine Zufallsvariable  $X$  eine abhängige Variable. Die zugehörige unabhängige Variable ist ein Elementarereignis  $\omega$  eines W-Raumes  $(\Omega, P)$ .

2. Das Wort Zufallsvariable erinnert auch an die bei mehrfacher Wiederholung des Experimentes variierenden Ergebnisse  $x_1, x_2, x_3 \dots$ . Häufig konstruiert man erst bei der mathematischen Modellierung den zugrundeliegenden W-Raum  $(\Omega, P)$  und die Abbildung  $X : \Omega \rightarrow M$ .

3. Zufallsvariable heißt auf englisch *random variable*, französisch *variable aléatoire*.

#### 1.6.2 Bsp. (Augensumme)

Man würfelt mit zwei fairen Würfeln. Die Zufallsvariable sei die erzielte Augensumme. Eigentlich interessiert nur der Ergebnisraum  $\mathcal{X} := \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$  und das W-Maß auf  $\mathcal{X}$ , daß dieses Zufallsexperiment modelliert. Einfacher ist aber das folgende Vorgehen:

1. Zwei unabhängige faire Würfel oder zweimaliges unabhängiges Würfeln beschreibt man durch dem Produktraum

$$\begin{aligned} \Omega &:= \{1, \dots, 6\} \times \{1, \dots, 6\} \\ &= \{(i, j) \mid i, j = 1, \dots, 6\} \end{aligned} \tag{1.6.1}$$

und der Laplace-Wahrscheinlichkeit auf  $\Omega$ . Wir bilden die Teilmengen

$$A_k := \{(i, j) \in \Omega \mid i + j = k\} \quad \text{für } k = 2, 3, \dots, 12.$$

Die Anzahl der Elemente ist

$$|A_2| = 1, |A_3| = 2, |A_4| = 3, \dots, |A_{12}| = 1.$$

Die Anzahl ist symmetrisch zu  $k = 7$ , da sich bei einem Würfel die Augen auf gegenüberliegenden Seiten zu 7 ergänzen.:

$$|A_{7-k}| = |A_{7+k}| \quad \text{für } k = 1, 2, \dots, 5.$$

Die entsprechenden Laplace-Wahrscheinlichkeiten sind:

$$\begin{aligned} P(A_2) &= \frac{1}{36}, & P(A_3) &= \frac{2}{36}, & P(A_4) &= \frac{3}{36}, \\ P(A_5) &= \frac{4}{36}, & P(A_6) &= \frac{5}{36}, & P(A_7) &= \frac{6}{36}, \\ P(A_8) &= \frac{5}{36}, & P(A_9) &= \frac{4}{36}, & P(A_{10}) &= \frac{3}{36}, \\ P(A_{11}) &= \frac{2}{36}, & P(A_{12}) &= \frac{1}{36}. \end{aligned}$$

Die gesuchte W-Funktion  $p_{\mathcal{X}}$  auf  $\mathcal{X} := \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$  ist also

$x$	2	3	4	5	6	7	...	11	12
$p_{\mathcal{X}}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	...	$\frac{2}{36}$	$\frac{1}{36}$

Die Wahrscheinlichkeit einer Teilmenge  $B \subset \mathcal{X}$  ist

$$P_{\mathcal{X}}(B) := \sum_{x \in B} p_{\mathcal{X}}(x) = P(\{(i, j) \mid i + j \in B\}).$$

2. Das obige vorgehen ist typisch. Man kann es mit Hilfe der Zufallsvariablen (Augensumme)

$$X : \Omega \rightarrow \mathbb{N} \quad X : (i, j) \mapsto i + j.$$

kürzer schreiben:

$(\Omega, P)$  werde wie in Gleichung 1.6.1 erklärt. Es ist  $\mathcal{X} := X(\Omega)$  die Wertemenge von  $X$ . Für eine Teilmenge  $B \subset \mathcal{X}$  ist

$$P_{\mathcal{X}}(B) = P(\{\omega \mid \omega \in \Omega, X(\omega) \in B\}) \tag{1.6.2}$$

für die rechte Seite schreiben wir kürzer

$$= P\{X \in B\}$$

Das durch Gleichung 1.6.2 erklärte W-Maß auf  $\mathcal{X}$  heißt die *Verteilung* der Zufallsvariablen  $X$ .

Wir haben in Gleichung 1.6.2 die Bezeichnung für das W-Maß von  $P_{\mathcal{X}}$  in  $P_X$  geändert, was sinnvoller ist.

Die Wahrscheinlichkeit eine 7 zu würfeln ist nun

$$P_X\{7\} := P\{X = 7\} = \frac{6}{36}.$$

und die Wahrscheinlichkeit, mindestens 7 Augen zu würfeln, ist  $P(\{\geq 7\}) = \frac{21}{36}$ .

Es führt auch nicht zu Schwierigkeiten, wenn man die Verteilung von  $X$  auf einer größeren Menge erklärt; z. Bsp auf  $M := \{1, 2, 3, \dots, 12\}$ . Dann ist eben  $P_X(\{1\}) = P(X = 1) = P(\emptyset) = 0$ .

#### 1.6.3 Satz (Verteilung einer Zufallsvariablen)

Es seien  $(\Omega, P)$  ein W-Raum,  $M$  eine endliche Menge und  $X$  eine  $M$ -wertige Zufallsvariable.

(i) Dann erhält man ein W-Maß  $P_X$  auf  $M$  durch die Vorschrift:

$$P_X(B) := P(\{\omega \mid \omega \in \Omega, X(\omega) \in B\})$$

für  $B \in 2^M$ . Man schreibt dies kürzer als  $P_X(B) = P(X \in B)$ .

$P_X$  heißt die **Verteilung** der Zufallsvariablen  $X$  oder das **Bild** von  $P$  unter  $X$  oder auch das durch  $X$  und  $P$  **induzierte W-Maß**.

(ii) Bezeichnet man wie üblich die Umkehrabbildung auf der Potenzmenge mit

$$\begin{aligned} X^{-1} : 2^M &\rightarrow 2^\Omega, \\ X^{-1} : B &\mapsto \{\omega \in \Omega \mid X(\omega) \in B\}, \end{aligned}$$

so gilt  $P_X := P \circ X^{-1}$ ; d.h.

$$P_X(B) = P(X^{-1}(B)) \quad \text{für } B \in 2^M.$$

**Anmerkung.** Bei einer reellen Zufallsvariablen  $X(\Omega, P) \rightarrow \mathbb{R}$  verwechselt man die Verteilung  $P_X$  nicht mit der kumulativen Verteilung. Für reelle  $X$  definiert man die *kumulative Verteilung*  $F_X$  durch

$$F_X : \mathbb{R} \rightarrow [0, 1], \quad F_X(t) := P(X \leq t) = \sum_{\substack{\omega \in \Omega \\ X(\omega) \leq t}} p_\omega.$$

Die kumulative Verteilung ist eine reelle Funktion einer reellen Variablen. Für endliche  $\Omega$  ist die kumulative Verteilung eine monoton wachsende rechtsseitig stetige Treppenfunktion.

Zwischen der Verteilung  $P_X$  und der kumulativen Verteilung  $F_X$  bestehen die folgenden Beziehungen:

$$\begin{aligned} F_X(t) &= P_X((-\infty, t] \cap X(\Omega)), \\ P((a, b] \cap X(\Omega)) &= F_X(b) - F_X(a). \end{aligned}$$

### 1.6.4 Festst. (Funktionen von Zufallsvariablen)

Gegeben seien ein W-Raum  $(\Omega, P)$ , endliche Mengen  $M, \Omega'$  und Zufallsvariablen

$$X : \Omega \rightarrow M, \quad Y : \Omega \rightarrow \Omega'$$

und eine Abbildung  $Z : \Omega' \rightarrow M$  so, daß

$$X = Z \circ Y$$

ist. Man bilde den W-Raum  $(\Omega', P_Y)$ . Dann haben die Zufallsvariablen

$$X : (\Omega, P) \rightarrow M \quad \text{und} \quad Z : (\Omega', P_Y) \rightarrow M$$

die gleiche Verteilung:  $(M, P_X) = (M, P_Z)$  oder kurz

$$(P_Y)_Z = P_{Z \circ Y}. \quad (1.6.3)$$

**Anmerkung.** 1. Eine andere Bezeichnung für das Bildmaß von  $P$  unter  $X$  ist

$$\mathcal{L}(X|P) = \mathcal{L}(X) := P_X.$$

Der Buchstabe  $\mathcal{L}$  steht für engl. *law* = Verteilung. Gleichung 1.6.3 liest sich in dieser Bezeichnung als  $\mathcal{L}(Z \circ Y|P) = \mathcal{L}(Z|\mathcal{L}(Y|P))$ .

2. Andere Autoren ([2]) nennen das Bildmaß  $P_X$  das von  $X$  induzierte Maß und bezeichnen es mit  $P \star X$ . Gleichung 1.6.3 liest sich in dieser Bezeichnung als  $(P \star Y) \star Z = P \star (Z \circ Y)$ .

3. Die obige Feststellung 1.6.4 besagt, daß man dasselbe Experiment mit unterschiedlichen Grundräumen und Zufallsvariablen modellieren kann.

### 1.6.5 Bem. (W-Maß $\Leftrightarrow$ Verteilung einer Z-Var.)

(i) Verteilungen von Zufallsvariablen und W-Maße sind zueinander äquivalente Konzepte.

- Die Verteilung einer Zufallsvariablen ist ein W-Maß.
- Zu jedem W-Maß  $P'$  auf einer endlichen Menge  $\Omega'$  gibt es eine Zufallsvariable  $X$  mit Bildmenge  $\Omega'$  und der Verteilung  $P'$ , z. B.  $\Omega := \Omega', P := P'$  und  $X := \text{id}_\Omega$ . Diese Zufallsvariable hat die Verteilung  $P_X = P'$ .

(ii) Aus Feststellung 1.6.4 folgt, daß  $X$  nicht eindeutig wählbar ist.

Darauf kommt es aber auch nicht an. Letztendlich will man die Wahrscheinlichkeit eines Ereignisses  $B \in 2^{\Omega'}$  bestimmen. Die Zufallsvariable  $X : (\Omega, P) \rightarrow \Omega'$  mit Verteilung  $P_X = P'$  kann dabei helfen, wenn man das Maß  $P\{X \in B\}$  leichter ermitteln kann. Man denke an das Beispiel 1.6.2 der Augensumme von zwei Würfeln, indem die Wahrscheinlichkeit für die Augensumme auf die einfachere Laplace-Wahrscheinlichkeit zurückgeführt wird.

(iii) Daher wird  $X : (\Omega, P) \rightarrow \Omega'$  später oft gar nicht explizit angegeben.

Zum Beispiel reicht die Angabe „Sei  $X$  eine binomialverteilte Zufallsvariable mit den Parametern  $n$  und  $p$ “ meist völlig aus. Dies bedeutet, daß „ $X$  die Wertemenge  $\{0, \dots, n\}$  hat und daß  $P_X = \mathcal{B}_{n,p}$  ist. Die explizite Angabe von  $(\Omega, P)$  und die Abbildungsvorschrift für  $X$  braucht man nur solange, bis man erkannt hat, daß  $X$  binomialverteilt mit diesen Parametern ist. Danach ist es zumeist völlig unwichtig, ob die Zufallsvariable  $X$  durch Würfeln oder zufälliges Ziehen von Karten oder sonstwie realisiert wird.

Andererseits kann aber ein gute Wahl eines speziellen  $X : (\Omega, P) \rightarrow \Omega'$  die Rechnung erleichtern. Resultate, die nur von der Verteilung  $P_X$  abhängen, gelten dann für alle Zufallsvariablen mit der gleichen Verteilung.

(iv) Ob man mit W-Räumen oder Zufallsvariablen argumentiert, ist Geschmackssache. Zufallsvariable sind oft anschaulicher und die Formulierungen mit Zufallsvariablen sind dichter am Problem. Das ist aber nicht unbedingt ein Vorteil!

Ein anderer Grund ist, Zufallsvariable kann man wie übliche Variable leicht in Formeln einsetzen, Dagegen ist das Hantieren mit Mengen und Maßen für den Nichtmathematiker gewöhnungsbedürftig.

## 1.7 Hypergeometrische Verteilung

### 1.7.1 Bsp. (Urnenmodell für hypergeom. Vert.)

1. Aus einer Urne mit  $N$  unterscheidbaren Kugeln wird eine Stichprobe vom Umfang  $n$  ohne Wiederholung entnommen. Die Urne enthalte zwei Sorten von Kugeln,  $K$  schwarze und  $N - K$  rote. Die Farbe soll keinen Einfluß auf Wahrscheinlichkeit einer Stichproben haben, d.h., alle Stichproben sind gleichwahrscheinlich.

Die Zufallsvariable  $X$  mit Werten in

$$\Omega := \{k \in \mathbb{N} \mid \max(0, n - (N - K)) \leq k \leq \min(K, n)\}$$

gibt die Anzahl der schwarzen Kugeln in der Stichprobe an.

Man überlege sich, daß  $\Omega$  die genaue Wertemenge von  $X$  ist.

Die Verteilung von  $X$  heiße die **hypergeometrische Verteilung** auf  $\Omega$ .

2. Für den Stichprobenraum kommen zwei Typen in Frage ( $\rightarrow$  Satz 1.2.4)

$\Omega_{\text{ord}}$ : Geordnete Stichproben vom Umfang  $n$  aus  $N$  ohne Wiederholung, Anzahl ist  $(N)_n$ .

$\Omega_{\text{unord}}$ : Ungeordnete Stichproben vom Umfang  $n$  aus  $N$  ohne Wiederholung, Anzahl ist  $\binom{N}{n}$ .

Man versehe  $\Omega_{\text{ord}}$  bzw.  $\Omega_{\text{unord}}$  jeweils mit der Laplace-Wahrscheinlichkeit  $P_{\text{ord}}$  bzw.  $P_{\text{unord}}$ .

3. Wir betrachten für  $X$  die beiden Fälle

$$\begin{aligned} X_{\text{ord}} &: \Omega_{\text{ord}} \rightarrow \Omega \\ X_{\text{unord}} &: \Omega_{\text{unord}} \rightarrow \Omega \end{aligned}$$

und zeigen, daß beide die gleiche Verteilung haben.

4. Es sei  $Y : \Omega_{\text{ord}} \rightarrow \Omega_{\text{unord}}$  die Abbildung, die jeder geordneten Stichprobe die entsprechende ungeordnete Stichprobe zuordnet. Jedes ungeordnete Stichprobe vom Umfang  $n$  hat  $n!$  mögliche Anordnungen, d.h.

$$|Y^{-1}\{\omega\}| = n! \quad \text{für alle } \omega \in \Omega_{\text{unord}}.$$

Da  $\Omega_{\text{ord}}$  die Laplace-Wahrscheinlichkeit hat ist die Verteilung  $P_Y$  die Gleichverteilung auf  $\Omega_{\text{unord}}$ . Aus

$$X_{\text{ord}} = X_{\text{unord}} \circ Y.$$

folgt nach Feststellung 1.6.4, daß  $X_{\text{ord}}$  und  $X_{\text{unord}}$  die gleiche Verteilung haben.

5. Die Verteilung von  $X_{\text{unord}}$  ist etwas einfacher zu berechnen, als die von  $X_{\text{ord}}$ . Wir bestimmen dazu die Anzahl von  $\{X_{\text{unord}} = k\} \subset \Omega_{\text{unord}}$ :

Jede Stichprobe zerfällt in eine Stichprobe aus den schwarzen Kugeln und eine Stichprobe aus den roten Kugeln. Es gibt

$\binom{K}{k}$  Möglichkeiten, aus  $K$  schwarzen Kugeln  $k$  auszuwählen.

$\binom{N-K}{n-k}$  Möglichkeiten, aus  $N - K$  roten Kugeln  $n - k$  auszuwählen.

Es tritt jede Kombination von *schwarzen* Stichproben mit *roten* Stichproben auf. Nach dem Abzählprinzip

1.2.2 ist die Anzahl von  $\{X_{\text{unord}} = k\}$  das Produkt der Möglichkeiten:

$$|\{X_{\text{unord}} = k\}| = \binom{K}{k} \binom{N-K}{n-k}.$$

Für die Laplace-Wahrscheinlichkeit auf  $\Omega_{\text{unord}}$  folgt:

$$P_{\text{unord}}\{X_{\text{unord}} = k\} = \frac{|\{X_{\text{unord}} = k\}|}{|\Omega_{\text{unord}}|} = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}.$$

### 1.7.2 Bez. (hypergeometrische Verteilung)

Es seien  $N, K, n \in \mathbb{N}$  mit

$$0 \leq K \leq N, \quad 1 \leq n \leq N$$

und

$$\Omega := \{k \in \mathbb{N} \mid \max(0, n - (N - K)) \leq k \leq \min(K, n)\}.$$

Das durch

$$h_{N,K,n}(k) := \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

eindeutig bestimmte W-Maß auf  $\Omega$  heißt die **hypergeometrische Verteilung** mit den Parametern  $N, K, n$  und wird mit  $\mathcal{H}_{N,K,n}$  bezeichnet. Die zugehörige W-Funktion ist  $h_{N,K,n}$ .

### 1.7.3 Bem. (Hypergeom. Vert. auf $\{0, 1, \dots, n\}$ )

Mitunter wird die hypergeometrische Verteilung auch auf dem größeren Raum

$$\tilde{\Omega} := \{0, 1, \dots, n\}$$

erklärt. Beachtet man, daß nach Definition

$$\begin{aligned} \binom{K}{k} &= 0 \quad \text{für } k > K, \\ \binom{N-K}{n-k} &= 0 \quad \text{für } k < n - (N - K) \end{aligned}$$

ist, so gilt hier die gleiche Formel

$$\mathcal{H}_{N,K,n}\{k\} := \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

für  $k \in \{0, \dots, n\}$ .

### 1.7.4 Bsp. (Herleitung mit $\Omega_{\text{ord}}$ )

Wir berechnen direkt die Verteilung der Zufallsvariablen:

$$X := X_{\text{ord}} : \Omega_{\text{ord}} \rightarrow \Omega$$

( $\rightarrow$  Beispiel 1.7.1 Punkt 3). Der Raum  $\Omega_{\text{ord}}$  der geordneten Stichproben vom Umfang  $n$  aus  $N$  Elementen hat nach Voraussetzung die Laplace-Wahrscheinlichkeit. Die Anzahl der Stichproben ist  $(N)_n$  ( $\rightarrow$  Satz 1.2.4).

Wir zählen, wieviele dieser Stichproben  $k$  schwarze Kugeln enthalten. Dazu wählen wir zunächst geordnete Stichproben von  $k$  schwarzen und  $n - k$  roten Kugeln. Dann wählen wir eine  $k$ -elementige Teilmenge aus  $\{1, \dots, n\}$ , die die Plätze angibt, auf die schwarzen Kugeln der Reihe nach gelegt werden. Die weißen Kugeln kommen der Reihe nach auf die verbleibenden Plätze. So erhalten wir alle geordneten Stichproben vom Umfang  $n$  mit  $k$  schwarzen Kugeln. Es gibt

$(K)_k$  geordnete Stichproben vom Umfang  $k$  aus den  $K$  schwarzen Kugeln.

$(N-K)_{n-k}$  geordnete Stichproben vom Umfang  $n-k$  aus den  $N-K$  roten Kugeln.

$\binom{n}{k}$  Möglichkeiten, die Plätze für die  $k$  schwarzen Kugeln zu wählen.

Nach dem Abzählprinzip 1.2.2 ist die Anzahl von  $\{X = k\}$  das Produkt der Möglichkeiten:

$$|\{X = k\}| = (K)_k (N-K)_{n-k} \binom{n}{k}.$$

Für die Laplace-Wahrscheinlichkeit auf  $\Omega_{\text{unord}}$  folgt:

$$\begin{aligned} P\{X = k\} &= \frac{|\{X = k\}|}{|\Omega_{\text{ord}}|} = \frac{(K)_k (N-K)_{n-k} \binom{n}{k}}{(N)_n} \\ &= \mathcal{H}_{N,K,n}\{k\}. \end{aligned}$$

Das Herleiten der hypergeometrischen Verteilung mit Hilfe geordneter Stichproben ist zwar etwas aufwendiger. Sie bietet dafür einen anderen Vorteil: Man kann die hypergeometrisch verteilte Zufallsvariable  $X_{\text{ord}}$  als Summe von  $n$  gleichverteilten Bernoulli-Variablen schreiben. Diese Darstellung ist hilfreich bei der Berechnung des  $\rightarrow$  Mittelwertes der hypergeometrischen Verteilung.

**1.7.5 Bsp.**  $N$  Kugeln bestehen wieder aus  $K$  schwarzen und  $N-K$  roten Kugeln.

1. Auf dem Raum  $\Omega_{\text{ord}}$  der geordneten Stichproben ohne Wiederholung vom Umfang  $n$  aus den  $N$  Kugeln bilde man für  $(i = 1, \dots, n)$  die Zufallsvariablen

$$\begin{aligned} V_i &: \Omega_{\text{ord}} \rightarrow \{0, 1\}, \\ V_i &: \sigma \mapsto \begin{cases} 1 & \text{wenn } \sigma_i \text{ schwarz,} \\ 0 & \text{wenn } \sigma_i \text{ rot.} \end{cases} \end{aligned}$$

2. Die Zufallsvariable ( $\rightarrow$  Beispiel 1.7.1 Punkt 3)

$$X_{\text{ord}} := \sum_{i=1}^n V_i$$

gibt die Anzahl der schwarzen Kugeln in der geordneten Stichprobe an.

3. Wie man leicht sieht, ist jedes  $V_i$  einfach Bernoulli-verteilt mit Erfolgswahrscheinlichkeit  $p := \frac{K}{N}$ . Dazu berechnen wir die Anzahl der Elemente von  $\{V_i = 1\}$ . Es gibt

$K$  mögliche schwarze Kugeln, die man auf den  $i$ -ten Platz legen kann.

$(N-1)_{n-1}$  Möglichkeiten die Kugeln für die restlichen  $n-1$  Plätze zu wählen.

Da  $\Omega_{\text{ord}}$  mit Laplace-Wahrscheinlichkeit versehen ist, folgt

$$P\{V_i = 1\} = \frac{K(N-1)_{n-1}}{(N)_n} = \frac{K}{N}.$$

Die  $V_1, \dots, V_n$  sind also gleichverteilte Bernoulli-Variablen.

4. Da die Summe  $X$  hypergeometrisch und nicht binomialverteilt ist, beschreibt das Tupel  $(V_1, \dots, V_n)$  aber kein  $n$ -faches Bernoulli-Experiment.

Da die Stichprobe ohne zurücklegen gezogen wird, beeinflussen sich die Ergebnisse  $V_i$  gegenseitig. Z. Bsp. in dem Fall  $N = n = 2$  und  $K = 1$  kann man aus dem Wert von  $V_1$  auf den Wert von  $V_2$  schließen!

Um die im obigen Beispiel in Punkt 4 aufgetretene Problematik besser zu verstehen führen wir die  $\rightarrow$  gemeinsame Verteilung und die  $\rightarrow$  Marginalverteilung von Tupeln von Zufallsvariablen ein.

Wir konstruieren noch ein weiteres Modell einer Zufallsvariablen mit hypergeometrischer Verteilung.

**1.7.6 Bsp.** Die  $N$  Kugeln bestehen wieder aus  $K$  schwarzen und  $N-K$  roten Kugeln.

Man realisiere eine geordnete Stichprobe vom Umfang  $n$  aus  $N$  nummerierten Kugeln, indem man die  $N$  Kugeln permutiert und dann die jeweils ersten  $n$  Stück auswählt. Der Grundraum

$$\Omega_{\text{perm}} := \{(\sigma_1, \dots, \sigma_n) \in \{1, \dots, N\}^N \mid \sigma_i \neq \sigma_j \ (i \neq j)\}$$

der Permutationen habe die Laplace-Wahrscheinlichkeit.

Die Zufallsvariable

$$X_{\text{perm}} : \Omega_{\text{perm}} \rightarrow \Omega$$

gibt die Anzahl der schwarzen Kugeln in  $(\sigma_1, \dots, \sigma_n)$  an.

**Aufgabe:** 1. Man zeige, daß  $X_{\text{perm}}$  die gleiche Verteilung hat, wie Zufallsvariable  $X_{\text{ord}} : \Omega_{\text{ord}} \rightarrow \Omega$  (siehe Beispiel 1.7.1 Punkt 3).  $X_{\text{perm}}$  ist folglich hypergeometrisch verteilt.

2. Man berechne direkt die Wahrscheinlichkeit von  $\{W = k\}$ .

**1.7.7 Bsp.** Man kann als Grundraum auch den Raum  $\Omega_{sr}$  aller  $n$ -Tupel

$$\omega := (\omega_1, \dots, \omega_n) \in \{s, r\}^n$$

von *optisch nicht unterscheidbaren* Kugeln nehmen, von denen höchstens  $K$  schwarz ( $= s$ ) und höchstens  $N-K$  rot ( $= r$ ) sind. Die Zufallsvariable

$$X_{sr} : \Omega_{sr} \rightarrow \Omega \subset \mathbb{N}_0$$

gebe die Anzahl der schwarzen Kugeln in einem  $n$ -Tupel an. Um die Wahrscheinlichkeit eines  $n$ -Tupels  $\omega \in \Omega_{sr}$  zu bestimmen, nummeriere man  $K$  schwarze und  $N-K$  rote Kugeln, um sie unterscheidbar zu machen. D.h. man führe eine Zufallsvariable

$$W : \Omega_{\text{ord}} \rightarrow \Omega_{sr}$$

ein, die *diese Nummern wieder entfernt* und versee  $\Omega_{sr}$  mit der Bildverteilung  $(P_{\text{ord}})_W$ . Da

$$X_{\text{ord}} = X_{sr} \circ W$$

ist, hat  $X_{sr}$  die gleiche Verteilung wie  $X_{\text{ord}}$ , d.h.  $X_{sr}$  ist hypergeometrisch verteilt.

**Aufgabe:** Man zeige: Für  $\omega \in \Omega_{sr}$  ist

$$(P_{\text{ord}})_W\{\omega\} = P_{\text{ord}}(W^{-1}\{\omega\}) = \frac{(K)_k (N-K)_{n-k}}{(N)_n}$$

wobei  $k := X_{sr}$  die Anzahl der schwarzen Kugeln in  $\omega$  ist. Man berechne nun direkt die Verteilung von  $X_{sr}$ .

### 1.8 Gemeinsame Verteilung von Z-Var.

#### 1.8.1 Bez. (Gemeinsame Verteilung von Z-Var.)

Es seien  $(\Omega, P)$  ein endlicher W-Raum,  $M_1, M_2, \dots, M_n$  endliche Mengen und  $X_\nu : \Omega \rightarrow M_\nu$  Zufallsvariable. Man fasse die Zufallsvariablen  $X_\nu$  zu einer Zufallsvariablen  $X$  mit Wertevorrat  $M := M_1 \times M_2 \times \dots \times M_n$  zusammen:

$$X = (X_1, X_2, \dots, X_n) : \Omega \rightarrow M,$$

$$X : \omega \mapsto (X_1(\omega), X_2(\omega), \dots, X_n(\omega)).$$

Die Verteilung  $P_X$  nennt man die *gemeinsame Verteilung* von  $X_1, X_2, \dots, X_n$ . Sie ist durch die W-Funktion

$$p_X(x_1, x_2, \dots, x_n) = P\{X = (x_1, x_2, \dots, x_n)\}$$

für  $(x_1, x_2, \dots, x_n) \in M$  eindeutig bestimmt.

Man kann dies auch als  $p_X(x_1, x_2, \dots, x_n) = P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$  schreiben.

#### 1.8.2 Bem. (Assoziativgesetz für gemeins. Vert.)

Es seien  $(\Omega, P)$  ein endlicher W-Raum,  $M_1, M_2, \dots, M_n$  endliche Mengen,  $X_\nu : \Omega \rightarrow M_\nu$  Zufallsvariable und  $X := (X_1, X_2, \dots, X_n)$ . Für  $1 = k_0 < k_1 < k_2 < \dots < k_l = n$  bilde man

$$Y_\lambda := (X_{k_{\lambda-1}}, \dots, X_{k_\lambda}) \quad \text{für } \lambda = 1, 2, \dots, l-1,$$

$Y_l = (X_{k_{l-1}}, \dots, X_n)$  und  $Y := (Y_1, Y_2, \dots, Y_l)$ . Dann haben  $X$  und  $Y$  die gleiche gemeinsame Verteilung:

$$P_X = P_Y.$$

#### 1.8.3 Bsp. (Indikatorvariable)

1. Ein Zufallexperiment werde durch den W-Raum  $(\Omega, P)$  beschrieben. Für  $A \in 2^\Omega$  ist die Indikatorvariable (charakteristische Funktion)  $\mathbb{1}_A$  eine Zufallsvariable mit Werten in der Menge  $\{0, 1\}$ . Die Indikatorvariable  $\mathbb{1}_A$  hat die W-Funktion

$$p := P\{\mathbb{1}_A = 1\} = P(A),$$

$$1 - p = P\{\mathbb{1}_A = 0\} = P(A^c).$$

Die Bildverteilung  $P_{\mathbb{1}_A}$  der Indikatorvariablen ist die *einfache Bernoulli-Verteilung mit Parameter  $p := P(A)$*  ( $\rightarrow$  Bezeichnung 1.5.1).

2. Man führe das Experiment  $n$ -mal „unabhängig“ durch. Dies wird durch den W-Raum  $(\Omega^n, P^{\otimes n})$  beschrieben. Für  $A \in 2^\Omega$  sei

$$A_\nu := \{\omega \in \Omega^n \mid \omega_\nu \in A\}$$

das Ereignis, daß im  $\nu$ -ten Experiment  $A$  eintritt. Die Zufallsvariable

$$X := \sum_{\nu=1}^n \mathbb{1}_{A_\nu} : \Omega^n \rightarrow \{0, 1, 2, \dots, n\} \quad (1.8.1)$$

gibt an, wie oft das Ereignis  $A$  bei den  $n$ -Versuchen eingetreten ist. Die gemeinsame Verteilung der Indikatorvariablen  $\mathbb{1}_{A_1}, \mathbb{1}_{A_2}, \dots, \mathbb{1}_{A_n}$  ist die  $n$ -fache Bernoulli-Verteilung auf  $\{0, 1\}^n$ , denn es gilt ( $\rightarrow$  Bezeichnung 1.5.1 Gleichung (1.5.1)):

$$P^{\otimes n}\{\mathbb{1}_{A_1} = x_1, \dots, \mathbb{1}_{A_n} = x_n\} = p^k(1-p)^{n-k},$$

wobei  $k = X(x_1, \dots, x_n)$  die Anzahl der Einsen in dem Tupel  $(x_1, \dots, x_n)$  ist.

Also ist die Verteilung  $P_X$  die Binomialverteilung mit den Parametern  $n$  und  $p = P(A)$  ( $\rightarrow$  Bezeichnung 1.5.2).

**Anmerkung.** Die in Gleichung 1.8.1 definierte Zufallsvariable hat also die gleiche Verteilung wie das in Bezeichnung 1.5.2 beschriebene Experiment. Dies ist ein gutes Beispiel für die in Bemerkung 1.6.5 erläuterte Verwendung von Zufallsvariablen.

#### 1.8.4 Bez. (Marginalverteilung)

Es seien  $\Omega_\nu$  Mengen und  $\Omega := \Omega_1 \times \dots \times \Omega_n$ . Wir bezeichnen die Projektion von  $\Omega$  auf die  $\nu$ -te Komponente, ( $\nu = 1, \dots, n$ ), mit

$$\text{pr}_\nu : \Omega \rightarrow \Omega_\nu, \quad \text{pr}_\nu : (\omega_1, \omega_2, \dots, \omega_n) \mapsto \omega_\nu.$$

Man spricht auch kurz von den Koordinatenvariablen.

(i) Ist  $P$  ein W-Maß auf dem Produktraum, so heißen die Verteilungen der Projektionen  $\text{pr}_\nu$  die *eindimensionalen Randverteilungen* oder *Marginalverteilungen* von  $P$ .

(ii) Ist  $I = \{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$  eine  $k$ -elementige Teilmenge, so nennt man die gemeinsame Verteilung von  $(\text{pr}_{i_1}, \text{pr}_{i_2}, \dots, \text{pr}_{i_k})$  eine *k-dimensionale Rand* oder *Marginalverteilung* von  $P$ .

(iii) Ist  $P_X$  die gemeinsame Verteilung der Zufallsvariablen  $X = (X_1, X_2, \dots, X_n)$  und  $\{i_1 < \dots < i_k\} \subset \{1, \dots, n\}$ , dann ist die entsprechende  $k$ -dimensionale Randverteilung von  $P_X$  die gemeinsame Verteilung von  $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$  ( $\rightarrow$  Feststellung 1.6.4). Daher heißt die gemeinsame Verteilung von  $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$  auch eine *k-dimensionale Randverteilung* der Zufallsvariablen  $X = (X_1, X_2, \dots, X_n)$ .

**Anmerkung.** 1. Eine Verteilung  $P$  auf einem Produktraum  $\Omega = \{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$  läßt sich als Matrix schreiben, wenn man die Werte  $P\{(\mu, \nu)\}$  wie eine Matrix anordnet. Die Randverteilung  $P_{\text{pr}_1}$  ist dann die Zeilensumme und  $P_{\text{pr}_2}$  die Spaltensumme.

2. I. a. ist  $P$  verschieden von dem Produkt seiner Marginalverteilungen.

**Bsp.** Hier ein einfaches Beispiel für die obige Situation. Das Experiment ist das Werfen eines fairen Würfels. Also  $\Omega := \{1, 2, 3, 4, 5, 6\}$  mit Laplace-Wahrscheinlichkeit. Die Zufallsvariablen seien gerade Augenzahl bzw Augenzahl mindestens vier:

$$X_1 := \mathbb{1}_{\{2,3,6\}} \quad \text{und} \quad X_2 := \mathbb{1}_{\{4,5,6\}}.$$

$P_{X_1}, P_{X_2}$  ist jeweils die Bernoulli-Verteilung mit Parameter  $\frac{1}{2}$ :

$$P_{X_\nu}(1) = P_{X_\nu}(0) = \frac{1}{2} \quad \text{für } \nu = 1, 2.$$

Die Werte der gemeinsamen Verteilung geben wir als Matrix an und tragen auch die beiden Marginalverteilungen ein:

	1	0	$p_{X_1}$
1	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{2}$
$p_{X_2}$	$\frac{1}{2}$	$\frac{1}{2}$	

Für das Produkt der Randverteilungen gilt aber

$$(P_{X_1} \otimes P_{X_2})\{(\mu, \nu)\} = \frac{1}{4} \quad \text{für } \mu, \nu = 1, 2.$$

**1.8.5 Bsp. (Randverteilung der Multinomial-V.)**

Die Multinomialverteilung  $\mathcal{M}_{n;p_1,\dots,p_m}$  auf  $\Omega_m = \{0, 1, \dots, n\}^m$  hat als  $i$ -te Randverteilung die Binomialverteilung  $\mathcal{B}_{n,p_i}$ . ( $\rightarrow$  Bemerkung 1.5.5) und Bezeichnung 1.5.2).

Wir leiten die Behauptung auf zwei Weisen her; der zweite Beweis folgt in Beispiel 1.8.9

Zur Abkürzung sei  $P := \mathcal{M}_{n;p_1,\dots,p_n}(k_1, \dots, k_m)$ . Mit Hilfe der Multinomialformel ( $\rightarrow$  Bezeichnung 1.2.14 Gleichung (1.2.1)) berechnen wir direkt aus der Definition der Multinomialverteilung den Wert

$$\begin{aligned} P\{\text{pr}_m = k\} &= n! \left( \sum_{\substack{k_1+\dots+k_{m-1} \\ = n-k}} \prod_{\mu=1}^{m-1} \frac{p_\mu^{k_\mu}}{k_\mu!} \right) \frac{p_m^k}{k!} \\ &= (p_1 + \dots + p_{m-1})^{n-k} \frac{n!}{(n-k)!} \frac{p_m^k}{k!} \\ &= \binom{n}{k} (1 - p_m)^{n-k} p_m^k = \mathcal{B}_{n,p_m}\{k\}. \end{aligned}$$

Aus der Symmetrie der Formel für die Multinomialverteilung folgt entsprechend  $P\{\text{pr}_i = k\} = \mathcal{B}_{n,p_i}\{k\}$  für  $i = 1, \dots, m$ .

**1.8.6 Bez.** Wir sagen, daß eine Abbildung  $f : \Omega \rightarrow M$  nur von der  $\nu$ -ten Koordinate abhängt, wenn es eine Abbildung

$$\tilde{f} : \Omega_\nu \rightarrow M \quad \text{gibt, so daß } f = \tilde{f} \circ \text{pr}_\nu$$

ist.

Man sagt hierfür auch,  $f$  läßt sich über  $\text{pr}_\nu$  faktorisieren.

**1.8.7 Satz (von je einer Koordinate abh. ZVn)**

Es seien  $(\Omega_\nu, P_\nu)$ ,  $(\nu = 1, \dots, n)$  endliche W-Räume und  $(\Omega, P) := (\Omega_1 \times \dots \times \Omega_n, P_1 \otimes \dots \otimes P_n)$  der Produktraum. Für Zufallsvariable  $X_\nu$ ,  $(\nu = 1, \dots, n)$ , die jeweils nur von der  $\nu$ -ten Koordinate abhängen, ist die gemeinsame Verteilung das Produkt der Verteilung der  $X_\nu$ :

Für dieses  $X := (X_1, X_2, \dots, X_n)$  ist

$$P_X = P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}.$$

**1.8.8 Bez. (Standardmodell für unabh. ZV.)**

Hat man  $n$  Zufallsexperimente, die durch die W-Räume  $(\Omega_1, P_1), \dots, (\Omega_n, P_n)$  beschrieben werden, so modelliert das Produktmaß  $P = P_1 \otimes \dots \otimes P_n$  die gemeinsame Durchführung dieser Experimente ohne gegenseitige Beeinflussung. Hängen die Zufallsvariablen  $X_\nu, (\nu = 1, \dots, n)$  jeweils nur vom Ausgang des  $\nu$ -ten Experimente ab, haben sie keinen Einfluß aufeinander. Die Verteilung der Zufallsvariablen  $X := (X_1, \dots, X_n)$  beschreibt das Ergebnis von  $n$  Zufallsexperimenten, die sich gegenseitig nicht beeinflussen. D.h. das W-Maß  $P_X$  des Zufallsexperimentes  $X$  ist das Produkt der W-Maße  $P_{X_\nu}$  der Einzelexperimente.

Der obige Satz 1.8.7 beschreibt einen wichtigen Spezialfall von unabhängigen Zufallsvariablen und ist zugleich das Standardmodell zur Konstruktion  $\rightarrow$  unabhängiger Zufallsvariabler.

Wir verwenden die allgemeine Idee von Bezeichnung 1.8.8 für eine zweite Herleitung der Randverteilung der Multinomialverteilung ( $\rightarrow$  Beispiel 1.8.5). Wir wählen dazu ein  $n$ -Tupel  $X = (X_1, \dots, X_m)$  von Zufallsvariablen  $X : (\Omega, P) \rightarrow \Omega_m$  die multinomial verteilt sind. Nach Bezeichnung 1.8.4 (iii) ist die  $i$ -te Randverteilung von  $P_X$  gleich  $P_{X_i}$ . Wir zeigen, daß  $P_{X_i}$  binomial verteilt ist.

**1.8.9 Bsp. (Randverteilung der Multinomial-V.)**

Die Multinomialverteilung  $\mathcal{M}_{n;p_1,\dots,p_m}$  auf  $\Omega_m = \{0, 1, \dots, n\}^m$  hat als  $i$ -te Randverteilung die Binomialverteilung  $\mathcal{B}_{n,p_i}$ . ( $\rightarrow$  Bemerkung 1.5.5) und Bezeichnung 1.5.2).

Dazu sei  $\Omega = \{\omega_1, \dots, \omega_m\}$  ein Grundraum mit  $m$  Elementen und  $P$  das W-Maß mit  $P\{\omega_i\} = p_i$  für  $i = 1, \dots, m$ . Man bilde das Produkt  $(\Omega^n, P^{\otimes n})$  und für  $i = 1, \dots, m$  die Zufallsvariablen

$$X_i : \Omega^n \rightarrow \{0, \dots, n\}$$

$$X_i : (y_1, \dots, y_n) \mapsto \sum_{j=1}^n \mathbb{1}_{\{\omega_i\}}(y_j).$$

$X_i$  beschreibt, wie oft das Element  $\omega_i$  unter den  $y_1, \dots, y_n$  vorkommt. Nach Bezeichnung 1.5.4 hat das  $m$ -Tupel  $X = (X_1, \dots, X_m)$  die Verteilung  $\mathcal{M}_{n;p_1,\dots,p_m}$ .

Nach Konstruktion ist die Zufallsvariable

$$Y_j : (x_1, \dots, x_n) \mapsto \mathbb{1}_{\{\omega_j\}}(x_j)$$

einfach Bernouli-verteilt mit Erfolgswahrscheinlichkeit  $p_i$  und hängt nur von der  $j$ -ten Komponente ab. Nach Satz 1.8.7 ist das  $n$ -Tupel  $Y = (Y_1, \dots, Y_n)$   $n$ -fach Bernoulli-verteilt mit dem Parameter  $p_i$  und folglich ( $\rightarrow$  Bezeichnung 1.5.1 (4)) ist die Summe

$$X_i := \sum_{j=1}^n Y_j$$

binomialverteilt mit den Parametern  $n$  und  $p_i$ .

### 1.9 Unabhängige Zufallsvariable

#### 1.9.1 Ziel (Unabhängigkeit von Zufallsvar.)

Man hat ein Zufallsexperiment  $E$  mit dem  $W$ -Raum  $(\Omega, P)$  und  $n$  Zufallsvariable  $X_\nu : \Omega \rightarrow M_\nu, (\nu = 1, \dots, n)$ . Wir führen nun die folgenden beiden Experimente durch:

- Wir führen das Experiment  $E$   $n$ -mal so durch, daß sich die einzelnen Durchführungen nicht beeinflussen und beobachten der Reihe nach die Elementarereignisse  $\omega_1, \dots, \omega_n$ . Bei der ersten Durchführung notieren den Wert  $X_1(\omega_1)$ , bei der zweiten Durchführung  $X_2(\omega_2)$  und beim  $n$ -ten mal den Wert  $X_n(\omega_n)$ . Nach Satz 1.8.7 ist die Verteilung der so gebildeten Zufallsvariablen

$$(\omega_1, \dots, \omega_n) \mapsto (X_1(\omega_1), \dots, X_n(\omega_n))$$

das Produkt  $P_{X_1} \otimes \dots \otimes P_{X_n}$  der Verteilung der Komponenten.

- Wir führen das Experiment  $E$  durch. Wenn das Ergebnis  $\omega \in \Omega$  ist, so notieren wir das  $n$ -Tupel  $X(\omega) := (X_1(\omega), \dots, X_n(\omega))$ .

Wir nennen die Zufallsvariablen  $X_1, \dots, X_n$  *unabhängig*, wenn sich bei der zweiten Versuchsanordnung dieselbe gemeinsame Verteilung ergibt wie bei der ersten Versuchsanordnung, d.h.

$$P_X = P_{X_1} \otimes \dots \otimes P_{X_n}$$

**Bsp.** Hier ein einfaches Beispiel für die obige Situation. Das Experiment ist das Werfen eines fairen Würfels. Also  $\Omega := \{1, 2, 3, 4, 5, 6\}$  mit Laplace-Wahrscheinlichkeit. Die Zufallsvariablen seien gerade Augenzahl bzw Augenzahl durch drei teilbar:

$$X_1 := \mathbb{1}_{\{2,4,6\}} \quad \text{und} \quad X_2 := \mathbb{1}_{\{3,6\}}.$$

$P_{X_1}$  ist die Bernoulli-Verteilung mit Parameter  $\frac{1}{2}$ :

$$P_{X_1}(1) = P_{X_1}(0) = \frac{1}{2}.$$

$P_{X_2}$  ist die Bernoulli-Verteilung mit Parameter  $\frac{1}{3}$ :

$$P_{X_2}(1) = \frac{1}{3}, \quad P_{X_2}(0) = \frac{2}{3}.$$

Die Werte der gemeinsamen Verteilung geben wir als Matrix an und tragen die Marginalverteilungen  $P_{X_1}$  und  $P_{X_2}$  ein:

	1	0	$P_{X_1}$
1	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{2}$
0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{2}$
$P_{X_2}$	$\frac{1}{3}$	$\frac{2}{3}$	

Man prüft leicht nach, daß das Produkt  $P_{X_1} \otimes P_{X_2}$  dieselbe Tabelle ergibt.

Nach Feststellung 1.10.2 (b) reicht es in diesem Fall, die folgende Beziehung zu prüfen:

$$\begin{aligned} \frac{1}{2} \cdot \frac{1}{3} &= P(X_1 = 1)P(X_2 = 1) \\ &\stackrel{!}{=} P(X_1 = 1, X_2 = 1) = P(\{6\}) = \frac{1}{6}. \end{aligned}$$

Die restlichen Werte der Tabelle folgen hieraus.

#### 1.9.2 Def. (Unabhängigkeit von Zufallsvar.)

Es seien  $(\Omega, P)$  ein endlicher  $W$ -Raum,  $M_\nu$  endliche Mengen und  $X_\nu : \Omega \rightarrow M_\nu, (\nu = 1, 2, \dots, n)$ , Zufallsvariable.

Das Tupel  $X := (X_1, X_2, \dots, X_n)$  heißt unabhängig wenn die gemeinsame Verteilung  $P_X$  gleich dem Produkt der Verteilungen  $P_{X_1}, P_{X_2}, \dots, P_{X_n}$  ist:

$$P_X = P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}.$$

**Anmerkung.** Zur Betonung oder Unterscheidung sagt man auch **stochastisch** unabhängig.

Sind die Zufallsvariablen  $X_1, \dots, X_n$  nicht unabhängig, so heißen sie (stochastisch) abhängig.

#### 1.9.3 Satz (Unabhängigkeit von Zufallsvar.)

Es seien  $(\Omega, P)$  ein endlicher  $W$ -Raum,  $M_\nu$  endliche Mengen und  $X_\nu : \Omega \rightarrow M_\nu, (\nu = 1, 2, \dots, n)$ , Zufallsvariable. Die folgenden Aussagen sind äquivalent:

- (a) Die Familie  $X := (X_1, X_2, \dots, X_n)$  ist unabhängig.
- (b) Für beliebige  $B_\nu \in M_\nu, (\nu = 1, 2, \dots, n)$  gilt:

$$\begin{aligned} P\{X \in B_1 \times B_2 \times \dots \times B_n\} \\ = P\{X_1 \in B_1\} \cdot P\{X_2 \in B_2\} \cdot \dots \cdot P\{X_n \in B_n\}. \end{aligned}$$

Man kann die Bedingung auch in der Form

$$\begin{aligned} P\{X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n\} \\ = P\{X_1 \in B_1\} \cdot P\{X_2 \in B_2\} \cdot \dots \cdot P\{X_n \in B_n\}. \end{aligned}$$

schreiben.

- (c) Für  $(x_1, \dots, x_n) \in M_1 \times \dots \times M_n$  gilt

$$\begin{aligned} P\{X = (x_1, \dots, x_n)\} \\ = P\{X_1 = x_1\} \cdot P\{X_2 = x_2\} \cdot \dots \cdot P\{X_n = x_n\}. \end{aligned}$$

(d) Die Wahrscheinlichkeitsfunktion  $p_X$  der gemeinsamen Verteilung von  $X := (X_1, X_2, \dots, X_n)$  hat Produktgestalt: d.h. es gibt Funktionen  $p_\nu : M_\nu \rightarrow [0, 1]$  derart, daß

$$p_X(x_1, \dots, x_n) = p_1(x_1) \cdot \dots \cdot p_n(x_n)$$

für  $(x_1, \dots, x_n) \in M_1 \times \dots \times M_n$  gilt.

**Anmerkung.** In der Literatur wird häufig Satz 1.9.3 als Definition für die Unabhängigkeit von Zufallsvariablen gewählt und später die Äquivalenz zu Definition 1.9.2 gezeigt.

#### 1.9.4 Bem.

Es seien  $X_\nu : (\Omega, P) \rightarrow M_\nu, (\nu = 1, 2, \dots, n)$  unabhängige Zufallsvariable.

- (i) Ist  $\sigma$  eine Permutation von  $\{1, 2, \dots, n\}$  so sind auch  $X_{\sigma_1}, X_{\sigma_2}, \dots, X_{\sigma_n}$  unabhängig.

Man kann also auch von einer Familie  $(X_i)_{i \in I}$  von unabhängigen Zufallsvariablen sprechen, wenn  $I$  eine endliche Indexmenge ist.

- (ii) Ist  $I \subset \{1, 2, \dots, n\}$  eine Indexmenge, so ist die Teilfamilie  $(X_i)_{i \in I}$  ebenfalls unabhängig.

(iii) Für eine disjunkte Zerlegung  $I_1 \dot{\cup} \dots \dot{\cup} I_k = \{1, 2, \dots, n\}$  bilde man die Zufallsvariablen  $Y_\varkappa := (X_i)_{i \in I_\varkappa} (\varkappa = 1, \dots, k)$ . Dann sind  $Y_1, Y_2, \dots, Y_k$  unabhängig

**1.9.5 Satz (Funktionen unabhängiger ZV.)**

Es seien  $X_\nu : (\Omega, P) \rightarrow M_\nu$ ,  $(\nu = 1, 2, \dots, n)$  unabhängige Zufallsvariable. Ferner sei  $\{1, 2, \dots, n\}$  die disjunkte Vereinigung von Indexmengen  $I_1, I_2, \dots, I_k$  und

$$f_\varkappa : \prod_{i \in I_\varkappa} M_i \rightarrow N_\varkappa \quad \text{für } \varkappa = 1, 2, \dots, k$$

Abbildungen in endliche Mengen  $N_\varkappa$ . Man bilde die Zufallsvariablen

$$Z_\varkappa := f_\varkappa \circ (X_i)_{i \in I_\varkappa} \quad \text{für } \varkappa = 1, 2, \dots, k.$$

Dann sind  $Z_1, Z_2, \dots, Z_n$  unabhängig.

**Speziell gilt:** Sind die Zufallsvariablen  $X_\nu : (\Omega, P) \rightarrow M_\nu$ ,  $(\nu = 1, 2, \dots, n)$  unabhängig, und sind  $f_\nu : M_\nu \rightarrow N_\nu$  beliebige Abbildungen,  $N_\nu$  endlich, so sind die Zufallsvariablen

$$Z_\nu := f_\nu \circ X_\nu \quad (\nu = 1, 2, \dots, n)$$

unabhängig.

**Anmerkung.** (i) Satz 1.9.5 gilt nicht mehr, wenn die  $I_i$  nicht disjunkt sind.

(ii) Auch ist für eine unabhängige  $n$ -Tupel reeller Zufallsvariablen  $X := (X_1, X_2, \dots, X_n)$  und eine invertierbare  $n \times n$ -Matrix  $A$  ist das  $n$ -Tupel  $AX$  i. a. nicht unabhängig.

beginbsp Zzweimaliger fairer Münzwurf. Es seien  $pr_1, pr_2 : \{0, 1\}^2 \rightarrow \{0, 1\}$  die beiden Ergebnisse.  $pr_1, pr_2$  sind unabhängige Zufallsvariable. Dagegen sind die Zufallsvariablen

$$X_1 := pr_1 + pr_2 \quad \text{und} \quad X_2 := pr_1 - pr_2$$

nicht unabhängig. So folgt aus  $X_1 = 2$ , daß  $X_1 = 0$  ist oder aus  $X_2 = \pm 1$  folgt  $X_1 = 1$ . Andererseits sind

$$pr_1 = \frac{1}{2}(X_1 + X_2) \quad \text{und} \quad pr_2 = \frac{1}{2}(X_1 - X_2)$$

unabhängig, obwohl  $X_1, X_2$  abhängig sind.

**1.10 Unabhängige Ereignisse**

**1.10.1 Bez. (Unabh. Ereignisse)**

Endlich viele Ereignisse  $A_1, \dots, A_n$  eines W-Raumes  $(\Omega, P)$  heißen (stochastisch) unabhängig, wenn die Indikatorfunktionen  $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$  unabhängig sind.

**1.10.2 Festst. (unabhängige Ereignisse)**

Für die Unabhängigkeit der Ereignisse  $A_1, A_2, \dots, A_n$  reicht es, eine der folgenden einfacheren Bedingungen (b) oder (c) nachzuprüfen.

Die folgenden Bedingungen sind äquivalent:

(a)  $A_1, A_2, \dots, A_n \in 2^\Omega$  sind unabhängig.

(b) für jede nichtleere Teilmenge  $I \subset \{1, 2, \dots, n\}$  gilt:

$$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i).$$

(c) Für jede nichtleere Teilmenge  $I \subset \{1, 2, \dots, n\}$  gilt

$$P(\bigcap_{i \in I} A_i \cap \bigcap_{j \in I^c} A_j^c) = \prod_{i \in I} P(A_i) \cdot \prod_{j \in I^c} P(A_j^c).$$

Damit eine Familie  $A_1, \dots, A_n$  unabhängig ist, muß man in Feststellung 1.10.2 die Produktformel (b) bzw. (c) für alle Teilfamilien  $I$  prüfen.

**1.10.3 Bsp.** Eine faire Münze wird dreimal geworfen. Der Ergebnisraum ist  $\{0, 1\}^3$  mit Laplace-Wahrscheinlichkeit.

$$\begin{aligned} A_1 &:= \{(\omega_1, \omega_2, \omega_3) \mid \omega_1 + \omega_2 + \omega_3 \geq 2\} \\ A_2 &:= \{(\omega_1, \omega_2, \omega_3) \mid \omega_1 = 1\} \\ A_3 &:= \{(\omega_1, \omega_2, \omega_3) \mid \omega_2 = \omega_3\} \end{aligned}$$

Es ist  $P(A_i) = \frac{1}{2}$  für  $i = 1, 2, 3$  und

$$P(A_1 \cap A_2 \cap A_3) = P\{(1, 1, 1)\} = \frac{1}{8},$$

aber

$$\begin{aligned} P(A_1 \cap A_2) &= P\{(1, 1, 1), (1, 1, 0), (1, 0, 1)\} \\ &= \frac{3}{8} \neq P(A_1)P(A_2). \end{aligned}$$

**Anmerkung. (unvereinbare Ereignisse sind abhängig)**

1. Wir nennen Ereignisse  $A_1, \dots, A_n$  unvereinbar, wenn  $\bigcap_{i=1}^n A_i = \emptyset$  und mindestens ein  $P(A_i) \neq 0$ . Unvereinbare Ereignisse sind abhängig und nicht etwa unabhängig!

2. Der Begriff der paarweisen Unabhängigkeit ist nicht weiter wichtig. Im folgende Beispiel werden drei paarweise unabhängige Ereignisse konstruiert, die zusammen unvereinbar sind.

**1.10.4 Bem. ( $\triangleleft$  paarweise unabh. Ereignisse)**

Für  $n \geq 3$  können Ereignisse  $A_1, \dots, A_n$  auch dann abhängig sein, wenn je zwei der Ereignisse unabhängig sind (sogenannte paarweise Unabhängigkeit) Beispiel: Fairer Würfel wird zweimal unabhängig voneinander geworfen. D.h.  $\Omega := \{1, \dots, 6\}^2$  mit Laplace-Wahrscheinlichkeit.

$$\begin{aligned} A_1 &:= \{(x_1, x_2) \mid x_1 \text{ gerade}\} \\ A_2 &:= \{(x_1, x_2) \mid x_2 \text{ gerade}\} \\ A_3 &:= \{(x_1, x_2) \mid x_1 + x_2 \text{ ungerade}\} \end{aligned}$$

Es ist

$$P(A_i) = \frac{1}{2} \quad \text{für } i = 1, 2, 3$$

$$P(A_i \cap A_j) = \frac{1}{4} \quad \text{für } i, j = 1, 2, 3, i \neq j.$$

Die Ereignisse sind also paarweise stochastisch unabhängig. Sie sind aber nicht unabhängig, da

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0$$

ist.

**Anmerkung.** Dass Beispiel der Ereignisse  $A_1$  und  $A_3$  zeigt noch folgendes: Zwei Ereignisse können sehrwohl stochastisch unabhängig sein, obwohl eins *mitbestimmt*, ob das andere eintritt ( $\rightarrow$  Bedingte Wahrscheinlichkeit)

Aus Satz 1.9.5 und den Regeln aus Bemerkung 1.3.1 ergibt sich das folgende Resultat:

### 1.10.5 Satz (Rechenregeln: unabh. Ereignisse)

Es seien  $A_1, \dots, A_n$  unabhängige Ereignisse. Für eine Zerlegung  $I_1 \dot{\cup} \dots \dot{\cup} I_k = \{1, 2, \dots, n\}$  bilde man mittels mengentheoretischer Operationen (Durchschnitt, Vereinigung, Komplement) aus den Familien  $\{A_i \mid i \in I_\varkappa\}$ , ( $\varkappa = 1, \dots, k$ ) neue Mengen  $B_{I_1}, \dots, B_{I_k}$ .

Dann sind auch  $B_{I_1}, \dots, B_{I_k}$  unabhängig.

Beispiel: Sind  $A, B, C, D, E, F$  unabhängig, so sind auch  $A \cap B$ ,  $C \cup (D \setminus E)$ ,  $F^c$  unabhängig.

### 1.11 Bedingte Wahrscheinlichkeit

**1.11.1 Bsp.** Beim Wurf mit einem fairen Würfel ist **unter der Annahme**, daß eine **gerade** Augenzahl gewürfelt wurde, anschaulich klar, daß

- die Wahrscheinlichkeit von  $\{1, 3, 5\}$  gleich  $0$ ,
- die Wahrscheinlichkeit von  $\{2, 4, 6\}$  gleich  $1$

ist, und daß  $\{2\}$ ,  $\{4\}$ ,  $\{6\}$  die gleiche Wahrscheinlichkeit haben. Also ist, **unter der obigen Annahme**, die Wahrscheinlichkeit von  $\{2\}$  gleich  $\frac{1}{3}$  u.s.w.

Wir wollen die intuitive Wahl der Wahrscheinlichkeiten im obigen Beispiel 1.11.1 mit Überlegungen zur relativen Häufigkeit in einer langen Folge von Experimenten untermauern. Wir lösen uns für diese Diskussion vom konkreten Beispiel des Würfels.

#### 1.11.2 Ziel (Häufigkeit bei bedingten Exprmt.n.)

Das folgende Überlegung beruht auf Plausibilität, sie ist so nicht mathematisch beweisbar, sondern dient als Motivation für die anschließende Definition 1.11.3

Ein Zufallsexperiment  $E$  wird mit einem W-Raum  $(\Omega, P)$  modelliert. Wir betrachten zwei Ereignisse  $A, B \in 2^\Omega$  mit  $A \subset B$  und nehmen an, daß in einer Serie von Zufallsexperimenten, die sich gegenseitig nicht beeinflussen, für hinreichend großes  $n$  die relative Häufigkeiten  $R_n(A) \approx P(A)$  bzw.  $R_n(B) \approx P(B)$  seien. Weiterhin sei  $P(B) > 0$ .

Eine Serie von Zufallsexperimenten ergebe die Beobachtungen  $x_1, x_2, \dots, x_N$ . Aus diesen wähle man die Teilserie derjenigen Beobachtungen aus, die in  $B$  liegen:

$$\tilde{x}_1, \dots, \tilde{x}_m \in B.$$

Für großes  $N$  werden die die relativen Häufigkeiten  $R_N(B) = \frac{m}{N}$  sehr dicht an  $P(B) > 0$  liegen. Auch wird  $R_N(A) \approx P(A)$  sein.

Die relative Häufigkeit für das Ereignis  $A$  in der ausgewählten Teilserie ist

$$\begin{aligned} \tilde{R}_m(A) &= \frac{1}{m} \sum_{\mu=1}^m \mathbb{1}_A(\tilde{x}_\mu) = \frac{N}{m} \frac{1}{N} \sum_{\nu=1}^N \mathbb{1}_A(x_\nu) \\ &= \frac{R_N(A)}{R_N(B)} \approx \frac{P(A)}{P(B)}. \end{aligned}$$

#### 1.11.3 Def. (Bedingte Wahrscheinlichkeit)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $B \in 2^\Omega$  mit  $P(B) > 0$ .

(i) Man definiert ein W-Maß  $P(\cdot|B)$  auf  $2^B$  durch die Vorschrift

$$P(A|B) := \frac{P(A)}{P(B)} \quad \text{für } A \in 2^B.$$

$P(A|B)$  heißt die bedingte Wahrscheinlichkeit von  $A$  gegeben  $B$ .

(ii) Oft ist es zweckmäßig, die bedingte Wahrscheinlichkeit als W-Maß auf dem ursprünglichen Grundraum  $\Omega$  aufzufassen. Man setzt dann

$$P(C|B) := P(C \cap B|B) \quad \text{für } C \in 2^\Omega.$$

Streng genommen handelt es sich hierbei um das Bild des W-Maßes  $P(\cdot|B)$  unter der Zufallsvariablen  $\iota : B \rightarrow \Omega$ , wobei  $\iota$  die Einbettung der Teilmenge  $B$  in  $\Omega$  ist. Man identifiziert in diesem Fall das W-Maß  $P(\cdot|B)$  mit seinem Bildmaß unter  $\iota$ .

(iii) Man definiert praktischer Weise die bedingte Wahrscheinlichkeit auch im Fall  $P(B) = 0$  und setzt  $P(C|B) = 0$ . Auf diese Weise erreicht man, daß die folgenden Formeln auch in diesem Fall sinnvoll bleiben. Natürlich ist in diesem Fall  $P(\cdot|B) \equiv 0$  kein W-Maß.

**Anmerkung.** Man kann die Definition 1.11.3 auch etwas simpler auffassen: Für ein  $B \in 2^\Omega$  ist die Einschränkung  $P|_{2^B}$  ein Maß auf  $B$ . Wenn  $P(B) > 0$  ist, kann man die Einschränkung normieren, so daß man ein W-Maß erhält. Man bezeichnet

$$P(\cdot|B) := \frac{1}{P(B)} P|_{2^B}.$$

$(B, P(\cdot|B))$  entspricht der Bildung eines Unterraumes bei anderen mathematischen Strukturen. Die dann anschließende Frage, wie man einen W-Raum aus „disjunkten“ Unterräumen rekonstruiert, findet für endliche W-Räume eine ganz einfache Antwort ( $\rightarrow$  Abschnitt 1.12)

**1.11.4 Bsp.** Der Spielleiter wirft verdeckt zwei faire Würfel und verkündet „Augensumme ist  $\geq 8$ “. Frage: Wie groß ist die bedingte Wahrscheinlichkeit, daß mindestens eine 6 geworfen wurde?

Antwort: Unter der Voraussetzung, daß der Spielleiter vorher festgelegt hatte, daß er nur eine der Informationen „Augensumme  $< 8$  oder  $\geq 8$ “, bekanntgegeben wird, ist die bedingte Wahrscheinlichkeit für mindestens eine 6 unter den Augen  $\frac{3}{5}$ .

Da die Spielregel vorher bekannt ist, kann man aus der Angabe des Spielleiters keine weiteren Informationen ziehen. Was anderes wäre es, wenn der Spielleiter auch statt der 8 eine andere mögliche Zahl wählen dürfte!

Der Grundraum ist  $\{1, \dots, 6\}^2$  mit Laplace-Wahrscheinlichkeit. Es gibt die Ereignisse

$$\begin{aligned} A_6 &:= \{(1, 6), \dots, (6, 6), (6, 5), \dots, (6, 1)\}, \\ B_8 &:= \left\{ \begin{array}{ccccc} \cdot & \cdot & \cdot & \cdot & (2, 6) \\ \cdot & \cdot & \cdot & (3, 5) & (3, 6) \\ \cdot & \cdot & (4, 4) & (4, 5) & (4, 6) \\ \cdot & (5, 3) & (5, 4) & (5, 5) & (5, 6) \\ (6, 2) & (6, 3) & (6, 4) & (6, 5) & (6, 6) \end{array} \right\} \end{aligned}$$

Offensichtlich ist  $|B_8| = 15$  und  $|A_6 \cap B_8| = 9$  und somit

$$P(A_6 | B_8) = \frac{9}{36} / \frac{15}{36} = \frac{3}{5}.$$

**1.11.5 Bsp.** Eine Urne enthalte  $K$  schwarze und  $N - K$  rote Kugeln. Vom Spielleiter werden zunächst zufällig aber verdeckt  $m$  Kugeln entnommen und nichts über deren Farben mitgeteilt. Dann werden öffentlich  $n$  Kugeln gezogen, wobei natürlich  $n \leq N - m$  ist.

Was ist die Wahrscheinlichkeit, daß unter den  $n$  öffentlich gezogen Kugeln  $k$  Stück schwarz sind.

Antwort. Die Anzahl der schwarzen Kugeln unter den  $n$  öffentlich gezogenen Kugeln ist hypergeometrisch verteilt mit den Parametern  $N, K, n$ . Die Wahrscheinlichkeit für  $k$  schwarze Kugeln ist

$$\mathcal{H}_{N,K,n}\{k\} = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Die vorhergehende *verdeckte* Entnahme von  $m$  Kugeln hat keinen Einfluss auf die Wahrscheinlichkeitsverteilung der nachfolgenden  $n$  Kugeln.

Ein schnelles Argument hierfür ist: Die Wahrscheinlichkeit einer geordneten Stichprobe, in der zuerst  $m$  Kugeln kommen und dann eine Anordnung von  $n$  Kugeln, von denen  $k$  schwarz sind, hat die gleiche Wahrscheinlichkeit wie eine Stichprobe, in der die besagten  $n$  Kugeln vor den  $m$  Kugeln kommen.

Man kann den Vorgang auch als  $m + n$ -stufiges Zufallsexperiment modellieren und die (bedingte) Wahrscheinlichkeit dafür berechnen, daß  $k$  schwarze Kugeln gezogen werden, wenn zuvor irgendwelche  $m$  Kugeln gezogen wurden ( $\rightarrow$  Abschnitt 1.13 Mehrstufige Experimente).

**1.11.6 Satz (von der totalen Wahrscheinlichkeit)**  
Es seien  $(\Omega, P)$  ein endlicher W-Raum und

$$\Omega = B_1 \dot{\cup} \dots \dot{\cup} B_n$$

eine Zerlegung von  $\Omega$  in disjunkte Teilmengen  $B_\nu \in 2^\Omega$ .

(i) Dann gilt die **Formel von der totalen Wahrscheinlichkeit**

$$P(A) = \sum_{\nu=1}^n P(A|B_\nu) \cdot P(B_\nu) \quad \text{für } A \in 2^\Omega.$$

(ii) Im Fall  $P(A) > 0$  gilt für  $k = 1, \dots, n$  die **Formel von Bayes**:

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{\nu=1}^n P(A|B_\nu) \cdot P(B_\nu)}, \quad (1.11.1)$$

**Anmerkung.** 1. Die Formel von Bayes folgt aus der Formel (i) von der totalen Wahrscheinlichkeit und der Beziehung

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A|B_k)P(B_k)}{P(A)}$$

2. Ein Spezialfall von Satz 1.11.6 (i) ist die Formel

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

D.h. die Wahrscheinlichkeit von  $A$  ergibt sich aus den Wahrscheinlichkeiten  $P(B), P(B^c)$  und den bedingten Wahrscheinlichkeiten  $P(A|B), P(A|B^c)$ .

Man beachte die Konvention:  $P(B) = 0 \Rightarrow P(A|B) = 0$ .

3. Genauso erhält man als Spezialfall von Satz 1.11.6 (ii) die Formel

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}.$$

**1.11.7 Bsp.** Eine Person wird ohne *besonderen* Verdacht auf HIV getestet, etwa anlässlich einer Operation. Überraschend fällt der Test positiv aus. Wie ist das Ergebnis zu bewerten?

Die derzeit üblichen Tests (Stand 2001) erkennen Erkrankte mit einer Wahrscheinlichkeit von 0,998 und liefern für Gesunde mit einer Wahrscheinlichkeit von 0,99 ein negatives Resultat. Es gibt in Deutschland ca. 50.000 Erkrankte bei  $80 \cdot 10^6$  Einwohnern. Es bezeichne

$K$  Ereignis, daß irgend eine Person an HIV erkrankt ist,

$T_{\text{pos}}$  Ereignis, daß irgendeine Person positiv getestet wird.

Geht man also aus, daß die Person einem *durchschnittlichen* Erkrankungsrisiko unterliegt, so gilt

$$P(K) = \frac{50.000}{80 \cdot 10^6} \approx 6.25 \cdot 10^{-4}.$$

Bekannt sind die bedingten Wahrscheinlichkeiten

$$P(T_{\text{pos}} | K) = 0,998$$

$$P(T_{\text{pos}} | K^c) = 1 - 0,99 = 0,01.$$

Also ist

$$P(K|T_{\text{pos}}) = \frac{P(T_{\text{pos}}|K)P(K)}{P(T_{\text{pos}}|K)P(K) + P(T_{\text{pos}}|K^c)P(K^c)}$$

$$= \frac{0,998 \cdot 6.25 \cdot 10^{-4}}{0,998 \cdot 6.25 \cdot 10^{-4} + 0,01 \cdot (1 - 6.25 \cdot 10^{-4})}$$

$$\approx 0,06.$$

Die Wahrscheinlichkeit, daß die Person tatsächlich erkrankt ist, wenn sie positiv getestet wurde, beträgt 6%.

Gehört die Person dagegen zu einer Risikogruppe  $A$  mit  $P(K_A) = 0,01$  so ist die bedingte Wahrscheinlichkeit, daß diese Person tatsächlich krank ist

$$P(K_A|T_{\text{pos}}) = \frac{P(T_{\text{pos}}|K_A)P(K_A)}{P(T_{\text{pos}}|K_A)P(K_A) + P(T_{\text{pos}}|K_A^c)P(K_A^c)}$$

$$= \frac{0,998 \cdot 0,01}{0,998 \cdot 0,01 + 0,01 \cdot 0,99} \approx 50\%.$$

Das Beispiel zeigt, daß ein allgemeines Screening für seltene Krankheiten, das nicht auf Risikogruppen beschränkt ist, selbst bei Verwendung zuverlässiger Tests, von eingeschränktem Wert ist.

**1.11.8 Satz (Multiplikationsformel)**

In einem endlichen W-Raum gilt für  $A_1, \dots, A_n \in 2^\Omega$

$$P\left(\bigcap_{\nu=1}^n A_\nu\right) = P(A_1) \cdot \prod_{\nu=2}^n P(A_\nu | \bigcap_{\mu=1}^{\nu-1} A_\mu)$$

$$= P(A_1) \cdot P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots$$

$$\cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

**Bsp. (Geburtstagsproblem)** Die Wahrscheinlichkeit, daß von  $n$  zufällig nacheinander ausgewählten Personen keine zwei am selben Tag Geburtstag haben, ist

$$\frac{(365)_n}{365^n}.$$

Dabei wurde das Jahr zu 365 Tagen ohne Schaltjahre gerechnet. Alle Tage treten mit gleicher Wahrscheinlichkeit als Geburtstage auf.

Der zugrunde liegende W-Raum ist also  $(\Omega^n, P^{\otimes n})$ , wobei  $\Omega = \{1, \dots, 365\}$  und  $P$  die Laplace-Wahrscheinlichkeit auf  $\Omega$  ist. Man setze  $A_k$  das Ereignis, daß die  $(k+1)$ -te Person an einem anderen Tag Geburtstag hat, als die davor gewählten und wende Satz 1.11.8 an.

Anderes Model, daß zur gleichen Lösung führt: Wahrscheinlichkeit, daß eine geordnete Stichprobe mit Wiederholung vom Umfang  $n$  aus der Menge  $\{1, \dots, 365\}$  keine zwei gleichen Elemente enthält.

**1.11.9 Bem. (bed. Wahrsch. und Unabhängigkeit)**

Gegeben seien zwei Ereignisse  $A, B$  eines  $W$ -Raumes  $(\Omega, P)$  und es sei  $P(B) > 0$ .

$A, B$  sind dann und nur dann unabhängig, wenn  $P(A|B) = P(A)$  gilt.

**1.12 Zusammengesetzte W-Maße****1.12.1 Satz (zusammengesetzte W-Maße)**

Es sei  $(\Omega_0, P_0) = \{\omega_{01}, \dots, \omega_{0n}\}$  ein  $W$ -Raum mit  $n$  Elementen und dem  $W$ -Maß  $P_0$ . Weiterhin seien  $(\Omega_{11}, P_{11}), \dots, (\Omega_{1n}, P_{1n})$  endliche  $W$ -Räume und

$$\Omega := \Omega_{11} \dot{\cup} \dots \dot{\cup} \Omega_{1n}$$

die disjunkte Vereinigung der  $\Omega_{11}, \dots, \Omega_{1n}$ .

Oder anders ausgedrückt:  $\Omega$  hat eine Zerlegung in  $n$  paarweise disjunkte Mengen  $\Omega_{11}, \dots, \Omega_{1n} \in 2^\Omega$ .

Zur Abkürzung setze man  $p_{0\nu} := P_0\{\omega_{0\nu}\}$ .

Dann gibt es genau ein  $W$ -Maß  $P$  auf  $\Omega$  derart, daß für  $\nu = 1, \dots, n$  gilt:

$$P(\Omega_{1\nu}) = p_{0\nu}, \quad (1.12.1)$$

und falls  $p_{0\nu} > 0$  ist:

$$P(A|\Omega_{1\nu}) = P_{1\nu}(A \cap \Omega_{1\nu}) \quad \text{für } A \in 2^\Omega. \quad (1.12.2)$$

Beachte: Wenn  $P(\Omega_{1\nu}) = p_{0\nu} = 0$  ist, so ist definitionsgemäß  $P(A|\Omega_{1\nu}) = 0$ .

**Anmerkung. (Zusammengesetzte W-Maße)** 1. Aus der Formel von der totalen Wahrscheinlichkeit ( $\rightarrow$  Satz 1.11.6 (i)) und den Eigenschaften (1.12.1) und (1.12.2) folgt für das zusammengesetzte  $W$ -Maß  $P$  die Formel:

$$P(A) = \sum_{\nu=1}^n p_{0\nu} P_\nu(A \cap \Omega_{1\nu}) \quad \text{für } A \in 2^\Omega. \quad (1.12.3)$$

Dadurch ist  $P$  eindeutig bestimmt.

2. Zum Beweis der Existenz definiert man  $P$  durch die Formel (1.12.3) und prüft nach, daß diese Funktion  $P$  ein  $W$ -Maß mit den Eigenschaften (1.12.1) und (1.12.2) ist.

3. Speziell folgt aus Gleichung (1.12.2) die Produktformel

$$P(A) = p_{0\nu} P_{1\nu}(A) \quad \text{für } A \in 2^{\Omega_{1\nu}}.$$

4. Im Spezialfall, daß die Räume  $(\Omega_{1\nu}, P_{1\nu})$  alle gleich sind:

$$(\tilde{\Omega}, \tilde{P}) := (\Omega_{11}, P_{11}) = \dots = (\Omega_{1n}, P_{1n})$$

kann man  $\Omega$  mit  $\Omega_0 \times \tilde{\Omega}$  folgendermaßen identifizieren: Für  $\omega \in \Omega_{1\nu}$  sei  $\tilde{\omega}$  das entsprechende Element in  $\tilde{\Omega}$ . Dann ergibt die Zuordnung

$$\omega \mapsto (\omega_{0\nu}, \tilde{\omega})$$

eine Bijektion der disjunkten Vereinigung  $\Omega := \Omega_{11} \dot{\cup} \dots \dot{\cup} \Omega_{1n}$  mit dem Produktraum  $\Omega_0 \times \tilde{\Omega}$ . Unter dieser Identifizierung ist das zusammengesetzte  $W$ -Maß  $P$  gerade das Produktmaß  $P_0 \otimes \tilde{P}$ .

Dies sieht man folgendermaßen: Für  $\omega \in \Omega_{1\nu}$  folgt aus den Bedingungen (1.12.1) und (1.12.2), daß

$$\begin{aligned} P\{\omega\} &= P(\{\omega\} | \Omega_{1\nu}) \cdot P(\Omega_{1\nu}) = P_{1\nu}\{\omega\} \cdot P_0\{\omega_{0\nu}\} \\ &= P_0 \otimes \tilde{P}\{(\omega_{0\nu}, \tilde{\omega})\} \end{aligned}$$

ist.

5. Auch sonst ist es oft vorteilhaft, die Elemente von  $\Omega := \Omega_{11} \dot{\cup} \dots \dot{\cup} \Omega_{1n}$  etwas aufwendiger als Paare

$$(\omega_{0\nu}, \omega) \quad \text{für } \omega \in \Omega_{1\nu}$$

zu schreiben. Man denke an gekoppelte Experimente, wenn das erste das Ergebnis  $\omega_{0\nu}$  ist, führe man das Experiment mit dem Ergebnisraum  $\Omega_{1\nu}$  aus. Man erhält nun die Gesamtbeobachtung  $(\omega_{0\nu}, \omega)$  mit  $\omega \in \Omega_{1\nu}$ . Für die Wahrscheinlichkeit eines Elementarereignisses folgt aus Gleichung (1.12.2) die **Pfadformel**:

$$P\{(\omega_{0\nu}, \omega)\} = P_0\{\omega_{0\nu}\} P_{1\nu}\{\omega\}. \quad (1.12.4)$$

Wir benutzen diese Schreibweise in dem folgenden Beispiel:

**1.12.2 Bsp. (zusammengesetzte W-Maße)**

1. Wir haben drei Urnen, eine weiße Urne, eine schwarze und eine rote. Alle drei enthalten jeweils eine Anzahl von schwarzen Kugeln und roten Kugeln.

- weiße Urne: drei schwarze, eine rote
- schwarze Urne: vier schwarze, eine rote
- rote Urne: drei schwarze, zwei rote

Zuerst wird aus der weißen Urne eine Kugel gezogen. War diese schwarz, wird anschließend aus der schwarzen Urne eine Kugel gezogen, anderenfalls aus der roten Urne. Der Ereignisraum besteht also aus den Tupeln

$$\Omega := \{(s, s), (s, r), (r, s), (r, r)\}$$

Gesucht ist das W-Maß  $P$  auf  $\Omega$ , das die eingangs beschriebene Zugreihenfolge modelliert.

2. Wir bestimmen  $P$  mit Hilfe von Satz 1.12.1 über zusammengesetzte W-Maße, Bekannt sind die W-Räume:

$(\Omega_0, P_0)$ : Ziehen einer Kugel aus der weißen Urne

$$p_{01} := P_0\{s\} = \frac{3}{4}, \quad p_{02} := P_0\{r\} = \frac{1}{4};$$

$(\Omega_{11}, P_{11})$ : Ziehen einer Kugel aus der schwarzen Urne:

$$P_{11}\{s\} = \frac{4}{5}, \quad P_{11}\{r\} = \frac{1}{5};$$

$(\Omega_{12}, P_{12})$ : Ziehen einer Kugel aus der roten Urne:

$$P_{12}\{s\} = \frac{3}{5}, \quad P_{12}\{r\} = \frac{2}{5};$$

Wir können  $\Omega$  als disjunkte Vereinigung von  $\Omega_{11}$  und  $\Omega_{12}$  auffassen, indem wir identifizieren:

$$\Omega_{11} \cong \{(s, s), (s, r)\}, \quad \Omega_{12} \cong \{(r, s), (r, r)\}. \quad (1.12.5)$$

Durch diese Bezeichnung unterscheiden wir eine schwarze Kugel  $(s, s)$  in der schwarzen Urne von einer schwarzen Kugel  $(r, s)$  in der roten Urne.

Die Elementarereignisse  $\omega \in \Omega$  haben die folgenden bedingten Wahrscheinlichkeiten  $P(\{\omega\}|\Omega_{1i})$ , ( $i = 1, 2$ ), die wir in einer Tabelle eintragen:

$\omega$	$P(\{\omega\} \Omega_{11})$	$P(\{\omega\} \Omega_{12})$	
$(s, s)$	$P_{11}\{(s, s)\}$	0	
$(s, r)$	$P_{11}\{(s, r)\}$	0	(1.12.6)
$(r, s)$	0	$P_{12}\{(r, s)\}$	
$(r, r)$	0	$P_{12}\{(r, r)\}$	

Durch passende Wahl und Addition dieser Werte berechnet man

$$P(A|\Omega_{1i}) = P_{1i}(A \cap \Omega_{1i}) \quad \text{für } A \in 2^\Omega \text{ und } i = 1, 2.$$

3. So erhält man für die Ereignisse

$$\begin{aligned} S &:= \{\text{zweite Kugel schwarz}\} \\ R &:= \{\text{zweite Kugel rot}\} \end{aligned}$$

die bedingten Wahrscheinlichkeiten

$$\begin{aligned} P(S|\Omega_{11}) &= P_{11}\{(s, s)\}, & P(R|\Omega_{11}) &= P_{11}\{(s, r)\}, \\ P(S|\Omega_{12}) &= P_{12}\{(r, s)\}, & P(R|\Omega_{12}) &= P_{12}\{(r, r)\}. \end{aligned}$$

Mit der Formel (1.12.3) von der totalen Wahrscheinlichkeit folgt:

$$\begin{aligned} P(S) &= p_{01}P(S|\Omega_{11}) + p_{02}P(S|\Omega_{12}) = \frac{3}{4} \cdot \frac{4}{5} + \frac{1}{4} \cdot \frac{3}{5} = \frac{3}{4}, \\ P(R) &= p_{01}P(R|\Omega_{11}) + p_{02}P(R|\Omega_{12}) = \frac{3}{4} \cdot \frac{1}{5} + \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{4}. \end{aligned}$$

4. Mit der Formel von Bayes ( $\rightarrow$  Satz 1.11.6 (ii)) erhält man die Antwort auf die folgende Frage: Wie groß ist die bedingte Wahrscheinlichkeit, daß wir aus der schwarzen bzw. der roten Urne gezogen haben, gegeben, daß die zweite Kugel schwarz ist:

$$\begin{aligned} P(\Omega_{11} | S) &= \frac{P(S | \Omega_{11})P(\Omega_{11})}{P(S)} = \frac{\frac{4}{5} \cdot \frac{3}{4}}{\frac{3}{4}} = \frac{4}{5}, \\ P(\Omega_{12} | S) &= \frac{P(S | \Omega_{12})P(\Omega_{12})}{P(S)} = \frac{\frac{3}{5} \cdot \frac{1}{4}}{\frac{3}{4}} = \frac{1}{5}, \end{aligned}$$

**Anmerkung. (a priori und a posteriori-Verteilung)**  $P(\Omega_{11} | S)$  ist die bedingte Wahrscheinlichkeit, daß beim ersten mal eine schwarze Kugel gezogen wurde gegeben, daß auch beim zweiten Zug eine schwarze Kugel gezogen wird. Genauso ist  $P(\Omega_{12} | S)$  die bedingte Wahrscheinlichkeit, daß beim ersten mal eine rote Kugel gezogen wurde gegeben, daß beim zweiten Zug eine schwarze Kugel gezogen wird. Hierfür ist die folgende Sprechweise üblich:

Die Wahrscheinlichkeiten  $p_{01} = \frac{3}{4}$  und  $p_{02} = \frac{1}{4}$  für eine schwarze bzw. der rote Kugel beim ersten Zug nennt man die *a priori* Wahrscheinlichkeiten. Die Information, daß beim zweiten Zug die aus einer der beiden Urnen gezogene Kugel schwarz ist, *verschiebt* die apriori Wahrscheinlichkeiten zu den *a posteriori* Wahrscheinlichkeiten  $P(\Omega_{11} | S) = \frac{4}{5}$  bzw.  $P(\Omega_{12} | S) = \frac{1}{5}$ .

Für die schwarze Urne ist die a posteriori Wahrscheinlichkeit gegeben  $S$  größer als die a priori Wahrscheinlichkeit, da bei ihr das Verhältnis der schwarzen Kugeln zu den roten größer ist, als bei der roten Urne.

### 1.13 Mehrstufige Experimente

Oft wird ein mehrstufiges Zufallsexperiment durch einen Ereignisbaum beschrieben.

Ein **Baum** ist ein zusammenhängender gerichteter Graph, der zwischen je zwei Ecken genau eine Kante enthält. Wir machen von den folgenden einfachen Eigenschaften von Bäumen Gebrauch:

- Es gibt genau einen Knoten, genannt *Wurzel*, mit Eingangsgrad 0, d.h. es führen keine Kanten hinein. Alle anderen Knoten haben den Eingangsgrad 1, d.h. es führt genau eine Kante hinein.

- Die Knoten mit Ausgangsgrad 0 – d.h. es führen keine Kanten hinaus – heißen die *Endknoten* des Baumes. Alle anderen Knoten heißen *Verzweigungsknoten*.

- Ein *Pfad* ist eine Folge von Kanten, so daß der Endknoten einer Kante der Anfangsknoten der nächsten ist.

In einem Baum gibt es zu jedem Knoten genau einen Pfad von der Wurzel zu diesem Knoten. Die *Tiefe* eines Knotens ist die Länge dieses Pfades.

- Die Zuordnung von Knoten zu ihren Pfaden ergibt eine Bijektion der Menge der Endknoten auf die Menge der maximalen Pfade des Baumes.

Ein Ereignisbaum dient dazu, auf der Menge der Endknoten oder gleichbedeutend der Menge der maximalen Pfade ein W-Maß mit geeigneten Eigenschaften zu konstruieren.

#### 1.13.1 Bez. (Ereignisbaum)

(i) Ausgehend von einem Knoten der 0-ten Stufe läuft je eine Kante, zu den die Ausgänge der 0-ten Stufe repräsentierenden Knoten, wobei jede Kante mit der entsprechenden **Übergangswahrscheinlichkeit** oder *Eintrittswahrscheinlichkeit* des entsprechenden Elementes der 0-ten Stufe versehen wird.

(ii) Jeder dieser Knoten der 1-ten Stufe ist entweder ein Endknoten oder von ihm laufen wieder Kanten zu Knoten 2-ter Stufe, die die Ausgänge der ersten Stufe beschreiben. Hier werden die Kanten mit der **Übergangswahrscheinlichkeit** oder *Eintrittswahrscheinlichkeit* für das Ergebnis in der ersten Stufe versehen; usw.

(iii) Die Übergangswahrscheinlichkeiten der Kanten, die von demselben Knoten ausgehen, müssen sich zu Eins addieren.

**Anmerkung.** In einem mehrstufigen Zufallsexperiment sind die Elementarereignisse die Endknoten (Blätter) oder gleichbedeutend die Pfade von der Wurzel zu den Endknoten.

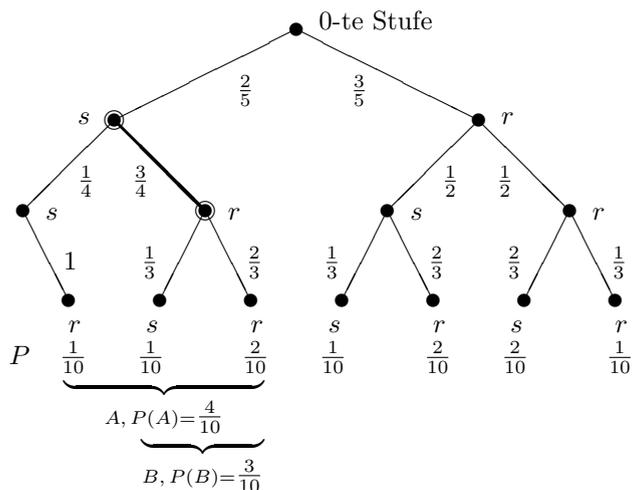
In einer Aufgabenstellung ist zumeist die Wahrscheinlichkeitsverteilung auf den Endknoten gefragt. Die Umformulierung in die entsprechenden Pfade ermöglicht es, die gesuchten Wahrscheinlichkeiten stufenweise zu berechnen:

Es gilt die **Pfadregel**: Die Wahrscheinlichkeit eines Pfades ist das Produkt der Übergangswahrscheinlichkeiten seiner Kanten. Für den Fall eines zweistufigen Experiments siehe Gleichung (1.12.4) und allgemein Folgerung 1.13.4.

Wir veranschaulichen die Bezeichnung 1.13.1 mit folgenden Beispiel eines dreistufigen Ereignisbaumes:

#### 1.13.2 Bsp. (Ereignisbaum)

Aus einer Urne mit 2 schwarzen und 3 roten Kugeln werden 3 Kugeln ohne Zurücklegen gezogen. Dies ergibt den folgenden Ereignisbaum:



Die Übergangswahrscheinlichkeit ist der jeweilige Anteil von schwarzen bzw. roten Kugeln in der Urne. Das Produkt dieser Anteile ergibt die Wahrscheinlichkeit einer Zugreihenfolge. In der untersten Zeile sind die Wahrscheinlichkeiten der Endknoten eingetragen.

Man beobachtet die folgenden Zusammenhänge: Für die angegebenen Teilmengen  $A, B$  ist

- die Wahrscheinlichkeit  $P(A) = \frac{2}{5}$  gleich der Übergangswahrscheinlichkeit der obersten linken Kante, die zu allen Endknoten der Menge  $A$  führt.

- die bedingte Wahrscheinlichkeit  $P(B|A) = \frac{3/10}{4/10} = \frac{3}{4}$ . Letzteres ist die Übergangswahrscheinlichkeit der Kante  $(s, r)$  (fett gezeichnet), entlang der sich die kleinere Menge  $B$  von der größeren Menge  $A$  *abspaltet*.

Die für  $A$  und  $B$  beobachteten Eigenschaften gelten für alle Teilmengen der Endknoten dieser Bauart und werden das W-Maß auf den Endknoten eindeutig charakterisieren ( $\rightarrow$  Satz 1.13.3).

**Aufgabe.** In dem Beispiel ist die Wahrscheinlichkeit der Pfade  $(s, s, r)$ ,  $(s, r, s)$ ,  $(r, s, s)$  immer gleich  $\frac{1}{10}$ . Man überlege sich, daß die Wahrscheinlichkeit eines Pfades nur von der Anzahl der schwarzen und roten Kugeln in dem Pfad und nicht von deren Reihenfolge abhängt.

**Aufgabe.** Aus einer Urne mit 2 schwarzen und 3 roten Kugeln werden Kugeln ohne zurücklegen gezogen. Man zeichne den Ereignisbaum für das folgende Experiment:

1. Es werden solange Kugeln entnommen, bis die zweite schwarze Kugel gezogen wurde.
2. Es werden solange Kugeln entnommen, bis die Urne leer ist.

Man berechne die Wahrscheinlichkeiten der einzelnen Pfade mit der Pfadregel. Wie groß ist in beiden Fällen die Wahrscheinlichkeit, daß in drei Zügen beide schwarzen Kugeln gezogen wurden? ( $\rightarrow$  Feststellung 1.13.5)

Der folgende Satz 1.13.3 besagt, daß es zu jedem Ereignisbaum, genau ein W-Maß auf den Raum der Pfade gibt, so daß die bedingte Wahrscheinlichkeit „der Menge der Pfade, die durch eine Kante laufen“, gegeben „die Menge der Pfade, die durch den Anfangspunkt der Kante laufen“, gerade die vorgegebene Übergangswahrscheinlichkeit dieser Kante ist.

#### 1.13.3 Satz (Mehrstufige Experimente)

Gegeben sei ein Ereignisbaum mit Wurzel  $a_0$ . Die Übergangswahrscheinlichkeiten der Kante mit Anfangspunkt  $a$  und Endpunkt  $b$  sei  $p(b|a)$ . Die Übergangswahrscheinlichkeiten der Kanten, die von einem Knoten ausgehen addieren sich zu 1.

Es sei  $\Omega$  der Raum der Endknoten. Für einen Knoten  $a$  sei  $\Omega_a$  die Menge der Endknoten, deren Pfad vom Ursprung zu dem Endknoten über den Knoten  $a$  führt.

Es gibt genau eine W-Maß  $P$  auf  $\Omega$  mit den folgenden beiden Eigenschaften:

(i) Für einen Knoten  $a$  erster Stufe ist

$$P(\Omega_a) = p(a|a_0).$$

(ii) Für eine Kante  $(a, b)$  des Baumes ist

$$P(\Omega_b|\Omega_a) = p(b|a).$$

**Beweisskizze.** Für einen Ereignisbaum sei der Grundraum  $\Omega$  die Menge der Endknoten. Zu jedem Endknoten gibt es genau einen Pfad von der Wurzel zu diesem Endknoten.

Der allgemeine Fall folgt aus dem ein- und zweistufigen Fall durch vollständige Induktion über die Länge  $l$  des längsten Pfades des Baumes zu einem Endknoten. Den Induktionsschritt „ $l \Rightarrow l + 1$ “ führen wir ab  $l = 2$  durch. Zuvor behandeln wir die dabei benötigten Fälle  $l = 0, 1, 2$ :

$l = 0$ : Für den Zweck des Beweises vereinbaren wir noch, daß im Falle eines trivialen Baumes, d.h. eines Baumes, der nur aus der Wurzel besteht und keine Kanten hat, die Wurzel ein Endknoten mit Wahrscheinlichkeit 1 ist.

$l = 1$ : Hierzu gibt es offensichtlich eine W-Raum  $(\Omega_0, P_0)$ , wobei  $P_0\{\omega_{0\nu}\}$  die Übergangswahrscheinlichkeit der betreffenden Kante ist.

$l = 2$ : Der Satz 1.12.1 zeigt für einen zweistufigen Baum die Existenz und Eindeutigkeit des W-Maßes auf den Endknoten.

**Induktionsschritt:** Für einen Baum der Länge  $l + 1$  entferne man die Wurzel und die von der Wurzel ausgehenden Kanten, die wir mit  $\omega_{01}, \dots, \omega_{0n}$  bezeichnen. Diese bilden einen Baum der Länge 1.

Die bisherigen Knoten erster Stufe sind nun die Wurzeln für eine Familie von  $n$  disjunkten Bäumen der Länge  $\leq l$ , die zusammen dieselben Endknoten haben, wie der vorgegebene Baum. Der Raum  $\Omega$  der Endknoten hat also eine disjunkte Zerlegung

$$\Omega := \Omega_{11} \dot{\cup} \dots \dot{\cup} \Omega_{1n}$$

in die Endknoten der gebildeten Teilbäume. Nach Induktionsannahme gibt es auf jedem der Räume  $\Omega_{1\nu}$  ein W-Maß  $P_{1\nu}$  mit den Eigenschaften (i) und (ii).

Nach Satz 1.12.1 gibt es auf  $\Omega$  genau ein W-Maß mit den Eigenschaften (1.12.1) und (1.12.2). Man prüfe nun nach, daß  $P$  die geforderten Eigenschaften (i) und (ii) im Bezug auf den ursprünglichen Baum hat.

**Anmerkung. (Bezeichnung der Übergangswahrscheinlichkeiten)** Manchmal schreibt man für eine Kante  $(a_{\nu-1}, a_\nu)$ , deren Anfang die Tiefe  $\nu - 1$  hat, die Übergangswahrscheinlichkeit auch suggestiver in der Form

$$p(a_\nu|a_0, \dots, a_{\nu-1}) := p(a_\nu|a_{\nu-1}),$$

wobei  $(a_0, \dots, a_{\nu-1}, a_\nu)$  der Pfad von der Wurzel zu  $a_\nu$  ist.

Man berechnet das W-Maß eines Pfades mit der Pfadregel:

#### 1.13.4 Folg. (Pfadregel)

Für einen Endknoten  $\omega \in \Omega$  sei  $(a_0, \dots, a_k = \omega)$  der Pfad vom Ursprung zu  $\omega$ . Dann ist

$$P\{\omega\} = p(a_1|a_0)p(a_2|a_1) \cdots p(a_k|a_{k-1}).$$

Ebenso folgt für einen inneren Knoten  $a_k$  mit Pfad  $(a_0, \dots, a_k)$ , daß

$$P(\Omega_{a_k}) = p(a_1|a_0)p(a_2|a_1) \cdots p(a_k|a_{k-1})$$

ist

**Beweis. (der Pfadregel)** Da

$$\{\omega\} = \bigcap_{\kappa=1}^k \Omega_{a_\kappa}$$

ist, folgt aus der Multiplikationsregel ( $\rightarrow$  Satz 1.11.8)

$$\begin{aligned} P\{\omega\} &= P(\Omega_{a_1})P(\Omega_{a_2}|\Omega_{a_1}) \cdots P(\Omega_{a_k}|\Omega_{a_{k-1}}) \\ &= p(a_1|a_0)p(a_2|a_1) \cdots p(a_k|a_{k-1}). \end{aligned}$$

**Anmerkung. (Fortgesetzte mehrstufige Experimente)**

Wenn man ein mehrstufiges Zufallsexperiment  $E_0$  noch durch einige weitere anschließende Zufallsexperimente zu einem Experiment  $E$  ergänzt, werden sich die Wahrscheinlichkeiten der zuvor gewonnenen Beobachtungen nicht ändern. Sie werden aber in die Wahrscheinlichkeiten des Gesamtexperimentes  $E$  eingehen.

Die folgende Feststellung besagt, wie man aus der zu  $E$  gehörenden W-Maß  $P$  das zu  $E_0$  gehörende W-Maß  $P_0$  berechnet:

Man erhält die Wahrscheinlichkeit eines Ereignisses  $\omega_0$  des Zufallsexperimentes  $E_0$ , wenn man die Wahrscheinlichkeiten aller möglichen Ereignisse des umfangreicheren Experimentes  $E$ , bei deren Gewinnung  $\omega_0$  als Zwischenschritt vorkommt, aufaddiert.

Diese Überlegung ist oft hilfreich. Man denke an ein Zufallsspiel, das beendet wird, sobald einer der Spieler eine vorgegebene Anzahl  $k$  von gewonnenen Spielen erreicht hat. Ein Spielverlauf besteht aus mindestens  $k$  und höchstens  $2k - 1$  Runden. Die Formeln werden meist übersichtlicher, wenn man in Gedanken immer  $2k - 1$  Runden spielt und dann die Spielverläufe zusammenfaßt, in denen jeweils ein Spieler mehr Runden gewonnen hat als der andere.

Formuliert man diese Regel in der Sprache der Ereignisbäume und Zufallsvariablen, so erhält man die folgende Feststellung.

#### 1.13.5 Festst. (Fortgesetzte mehrstufige Exp.)

Es sei  $T_0$  ein Ereignisbaum. Man bilde einen größeren Ereignisbaum  $T$  indem man an einige der Endknoten von  $T_0$  jeweils einen beliebigen Ereignisbaum anfügt.

Es seien  $\Omega_0$  die Endknoten von  $T_0$  und  $\Omega$  die Endknoten von  $T$ . Nach Konstruktion gibt es zu jedem Endknoten  $\omega \in \Omega$  genau einen Endknoten  $\omega_0 \in \Omega_0$  derart, daß der Pfad von der Wurzel nach  $\omega$  durch  $\omega_0$  geht. Auf diese Weise erklärt man eine Abbildung

$$X : \Omega \rightarrow \Omega_0 \quad \text{durch } \omega \mapsto \omega_0.$$

Es seien  $(\Omega_0, P_0)$  und  $(\Omega, P)$  die durch Satz 1.13.3 eindeutig bestimmten W-Räume.

Dann ist die Bildverteilung  $P_X = P_0$ .

#### 1.13.6 Bsp. (Verteilung der Asse beim Skat)

Entgegen der übliche Weise kann man sich vorstellen, daß jeder der Spieler der Reihe nach zehn Karten der 32 Karten auf einmal bekommt und die verbleibenden zwei in den „Skat“ kommen. Bei einem gut gemischten Kartenspiel ist die Reihenfolge in der die Karten ausgegeben werden irrelevant. Man vergleiche die Diskussion zur Herleitung der hypergeometrischen Verteilung ( $\rightarrow$  Abschnitt 1.7).

Wir modellieren dies als 3-stufiges Experiment, bei dem alle möglichen Zuteilungen von Karten mit der gleichen Anzahl von Assen für den jeweiligen Spieler zu einem Knoten zusammengefaßt sind, d.h. wir unterscheiden nur zwei Typen von Karten: *Asse* und *nicht-Asse*. Man kann den Baum verschieden weit ausführen.

– Sobald die vier Asse verteilt, ist ein Endknoten erreicht. Die Endknoten haben unterschiedliche Tiefe.

Es gibt also 5 Knoten erster Stufe für die Fälle, daß der erste Spieler  $k_1 = 0, \dots, 4$  Asse erhalten hat. Von dem Knoten 0 zweigen fünf Kanten ab, vom nächsten vier Kanten usw. Der Knoten 4 ist bereits ein Endknoten. Und so fort.

– Man zeichnet für jeden der drei Spieler die auf Grund der vorangehenden Verteilung der Asse verbleibenden Möglichkeiten ein. Alle Endknoten haben die Tiefe drei.

D.h., von dem Knoten 4 erster Stufe gibt es genau eine Kante zu dem Knoten (4, 0) und von dort eine Kante zu einem Knoten (4, 0, 0). Die Übergangswahrscheinlichkeiten für diese Kanten sind  $p(0|4) = 1$  und  $p(0|4, 0) = 1$ .

– Man zeichnet für jeden der drei Spieler je fünf Knoten für die Fälle  $k_1 = 0, \dots, 4$  Asse. Auf die Weise entsteht ein  $\rightarrow$  Produktbaum.

Auf Grund der bereits vergebenen Asse erhalten dann die unmöglichen Kanten die Übergangswahrscheinlichkeit 0.

Nach Feststellung 1.13.5 führen alle Möglichkeiten zum gleichen Ergebnis.

Die Übergangswahrscheinlichkeit für die Anzahl Asse ist hypergeometrisch verteilt, wobei aber die Parameter von dem vorangehenden Pfad abhängen.

Es sei  $\Omega$  die Menge der möglichen Pfade dieses Baumes und  $X_i : \Omega \rightarrow \{0, \dots, 4\}$  die Anzahl der Asse für den  $i$ -ten Spieler. Für die Übergangswahrscheinlichkeiten folgt dann

$$\begin{aligned} p(X_1 = k_1) &= \mathcal{H}_{32,4,10}(k_1), \\ p(X_2 = k_2 | X_1 = k_1) &= \mathcal{H}_{22,4-k_1,10}(k_2), \\ p(X_3 = k_3 | X_1 = k_1, X_2 = k_2) &= \mathcal{H}_{12,4-k_1-k_2,10}(k_3), \end{aligned}$$

Offensichtlich sind die Regeln für einen Ereignisbaum erfüllt. Es sei  $P$  die Wahrscheinlichkeitsverteilung auf den Pfaden zu den Endknoten.

Für die Wahrscheinlichkeit des Pfades (1, 1, 1), daß jeder Spieler genau ein Ass bekommt, erhält man mit der Pfadregel ( $\rightarrow$  Folgerung 1.13.4)

$$P\{(1, 1, 1)\} = \frac{\binom{4}{1} \binom{28}{9}}{\binom{32}{10}} \cdot \frac{\binom{3}{1} \binom{19}{9}}{\binom{22}{10}} \cdot \frac{\binom{2}{1} \binom{10}{9}}{\binom{12}{10}}$$

Die meisten Fakultäten kürzen sich weg:

$$= 10^3 \frac{4!2!}{32 \cdot 31 \cdot 30 \cdot 29} \approx 0,056.$$

**Bsp. (Produktbaum)** Das Produkt  $\Omega_1 \times \dots \times \Omega_n$  von endlichen Mengen kann man auch als Baum ansehen:

Mit dem leeren Tupel  $()$  bezeichne man die Wurzel, Die Knoten des Baumes sind die  $k$ -Tupel  $(x_1, \dots, x_k) \in \Omega_1 \times \dots \times \Omega_k$  mit  $k = 1, \dots, n$ . Schreibt man Pfade  $(( ), (x_1), (x_1, x_2), \dots, (x_1, \dots, x_k))$  vereinfacht als  $(x_1, \dots, x_k)$  so kann man die  $k$ -Tupel auch als Pfade lesen. Die  $n$ -Tupel sind die Endknoten oder die maximalen Pfade des Baumes.

Wir formulieren den Satz 1.13.3 noch einmal für den wichtigen Spezialfall der Produktbäume:

### 1.13.7 Folg. (Produktbäume)

Gegeben sei ein  $n$ -stufiges Experiment, ( $n \geq 2$ ). Das  $\nu$ -te Telexperiment habe die endliche Ergebnismenge  $\Omega_\nu$ ,

( $\nu = 1, \dots, n$ ). Man schreibe den Ereignisbaum als Produktbaum

$$\Omega := \Omega_1 \times \dots \times \Omega_n$$

und bezeichne die Ergebnisse des  $\nu$ -ten Experiments mit  $X_\nu : \Omega \rightarrow \Omega_\nu$ ; d.h.  $X_\nu$  ist die  $\nu$ -te Projektion  $\text{pr}_\nu$ .

Gegeben seien für jede Kante eines Pfades  $(x_\nu, x_{\nu-1}, \dots, x_1)$  die Übergangswahrscheinlichkeit  $p(x_\nu | x_{\nu-1}, \dots, x_1)$ , so daß  $0 \leq p(x_\nu | x_{\nu-1}, \dots, x_1) \leq 1$  und

$$\sum_{x \in \Omega_\nu} p(x | x_{\nu-1}, \dots, x_1) = 1$$

ist. Im Falle  $\nu = 1$  schreiben wir die Übergangswahrscheinlichkeiten als  $p(x_1)$  statt  $p(x_1 | ())$ .

Dann gibt es genau ein W-Maß  $P$  auf  $\Omega$  mit folgenden Eigenschaften:

(i) Für  $x_1 \in \Omega_1$  gilt

$$\begin{aligned} P\{X_1 = x_1\} &= P(\{x_1\} \times \Omega_2 \times \dots \times \Omega_n) \\ &= p(x_1). \end{aligned}$$

(ii) Für  $(x_1, \dots, x_\nu) \in \Omega_1 \times \dots \times \Omega_\nu$ , ( $\nu = 2, \dots, n$ ), gilt<sup>3</sup>

$$\begin{aligned} &P(X_\nu = x_\nu | X_1 = x_1, \dots, X_{\nu-1} = x_{\nu-1}) \\ &= P\left(\prod_{j=1}^{\nu-1} \Omega_j \times \{x_\nu\} \times \prod_{j=\nu+1}^n \Omega_j \mid \{(x_1, \dots, x_{\nu-1})\} \times \prod_{j=\nu+1}^n \Omega_j\right) \\ &= p(x_\nu | x_{\nu-1}, \dots, x_1). \end{aligned}$$

Dieses Maß wird mit der Pfadregel berechnet:

$$P\{\omega\} = p(x_1)p(x_2|x_2) \cdots p(x_n | x_1, \dots, x_{n-1})$$

für  $\omega = (x_1, \dots, x_n) \in \Omega$ .

<sup>3</sup>Man schreibt kurz

$$P(X = x | Y = y) := P(\{X = x\} | \{Y = y\}).$$

### 1.14 Polya's Urnenmodell

**Anmerkung. (Polya's Urnenmodelle)** Polya dachte bei dem folgenden Beispiel 1.14.1 an die Ausbreitung einer Infektion. Er wählte ein Urnenmodell, bei dem die Kugeln die infizierten bzw. die immunisierten Individuen innerhalb einer Population repräsentieren. Die Anzahlen der infizierten und der immunisierten wachsen nach einer Zufallsregel. Man kann aber auch an die Modellierung von „Seilschaften“ denken.

Das Modell kann man variieren, indem man statt nur einer zusätzlichen Kugel mehrere Kugeln der gleichen Farbe zurücklegt oder auch wegnimmt. Die so entstehenden Verteilungen für die Anzahl der gezogenen schwarzen Kugeln heißen Polya-Verteilungen. Spezialfälle sind die Binomialverteilung (keine zusätzliche Kugel) und die hypergeometrische Verteilung (kein Zurücklegen).

Wir untersuchen als Beispiel den Fall, daß jeweils die gezogene Kugel und eine zusätzliche Kugel der gleichen Farbe zurückgelegt wird.

#### 1.14.1 Bsp. (Polya's Urnenmodell)

Einer Urne enthält  $K$  schwarze und  $N - K$  rote Kugeln. Es wird eine Kugel gezogen und zusammen mit einer weiteren neuen Kugel derselben Farbe in die Urne zurückgelegt und so fort. Die Anzahl der Kugeln wächst also nach jedem Zug und die Zusammensetzung der Urne ändert sich zufällig.

1. Bei einer Zugfolge der Länge  $n$  seien  $k$  schwarze und  $n - k$  rote Kugeln gezogen worden. Nach dem Zurücklegen enthält die Urne dann  $N + n$  Kugeln, davon sind

$$K + k \text{ schwarz und } N + n - (K + k) \text{ rot.}$$

woraus sich die Übergangswahrscheinlichkeiten nach dieser Zugfolge für den nächsten Zug ergeben:

Die Wahrscheinlichkeit nun eine schwarze Kugel zu ziehen ist  $\frac{K+k}{N+n}$ .

Die Wahrscheinlichkeit nun eine rote Kugel zu ziehen ist  $\frac{N+n-(K+k)}{N+n}$ .

Führt man diese Zufallsexperiment  $n$ -mal durch, so erhält einen Ereignisbaum und somit ein  $W$ -Maß  $P_n$  auf dem Raum  $\Omega_n$  der Endknoten. Nimmt man statt der Endknoten die zugehörigen Pfade, so ist in diesem Fall  $\Omega_n = \{s, r\}^n$ .

2. Da die Übergangswahrscheinlichkeiten in jedem Schritt nicht von der Reihenfolge abhängen, in der rote oder schwarze Kugeln zuvor gezogen wurden, sondern nur von der Anzahl der bisher gezogenen schwarzen Kugeln, bieten sich die folgende Bezeichnungen an:

$$Z_i := \begin{cases} 1 & i\text{-te Kugel schwarz,} \\ 0 & i\text{-te Kugel rot.} \end{cases}$$

Die Anzahl der bis zum  $j$ -ten Schritt gezogenen schwarzen Kugeln ist

$$X_j := \sum_{i=1}^j Z_i.$$

3. Die Wahrscheinlichkeit eines Pfades  $(z_1, \dots, z_n) \in$

$\Omega_n$  mit  $X(z_1, \dots, z_n) = \sum_{i=1}^n Z_i(z_i) = k$  ist:

$$\begin{aligned} P_n(Z_1 = z_1, \dots, Z_n = z_n) & \quad (*) \\ &= \frac{K(K+1)\cdots(K+k-1)\cdot(N-K)(N-K+1)\cdots(N-K+(n-k)-1)}{N(N+1)\cdots(N+n-1)} \\ &= \frac{(K+k-1)_k((N-K)+(n-k)-1)_{n-k}}{(N+n-1)_n} \end{aligned}$$

Die Gleichung (\*) folgt aus der Pfadregel ( $\rightarrow$  Folgerung 1.13.4) durch Induktion über  $n$ .

4. Die Wahrscheinlichkeit, daß nach  $n$  Schritten  $k$  schwarze Kugeln gezogen wurden ist

$$\begin{aligned} P_n\{X_n = k\} & \quad (**) \\ &= \sum_{\substack{(z_1, \dots, z_n) \in \{s, r\}^n \\ X(z_1, \dots, z_n) = k}} P\{Z_1 = z_1, \dots, Z_n = z_n\} \\ &= \binom{n}{k} \frac{(K+k-1)_k((N-K)+(n-k)-1)_{n-k}}{(N+n-1)_n} \\ &= \binom{K+k-1}{k} \frac{((N-K)+(n-k)-1)}{n-k} / \binom{N+n-1}{n} \end{aligned}$$

Man kann diese Gleichung (\*\*) auch folgendermaßen deuten: In  $N$  Kästen, von denen  $K$  schwarz sind, werden  $n$  nicht unterscheidbare Kugeln mit Wiederholung gelegt. Auf dem Raum der Belegungen wähle man die Gleichverteilung. Die Zufallsvariable  $X_n$  gibt an, wieviele Kugeln in den schwarzen Kästen liegen.

5. Im Falle  $N = 2$  und  $K = 1$  ist

$$P_n\{X_n = k\} = \frac{1}{(n+1)} \quad \text{für } k = 0, \dots, n.$$

Die Zufallsvariable  $X_n$  beschreibt in diesem Fall ein komplizierte Methode, die Gleichverteilung zu erzeugen.

### 1.15 Erwartungswert

#### 1.15.1 Bsp.

Ihnen wird folgendes Spiel angeboten. Sie dürfen eine Münze vier mal werfen. Wenn in jedem der vier Würfe Zahl fällt, gewinnen Sie 20 Euro. Erscheint in genau drei Würfeln Zahl, so erhalten Sie 10 Euro. Sie müssen aber bei jedem Spiel einen Einsatz von 4 Euro zahlen. Würden Sie zustimmen, dieses Spiel einen ganzen Abend zu spielen?

Wenn Sie darauf eingehen, werden Sie ab und zu mal 6 Euro oder sogar 16 Euro gewinnen und in vielen anderen Runden 4 Euro verlieren. Die Frage lautet also, welchen *durchschnittlichen* Gewinn oder Verlust Sie bei den vielen Runden, die im Laufe des Abend gespielt werden, pro Runde zu erwarten haben.

Ob Sie in einer Runde etwas gewinnen oder verlieren, hängt vom Ausgang eines Zufallsexperiments ab. Wenn man unterstellt, daß es sich eine faire Münze handelt, wird man das Zufallsexperiment durch den Ereignisraum  $\Omega := \{1, 0\}^4$  mit der Laplace-Wahrscheinlichkeit beschreiben. Dabei stehe 1 für Zahl und 0 für Kopf. Die Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  mit

$$X(\omega) := \begin{cases} -4 & \text{für } \sum_{i=1}^4 \omega_i < 3, \\ 6 & \text{für } \sum_{i=1}^4 \omega_i = 3. \\ 16 & \text{für } \sum_{i=1}^4 \omega_i = 4 \end{cases}$$

gibt an, welchen Gewinn oder Verlust Sie in einem Spiel mit dem Ergebnis  $\omega = (\omega_1, \dots, \omega_4) \in \{1, 0\}^4$  machen. Die drei Fälle, die eintreten können, haben die Wahrscheinlichkeiten

$$P\{X = k\} := \begin{cases} \frac{11}{16} & \text{für } k = -4 \\ \frac{4}{16} & \text{für } k = 6 \\ \frac{1}{16} & \text{für } k = 16. \end{cases}$$

Bei einer großen Anzahl  $n$  von Spielen „erwartet“ man, daß die relativen Häufigkeiten in der die Verluste oder Gewinne eintreten, ungefähr gleich den angegebenen Wahrscheinlichkeiten sein werden:

$$\begin{aligned} R_n\{X = -4\} &\approx \frac{11}{16}, \\ R_n\{X = 6\} &\approx \frac{4}{16}, \\ R_n\{X = 16\} &\approx \frac{1}{16}. \end{aligned}$$

Die Bilanz bei  $n$  Spielen ist also voraussichtlich

$$\begin{aligned} n \cdot (-4R_n\{X = -4\} + 6R_n\{X = 6\} + 16R_n\{X = 16\}) \\ \approx n \left( -4 \cdot \frac{11}{16} + 6 \cdot \frac{4}{16} + 16 \cdot \frac{1}{16} \right) \\ = n \cdot \frac{-1}{4}. \end{aligned}$$

Bei dem Spiel wird man also im Schnitt  $\frac{1}{4}$  Euro pro Runde verlieren.

Entsprechende Überlegungen für eine beliebige **reellwertige Zufallsvariable** führen auf die folgende Definition:

#### 1.15.2 Def. (Erwartungswert)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $X : \Omega \rightarrow \mathbb{R}$  eine reellwertige Zufallsvariable mit der endlichen Wertemenge  $X(\Omega) \subset \mathbb{R}$ . Dann heißt die Zahl

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

der Erwartungswert der Zufallsvariablen  $X$ .

Zur Klarstellung, welches W-Maß verwendet wird, schreibt man auch  $E_P(X)$ .

#### 1.15.3 Festst. (Formeln: Erwartungswert)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $X, Y : \Omega \rightarrow \mathbb{R}$  reellwertige Zufallsvariable.

(i) Der Erwartungswert einer Zufallsvariablen hängt nur von ihrer Verteilung ab:

$$E(X) := \sum_{x \in X(\Omega)} x \cdot P_X\{x\}.$$

Man spricht deshalb auch von dem Erwartungswert eines W-Maßes auf einer endlichen Teilmenge von  $\mathbb{R}$ .

(ii) Häufig ist es praktisch, den Erwartungswert als Summe über die Elemente von  $\Omega$  auszudrücken:

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P\{\omega\}.$$

**Anmerkung.** Die Formel Feststellung 1.15.3 (i) erlaubt es, bei der Berechnung des Erwartungswertes die Zufallsvariable durch eine andere mit der gleichen Verteilung zu ersetzen.

**Anmerkung. (geordneter Vektorraum der ZVn)** Die reellen Zufallsvariablen  $X, Y$  über einem W-Raum  $(\Omega, P)$  bilden einen **geordneten reellen Vektorraum** mit der Addition

$$(X + Y)(\omega) := X(\omega) + Y(\omega),$$

und der Multiplikation

$$(aX)(\omega) := aX(\omega) \quad \text{für } a \in \mathbb{R}.$$

Die Ordnung  $X \leq Y$  bedeutet, daß punktweise  $X(\omega) \leq Y(\omega)$  für alle  $\omega \in \Omega$  ist.

#### 1.15.4 Festst. (Erwartungswert linear, monoton)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $X, Y : \Omega \rightarrow \mathbb{R}$  reellwertige Zufallsvariable.

(i) Die Bildung des Erwartungswertes ist **linear**:

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) \\ E(\lambda X) &= \lambda E(X) \quad \text{für } \lambda \in \mathbb{R}. \end{aligned}$$

(ii) Die Bildung des Erwartungswertes ist **monoton wachsend**:

$$\text{aus } X \leq Y \text{ folgt } E(X) \leq E(Y).$$

(iii) Der Erwartungswert einer Indikatorfunktion ist gleich der Wahrscheinlichkeit des zugehörigen Ereignisses:

$$E(\mathbb{1}_A) = P(A) \quad \text{für } A \in 2^\Omega.$$

#### 1.15.5 Bsp. (Serien)

Es sei  $(Z_0, Z_1, \dots, Z_n)$  eine Tupel von  $n + 1$ -fach Bernoulli-verteilten Zufallsvariablen mit Parameter  $p \in [0, 1]$ . Dann ist jeder der Zufallsvariablen

$$S_i := \mathbb{1}_{\{Z_i \neq Z_{i-1}\}} \quad \text{für } i = 1, \dots, n$$

Bernoulli-verteilt mit Parameter  $2p(1 - p)$ . Um dies einzusehen, bilden wir die gemeinsame Verteilung von  $S_{i-1}, S_i$ :

$Z_{i-1} \setminus Z_i$	1	0
1	$p^2$	$p(1 - p)$
0	$(1 - p)p$	$(1 - p)^2$

Also ist  $P\{S_i = 1\} = 2p(1 - p)$ . Wenn  $S_i = 1$  ist, so wird an der  $i$ -ten Stelle eine neue Serie von gleichen Werten eingeleitet. Die Zufallsvariable

$$Y := \sum_{i=1}^n S_i$$

beschreibt die Anzahl der Serien in einem Ergebnistupel  $(z_0, z_1, \dots, z_n)$ , wobei aber die von  $z_0$  eingeleitete Serie nicht mitzählt. Die Anzahl der Serien in dem Tupel ist  $X := Y + 1$  und der Erwartungswert für die Anzahl von Serien ist ( $\rightarrow$  Feststellung 1.15.4)

$$E(X) = E(1) + E(Y) = 1 + n \cdot 2p(1 - p).$$

Im Falle eines  $n + 1$ -maligen fairen Münzwurf ist die erwartete Anzahl von Serien gleich  $1 + \frac{n}{2}$ .

**Anmerkung.** Da der Erwartungswert einer Zufallsvariablen  $X$  nur von der Verteilung  $P_X$  abhängt, erhält man aus Feststellung 1.6.4 die Transformationsformel für den Erwartungswert:

**1.15.6 Festst. (Transformationsformel)**

Gegeben seien ein endlicher  $W$ -Raum  $(\Omega, P)$ , eine endliche Menge  $\Omega'$  und eine Abbildung  $Y : \Omega \rightarrow \Omega'$ . Man versehe  $\Omega'$  mit der Bildverteilung  $P_Y$ .

Wenn für die Zufallsvariablen

$$X : \Omega \rightarrow \mathbb{R} \quad \text{und} \quad Z : \Omega' \rightarrow \mathbb{R}$$

die Beziehung  $X = Z \circ Y$  besteht, dann gilt

$$E_P(X) = E_{P_Y}(Z).$$

Man kann dies Ergebnis folgendermaßen schreiben ( $\rightarrow$  Feststellung 1.6.4)

$$E_P(X) = \sum_{x \in X(\Omega)} x P_X\{x\} = \sum_{z \in Z(\Omega')} z (P_Y)_Z\{z\} = E_{P_Y}(Z).$$

**1.15.7 Folg. (Funktionen von ZVn)**

Es seien  $(\Omega, P)$  ein endlicher  $W$ -Raum und  $X : \Omega \rightarrow \mathbb{R}$  eine reellwertige Zufallsvariable. Für eine reelle Funktion  $f : X(\Omega) \rightarrow \mathbb{R}$  gilt

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

**1.15.8 Bsp. (Erwartungswerte)**

Es seien  $(\Omega, P)$  ein endlicher  $W$ -Raum und  $X : \Omega \rightarrow \mathbb{R}$  eine reellwertige Zufallsvariable.

1. Es sei  $X$  **Laplace-verteilt**,  $n := |\Omega|$ , und  $\Omega = \{\omega_1, \dots, \omega_n\}$ . Dann gilt

$$E(X) = \frac{1}{n} \sum_{\nu=1}^n X(\omega_\nu).$$

Der Erwartungswert einer Laplace-verteliten Zufallsvariablen  $X$  ist das *arithmetische Mittel* der Werte von  $X$ .

2. Es sei  $X$  **Bernoulli-verteilt** mit dem Parameter  $p$ . Dann gilt

$$E(X) = 1 \cdot P_X\{1\} + 0 \cdot P\{0\} = 1 \cdot p + 0 \cdot (1 - p) = p.$$

3. Es sei  $X$  **binomialverteilt** mit den Parametern  $n$  und  $p$ . Dann ist

$$E(X) = np.$$

**Beweis.** Für den Beweis benutzen wir das Standardmodell eines Bernoulli-Experimentes ( $\rightarrow$  Bezeichnung 1.8.8). Für einen rechnerischen Beweis  $\rightarrow$  Beispiel 1.15.9.

Es sei  $\Omega := \{1, 0\}$  mit  $W$ -Maß  $P\{1\} := p, P\{0\} = 1 - p$ . Auf  $(\Omega^n, P^{\otimes n})$  gibt die Zufallsvariablen

$$X(\omega) = \sum_{i=1}^n \text{pr}_i(\omega) = \sum_{i=1}^n \omega_i$$

die Anzahl der Einsen in dem  $n$ -Tupel  $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$  an.

Das Tupel  $(\text{pr}_1, \dots, \text{pr}_n)$  ist  $n$ -fach Bernoulli-verteilt mit dem Parameter  $p$  und folglich ist die Summe  $X$  binomialverteilt mit dem Parametern  $n$  und  $p$  ( $\rightarrow$  Beispiel 1.8.3). Nach Feststellung 1.15.4 folgt nun

$$E(X) = \sum_{i=1}^n E(\text{pr}_i) = np.$$

4. Es sei  $X$  **hypergeometrisch** verteilt mit den Parametern  $N, K$  und  $n$ . Dann gilt

$$E(X) = n \frac{K}{N}.$$

**Beweis.** Für den Beweis benutzen wir ein Urnenmodell mit geordneten Stichproben vom Umfang  $n$  ohne Wiederholung ( $\rightarrow$  Beispiel 1.7.5). Für einen rechnerischen Beweis  $\rightarrow$  Beispiel 1.15.9.

Eine Urne mit  $N$  Kugeln enthalte  $K$  schwarze und  $N - K$  rote Kugeln. Auf dem Raum  $\Omega_{\text{ord}}$  der geordneten Stichproben  $\sigma = (\sigma_1, \dots, \sigma_n)$  ohne Wiederholung vom Umfang  $n$  aus den  $N$  Kugeln bilde man für  $(i = 1, \dots, n)$  die Zufallsvariablen

$$V_i : \Omega_{\text{ord}} \rightarrow \{0, 1\},$$

$$V_i : \sigma \mapsto \begin{cases} 1 & \text{wenn } \sigma_i \text{ schwarz,} \\ 0 & \text{wenn } \sigma_i \text{ rot.} \end{cases}$$

Jede der Zufallsvariablen  $V_i$  ist Bernoulli-verteilt mit Parameter  $p = \frac{K}{N}$ . Die Zufallsvariable ( $\rightarrow$  Beispiel 1.7.1 Punkt 3)

$$X = X_{\text{ord}} := \sum_{i=1}^n V_i$$

gibt die Anzahl der schwarzen Kugeln in der geordneten Stichprobe an.  $X$  ist hypergeometrisch verteilt mit den Parametern  $N, K, n$ . Nach Feststellung 1.15.4 folgt nun

$$E(X) = \sum_{i=1}^n E(V_i) = np.$$

**Anmerkung.** Man beachte, das bei gleichen Parametern  $n$  und  $p = \frac{K}{N}$  die Binomialverteilung und die hypergeometrische Verteilung den gleichen Erwartungswert haben.

Der Erwartungswert der schwarzen Kugeln in einer Stichprobe vom Umfang  $n$  aus einer Urne mit  $K$  schwarzen und  $N - K$  roten Kugeln ist in den beiden Fällen „Ziehen mit Zurücklegen“ und „Ziehen ohne Zurücklegen“ gleich  $n \frac{K}{N}$ .

**1.15.9 Bsp.** Man kann den Erwartungswert der Binomialverteilung und der hypergeometrischen Verteilung auch direkt aus den Formeln für die Verteilung berechnen:

1. Es sei  $X$  binomialverteilt mit den Parametern  $n, p$ :

$$P\{X = k\} = \mathcal{B}_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dann folgt

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{\varkappa=0}^{n-1} \binom{n-1}{\varkappa} p^{\varkappa} (1-p)^{(n-1)-\varkappa} \\ &= np(p + (1-p))^{n-1} = np. \end{aligned}$$

2. Es sei  $X$  hypergeometrisch verteilt mit den Parametern  $N, K, n$ . Aus

$$P\{X = k\} = \mathcal{H}_{N,K,n}(k) = \binom{N}{n}^{-1} \binom{K}{k} \binom{N-K}{n-k}$$

folgt

$$\begin{aligned} E(X) &= \binom{N}{n}^{-1} \sum_{k=0}^n k \binom{K}{k} \binom{N-K}{n-k} \\ &= \binom{N}{n}^{-1} K \sum_{k=1}^n \binom{K-1}{k-1} \binom{N-K}{n-k} \\ &= \binom{N}{n}^{-1} K \sum_{\varkappa=0}^{n-1} \binom{K-1}{\varkappa} \binom{(N-1)-(K-1)}{n-1-\varkappa} \\ &= \binom{N}{n}^{-1} K \binom{N-1}{n-1} = n \frac{K}{N}. \end{aligned}$$

Hierbei wurde die folgende Formel für Binomialkoeffizienten benutzt

$$\begin{aligned} &\binom{N-1}{n-1}^{-1} \sum_{\varkappa=0}^{n-1} \binom{K-1}{\varkappa} \binom{(N-1)-(K-1)}{(n-1)-\varkappa} \\ &= \sum_{\varkappa=0}^{n-1} \mathcal{H}_{N-1,K-1,n-1}(\varkappa) = 1. \end{aligned}$$

**1.15.10 Bem. (partielle Summation)**

Für eine Zufallsvariable  $X : (\Omega, P) \rightarrow \mathbb{N}_0$  ist die folgende Summationsmethode oft hilfreich.

Mit  $m := \max(X(\Omega))$  gilt

$$E(X) = \sum_{\nu=1}^m P\{X \geq \nu\}$$

**Beweis.**

$$\begin{aligned} E(X) &= \sum_{\nu=0}^m \nu P\{X = \nu\} \\ &= \sum_{\nu=1}^m \sum_{\mu=\nu}^m P\{X = \mu\} \quad (*) \\ &= \sum_{\nu=1}^m \sum_{\mu=\nu}^m P\{X = \mu\} \quad (**) \\ &= \sum_{\nu=1}^m P\{X \geq \nu\} \end{aligned}$$

Für die Umsummierung von (\*) zu (\*\*) schaue man sich das folgende Bild an:

$$\begin{array}{ccccccc} P\{X = 1\} & & & & & & \\ P\{X = 2\} & P\{X = 2\} & & & & & \\ P\{X = 3\} & P\{X = 3\} & P\{X = 3\} & & & & \\ \vdots & & & & & & \\ P\{X = m\} & P\{X = m\} & P\{X = m\} & \dots & P\{X = m\} & & \end{array}$$

In (\*) wird über die Zeilen summiert und in (\*\*) über die Spalten.

### 1.16 Bedingte Erwartung

#### 1.16.1 Def. (bedingte Erwartung)

Es seien  $(\Omega, P)$  ein W-Raum,  $A \in 2^\Omega$  und  $X : \Omega \rightarrow \mathbb{R}$ .

Unter der bedingten Erwartung von  $X$  gegeben  $A$  versteht man den Erwartungswert der Einschränkung  $X|_A$  von  $X$  auf die Menge  $A$  in bezug auf das bedingte W-Maß  $P(\cdot|A)$ .

Man bezeichnet den bedingten Erwartungswert von  $X$  gegeben  $A$  mit

$$E(X|A) := \sum_{x \in X(A)} x \cdot P(X = x|A) \tag{1.16.1}$$

und wenn  $P(A) > 0$  ist:

$$= \frac{1}{P(A)} \sum_{\omega \in A} X(\omega)P\{\omega\}. \tag{1.16.2}$$

$$\tag{1.16.3}$$

Da die bedingte Wahrscheinlichkeit  $P(\cdot|A)$  auf die Menge  $A$  konzentriert ist, gilt auch

$$P(X|A) := \sum_{x \in X(\Omega)} x \cdot P(X = x|A). \tag{1.16.4}$$

**Anmerkung.** Analog zur Formel von der totalen Wahrscheinlichkeit ( $\rightarrow$  Satz 1.11.6) gilt:

#### 1.16.2 Satz (totale Erwartungswert)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und

$$\Omega = B_1 \dot{\cup} \dots \dot{\cup} B_n$$

eine Zerlegung von  $\Omega$  in disjunkte Teilmengen  $B_\nu \in 2^\Omega$  mit  $P(B_\nu) > 0$ . Für eine reelle Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  gilt dann

$$E(X) = \sum_{\nu=1}^n E(X|B_\nu) \cdot P(B_\nu)$$

**Anmerkung.** Zur Vereinfachung vereinbaren wir die folgende

**Konvention:** Damit die Formel vom totalen Erwartungswert allgemein gilt, setze man im Fall  $P(B) = 0$  auch  $E(X|B) = 0$ . Dies paßt zu der Vereinbarung Definition 1.11.3 (iii).

Man beachte nur, daß im Fall  $P(B) = 0$  weder  $P(\cdot|B) \equiv 0$  ein W-Maß ist noch  $E(X|B) = 0$  ein „Erwartungswert“ ist.

Eine andere Möglichkeit ist es, in Fall  $P(B) = 0$  den Ausdruck  $E(X|B)$  undefiniert zu lassen, und zu vereinbaren, daß dann  $E(X|B)P(B) = 0$  sein soll. Da im Falle endlicher und auch diskreter W-Räume ein Ereignis  $B$  mit  $P(B) = 0$  mit Sicherheit nicht eintritt, ist es unwichtig, wie man in diesem Fall das Symbol  $E(X|B)$  interpretiert. Wir kommen im Fall stetiger Verteilungen auf diese Frage wieder zurück.

#### 1.16.3 Bem.

Es seien  $(\Omega, P)$  ein W-Raum,  $A \in 2^\Omega$ ,  $P(A) > 0$  und  $X : \Omega \rightarrow \mathbb{R}$ . Sind  $X$  und  $\mathbb{1}_A$  unabhängig, so gilt

$$E(X|A) = E(X)$$

#### 1.16.4 Bez. (bedingte Erwartung $E(X|Y)$ )

Es seien  $(\Omega, P)$  ein endlicher W-Raum,  $M$  eine endliche Menge und

$$X : (\Omega, P) \rightarrow \mathbb{R} \quad \text{und} \quad Y : (\Omega, P) \rightarrow M$$

Zufallsvariable. Man definiert die *bedingte Erwartung von  $X$  gegeben  $Y$*  als

$$E(X|Y) : (M, P_Y) \rightarrow \mathbb{R},$$

mit

$$E(X|Y) : y \mapsto \begin{cases} E(X | Y = y) & \text{falls } P\{Y = y\} > 0, \\ 0 & \text{sonst} \end{cases}$$

für  $y \in M$ . Man beachte, daß diese bedingte Erwartung eine reelle Zufallsvariable ist.

**Bsp. (bedingte Erwartung bzgl. einer ZV)**  $X$  gebe das Gewicht eines Bürgers abgerundet in kg an und  $Y$  die Körpergröße abgerundet in cm. Man versehe die Menge der Bürger mit der Laplace-Verteilung  $P$ . Dann ist

$E(X)$  das Durchschnittsgewicht der Bevölkerung.

$E(X | Y = y)$  das Durchschnittsgewicht der Bürger, die  $y$  cm groß sind.

$E(X|Y)$  das Durchschnittsgewicht als Funktion der Größe.

$P_Y$  beschreibt den Anteil der Bürger mit gegebener Körpergröße an der Bevölkerung.

Nun ist anschaulich klar, daß der Erwartungswert von  $E(X|Y)$  bezüglich  $P_Y$  gleich dem Durchschnittsgewicht  $E(X)$  der Bevölkerung ist.

Wir wollen die obige anschauliche Überlegung in eine allgemeingültige Aussage überführen ( $\rightarrow$  Feststellung 1.16.5).

**Anmerkung.** 1. Mit der Vereinbarung  $E(X|B) = 0$ , falls  $P(B) = 0$  ist, können wir die obige Definition kürzer schreiben:

$$E(X|Y) : y \mapsto E(X | Y = y) \quad \text{für } y \in M.$$

2. Ist  $Y : (\Omega, P) \rightarrow M$  eine Zufallsvariable, so bildet die Familie  $\{Y = y\}$ , ( $y \in M$ ), eine Zerlegung von  $\Omega$ , bei der jedoch  $P\{Y = y\} = 0$  möglich ist. Wenn  $Y$  nicht surjektiv ist, sind auch einige der Mengen  $\{Y = y\}$  leer. Auch das stört nicht weiter.

Aus dem Satz von der totalen Erwartung folgt nun:

#### 1.16.5 Festst. (Erwartungswert von $E(X|Y)$ )

Es seien  $(\Omega, P)$  ein endlicher W-Raum,  $M$  eine endliche Menge und

$$X : (\Omega, P) \rightarrow \mathbb{R} \quad \text{und} \quad Y : (\Omega, P) \rightarrow M$$

Zufallsvariable. Für den Erwartungswert der bedingten Erwartung  $E(X|Y)$  gilt dann

$$E(E(X|Y)) = E(X).$$

#### 1.16.6 Bem. ( $E(X|Y)$ für $X, Y$ unabhängig)

Es seien  $(\Omega, P)$  ein W-Raum,  $A \in 2^\Omega$  und  $X, Y : \Omega \rightarrow \mathbb{R}$ . Sind  $X$  und  $Y$  unabhängig, so gilt

$$E(X|Y) = E(X) \mathbb{1}_{\{y|P\{Y=y\}>0\}}$$

$$P(E(X|Y) - E(X)) = 1$$

Da es auf die Werte auf unmöglichen Ereignissen nicht ankommt, setzt man für unabhängige  $X, Y$  zur Vereinfachung  $E(X|Y) = E(X)$ .

**Anmerkung. (bedingte Verteilung)** Zur Berechnung der bedingten Erwartung  $E(X|A)$  ist es oft nützlich, zuerst die *bedingte Verteilung*

$$x \mapsto P(X = x | A)$$

von  $X$  gegeben  $A$  zu berechnen.

Zur Berechnung der bedingten Erwartung  $E(X|Y)$  benötigt man zuerst die *bedingte Verteilung*

$$x \mapsto P(X = x | Y = y)$$

von  $X$  gegeben  $Y = y$ .

**1.16.7 Bez. (bedingte Verteilung)**

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $X : \Omega \rightarrow \mathbb{R}$ .

1. Für  $A \in 2^\Omega$  nennt man

$$2^{X(\Omega)} \ni B \mapsto P(X \in B | A)$$

die bedingte Verteilung von  $X$  gegeben  $A$ . Für  $P(A) > 0$  ist die *bedingte Verteilung* ein W-Maß.

Man schreibt auch  $P_{X|A}(B) := P(X \in B | A)$ .

2. Für eine Zufallsvariable  $Y \rightarrow M$  und  $C \in 2^M$  heißt

$$2^{X(\Omega)} \ni B \mapsto P(X \in B | Y \in C)$$

die *bedingte Verteilung* von  $X$  gegeben  $Y \in C$ .

Man schreibt auch  $P_{X|Y}(B|C) := P(X \in B | Y \in C)$ .

**1.16.8 Bsp.**

Es seien  $X, Y$  unabhängige, binomialverteilte Zufallsvariable mit den Parametern  $K, p$  bzw.  $L, p$  und  $0 < p < 1$ . Dann gilt mit  $N := K + L$

- (i) Die bedingte Verteilung von  $X$  gegeben  $X + Y = n$  ist die *hypergeometrische Verteilung* mit den Parametern  $N, K, n$ :

$$P(X | X + Y = n) = \mathcal{H}_{N,K,n} \quad \text{für } n = 0, \dots, N.$$

- (ii) Der bedingte Erwartung von  $X$  gegeben  $X + Y$  ist

$$E(X | X + Y) : n \mapsto n \frac{K}{N} \quad \text{für } n = 0, \dots, N.$$

**Beweis.** Wir betrachten  $N$ -fach Bernoulli-verteilte Zufallsvariable  $Z_1, \dots, Z_N$ , wobei  $N := K + L$  ist. D.h., wir haben  $N$  unabhängige Bernoulli-Experimente mit der gleichen Erfolgswahrscheinlichkeit  $0 < p < 1$ . Die Zufallsvariablen

$$X := \sum_{i=1}^K Z_i, \quad Y := \sum_{i=K+1}^N Z_i.$$

geben die Anzahl der Erfolge unter den ersten  $K$  bzw. unter den restlichen  $L$  Experimenten an.  $X + Y$  ist die Gesamtanzahl der Erfolge. Nach Bezeichnung 1.5.1 haben  $X$  und  $Y$  die angegebene Verteilung.

Zunächst wollen wir das Ergebnis, daß  $X | \{X + Y = n\}$  hypergeometrisch verteilt ist, anschaulich interpretieren. Dabei sieht man auch, warum der Erfolgsparameter  $p$  nicht in das Ergebnis eingeht. Es muß nur  $0 < p < 1$  gelten.

Dies folgt aber aus dem obigen Modell: Gegeben sind nun  $X + Y = n$  Erfolge, die sich in  $X$  Erfolge in den ersten  $K$  Experimenten und  $Y$  Erfolge unter den restlichen  $L = N - K$  Experimenten aufteilen. Eine solche Aufteilung wird durch die Einschränkung  $X | \{X + Y = n\}$  beschrieben.

Da es sich um unabhängige Bernoulli-Experimente mit gleicher Erfolgswahrscheinlichkeit handelt, sind alle diese Aufteilungen gleichwahrscheinlich. Jede Aufteilung der  $n$  Erfolge entspricht also einer Stichprobe vom Umfang  $n$  aus einer Urne mit  $K$  schwarzen Kugeln und  $N - K$  roten Kugeln ( $\rightarrow$  Beispiel 1.7.1). Folglich sind die Aufteilungen hypergeometrisch verteilt mit den Parametern  $N, K, n$ .

Wen diese Argumentation nicht ganz überzeugt, der rechnet mit den Definitionen:

Da die  $Z_1, \dots, Z_N$   $N$ -fach Bernoulli-verteilt mit dem Parameter  $p$  sind, ist

- $X$  binomialverteilt mit den Parametern  $K, p$ ,
- $Y$  binomialverteilt mit den Parametern  $L, p$ ,
- $X + Y$  binomialverteilt mit den Parametern  $N, p$ .

Wir berechnen nun die bedingte Wahrscheinlichkeit  $P(X = k | X + Y = n)$  nach Definition 1.11.3:

$$\begin{aligned} P(X = k | X + Y = n) &= \frac{P\{X = k, X + Y = n\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k, Y = n - k\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k\} \cdot P\{Y = n - k\}}{P\{X + Y = n\}} \\ &= \frac{\mathcal{B}_{K,p} k \mathcal{B}_{N-K,p} n - k}{\mathcal{B}_{N,p} n} \\ &= \frac{\binom{K}{k} p^k (1-p)^{K-k} \cdot \binom{N-K}{n-k} p^{n-k} (1-p)^{(N-K)-(n-k)}}{\binom{N}{n} p^n (1-p)^{N-n}} \\ &= \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \\ &= \mathcal{H}_{N,K,n}\{k\}. \end{aligned}$$

Nach Beispiel 1.15.8 (4) ist

$$E(X | X + Y)(n) := E(X | X + Y = n) = n \frac{K}{N}$$

Zur Übung berechnen wir noch

$$\begin{aligned} E(E(X | X + Y)) &= \sum_{n=0}^N E(X | X + Y = n) P\{X + Y = n\} \\ &= \frac{K}{N} \sum_{n=0}^N n \mathcal{B}_{N,p}\{n\} \\ &= \frac{K}{N} N p = K p = E(X), \end{aligned}$$

wie es nach Feststellung 1.16.5 sein muß.

**Aufgabe** Es seien  $(X_1, X_2, X_3)$  multinomialverteilt mit den Parametern  $N; p_1, p_2, p_3$ , wobei  $p_1, p_2, p_3 \in (0, 1)$  und  $p_1 + p_2 + p_3 = 1$  ist ( $\rightarrow$  Beispiel 1.5.3)

Man zeige:  $X_1$  gegeben  $X_2 = n$  ist binomialverteilt mit den Parametern  $N - n, \frac{p_1}{1-p_2}$  und berechne die bedingte Erwartung  $E(X_1 | X_2 = n)$ .

Man gebe eine anschauliche Erklärung für das Resultat.

**1.16.9 Bsp. (Zufällige Anzahl der Summanden)**

Es seien  $X_1, \dots, X_k : (\Omega, P) \rightarrow \mathbb{R}$  Zufallsvariable mit gleichem Erwartungswert

$$E(X_i) = E(X_j) \quad \text{für } (i, j = 1, \dots, k)$$

und  $N : (\Omega, P) \rightarrow \{1, \dots, N\}$  eine von  $(X_1, \dots, X_k)$  unabhängige ganzzahlige Zufallsvariable.

Man bilde die Summe

$$S := X_1 + \dots + X_N$$

mit einer zufälligen Anzahl von Summanden. Dann gilt

$$E(S) = E(X_1) \cdot E(N).$$

**Beweis.** Da  $X := (X_1, \dots, X_k)$  und  $N$  unabhängig sind, sind für festes  $n \in \{1, \dots, k\}$  auch

$$S_n := X_1 + \dots + X_n$$

und  $N$  unabhängig ( $\rightarrow$  Satz 1.9.5). Also ist

$$\begin{aligned} P(S = s \mid N = n) &= P(\{S_n = s\} \mid \{N = n\}) \\ &= P\{S_n = s\}. \end{aligned}$$

Wir berechnen nun für  $n \in \{1, \dots, k\}$  die bedingte Erwartung

$$\begin{aligned} E(S \mid N = n) &= \sum_{s \in \Omega} s P(S = s \mid N = n) \\ &= \sum_{s \in \Omega} s P\{S_n = s\} \\ &= E(S_n) = E(X_1 + \dots + X_n) \\ &= E(X_1) + \dots + E(X_n) \end{aligned}$$

und da die  $E(X_1) = \dots = E(X_n)$  ist

$$= E(X_1) \cdot n.$$

Es gilt also  $E(S|N) = E(X_1)N$ . Nach Feststellung 1.16.5 ist

$$E(S) = E(E(S|N)) = E(X_1) \cdot E(N).$$

## 1.17 Erwartungswert des Produktes

### 1.17.1 Satz (Produkt zweier unabhängiger ZV.)

Gegeben sei ein endlicher W-Raum  $(\Omega, P)$ . Sind  $X, Y : \Omega \rightarrow \mathbb{R}$  unabhängige Zufallsvariable, so gilt

$$E(XY) = E(X) \cdot E(Y)$$

**Beweis.** Man setze  $M := X(\Omega)$ ,  $N := Y(\Omega)$ . Man zerlege den Raum  $\Omega$  mit Hilfe der Zufallsvariablen  $(X, Y)$

$$\Omega = \bigcup_{(x,y) \in M \times N} \{(X, Y) = (x, y)\}$$

Auf jeder der Mengen  $\{(X, Y) = (x, y)\}$  ist das Produkt  $XY = xy$  konstant und folglich  $E(XY \mid (X, Y) = (x, y)) = xy$ . Aus der Formel für den totalen Erwartungswert folgt nun

$$\begin{aligned} E(XY) &= \sum_{(x,y) \in M \times N} E(XY \mid (X, Y) = (x, y)) P\{(X, Y) = (x, y)\} \\ &= \sum_{(x,y) \in M \times N} xy P\{(X, Y) = (x, y)\} \end{aligned}$$

und da  $X, Y$  unabhängig sind gilt nach Satz 1.9.3 (c)

$$\begin{aligned} &= \sum_{(x,y) \in M \times N} xy P\{X = x\} \cdot P\{Y = y\} \\ &= \sum_{x \in M} \sum_{y \in N} x P\{X = x\} \cdot y P\{Y = y\} \\ &= \sum_{x \in M} x P\{X = x\} \cdot \sum_{y \in N} y P\{Y = y\} \\ &= E(X) \cdot E(Y). \end{aligned}$$

### 1.17.2 Folg. (Produkt unabhängiger ZVn)

Gegeben sei ein endlicher W-Raum  $(\Omega, P)$ . Sind  $X_i : \Omega \rightarrow \mathbb{R}$ , ( $i = 1, \dots, n$ ) unabhängige Zufallsvariable, so gilt

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i).$$

**Anmerkung.** Man kann die Aussage über den Erwartungswert des Produktes zweier Zufallsvariablen auch folgendermaßen formulieren:

### 1.17.3 Bem. (bed. Erwart.-wert des Produktes)

Gegeben sei ein endlicher W-Raum  $(\Omega, P)$  Für  $X, Y : \Omega \rightarrow \mathbb{R}$  gilt

$$E(XY \mid Y) = E(X|Y) \cdot Y.$$

Für unabhängige  $X, Y$ , ist  $P(E(X|Y) - E(X)) = 1$  ( $\rightarrow$  Bemerkung 1.16.6) und folglich  $E(E(X|Y)) = E(X)$ . Mit Feststellung 1.16.5 folgt nun

$$E(XY) = E(E(XY \mid Y)) = E(X) \cdot E(Y).$$

**Anmerkung.** Die Menge der reellen Zufallsvariablen auf einem endlichen W-Raum  $(\Omega, P)$  bilden einen reellen Vektorraum  $\mathbf{V}$  der Dimension  $n = |\Omega|$ . Ist  $\Omega = \{\omega_1, \dots, \omega_n\}$ , so bilden die Indikatorfunktionen

$$\mathbb{1}_{\{\omega_1\}}, \dots, \mathbb{1}_{\{\omega_n\}}$$

Eine Basis von  $\mathbf{V}$ . Man jede reelle Zufallsvariable  $X$  in der Form

$$X = \sum_{\nu=1}^n X(\omega_\nu) \mathbb{1}_{\{\omega_\nu\}}$$

schreiben. Bezeichnet man die Werte  $x_\nu := X(\omega_\nu)$  so gilt  $X = \sum_{\nu=1}^n x_\nu \mathbb{1}_{\{\omega_\nu\}}$ . Das Tupel  $(x_\nu)_{\nu=1}^n$  sind die Koordinaten von  $X \in \mathbf{V}$ .

Auf dem Vektorraum  $\mathbb{R}^n$  verwendet man gerne das *Standardskalarprodukt*

$$x \cdot y := \sum_{\nu=1}^n x_\nu y_\nu \quad \text{für } x = (x_\nu), y = (y_\nu) \in \mathbb{R}^n.$$

und die Norm

$$\|x\| = \sqrt{\sum_{\nu=1}^n x_\nu^2} = \sqrt{x \cdot x}$$

Im  $\mathbb{R}^3$  ist  $\|x - y\|$  der übliche *Euklidische* Abstand der Punkte mit den Koordinaten  $x$  und  $y$ . Die Norm ist also der Abstand zum Nullpunkt. Man kann die Euklidische Geometrie des dreidimensionalen Anschauungsraumes mit Hilfe des Standardskalarproduktes und der Norm beschreiben. So stehen zwei Vektoren  $x, y$  genau dann senkrecht aufeinander, wenn das Standardskalarprodukt  $x \cdot y = 0$  ist. Ersetzt man die drei Dimensionen durch  $n$ -Dimensionen so erhält man eine naheliegende Verallgemeinerung.

Auf dem Vektorraum  $V$  verwendet man ein anderes „Skalarprodukt“, daß besser mit dem  $W$ -Raum  $(\Omega, P)$  harmonisiert, aber sonst analoge Eigenschaften hat. Man nimmt den Erwartungswert des Produktes

$$E(XY) = \sum_{\nu=1}^n X(\omega_\nu)Y(\omega_\nu)P\{\omega_\nu\} \quad \text{für } X, Y \in V.$$

Damit kann man eigentlich genauso rechnen wie mit dem Standardskalarprodukt. An die Stelle der Norm tritt der Ausdruck

$$\sqrt{E(X^2)},$$

für den man kein eigenes Symbol einführt. Dieses Skalarprodukt ermöglicht es mit Zufallsvariablen geometrisch zu argumentieren.

Man beachte aber, daß aus  $\sqrt{E(X^2)} = 0$  i.a. nicht  $X = 0$  sondern nur  $P\{X = 0\} = 1$  folgt.

### 1.17.4 Festst. (Eigenschaften von $E(XY)$ )

Die Funktion

$$V \times V \ni (X, Y) \mapsto E(XY)$$

hat die folgenden Eigenschaften:

- (i) **linear** in der ersten Variablen und **symmetrisch**

$$E((\lambda_1 X_1 + \lambda_2 X_2)Y) = \lambda_1 E(X_1 Y) + \lambda_2 E(X_2 Y)$$

$$E(XY) = E(YX).$$

- (ii) **positiv semidefinit**

$$E(X^2) \geq 0 \quad \text{und} \quad E(X^2) = 0 \quad \Leftrightarrow \quad P\{X = 0\} = 1.$$

Man sagt kurz,  $E(XY)$  ist eine symmetrische positiv semidefinite Bilinearform.

Wenn  $P\{\omega_\nu\} > 0$ , ( $\nu = 1, \dots, n$ ), ist, dann ist die Funktion  $E(XY)$  **positiv definit**

$$E(X^2) \geq 0 \quad \text{und} \quad E(X^2) = 0 \quad \Leftrightarrow \quad X = 0.$$

- (iii) Es gilt die **Schwarzsche Ungleichung**

$$|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$$

und die **Dreiecksungleichung**

$$\sqrt{E((X + Y)^2)} \leq \sqrt{E(X^2)} + \sqrt{E(Y^2)}.$$

In der Schwarzschen Ungleichung gilt das Gleichheitszeichen genau dann, wenn es ein  $\lambda \in \mathbb{R}$  so gibt, daß

$$P\{X = \lambda Y\} = 1 \quad \text{oder} \quad P\{\lambda X = Y\} = 1 \quad (*)$$

ist. In der Dreiecksungleichung gilt das Gleichheitszeichen genau dann, wenn es ein  $\lambda \geq 0$  gibt, so daß  $(*)$  gilt.

## 1.18 Varianz und Kovarianz

**Anmerkung.** Man will in einer einzigen Zahl zusammenfassen, wie weit die Werte einer reellen Zufallsvariablen  $X$  um ihren Erwartungswert **streuen**. Dafür bieten sich im Prinzip verschiedene Abstandsbegriffe an, z. B. die **durchschnittliche Abweichung**:

$$E(|X - E(X)|) = \sum_{x \in X(\Omega)} |x - E(X)|P\{X = x\}. \quad (1.18.1)$$

Für die durchschnittliche Abweichung gibt es aber wenig handhabbare Formeln.

Wegen der guten Eigenschaften, die aus der formalen Analogie zum euklidischen Abstand herrühren, bevorzugt man die **Standardabweichung**. Häufig ist es praktischer mit dem Quadrat der Standardabweichung, genannt die **Varianz**, zu rechnen.

### 1.18.1 Def. (Standardabweichung und Varianz)

Es sei  $(\Omega, P)$  ein endlicher  $W$ -Raum. Für  $X : (\Omega, P) \rightarrow \mathbb{R}$  definiert man

$$\sigma(X) := \sqrt{E((X - E(X))^2)} \quad (\text{Standardabweichung}) \quad (1.18.2)$$

$$\text{Var}(X) := E((X - E(X))^2) \quad (\text{Varianz}) \quad (1.18.3)$$

Um Klammern zu sparen, schreiben wir  $EX := E(X)$ .

### 1.18.2 Bem. (Varianz eines $W$ -Maßes)

- 1. Für ein ein  $W$ -Maß  $P$  auf einer endlichen Teilmenge  $\Omega \subset \mathbb{R}$  definiert man den **Mittelwert**

$$\mu(P) := \sum_{\omega \in \Omega} \omega P\{\omega\}, \quad (1.18.4)$$

die **Standardabweichung**:

$$\sigma(P) := \left( \sum_{\omega \in \Omega} (\omega - \mu)^2 P\{\omega\} \right)^{1/2} \quad (1.18.5)$$

und die **Varianz**:

$$\text{Var}(P) := \sum_{\omega \in \Omega} (\omega - \mu)^2 P\{\omega\} \quad (1.18.6)$$

- 2. Es ist  $\mu = E(\text{id}_\Omega)$ ,  $\sigma(P) = \sigma(\text{id}_\Omega)$  und  $\text{Var}(P) = \text{Var}(\text{id}_\Omega)$ .

- 3. Erwartungswert, Standardabweichung und Varianz einer reellen Zufallsvariablen  $X$  hängen nur von der Verteilung  $P_X$  ab:

$$E(X) = \mu(P_X), \quad \sigma(X) = \sigma(P_X), \quad \text{Var}(X) = \text{Var}(P_X).$$

### 1.18.3 Festst. (Rechenregeln: Varianz)

Es seien  $(\Omega, P)$  ein endlicher  $W$ -Raum und  $X, X_1, \dots, X_n$  reelle Zufallsvariable auf  $\Omega$ . Es gilt

$$(i) \text{Var}(\lambda X + b) = \lambda^2 \text{Var}(X) \quad \text{für } \lambda, a \in \mathbb{R}.$$

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

- (iii) Sind  $X_1, \dots, X_n$  unabhängig, so gilt die Gleichung von Bienaymé:

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n). \quad (1.18.7)$$

Die Gleichung von Bienaymé (1.18.7) gilt allgemeiner, wenn  $X_1, \dots, X_n \rightarrow$  unkorreliert sind. Es reicht z.B., daß sie paarweise unabhängig sind.

**Beweis.** (i) Es reicht, die Aussage für  $n = 2$  zu zeigen.

$$\begin{aligned} \text{Var}(X_1 + X_2) &= E(((X_1 - EX_1) + (X_2 - EX_2))^2) \\ &= E((X_1 - EX_1)^2) + E((X_1 - EX_1) \cdot (X_2 - EX_2)) \\ &\quad + E((X_2 - EX_2)^2). \end{aligned}$$

Da  $((X_1 - EX_1)$  und  $(X_2 - EX_2)$  unabhängig sind, folgt nach Satz 1.17.1

$$\begin{aligned} \text{Var}(X_1 + X_2) &= \text{Var}(X_1) + E(X_1 - EX_1) \cdot E(X_2 - EX_2) \\ &\quad + \text{Var}(X_2) = \text{Var}(X_1) + \text{Var}(X_2). \end{aligned}$$

### 1.18.4 Bsp. (Varianz einiger Verteilungen)

(i) Ist  $X : (\Omega, P) \rightarrow \{1, \dots, n\}$  Laplace-verteilt, so gilt

$$\text{Var}(X) = \frac{n^2 - 1}{12}.$$

(ii) Für eine Bernoulli-verteilte Zufallsvariable  $X$  mit Erfolgsparameter  $p$  gilt

$$\text{Var}(X) = p(1 - p).$$

(iii) Für eine Zufallsvariable  $X$ , die binomialverteilt mit den Parametern  $n, p$  ist, gilt

$$\text{Var}(X) = np(1 - p).$$

(iv) Für eine Zufallsvariable  $X$ , die hypergeometrisch verteilt ist mit den Parametern  $N, K, n$  ist, gilt

$$\text{Var}(X) = \frac{N - n}{N - 1} n \frac{K}{N} \left(1 - \frac{K}{N}\right).$$

**Anmerkung.** In einer Urne mit  $K$  schwarzen Kugeln und  $N - K$  roten Kugeln ist  $p := \frac{K}{N}$  die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen. Eine Stichprobe vom Umfang  $n$  mit Wiederholung aus der Urne ist die Anzahl der schwarzen Kugeln binomialverteilt mit Parameter  $n, p$  und in einer Stichprobe ohne Wiederholung ist die Anzahl der schwarzen Kugeln hypergeometrisch verteilt mit den Parametern  $N, K, n$ . In beiden Fällen hat man den gleichen Erwartungswert  $np = n \frac{K}{N}$ , aber die Varianz ist im hypergeometrischen Fall kleiner:

$$np(1 - p) < \frac{N - n}{N - 1} np(1 - p) \quad \text{für } n = 2, \dots, N.$$

Im Extremfall  $n = N$  ist die Varianz der hypergeometrischen Verteilung 0, was klar ist, da man in diesem Fall mit Sicherheit alle schwarzen Kugeln zieht.

**Beweis.** (i) Ist  $X : (\Omega, P) \rightarrow \{1, \dots, n\}$  Laplace-verteilt, so gilt

$$\begin{aligned} E(X^2) &= \sum_{\nu=1}^n \nu^2 \frac{1}{n} \\ &= \frac{n(n+1)(2n+1)}{6} \frac{1}{n} = \frac{(n+1)(2n+1)}{6}. \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 \\ &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{n^2 - 1}{12}. \end{aligned}$$

(ii) Da  $X^2 = X$  ist, folgt aus Feststellung 1.18.3 (ii)

$$\text{Var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p).$$

(iii) Für den Beweis benutzen wir das Standardmodell eines Bernoulli-Experimentes ( $\rightarrow$  Bezeichnung 1.8.8).

Es sei  $\Omega := \{1, 0\}$  mit W-Maß  $P\{1\} := p, P\{0\} = 1 - p$ . Auf  $(\Omega^n, P^{\otimes n})$  gibt die Zufallsvariablen

$$X(\omega) = \sum_{i=1}^n \text{pr}_i(\omega) = \sum_{i=1}^n \omega_i$$

die Anzahl der Einsen in dem  $n$ -Tupel  $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$  an.

Das Tupel  $(\text{pr}_1, \dots, \text{pr}_n)$  ist  $n$ -fach Bernoulli-verteilt mit dem Parameter  $p$  und folglich ist die Summe  $X$  binomialverteilt mit dem Parametern  $n$  und  $p$  ( $\rightarrow$  Beispiel 1.8.3). Aus Feststellung 1.18.3 (iii) folgt nun

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(\text{pr}_i) = np(1 - p).$$

(iv)  $\rightarrow$  Übung.

### 1.18.5 Bsp. (Fixpunkte einer Permutation)

Wir benutzen die Notation aus Beispiel 1.3.7  $(\Omega, P)$  sei der W-Raum der Permutationen von  $\{1, \dots, n\}$  mit der Laplace-Wahrscheinlichkeit. Für  $\nu = 1, \dots, n$  sei

$$A_\nu = \{(\sigma_1, \dots, \sigma_n) \in \Omega \mid \sigma_\nu = \nu\}.$$

Es ist

$$\begin{aligned} P(A_\nu) &= \frac{(n-1)!}{n!} = \frac{1}{n}, \\ P(A_\mu \cap A_\nu) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \quad (\mu \neq \nu). \end{aligned}$$

Die Zufallsvariable

$$X := \sum_{\nu=1}^n \mathbb{1}_{A_\nu}$$

beschreibt die Anzahl der Fixpunkte einer Permutation. Der Erwartungswert von  $X$  ist

$$E(X) = \sum_{\nu=1}^n E(\mathbb{1}_{A_\nu}) = n \frac{1}{n} = 1.$$

Wir berechnen die Varianz mit der Formel  $\text{Var}(X) = E(X^2) - (E(X))^2$ . Es gilt

$$\begin{aligned} E(X^2) &= E\left(\sum_{\nu=1}^n (\mathbb{1}_{A_\nu}^2)\right) \\ &= \sum_{\nu=1}^n E(\mathbb{1}_{A_\nu}^2) + 2 \sum_{1 \leq \mu < \nu \leq n} E(\mathbb{1}_{A_\mu} \mathbb{1}_{A_\nu}) \\ &= \sum_{\nu=1}^n P(A_\nu) + 2 \sum_{1 \leq \mu < \nu \leq n} P(A_\mu \cap A_\nu) \\ &= n \frac{1}{n} + 2 \binom{n}{2} \frac{1}{n(n-1)} = 1 + 1 = 2. \end{aligned}$$

Also ist  $\text{Var}(X) = 2 - 1 = 1$ . Weder Erwartungswert noch Varianz hängen von  $n$  ab.

**Anmerkung.** Die Varianz erhält man, wenn man in der symmetrischen Bilinearform

$$(X, Y) \mapsto E((X - EX) \cdot (Y - EY))$$

$X = Y$  setzt. Diese Bilinearform heißt die Kovarianz von  $X$ .

### 1.18.6 Bez. (Kovarianz)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $X, Y$  reelle Zufallsvariable auf  $(\Omega, P)$ .

(i) Die Kovarianz von  $X$  und  $Y$  bezeichnet man mit

$$\text{Cov}(X, Y) := E((X - EX) \cdot (Y - EY)).$$

Die Kovarianz  $\text{Cov}(X, Y)$  hängt nur von der *gemeinsamen Verteilung* von  $(X, Y)$  ab.

(ii) Es gilt

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

**1.18.7 Festst. (Regeln: Kovarianz)**

(i) Die Kovarianz ist eine symmetrische, positiv semi-definite Bilinearform. Es gilt also:

$$\begin{aligned} \text{Cov}(\lambda_1 X_1 + \lambda_2 X_2, Y) &= \lambda_1 \text{Cov}(X_1, Y) + \lambda_2 \text{Cov}(X_2, Y), \\ \text{Cov}(X, Y) &= \text{Cov}(Y, X) \\ \text{Cov}(X, X) &= \text{Var}(X) \geq 0. \end{aligned}$$

(ii) Es gilt die Schwarzsche Ungleichung

$$|\text{Cov}(X, Y)| \leq \sigma(X) \sigma(Y).$$

(iii) und die Dreiecksungleichung

$$\sigma(X + Y) \leq \sigma(X) + \sigma(Y).$$

**1.18.8 Bem. (Summen unkorrelierter ZVn)**

Für reelle Zufallsvariable  $X_1, \dots, X_n : (\Omega, P) \rightarrow \mathbb{R}$  gilt:

(i) Es gilt

$$\text{Var}\left(\sum_{\nu=1}^n X_\nu\right) = \sum_{\nu=1}^n \text{Var}(X_\nu) + 2 \sum_{1 \leq \mu < \nu \leq n} \text{Cov}(X_\mu, X_\nu).$$

(ii) Sind  $X_1, \dots, X_n$  paarweise unkorreliert; d.h.

$$\text{Cov}(X_\mu, X_\nu) = 0 \quad \text{für } \mu \neq \nu,$$

dann gilt die Gleichung von Bienaymé:

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

(iii) Sind  $X_1, \dots, X_n$  paarweise unkorreliert, so gilt für die Standardabweichung des arithmetischen Mittels  $\frac{1}{n} \sum_{\nu=1}^n X_\nu$

$$\sigma\left(\frac{1}{n} \sum_{\nu=1}^n X_\nu\right) = \frac{\sigma(X_1)}{\sqrt{n}}.$$

Geometrisch gesehen, ist der Fall (ii) der Satz von Pythagoras: Das Verschwinden der Kovarianz bedeutet, daß  $(X_\mu - EX_\mu) \perp (X_\nu - EX_\nu)$  für  $\mu \neq \nu$  ist. Die Varianz ist das Quadrat der Länge von  $X_\nu - EX_\nu$ .

**1.18.9 Bsp. (Varianz von Serien)**

Wir benutzen die Bezeichnungen von Beispiel 1.15.5.  $(Z_0, Z_1, \dots, Z_n)$  sind  $n + 1$ -fach Bernoulli-verteilt mit Parameter  $p$ .

$$S_i := \mathbb{1}_{\{Z_i \neq Z_{i-1}\}} \quad \text{für } i = 1, \dots, n$$

sind Bernoulli-verteilt mit Parameter  $2p(1-p)$ . Da  $S_i, S_k$  für  $i \leq i+2 \leq k$  unabhängig sind ( $\rightarrow$  Satz 1.9.5), ist

$$\text{Cov}(S_i, S_k) = 0 \quad \text{für } i \leq i+2 \leq k.$$

Da

$$E(\mathbb{1}_{\{Z_i \neq Z_{i-1}\}} \mathbb{1}_{\{Z_{i+1} \neq Z_i\}}) = p^2(1-p) + p(1-p)^2$$

ist

$$\begin{aligned} \text{Cov}(S_i, S_{i+1}) &= E(S_i S_{i+1}) - E(S_i) E(S_{i+1}) \\ &= p^2(1-p) + p(1-p)^2 - 4p^2(1-p)^2 = p(1-p)(1-4p(1-p)) \end{aligned}$$

Da  $S_i^2 = S_i$  ist, folgt

$$\begin{aligned} \text{Var}(S_i) &= E(S_i^2) - (E(S_i))^2 \\ &= 2p(1-p) - 4p^2(1-p)^2 = 2p(1-p)(1-2p(1-p)) \end{aligned}$$

Mit Bemerkung 1.18.8 (i) erhalten wir nun die Varianz der Wechselsumme

$$\begin{aligned} \text{Var}(Y) &= \text{Var}\left(\sum_{i=1}^n S_i\right) \\ &= \sum_{i=1}^n \text{Var}(S_i) + \sum_{i=1}^{n-1} \text{Cov}(S_i, S_{i+1}) \\ &= 2np(1-p)(1-2p(1-p)) + 2(n-1)(p(1-p)(1-4p(1-p))). \end{aligned}$$

Da die Varianz translationsinvariant ist ( $\rightarrow$  Feststellung 1.18.3 (i)) ist die Varianz der Anzahl  $X = Y + 1$  der Serien

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

**Anmerkung.** Die Korrelation ist eine Kennzahl dafür, in wie weit  $X$  und  $Y$  einer linearen Relation genügen.

**1.18.10 Bez. (Korrelation)**

Für reelle Zufallsvariable  $X, Y : (\Omega, P) \rightarrow \mathbb{R}$  mit positiver Variation  $V \text{Var}(X) > 0, \text{Var}(Y) > 0$  definiert man die Korrelation

$$\text{cor}(X; Y) := \frac{\text{Cov}(X, Y)}{\sigma(X) \sigma(Y)}.$$

Aus der Schwarzschen Ungleichung für die Kovarianz ( $\rightarrow$  Feststellung 1.18.7 (ii)) folgt

$$-1 \leq \text{cor}(X, Y) \leq 1.$$

**1.18.11 Bem.**

Es seien  $X, Y : (\Omega, P) \rightarrow \mathbb{R}$  reelle Zufallsvariable.

(i) sind die folgenden Aussagen äquivalent:

(a)  $\text{cor}(X, Y) \in \{-1, 1\}$ .

(b) In der Schwarzschen Ungleichung für die Kovarianz gilt das Gleichheitszeichen:

$$|\text{Cov}(X, Y)| = \sigma(X) \sigma(Y).$$

(c) Es gibt ein  $\lambda \in \mathbb{R}$  so, daß

$$P(\lambda(X - EX) - (Y - EY)) = 1$$

oder

$$P(\lambda(Y - EY) - (X - EX)) = 1.$$

ist. D.h., die Zufallsvariablen  $X - EX$  und  $Y - EY$  sind „mit Wahrscheinlichkeit 1“ linear abhängig.

(d) Es gibt Konstanten  $\alpha, \beta, \gamma \in \mathbb{R}$  derart, daß

$$P(\alpha X + \beta Y + \gamma = 0) = 1$$

ist. D.h., die Punkte  $\{(X(\omega), Y(\omega) \mid \omega \in \Omega\}$  liegen mit „mit Wahrscheinlichkeit 1“ auf einer Geraden.

(ii) Sind  $X, Y$  unabhängig, so ist  $\text{cor}(X, Y) = 0$ , d.h., unabhängige Zufallsvariable sind linear unabhängig. Die Umkehrung gilt i.a. nicht!

**Anmerkung.** Im Feststellung 1.17.4 hatten wir gesehen, daß der Ausdruck

$$X \mapsto \sqrt{E(X^2)}$$

analoge Eigenschaften hat, wie der Euklidische Abstand. Die Standardabweichung gibt also den Abstand einer reellen Zufallsvariablen zu der Konstanten  $E(X)$  an. Fragt man umgekehrt nach der Konstanten, die den Abstand zu  $X$  minimiert, so erhält man den Erwartungswert  $E(X)$ .

### 1.18.12 Festst. (Minimaleigenschaft des EW)

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $X$  eine reelle Zufallsvariable auf  $\Omega$ . Dann gilt für  $a \in \mathbb{R}$ :

$$E((X - a)^2) = \text{Var}(X) + (E(X) - a)^2.$$

Der Abstand  $\sqrt{E((X - a)^2)}$  wird also minimal für  $a := E(X)$ .

**Beweis.**

$$\begin{aligned} E((X - a)^2) &= E(X^2 - 2aX + a^2) \\ &= E(X^2) - 2aEX + a^2 \\ &= E(X^2) - (EX)^2 + (a - EX)^2 \\ &= E((X - EX)^2) + (a - EX)^2 \\ &= \text{Var}(X) + (a - EX)^2. \end{aligned}$$

**Anmerkung.** Der Erwartungswert  $EX$  minimiert die durchschnittliche quadratische Abweichung  $E((X - a)^2)$  von einer Konstanten  $a$ . Sucht man statt dessen das Minimum der durchschnittlichen Abweichung  $E(|X - a|)$ , so wird man einen anderen „mittleren Wert“ von  $X$  geführt, den **Median**.

## 1.19 Bedingte Varianz

### 1.19.1 Bem. (bedingte Varianz)

Da  $E(X \mid Y = y) := E_{P(\cdot|\{Y=y\})}$  der Erwartungswert bezgl. der bedingten Verteilung  $P(\cdot|\{Y = y\})$  ist, gelten für  $E(X \mid Y = y)$  und somit auch für

$$\begin{aligned} \text{Var}(X \mid Y = y) &:= E_{P(\cdot|Y=y)}((X - E(X \mid Y = y))^2) \\ &= E((X - E(X \mid Y = y))^2 \mid Y = y). \end{aligned}$$

alle Rechenregeln einer Varianz.

(i) Man definiert die bedingte Varianz

$$\text{Var}(X \mid Y) = E((X - E(X|Y) \circ Y)^2 \mid Y).$$

(ii) Insbesondere gilt

$$\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E(X|Y)). \quad (1.19.1)$$

**Beweis.** Nach Feststellung 1.18.3 (ii) ist

$$E((X - EX)^2 \mid Y = y) = \text{Var}(X \mid Y = y) + (E(X \mid Y = y) - EX)^2$$

Mit der Gleichung vom totalen Erwartungswert ( $\rightarrow$  Feststellung 1.16.5) folgt nun

$$\begin{aligned} \text{Var}(X) &= E(E((X - EX)^2 \mid Y)) \\ &= E(\text{Var}(X|Y)) + E((E(X|Y) - EX)^2) \\ &= E(\text{Var}(X|Y)) + E((E(X|Y) - E(E(X|Y)))^2) \\ &= E(\text{Var}(X|Y)) + \text{Var}(E(X|Y)). \end{aligned}$$

### 1.19.2 Bsp. (Zufällige Anzahl der Summanden)

Es seien  $X_1, \dots, X_k$  unkorrelierte reelle Zufallsvariable mit gleichen Erwartungswert und gleicher Varianz. Mit einer von  $X_1, \dots, X_k$  unabhängigen Zufallsvariablen  $N$  mit Werten in  $\{1, \dots, k\}$  bilde man die Summe  $S := X_1 + \dots + X_N$ . Dann gilt

$$\text{Var}(S) = \text{Var}(X_1)E(N) + E(X_1)^2 \text{Var}(N).$$

**Beweis.** Nach Beispiel 1.16.9 ist  $E(S|N) = E(X_1)N$ . Da die  $X_1, \dots, X_n$  paarweise unkorreliert sind, folgt

$$\begin{aligned} \text{Var}(S \mid N = n) &= \text{Var}(S_n \mid N = n) = \text{Var}(S_n) \\ &= \text{Var}(X_1) + \dots + \text{Var}(X_n) = n \text{Var}(X_1) \end{aligned}$$

und somit  $\text{Var}(S|N) = \text{Var}(X_1)N$ . Nach Gleichung (1.19.1) ist

$$\begin{aligned} \text{Var}(S) &= E(\text{Var}(S|N)) + \text{Var}(E(S|N)) \\ &= \text{Var}(X_1)E(N) + (EX_1)^2 \text{Var}(N) \end{aligned}$$

### 1.19.3 Festst. (Minimal. der bed. Erwartung)

Gegeben seien ein endlicher W-Raum  $(\Omega, P)$ , eine endliche Menge  $M$  und Zufallsvariable

$$X : (\Omega, P) \rightarrow \mathbb{R} \quad \text{und} \quad Y : (\Omega, P) \rightarrow M.$$

Man suche unter allen Zufallsvariablen der Form

$$Z \circ Y, \quad \text{wobei } Z : M \rightarrow \mathbb{R},$$

diejenige, die die mittlere quadratische Abweichung

$$E((X - Z \circ Y)^2)$$

minimiert. Die Bestapproximation in diesem Sinne ist  $Z = E(X|Y)$ .

**Beweis.** Aus

$$\begin{aligned} E(X - Z \circ Y \mid Y = y) &= E(X - Z(y) \mid Y = y) \\ &= E(X \mid Y = y) - Z(y) \end{aligned}$$

folgt

$$E(X - Z \circ Y \mid Y) = E(X \mid Y) - Z.$$

Da nach Feststellung 1.18.3 (i)

$$\text{Var}(X - Z \circ Y \mid Y = y) = \text{Var}(X - Z(y) \mid Y = y) = \text{Var}(X \mid Y = y)$$

ist, folgt mit Feststellung 1.18.3 (ii)

$$\begin{aligned} E((X - Z \circ Y)^2 \mid Y = y) &= \text{Var}(X - Z \circ Y \mid Y = y) + (E(X - Z \circ Y \mid Y = y))^2 \\ &= \text{Var}(X \mid Y = y) + (E(X \mid Y = y) - Z(y))^2. \end{aligned}$$

Es gilt also

$$E((X - Z \circ Y)^2 \mid Y) = \text{Var}(X \mid Y) + (E(X \mid Y) - Z)^2$$

Mit der Gleichung vom totalen Erwartungswert ( $\rightarrow$  Feststellung 1.16.5) folgt nun

$$\begin{aligned} E((X - Z \circ Y)^2) &= E(E((X - Z \circ Y)^2 \mid Y)) \\ &= E(\text{Var}(X \mid Y)) + E((E(X \mid Y) - Z \circ Y)^2) \end{aligned}$$

Die rechte Seite wird minimal für  $Z = E(X \mid Y)$ .

## 1.20 Schwaches Gesetz der großen Zahl

**Anmerkung.** Die **Chebychev-Ungleichung** gibt eine einfache Abschätzung für die Wahrscheinlichkeiten der Abweichung einer Zufallsvariablen von ihrem Erwartungswert. Der Beweis wird übersichtlicher, wenn man die Chebychev-Ungleichung auf die **Markov-Ungleichung** zurückführt.

Die erstere heißt auch die Chebychev-Bienaymé-Ungleichung. Man findet die folgenden Schreibweisen: *Chebychev* = *Čebyšev* = *Tschebyschev*.

Diese beiden Ungleichungen sind trotz ihrer einfachen Bauart und Herleitung erstaunlich wirkungsvoll. Für spezielle W-Maße gibt es schärfere Abschätzungen, die aber mehr Voraussetzungen erfordern. Die Markov-Ungleichung und die Chebychev-Ungleichung gelten für **jedes** W-Maß

### 1.20.1 Lemma (Markov-Ungleichung)

Es sei  $(\Omega, P)$  ein endlicher W-Raum. Für  $X : (\Omega, P) \rightarrow [0, \infty)$  gilt

$$P\{X \geq c\} \leq \frac{E(X)}{c} \quad \text{für } c > 0.$$

**Beweis.** Für die Indikatorfunktion von  $\{X \geq c\}$  gilt

$$\mathbb{1}_{\{X \geq c\}} \leq \frac{X}{c}.$$

Aus der Monotonie und der Linearität des Erwartungswertes ( $\rightarrow$  Feststellung 1.15.4) folgt somit

$$P\{X \geq c\} = E(\mathbb{1}_{\{X \geq c\}}) \leq E\left(\frac{X}{c}\right) = \frac{E(X)}{c}.$$

**Anmerkung.** Etwas allgemeiner ist die folgende Variante der Markov-Ungleichung, die man ebenso beweist.

### 1.20.2 Lemma (Markov-Ungleichung)

Ist  $\varphi : [0, \infty) \rightarrow [0, \infty)$  monoton wachsend,  $0 < c$  mit  $\varphi(c) > 0$ , so gilt für jede reelle Zufallsvariable  $X$

$$P\{|X| \geq c\} \leq \frac{E(\varphi(|X|))}{\varphi(c)}.$$

### 1.20.3 Satz (Chebychev-Ungleichung)

Es sei  $(\Omega, P)$  ein endlicher W-Raum. Für eine Zufallsvariable  $X : (\Omega, P) \rightarrow \mathbb{R}$  gilt

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\text{Var}(X)}{\varepsilon^2} \quad \text{für } \varepsilon > 0.$$

**Anmerkung.** Man nennt die Konstante in der Chebychev-Ungleichung üblicherweise  $\varepsilon > 0$ , weil man die Ungleichung für *kleine*  $\varepsilon$  verwendet. Was im konkreten Fall *klein* bedeutet, hängt von der Streuung von  $X$  ab. Man schreibt die Chebychev-Ungleichung deshalb häufig in der Form ( $c > 0$ ):

$$P\{|X - E(X)| \geq c \sigma(X)\} \leq \frac{1}{c^2}. \quad (1.20.1)$$

oder

$$P\{|X - E(X)|^2 \geq c \text{Var}(X)\} \leq \frac{1}{c}. \quad (1.20.2)$$

**Bsp. (Chebychev-Ungl. für binomialverteilte ZV)** Aus der Chebychev-Ungleichung folgt für eine beliebige Zufallsvariable  $X$ :

$$P\{|X - E(X)| \geq 2 \sigma(X)\} \leq \frac{1}{4} = 0,25.$$

Für eine binomialverteilte Zufallsvariable  $X$  ergibt eine genauere Rechnung, daß für *große*  $n$

$$P\{|X - E(X)| \geq 2 \sigma(X)\} \approx 0,05$$

ist.

**Anmerkung.** Wendet man die Chebychev-Ungleichung auf das **arithmetische Mittel**  $\frac{1}{n} \sum_{\nu=1}^n X_\nu$  paarweise unkorrelierter

Zufallsvariablen  $X_1, \dots, X_n$  an, die alle den gleichen Erwartungswert und die gleiche Varianz  $\sigma^2$  haben so erhält man mit Bemerkung 1.18.8 (iii) den folgenden Satz. Man beachte, daß das arithmetische Mittel eine um den Faktor  $\frac{1}{n}$  kleinere Varianz hat

$$\text{Var}\left(\frac{1}{n} \sum_{\nu=1}^n X_\nu\right) = \frac{\sigma^2}{n}.$$

**1.20.4 Festst. (schwaches Gesetz der großen Zahl)**

Es seien  $(\Omega, P)$  ein endlicher W-Raum und  $X_1, \dots, X_n : (\Omega, P) \rightarrow \mathbb{R}$  paarweise unkorrelierte Zufallsvariable mit

$$\text{Var}(X_\nu) \leq M \quad \text{für } \nu = 1, \dots, n.$$

Dann gilt für alle  $\varepsilon > 0$

$$P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n (X_\nu - EX_\nu)\right| \geq \varepsilon\right\} \leq \frac{M}{n\varepsilon^2}.$$

Haben überdies  $X_1, \dots, X_n$  alle den gleichen Erwartungswert  $\mu := E(X_\nu)$ , ( $\nu = 1, \dots, n$ ) so gilt

$$P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n X_\nu - \mu\right| \geq \varepsilon\right\} \leq \frac{M}{n\varepsilon^2}. \tag{1.20.3}$$

**Anmerkung. (schwaches Gesetz der großen Zahl)** Der obige Feststellung 1.20.4 ist eigentlich eine Vorstufe des *schwachen Gesetzes der großen Zahl*. Dies erhält man, wenn man in Gleichung (1.20.3) die Anzahl  $n$  der Zufallsvariablen gegen  $\infty$  gehen läßt:

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n X_\nu - \mu\right| \geq \varepsilon\right\} = 0. \tag{1.20.4}$$

Im Spezialfall von Indikatorvariablen fallen die Begriffe *unkorreliert* und *unabhängig* zusammen. Man erhält somit das Ergebnis von Jakob Bernoulli ( $\approx 1865$ ):

**Satz. (J. Bernoulli)** Für eine Folge  $A_1, A_2, \dots$  unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit  $p = P(A_n)$  gilt

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n \mathbb{1}_{A_\nu} - p\right| \geq \varepsilon\right\} = 0. \tag{1.20.5}$$

Durch Übergang zu der komplementären Menge folgt.

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n \mathbb{1}_{A_\nu} - p\right| < \varepsilon\right\} = 1. \tag{1.20.6}$$

Deutet man dies als eine Folge unabhängiger Wiederholungen desselben Zufallsexperimentes, und die unabhängigen Ereignisse  $A_n$  als das Eintreten des Ergebnisses  $A$  im  $n$ -ten Versuch, so wird die relative Häufigkeit  $R_n(A)$ , mit der das Ereignis  $A$  eintritt, durch die folgende Formel gegeben:

$$R_n(A) = \frac{1}{n} \sum_{\nu=1}^n \mathbb{1}_{A_\nu}.$$

Die Gleichung (1.20.6) lautet nun

$$\lim_{n \rightarrow \infty} P\{|R_n(A) - P(A)| < \varepsilon\} = 1. \tag{1.20.7}$$

Die Wahrscheinlichkeit, daß sich in einer Folge unabhängiger Bernoulli-Experimente mit Parameter  $p$  die relative Häufigkeit  $R_n$  der Erfolge von dem Parameter  $p$  um weniger als ein beliebig kleines  $\varepsilon > 0$  unterscheidet, konvergiert gegen Eins. Die axiomatische W-Theorie enthält und bestätigt die empirisch Beobachtung, daß man die Wahrscheinlichkeit durch relative Häufigkeiten langer Ketten von *unabhängigen* Experimenten beliebig genau approximieren kann.

Wir wollen dies mit den Mitteln der bisher entwickelten Theorie endlicher W-Räume nochmal genauer formulieren:

**1.20.5 Festst. (G. der gr. Z. für rel. Häufigkeit.)**

Es sei  $(\Omega, P)$  ein endlicher W-Raum. Für  $A \in 2^\Omega$  gilt

$$P^{\otimes n} \left\{ (\omega_1, \dots, \omega_n) \in \Omega^n \mid \left| \frac{1}{n} \sum_{\nu=1}^n \mathbb{1}_A(\omega_\nu) - P(A) \right| \geq \varepsilon \right\} \leq \frac{1}{4n\varepsilon^2}.$$

**Beweis.** Die Zufallsvariablen ( $\nu = 1, \dots, n$ )

$$X_\nu : (\Omega^n, P^{\otimes n}) \rightarrow \{1, 0\},$$

$$X_\nu : (\omega_1, \dots, \omega_n) \mapsto \mathbb{1}_A(\omega_\nu)$$

sind unabhängig und Bernoulli-verteilt mit Parameter  $p = P(A)$ . Es ist  $E(X_\nu) = p$  und  $\text{Var}(X_\nu) = p(1-p) \leq \frac{1}{4}$ . Nach der Chebychev-Ungleichung ist

$$P^{\otimes n} \left\{ \left| \frac{1}{n} \sum_{\nu=1}^n X_\nu - p \right| \geq \varepsilon \right\} \leq \frac{1}{4n\varepsilon^2}.$$

**Anmerkung. (Problem)** Nur reicht die bisher entwickelte Theorie **endlicher** W-Räume nicht aus, um den Grenzübergang  $n \rightarrow \infty$  in den Gleichungen (1.20.4) – (1.20.7) zu vollziehen. Das Problem ist, daß es auf einem endlichen W-Raum niemals eine Folge paarweise unkorrelierter reeller Zufallsvariablen gibt.

**Beweis.** Das kann man folgendermaßen einsehen. Wenn  $|\Omega| = n$ , so hat der Vektorraum  $V$  der reellen Zufallsvariablen auf  $\Omega = \{\omega_1, \dots, \omega_n\}$  die Dimension  $n$ . Man kann ohne Einschränkung annehmen daß  $p_\nu := P\{\omega_\nu\} > 0$  ist. Dann ist  $E(XY)$  ein positiv definites Skalarprodukt auf  $V$  ( $\rightarrow$  Feststellung 1.17.4 (ii)). Eine Familie  $(Y_i)_{i \in I}$  von Vektoren  $Y_i \in V$ , die paarweise orthogonal sind

$$E(Y_i Y_j) = 0 \quad \text{für } i \neq j,$$

ist linear unabhängig. Folglich enthält die Familie höchstens  $n$  Elemente. Somit hat jede Familie paarweise unkorrelierter reeller Zufallsvariablen höchstens  $n$  Elemente.

**Anmerkung. (abzählbar viele Teilerperimente)** Die Lösung dieses Problems bieten unendliche W-Räume mit nichtdiskreten W-Maßen, wie man sie auch bei geometrische Wahrscheinlichkeitsaufgaben findet. Bevor diese allgemeineren W-Räume zur Verfügung stehen, verabreden wir folgende Konvention zum Umgang mit abzählbaren Folgen von Zufallsvariablen:

Wir betrachten eine Folge  $(E_n)_{n \in \mathbb{N}}$  von Zufallsexperimenten, von denen jedes nur endlich viele Ausgänge hat. Für jedes  $n \in \mathbb{N}$  bildet man einen endlichen W-Raum  $(\Omega_n, P_n)$ , der die ersten  $n$  Teilerperimente  $E_1, \dots, E_n$  modelliert. Die Elementarereignisse in  $\Omega_n$  sind die möglichen Ergebnisse  $\omega := (\omega_1, \dots, \omega_n)$  der ersten  $n$  Teilerperimente.

Das System  $\Omega_1, \dots, \Omega_n$  bildet einen Ereignisbaum, in dem die Knoten der Tiefe  $\nu$  die Menge  $\Omega_\nu$  sind, mit den Übergangswahrscheinlichkeiten

$$p(\omega_{\nu+1} \mid \omega_1, \dots, \omega_\nu) := \frac{P_{\nu+1}\{(\omega_1, \dots, \omega_{\nu+1})\}}{P_\nu\{(\omega_1, \dots, \omega_\nu)\}}.$$

( $\rightarrow$  Bezeichnung 1.13.1)

Zu jedem Elementarereignis  $(\omega_1, \dots, \omega_{n+1})$  in  $\Omega_{n+1}$  bildet man das entsprechende Elementarereignis  $(\omega_1, \dots, \omega_n)$  in  $\Omega_n$ , indem man das  $n+1$ -te Ergebnis wegläßt. Dies definiert eine **Projektion**  $\text{pr}_n : \Omega_{n+1} \rightarrow \Omega_n$  und es gilt

$$P_{n+1}\{\omega \mid \text{pr}_{n+1}(\omega) = (\omega_1, \dots, \omega_n)\} = P_n\{(\omega_1, \dots, \omega_n)\}.$$

D.h.  $P_n$  ist die Bildverteilung von  $\text{pr}_{n+1}$ . Induktiv folgt dann, daß  $P_n$  die Bildverteilung von  $\text{pr}_{n+1} \circ \text{pr}_{n+2} \circ \dots \circ \text{pr}_{n+k}$  ist. ( $\rightarrow$  Feststellung 1.13.5).

**Konvention.** Unter einer Folge  $(X_n)_n$  von reellen Zufallsvariablen verstehen wir das folgende: Die  $n$ -te Zufallsvariable ist eine reelle Funktion auf dem  $n$ -ten Raum:

$$X_n : (\Omega_n, P_n) \rightarrow \mathbb{R}.$$

Die Werte von  $X_n$  hängen nur von  $(\omega_1, \dots, \omega_n)$  ab.

Man bezeichnet dies als eine Folge von Zufallsvariablen auf einem projektiven System von W-Räumen.

**1.20.6 Bez. (Folgen von ZVn auf proj. System)**

1. Ein **projektives System** ist eine Folge von endlichen W-Räumen  $(\Omega_n, P_n)$ ,  $(n \in \mathbb{N})$ , die durch Zufallsvariablen

$$\text{pr}_n : \Omega_{n+1} \rightarrow \Omega_n$$

so verbunden sind, daß  $P_n$  die Bildverteilung von  $\text{pr}_n$  ist:

$$(P_{n+1})_{\text{pr}_n} = P_n.$$

2. Eine Folge von Zufallsvariablen des projektiven Systems

$$X_n := (\Omega_n, P_n) \rightarrow \mathbb{R} \quad \text{für } n \in \mathbb{N}$$

nennt man kurz eine Folge  $(X_n)_n$  von reellen Zufallsvariablen.

3. Man kann  $X_n$  als Zufallsvariable auf einem der höheren Räume  $\Omega_{n+k}$  betrachten, indem man die passenden Projektionen davorschaltet:

$$X_n^{(n+k)} := X_n \circ \text{pr}_{n+1} \circ \text{pr}_{n+2} \circ \dots \circ \text{pr}_{n+k}$$

Es sei  $X_n^n := X_n$ .

4.  $X_n^{(n+k)}$  hat die gleiche Verteilung wie  $X_n$ , also den gleichen Erwartungswert  $E(X_n)$  und die gleiche Varianz  $\text{Var}(X_n)$ .

5.  $X_1, \dots, X_n$  heißen unabhängig, wenn  $X_1^{(n)}, \dots, X_n^{(n)}$  unabhängig sind.

Dann sind natürlich auch  $X_1^{(n+k)}, \dots, X_n^{(n+k)}$  unabhängig ( $\rightarrow$  Definition 1.9.2 und Feststellung 1.6.4).

6. Die Folge  $(X_n)_n$  heißt unabhängig, wenn jeder endliche Abschnitt  $X_1, \dots, X_n$  unabhängig ist.

7.  $X_1, \dots, X_n$  heißen paarweise unkorreliert, wenn  $X_1^{(n)}, \dots, X_n^{(n)}$  paarweise unkorreliert sind.

Dann sind natürlich auch  $X_1^{(n+k)}, \dots, X_n^{(n+k)}$  paarweise unkorreliert.

8. Im allgemeinen läßt man die obigen Indizes weg und schreibt kurz  $X_n$  statt  $X_n^{(n+k)}$ . In diesem Sinne verstehen wir unter dem arithmetischen Mittel von  $X_1, \dots, X_n$  die Summe

$$S_n := \frac{1}{n} \sum_{\nu=1}^n X_\nu := \frac{1}{n} \sum_{\nu=1}^n X_\nu^{(n+k)} \quad (1.20.8)$$

für irgend ein  $k \in \mathbb{N}_0$ .

Genauer gesagt, bilden die arithmetischen Mittel eine Folge von Zufallsvariablen  $S_n : (\Omega_n, P_n) \rightarrow \mathbb{R}$  und die rechte Seite der Gleichung (1.20.8) definiert die entsprechende Variable  $S_n^{(n+k)}$ .

9. Ebenso lassen wir den Index  $n$ , der das Modell angibt, bei der Bildung der Wahrscheinlichkeit weg und schreiben kurz  $P$  statt  $P_n$ . Für ein Ereignis  $A \in 2^{\Omega_n}$  ist also

$$P(A) := P_n(A) = P_{n+k}\{\text{pr}_{n+1} \circ \dots \circ \text{pr}_{n+k} \in A\}$$

Aus Feststellung 1.20.4 folgt sofort:

**1.20.7 Satz (schwaches Gesetz der großen Zahl)**  
*Es seien  $(\Omega_n, P_n, \text{pr}_n)$  ein projektives System und  $(X_n)_{n \in \mathbb{N}}$  eine Folge paarweise unkorrelierter Zufallsvariabler auf dem System mit gleichen Erwartungswert*

$\mu := E(X_n)$  und beschränkter Varianz  $\text{Var}(X_n) \leq M < \infty$  für alle  $n \in \mathbb{N}$ . Dann gilt

$$P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n X_\nu - \mu\right| \geq \varepsilon\right\} \leq \frac{M}{\varepsilon^2 n} \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

**1.20.8 Bsp. (Ein vorteilhaftes Spiel?)**

Ein Spiel heißt fair, wenn der Erwartungswert des Gewinns gleich dem Erwartungswert des Verlustes ist. Ist der Erwartungswert des Gewinns größer als der Erwartungswert des Verlustes, so heißt das Spiel **vorteilhaft**. Das folgende Beispiel zeigt, daß man auch bei einem vorteilhaften mit *beliebig großer Wahrscheinlichkeit* sein gesamtes Kapital verlieren kann.

Der Spieler beginnt mit einem Startkapital  $K_0 = 1$  und hat nach  $n$  Runden das Kapital  $X_n$ . In jeder Runde wird eine faire Münze geworfen. Das Kapital in der  $n$ -ten Runde sei

$$X_n := \begin{cases} \frac{1}{2} X_n, & \text{wenn Kopf fällt,} \\ \frac{5}{3} X_n, & \text{wenn Zahl fällt.} \end{cases}$$

Setzt man

$$Y_n := \begin{cases} \frac{1}{2}, & \text{wenn Kopf fällt,} \\ \frac{5}{3}, & \text{wenn Zahl fällt,} \end{cases}$$

so gilt  $X_n = Y_1 \cdots Y_n K_0$ . Die  $(Y_n)_n$  sind unabhängig und haben alle den gleichen Erwartungswert

$$E(Y_n) = \frac{1}{2} P\{\text{Kopf}\} + \frac{3}{5} P\{\text{Zahl}\} = \left(\frac{1}{2} + \frac{3}{5}\right) \frac{1}{2} = \frac{13}{10}.$$

Nach Folgerung 1.17.2 ist

$$E(X_n) = E(Y_1) \cdots E(Y_n) K_0 = \left(\frac{13}{10}\right)^n K_0 \rightarrow \infty.$$

Das Spiel sieht also sehr vorteilhaft aus. Aber wie entwickelt sich das Kapital? Dazu bilden wir

$$\frac{1}{n} \log X_n = \frac{1}{n} \sum_{\nu=1}^n \log Y_\nu$$

und wenden hierauf das schwache Gesetz der großen Zahl an. Mit  $\mu = E(\log Y_n)$ ,  $\sigma^2 = \text{Var}(Y_n)$  für  $n \in \mathbb{N}$  gilt

$$P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n \log Y_\nu - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon} \rightarrow 0.$$

Nun ist aber  $\mu = \frac{1}{2}(\log \frac{1}{2} + \log \frac{3}{5}) < 0$ . Für  $\varepsilon := -\frac{1}{2}\mu$  folgt daher

$$\begin{aligned} P\{X_n \leq e^{n\mu/2}\} &= P\left\{\frac{1}{n} \log X_n \leq \frac{\mu}{2}\right\} \\ &= P\left\{\frac{1}{n} \log X_n - \mu \leq -\frac{\mu}{2}\right\} \\ &\geq P\left\{\left|\frac{1}{n} \log X_n - \mu\right| \leq -\frac{\mu}{2}\right\} \\ &= 1 - P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n \log Y_\nu - \mu\right| \geq \varepsilon\right\} \\ &\rightarrow 1 \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Da  $\mu < 0$  ist, strebt  $e^{n\mu/2}$  *exponentiell* schnell gegen 0 für  $n \rightarrow \infty$ . Für große  $n$  hat man also mit sehr großer Wahrscheinlichkeit sein Kapital nahezu verspielt. Der Erwartungswert  $E(X_n)$  ist dagegen sehr groß, da man mit wenn auch sehr kleiner Wahrscheinlichkeit enorme Gewinne machen kann.

**Anmerkung. (starkes Gesetz der großen Zahl)** Das starke Gesetz der großen Zahl können wir mit den bisherigen Hilfsmitteln nicht adäquat formulieren. Wir leiten schon mal das entscheidende Lemma her, aus dem das starke Gesetz der großen Zahl dann leicht folgt.

Als Beispiel denken wir uns ein und dasselbe Experiment  $E$  immer wieder unabhängig durchgeführt. Man modelliere  $E$  mit dem W-Raum  $(\Omega, P)$  und die  $n$ -malige Durchführung von  $E$  mit dem Produktraum  $(\Omega^n, P^{\otimes n})$ . Dann beschreibt das projektive System  $(\Omega^n, P^{\otimes n}, \text{pr}_n)$  die nicht abbrechende Folge der unabhängigen Wiederholungen von  $E$ . Für eine Zufallsvariable  $X : (\Omega, P) \rightarrow \mathbb{R}$  definiert

$$X_n : (\omega_1, \dots, \omega_n) \mapsto X(\omega_n)$$

eine Folge unabhängiger, identisch verteilter Zufallsvariablen ( $\rightarrow$  Satz 1.8.7) Für eine solche Folge kann man das schwache Gesetz der großen Zahl wesentlich verschärfen:

### 1.20.9 Lemma

Es seien  $(\Omega_n, P_n, \text{pr}_n)$  ein projektives System und  $(X_n)_{n \in \mathbb{N}}$  eine Folge unabhängiger, identisch verteilter Zufallsvariablen auf dem System. Es sei  $\mu := E(X_n)$ ,  $(n \in \mathbb{N})$ . Dann gibt es eine Konstante  $0 < C < \infty$  so, daß

$$P\left\{\left|\frac{1}{n} \sum_{\nu=1}^n X_\nu - \mu\right| \geq \varepsilon\right\} \leq \frac{C}{\varepsilon^4 n^2} \quad \text{für } n \rightarrow \infty.$$

Man beachte: Die Wahrscheinlichkeit geht wie  $n^{-2}$  gegen Null. Dagegen ist im schwachen Gesetz der großen Zahl die Konvergenzgeschwindigkeit wie  $n^{-1}$

**Beweis.** Zur Abkürzung setzen wir  $Y_n = X_n - \mu$ , so daß  $E(Y_n) = 0$  ist. Die  $(Y_n)_n$  sind ebenfalls unabhängig und identisch verteilt. Da für paarweise verschiedene  $i, j, k, l$

$$\begin{aligned} E(Y_i Y_j Y_k Y_l) &= E(Y_i^2 Y_j Y_k) = E(Y_i^3 Y_j) = 0, \\ E(Y_i^2 Y_j^2) &= E(Y_i^2) E(Y_j^2) \end{aligned}$$

gilt, folgt

$$\begin{aligned} E\left(\frac{1}{n} \sum_{\nu=1}^n Y_\nu\right)^4 &= \frac{1}{n^4} \sum_{1 \leq i, j, k, l \leq n} E(Y_i Y_j Y_k Y_l) \\ &= \frac{6}{n^4} \sum_{1 \leq i < j \leq n} E(Y_i^2 Y_j^2) + \frac{1}{n^4} \sum_{i=1}^n E(Y_i^4) \\ &= n(n-1) \frac{6}{n^4} (E(Y_1^2))^2 + n \frac{1}{n^4} E(Y_1^4) \\ &\leq \frac{C}{n^2}. \end{aligned}$$

Da die  $(Y_n)_n$  identisch verteilt sind, hängt die Konstante  $C$  nicht von  $n$  ab. Mit der Markov-Ungleichung erhalten wir hieraus

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{\nu=1}^n Y_\nu\right| \geq \varepsilon\right) &= P\left(\left(\frac{1}{n} \sum_{\nu=1}^n Y_\nu\right)^4 \geq \varepsilon^4\right) \\ &\leq \frac{E\left(\frac{1}{n} \sum_{\nu=1}^n Y_\nu\right)^4}{\varepsilon^4} \leq \frac{C}{\varepsilon^4 n^2}. \end{aligned}$$

## 2 Diskrete und stetige W-Räume

### 2.1 Uniforme und geometrische Verteilung

**Anmerkung.** Beim schwachen Gesetz der großen Zahl haben wir die Grenzen der bisher als Modelle betrachteten endlichen W-Räume erreicht. Die in Bezeichnung 1.20.6 eingeführten projektiven Systeme  $(\Omega_n, P_n, \text{pr}_n)$  endlicher W-Räume sind umständlich zu handhaben und nicht allgemein genug:

- Man müßte Produkte, Bildverteilungen, Bedingte Verteilungen ... für projektive Systeme erklären. Wie verhalten sich projektive Systeme von projektiven Systemen?
- Interpretiert in einem projektiven System den Parameter  $n \in \mathbb{N}$  als Zeitpunkte, in denen die Experimente durchgeführt werden, so könnte man auch an *kontinuierliche Zeitparameter*  $t \in \mathbb{R}$  denken.

Der elegante und auch anschauliche Ausweg sind *unendliche* W-Räume.

Wir wollen hier zwei einfache Beispiele unendlicher W-Räume mit anschaulichen Hilfsmitteln vorführen. Im Beispiel 2.1.2 erhalten wir eine stetige W-Verteilung auf dem Intervall  $[0, 1)$ , im zweiten Beispiel 2.1.3 einen diskretes W-Maß auf  $\mathbb{N}$ . Für beide Beispiele geben wir jeweils zwei Herleitungen an,

- als **projektive Limes** eines projektiven Systems endlicher W-Räume,
- als Grenzwerte einer Folge endlicher Verteilungen auf  $[0, 1]$  bzw. auf  $\mathbb{N}$ .

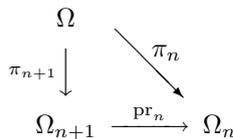
#### 2.1.1 Bem. (Projektiver Limes)

Es sei  $(\Omega_n, P_n, \text{pr}_n)$  ein projektives System endlicher W-Räume. Die Frage ist, kann man nicht ein W-Raum  $(\Omega, P)$  finden, derart daß die Räume  $(\Omega_n, P_n)$  Bilder von  $(\Omega, P)$  mit der Bildverteilung sind. D.h., es gibt eine Folge von Zufallsvariablen

$$\pi_n : (\Omega, P) \rightarrow (\Omega_n, P_n)$$

mit den folgenden Eigenschaften:

- (i) Das Diagramm



kommutiert für alle  $n \in \mathbb{N}$ ,

- (ii)  $P_n = P_{\pi_n}$ .

Häufig kommt man mit den beiden Eigenschaften (i) und (ii) aus. Zur Definition des projektiven Limes gehört noch die Eindeutigkeit, die wir der Vollständigkeit halber hier mit aufführen. Die Eindeutigkeit folgt aus einer *universellen Eigenschaft*. Zur Formulierung dieser universellen Eigenschaft benötigt man den Begriff der **meßbaren Abbildung**, den wir hier ohne Erklärung verwenden.

(iii) Man nennt  $(\Omega, P, \pi_n)$  den *projektiven Limes* des projektiven Systems  $(\Omega_n, P_n, \text{pr}_n)$ , wenn  $(\Omega, P, \pi_n)$  die folgende *universelle Eigenschaft* hat:

Ist  $(\tilde{\Omega}, \tilde{P}, \tilde{\pi}_n)$  ein weiterer W-Raum mit den Eigenschaften (i) und (ii), so gibt es eine eindeutig bestimmte meßbare Abbildung

$$\rho : \tilde{\Omega} \rightarrow \Omega \quad \text{mit} \quad \tilde{\pi}_n = \pi_n \circ \rho.$$

D.h.  $\Omega$  ist in diesem Sinne minimal.

#### 2.1.2 Bsp. (Münzwurf-Raum)

Das immerwährende Werfen einer fairen Münze beschreiben wir durch das projektive System

$$(\Omega_n, P_n, \text{pr}_n) := (\{0, 1\}^n, P^{\otimes n}, \text{pr}_n).$$

Dabei ist  $P$  die Laplace-Wahrscheinlichkeit auf  $\{0, 1\}$  und  $\text{pr}_n : \{0, 1\}^{n+1} \rightarrow \{0, 1\}^n$  die Projektion auf die ersten  $n$  Komponenten.

Als projektiver Limes  $(\Omega, P)$  bietet sich hier ein guter Bekannter aus der Analysis an, das Intervall  $[0, 1) \subset \mathbb{R}$ . Wir beschränken uns auf die Eigenschaften Bemerkung 2.1.1 (i) und (ii).

Die universelle Eigenschaft (iii) gilt auch. Wir zeigen sie aber nicht, da man für den Nachweis etwas Maßtheorie benötigt.

Damit die folgenden Formeln schön symmetrisch sind, lassen wir bei allen Intervalle den rechten Endpunkt weg, was aber nicht weiter wichtig ist. Die Wahrscheinlichkeit eines Teilintervalls  $[a, b) \subset [0, 1)$  ist seine Länge:

$$\mathcal{U}([a, b)) := b - a.$$

Für endliche Vereinigungen disjunkter Intervalle addiere man die Längen. Man bezeichnet diese Wahrscheinlichkeitsverteilung mit  $\mathcal{U} := \mathcal{U}_{[0,1)}$  und nennt sie die Gleichverteilung oder *uniforme* Verteilung auf  $[0, 1)$ .

Die Teilmengen

$$A_n := \left[ \frac{1}{2^n}, \frac{2}{2^n} \right) \cup \left[ \frac{3}{2^n}, \frac{4}{2^n} \right) \cup \dots \cup \left[ \frac{2^n - 1}{2^n}, \frac{2^n}{2^n} \right)$$

sind unabhängig, da für alle  $\{\nu_1, \dots, \nu_k\} \subset \{1, \dots, n\}$  gilt:

$$\mathcal{U}\left(\bigcap_{i=1}^k A_{\nu_i}\right) = 2^{-k} = \prod_{i=1}^k \mathcal{U}(A_{\nu_i}).$$

Die Indikatorfunktionen  $\mathbb{1}_{A_n}$ , ( $n \in \mathbb{N}$ ) sind  $n$ -fach Bernoulli-verteilt mit Parameter  $p = \frac{1}{2}$ . Es gilt also:

- (i) Die Bildverteilung von

$$\pi_n := (\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n})$$

ist die Produktverteilung  $P^{\otimes n}$  auf  $\{0, 1\}^n$ .

- (ii) Nach Konstruktion ist  $\pi_n = \text{pr}_n \circ \pi_{n+1}$ .

Auf  $([0, 1), \mathcal{U})$  gibt es eine abzählbare Familie  $(\mathbb{1}_{A_n})_n$  unabhängiger und identisch verteilter reeller Zufallsvariablen. Diese beschreiben den unendlichen Münzwurf.

**Anmerkung.** Man fragt sich, ob man nicht statt  $([0, 1), \mathcal{U})$  den unendlichen Produktraum  $\{0, 1\}^{\mathbb{N}}$  versehen mit einem *unendlichen Produktmaß*  $P^{\otimes \mathbb{N}}$  nehmen kann? In der Maßtheorie zeigt man, daß das geht und daß die W-Räume  $([0, 1), \mathcal{U})$  und  $(\{0, 1\}^{\mathbb{N}}, P^{\otimes \mathbb{N}})$  in einem noch zu präzisierenden Sinne isomorph sind. Beide Räume repräsentieren den sogenannten *projektiven Limes* des projektiven Systems:

$$\lim_{n \rightarrow \infty} \text{proj}(\{0, 1\}^n, P^{\otimes n}, \text{pr}_n),$$

was wir hier aber mangels Maßtheorie nicht beweisen können.

**2.1.3 Bsp. (geometrische Verteilung)**

In einem  $n$ -fachen Bernoulli-Experiment mit Erfolgswahrscheinlichkeit  $p > 0$  ist die Wahrscheinlichkeit, den ersten Erfolg genau im  $k + 1$ -ten Teilversuch zu haben, d.h. man hat zuvor  $k$  Mißerfolge

$$\mathcal{G}_{n;p}\{k\} := \begin{cases} p(1-p)^k & \text{für } k = 0, \dots, n-1, \\ (1-p)^n & \text{wenn gar kein Erfolg eintritt.} \end{cases}$$

Wir werden den letzteren Fall, daß kein Erfolg eintritt, mit  $k = n$  kodieren. Da nach der geometrischen Summenformel

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

gilt, ist  $\mathcal{G}_{n;p}$  ein W-Maß auf  $\Omega_n := (\{1, \dots, n, n+1\}, \mathcal{G}_{n;p})$ . Man nennt daher  $\mathcal{G}_{n;p}$  die **gestoppte  $n$ -te geometrische Verteilung**. „gestoppt“, da man ja weiter experimentieren kann, bis ein Erfolg eintritt. Mit der Projektion  $\text{pr}_n : \Omega_{n+1} \rightarrow \Omega_n$ ,

$$\text{pr}_n : k \mapsto \begin{cases} k & \text{für } k = 1, \dots, n, \\ n & \text{für } k = n+1, n+2. \end{cases}$$

erhält man ein projektives System  $(\Omega_n, \mathcal{G}_{n;p}, \text{pr}_n)$  von endlichen W-Räumen ( $\rightarrow$  Bezeichnung 1.20.6).

Den projektiven Limes kann man in diesem Fall sofort erraten, es ist der Raum  $\mathbb{N}$  der natürlichen Zahlen mit der **geometrischen Verteilung**

$$\mathcal{G}_p\{k\} = p(1-p)^k \quad \text{für } k \in \mathbb{N}_0.$$

Mit der geometrischen Reihe überprüft man

$$\sum_{k=0}^{\infty} P\{k\} = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1-(1-p)} = 1.$$

Die gestoppten W-Räume  $(\Omega_n, \mathcal{G}_{n;p})$  sind die Bilder von  $(\mathbb{N}, \mathcal{G}_p)$  unter der Projektion

$$\pi_n : \mathbb{N}_0 \rightarrow \{0, \dots, n\}$$

mit

$$\pi_n : k \mapsto \begin{cases} k & \text{für } k = 1, \dots, n, \\ n & \text{für } k = n+1, n+2, \dots \end{cases}$$

Das System  $(\mathbb{N}_0, \mathcal{G}_p, \pi_n)$  erfüllt also die Bedingungen (i) und (ii) von Bemerkung 2.1.1.

Man sieht leicht, daß auch die universelle Eigenschaft  $\rightarrow$  Bemerkung 2.1.1 (iii) erfüllt ist:

Ist  $(\tilde{\Omega}_P, \tilde{\pi}_n)$  ein weiteres System mit (i) und (ii), so bilde man  $\rho : \tilde{\Omega} \rightarrow \mathbb{N}$  mit

$$\rho(\tilde{\omega}) := \lim_{n \rightarrow \infty} \tilde{\pi}_n(\tilde{\omega}).$$

Man beachte, daß die Folge  $(\tilde{\pi}_n(\tilde{\omega}))_n$  monoton wächst und schließlich konstant ist.

**Anmerkung.** In der folgenden Feststellung verwenden wir bereits einige Eigenschaften diskreter W-Räume, die erst im folgenden Kapitel sauber definiert werden.

**2.1.4 Bez. (Geometrische Verteilung)**

(i) Die geometrische Verteilung auf  $\mathbb{N}_0$  mit dem Parameter  $p \in (0, 1)$  ist gegeben durch

$$\mathcal{G}_p\{k\} := p(1-p)^k \quad \text{für } k \in \mathbb{N}_0.$$

(ii) Eine Zufallsvariable  $X$  mit Werten in  $\mathbb{N}$  ist geometrisch verteilt mit Parameter  $p$ , ( $0 < p < 1$ ), wenn ihre Bildverteilung auf  $\mathbb{N}$  die geometrische Verteilung  $\mathcal{G}_p$  ist:

$$P\{X = k\} = p(1-p)^{k-1} \quad \text{für } k \in \mathbb{N}_0.$$

**2.2 Diskrete W-Räume**

**2.2.1 Def. (diskreter W-Raum)**

Es sei  $\Omega$  eine beliebige Menge und  $\{\omega_n \mid n \in \mathbb{N}\}$  eine abzählbare Teilmenge von  $\Omega$ . Jedem  $\omega_n$  sei eine Wahrscheinlichkeit  $p_n$  zugeordnet, derart daß

$$0 \leq p_n \leq 1 \quad \text{für } n \in \mathbb{N}, \tag{2.2.1}$$

$$\sum_{n=1}^{\infty} p_n = 1. \tag{2.2.2}$$

Man definiert dann die Wahrscheinlichkeit einer Teilmenge  $A \in 2^\Omega$ , indem man die Wahrscheinlichkeiten der  $\omega_n \in A$  aufaddiert.

$$P(A) := \sum_{n=1}^{\infty} p_n \mathbb{1}_A(\omega_n). \tag{2.2.3}$$

Man nennt dann  $P : 2^\Omega \rightarrow [0, 1]$  ein **diskretes W-Maß auf  $\Omega$** , daß auf die abzählbare Menge  $\{\omega_n \mid n \in \mathbb{N}\}$  **konzentriert** ist.

**Anmerkung. (Diskrete W-Räume)** 1. Eine konvergente Reihe mit positiven Summanden kann man beliebig umsortieren. Es kommt also auf die Reihenfolge der Punkte  $\omega_n$  nicht an. Man schreibt daher auch

$$\sum_{n \in \mathbb{N}} p_n := \sum_{n=1}^{\infty} p_n. \tag{2.2.4}$$

Da es auf die Reihenfolge nicht ankommt, kann man auch andere abzählbare Indexmengen  $I = \{i_n \mid n \in \mathbb{N}\}$  zulassen. Für die Summe schreibt man dann  $\sum_{i \in I} p_i := \sum_{n=1}^{\infty} p_{i_n}$ .

2. In den theoretischen Aussagen werden wir die abzählbar vielen  $\omega_n$  immer mit  $n = 1, 2, \dots$  durchnummerieren. Es können aber auch andere abzählbare Indexmengen auftreten, wie

– die ganzen Zahlen  $\mathbb{Z}$ . Diese kann man folgendermaßen abzählen:  $\mathbb{Z} = 0, 1, -1, 2, -2, 3, -3, \dots$  Man schreibt in diesem Fall für die Summe der Wahrscheinlichkeiten  $\sum_{n=-\infty}^{\infty} p_n$ .

– doppelt indizierte Punkte  $\omega_{m,n}$  mit  $m, n \in \mathbb{N}$ . Mit dem Cantorschen Diagonalverfahren kann man die Punkte folgendermaßen zeilenweise durchzählen:

$$\begin{matrix} \omega_{11} & & & & & \\ \omega_{12} & \omega_{21} & & & & \\ \omega_{13} & \omega_{22} & \omega_{31} & & & \\ \omega_{14} & \omega_{23} & \omega_{32} & \omega_{41} & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \end{matrix}$$

– Allgemeiner gilt: Wenn  $I, J$  abzählbar sind, so ist auch  $I \times J$  abzählbar.

3. Häufig wird  $\Omega = \{\omega_n \mid n \in \mathbb{N}\}$  sein. Es wird sich als praktisch erweisen, auch solche Fälle zuzulassen, wie  $\Omega = \mathbb{R}$  und  $\{\omega_1, \omega_2, \dots\} = \mathbb{Z}$ .

4. Eigentlich gehören die endlichen W-Räume ebenfalls zu den diskreten W-Räumen. Man sollte die Definition 2.2.1 eigentlich so formulieren:

**2.2.2 Def. (alternative Definition: diskrete W-Räume)**

Es sei  $\Omega$  eine beliebige Menge,  $I$  eine höchstens abzählbar unendliche Indexmenge und  $\{\omega_i \mid i \in I\}$  eine Teilmenge von  $\Omega$ . Jedem  $\omega_i$  sei eine Wahrscheinlichkeit  $p_i$  zugeordnet, derart daß

$$0 \leq p_i \leq 1 \quad \text{für } i \in I, \quad (2.2.5)$$

$$\sum_{i \in I} p_i = 1. \quad (2.2.6)$$

Man definiert dann die Wahrscheinlichkeit einer Teilmenge  $A \in 2^\Omega$ , indem man die Wahrscheinlichkeiten der  $\omega_i \in A$  aufaddiert.

$$P(A) := \sum_{i \in I \cap A} p_i \mathbb{1}_A(\omega_i). \quad (2.2.7)$$

Man nennt dann  $P : 2^\Omega \rightarrow [0, 1]$  ein diskretes W-Maß auf  $\Omega$ .

**Konvention:** Wir werden i. a. die einfachere Formulierung der Definition 2.2.1 benutzen und stillschweigens bemerken, daß alle Folgerungen natürlich auch für endliche W-Räume gelten.

**Anmerkung. (diskretes W-Maß ist  $\sigma$ -additiv)** Ein diskreter W-Raum erfüllt die beiden Axiome (i) und (ii) aus der Definition 1.1.5 eines endlichen W-Raumes. In einem unendlichen diskreten W-Raum  $(\Omega, P)$  kann man abzählbare paarweise disjunkte Teilmengen  $A_n \in 2^\Omega$ , ( $n \in \mathbb{N}$ ), betrachten und die Aussage in Feststellung 1.1.6 Gleichung (1.1.1) über die Additivität von  $P$  verschärfen.

**2.2.3 Satz (diskretes W-Maß ist  $\sigma$ -additiv)**

Es sei  $(\Omega, P)$  ein diskreter W-Raum. Dann gilt für jede Familie  $(A_n)_{n \in \mathbb{N}}$  paarweise disjunkter Mengen  $A_n \in 2^\Omega$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \quad (2.2.8)$$

Diese Eigenschaft heißt  $\sigma$ -Additivität.

**Anmerkung.** Mit dem Zusatz  $\sigma$  Bezeichnet man kurz Regeln über abzählbare Vereinigungen oder Summen. Ebenso ist das Kürzel  $\delta$  für abzählbare Durchschnitte oder Produkte im Gebrauch.

**Beweis.**  $P$  sei konzentriert auf die Menge  $\{\omega_i \mid i \in \mathbb{N}\}$ .

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{i=1}^{\infty} p_i \mathbb{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega_i)$$

Da die  $A_n$  paarweise disjunkt sind, gilt

$$= \sum_{i=1}^{\infty} p_i \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega_i)$$

wobei in der Summe  $\sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega_i)$  alle Summanden bis auf höchstens einen gleich 0 sind. Da alle Terme nichtnegativ sind, darf man die Summationsreihenfolge vertauschen:

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} p_i \mathbb{1}_{A_n}(\omega_i) \\ &= \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

**2.2.4 Festst. (Resultate aus Kap. 1.)**

1. Für einen abzählbar unendlichen diskreten W-Raum  $(\Omega, P)$ , wobei  $P$  auf  $\{\omega_n \mid n \in \mathbb{N}\}$  konzentriert ist, gelten alle Definitionen und Ergebnisse, die wir in Abschnitt 1 getroffen bzw. hergeleitet haben sinngemäß weiter, da für ihre Herleitung nur endliche Operationen – Vereinigung, Durchschnitt, Komplement von Mengen, Summe, Produkt, Quotient von Zahlen – verwendet werden; diese gelten mit der Definition 2.2.1, Gleichung (2.2.3) weiter.

2. Bei der Bildung des Erwartungswertes einer reellen Zufallsvariablen  $X : (\Omega, P) \rightarrow \mathbb{R}$  muß man immer die Zusatzvoraussetzung machen, daß die Reihe

$$E(|X|) := \sum_{n=1}^{\infty} |X(\omega_n)| P\{\omega_n\} \quad (2.2.9)$$

**absolut konvergent** ist. Wir schreiben hierfür kurz  $E(|X|) < \infty$ .

Dann konvergiert auch die Reihe

$$E(X) := \sum_{n=1}^{\infty} X(\omega_n) P\{\omega_n\} \quad (2.2.10)$$

und die Summanden dürfen **beliebig umsortiert** werden. D.h., man kann mit absolutkonvergenten Reihen bedenkenlos rechnen.

3. Wenn die Varianz auftritt, muß vorausgesetzt werden, daß  $E(|X|^2) < \infty$  ist. Aus der Schwarzischen Ungleichung folgt

$$E(|X|) = E(|1 \cdot X|) \leq \sqrt{E(1)} \sqrt{E(|X|^2)} < \infty.$$

Dann ist auch

$$\text{Var}(X) = E(X^2) - (E(X))^2 < \infty.$$

Die Kovarianz bildet man nur für reelle Zufallsvariablen  $X, Y$ , für die

$$E(|X|^2) < \infty \quad \text{und} \quad E(|Y|^2) < \infty$$

ist.

**2.2.5 Festst. (Erwartungswert der geomtr. V)**

Für eine geometrisch verteilte Zufallsvariable gilt:

(i) Der Erwartungswert ist

$$E(X) = \frac{1-p}{p} = \frac{1}{p} - 1.$$

(ii) Die Varianz ist

$$\text{Var}(X) = \frac{1-p}{p^2}$$

**Anmerkung.** Beim wiederholten Würfeln mit einem fairen Würfel hat man also *im Mittel*

$$E(X) = \left(\frac{1}{6}\right)^{-1} - 1 = 5$$

Fehlversuche, bis die erste 6 fällt. Die Varianz von  $X$  ist

$$\text{Var}(X) = \frac{5}{6} \left(\frac{1}{6}\right)^{-2} = 30.$$

Die Standardabweichung ist  $\sigma(X) = \sqrt{30}$ .

**Beweis.** (i) Da man Potenzreihen innerhalb Ihres Konvergenzbereiches gliedweise differenzieren kann, gilt

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kp(1-p)^k \\ &= -p(1-p) \frac{d}{dp} \sum_{k=0}^{\infty} (1-p)^k \\ &= -p(1-p) \frac{d}{dp} \frac{1}{1-(1-p)} = p(1-p) \frac{1}{p^2} \\ &= \frac{1}{p} - 1. \end{aligned}$$

(ii) Aus Feststellung 1.18.3 (ii) folgt

$$\begin{aligned} \text{Var}(X) &= E(X(X-1)) + E(X) - (E(X))^2 \\ E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1)p(1-p)^k \\ &= p(1-p)^2 \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &= p(1-p)^2 \frac{d^2}{dp^2} \sum_{k=0}^{\infty} (1-p)^k \\ &= p(1-p)^2 \frac{d^2}{dp^2} \frac{1}{1-(1-p)} \\ &= p(1-p)^2 \frac{2}{p^3} \\ &= \frac{2(1-p)^2}{p^2} \end{aligned}$$

Somit ist

$$\text{Var}(X) = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \frac{(1-p)^2}{p^2} = \frac{1-p}{p^2}.$$

**Bsp. (St. Peterburger Paradoxon)** Das Rechnen und argumentieren mit Zufallsvariablen  $X$ , für die  $E(|X|) = \infty$  ist, führt leicht zu paradoxen Ergebnissen. Hier ist ein Beispiel, das auf einen Artikel zurückgeht, den J. Bernoulli 1738 in der Zeitschrift der *St. Petersburger Akademie* publizierte.

In einem Glückspiel wird eine faire Münze geworfen. Fäll Kopf, so erhält der Spieler den doppelten Einsatz, anderenfalls ist der Einsatz verloren. Ein Spieler mit unbegrenztem Kapital (und Zeit) entscheidet sich für die folgende Strategie:

*Er setzt beim ersten Spiel 1 Euro; gewinnt er, so hört er auf zu spielen. Anderenfalls verdoppelt er seinen Einsatz. Dies macht er so lange, bis er das erste mal gewinnt.*

Gewinnt er beim  $n$ -ten Spiel, so hat er bis dahin

$$\sum_{\nu=1}^n 2^\nu = 2^n - 1$$

Euro gesetzt und  $2^n$  Euro gewonnen. Der Reingewinn ist 1 Euro. die Wahrscheinlichkeit, daß er erstmals im  $n$ -ten Spiel gewinnt ist

$$\mathcal{G}_{1/2}\{n-1\} = \frac{1}{2} \left(\frac{1}{2}\right)^{n-1} = 2^{-n}.$$

Die Wahrscheinlichkeit, daß er nie gewinnt ist also kleiner als  $2^{-n}$  für alle  $(n = 1, 2, \dots)$ , d.h. gleich 0.

Obwohl das Spiel fair ist, liefert die Methode mit Sicherheit den Reingewinn 1 Euro!

Das Problem liegt darin begründet, daß bei beliebig langer Spielzeit der Gewinn – ohne Berücksichtigung des Einsatzes – unendlichen Erwartungswert besitzt.

Der Grundraum  $\mathbb{N}$  mit der geometrischen Wahrscheinlichkeit  $\mathcal{G}_{1/2}$  beschreibt die Situation, daß der Gewinn im  $n$ -ten Spiel eintritt. Die Zufallsvariable

$$X : \mathbb{N} \rightarrow \mathbb{N} \quad \text{mit } X(n) = 2^n$$

beschreibt den Gewinn.  $X$  hat aber unendlichen Erwartungswert:

$$E(X) = \sum_{n=1}^{\infty} X(n)\mathcal{G}_{1/2}\{n\} = \sum_{n=1}^{\infty} 2^n 2^{-n} = \infty.$$

Ebenso hat das Eingesetzte Kapital  $Y = X - 1$  unendlichen Erwartungswert.

In einem Spiel ist der Erwartungswert des Gewinnes, sofern er endlich ist, der faire Preis, den man zu Beginn des Spieles setzen kann. Ist der Erwartungswert unendlich, so gibt es keinen sinnvollen fairen Preis. In dem obigen Spiel müßte der Spieler vorab unendlich viel Kapital als Einsatz zahlen!

**Anmerkung.** Das Produkt endlich vieler diskreter W-Räume ist wieder diskret:

### 2.2.6 Bez. (endl. Produkte diskreter WR.)

(i) Es seien  $(\Omega_i, P_i)$ ,  $(i = 1, \dots, k)$  endlich viele diskreter W-Räume. Die W-Maße seien jeweils auf die Mengen

$$M_i = \{\omega_n^{(i)} \mid n \in \mathbb{N}\}$$

konzentriert.

$$\Omega := \Omega_1 \times \dots \times \Omega_k.$$

Man erklärt das Produktmaß  $P := P_1 \otimes \dots \otimes P_k$  auf

$$\Omega := \Omega_1 \times \dots \times \Omega_k.$$

durch

$$P(A) = \sum_{(n_1, \dots, n_k) \in \mathbb{N}^k} p_{n_1} \dots p_{n_k} \mathbb{1}_A(\omega_{n_1}^{(1)}, \dots, \omega_{n_k}^{(k)}) \tag{2.2.11}$$

$$= \sum_{n_1=1}^{\infty} \dots \sum_{n_k=1}^{\infty} p_{n_1} \dots p_{n_k} \mathbb{1}_A(\omega_{n_1}^{(1)}, \dots, \omega_{n_k}^{(k)}) \tag{2.2.12}$$

für  $A \in 2^\Omega$ .

(ii) Aus Gleichung (2.2.12) folgt die Produkteigenschaft:

$$P(A_1 \times \dots \times A_k) = P_1(A_1) \dots P_k(A_k)$$

für  $A_i \in 2^{\Omega_i}$ ,  $(i = 1, \dots, k)$ .

(iii)  $M := M_1 \times \dots \times M_k$  eine abzählbare Teilmenge von  $\Omega_1 \times \dots \times \Omega_n$  und das Produktmaß ist auf  $M$  konzentriert.

(iv) (iv) Mit dieser Definition des Produktmaßes gelten alle Resultate aus Kapitel 1 unverändert fort.

### 2.2.7 Bez. (Bildverteilung einer ZV)

Es sei  $(\Omega, P)$  ein diskreter W-Raum und  $P$  sei konzentriert auf die abzählbare Menge  $\{\omega_n \mid n \in \mathbb{N}\}$ . Für eine Zufallsvariable  $X : (\Omega, P) \rightarrow M$  ist dann das Bildmaß

$$P_X(B) := P\{X \in B\} \quad \text{für } B \in 2^M$$

auf die **höchsten abzählbare** Menge  $\{\mathcal{X} := X(\omega_n) \mid n \in \mathbb{N}\}$  konzentriert; diese kann auch endlich sein. Also ist  $(M, P_X)$  wieder ein diskreter W-Raum.

### 2.2.8 Folg. (Bildverteilungen)

Mit dieser Definition des Bildmaßes ( $\rightarrow$  Bezeichnung 2.2.7) gelten alle Resultate aus Kapitel 1 weiter. Insbesondere gilt:

(i) Endlich viele Zufallsvariable  $X = (X_1, \dots, X_n)$  sind unabhängig, wenn ihre gemeinsame Verteilung  $P_X$  gleich dem Produkt ihrer Randverteilungen ist:

$$P_X = P_{X_1} \otimes \dots \otimes P_{X_n}.$$

(ii) Für den Erwartungswert einer reellen Zufallsvariablen  $(\Omega, P) \rightarrow \mathbb{R}$  mit  $E(|X|) < \infty$  gilt

$$E(X) = \sum_{x \in \mathcal{X}} x P_X\{x\}$$

**Anmerkung.** Daß das Bild eines diskreten  $W$ -Raumes wieder ein diskreter  $W$ -Raum ist, sieht sehr ansprechend aus, schränkt aber die Anwendbarkeit der diskreten  $W$ -Räume zur Modellbildung stark ein:

*Es gibt auf einem diskreten  $W$ -Raum keine abzählbare Folge  $(X_n)_n$  identisch verteilter unabhängiger Zufallsvariabler.*

Wir wollen das Begründung kurz skizzieren: Annahme,  $(X_n)_n$  sei eine abzählbare Folge identisch verteilter unabhängiger Zufallsvariabler. Man wähle  $B_n \in X_n(\Omega)$  so, daß  $P\{X_n\}(B_n) = p \in (0, 1)$  für  $n \in \mathbb{N}$ . Die Zufallsvariablen  $\mathbf{1}_{X_n^{-1}(B_n)}$  bilden eine abzählbare Folge von unabhängigen Bernoulli-Variablen mit Parameter  $p \in (0, 1)$ .

Ähnlich wie im Beispiel 2.1.2 des unendlichen Münzwurfs zeigt man, daß der projektive Limes dieses System das Intervall  $[0, 1)$  mit der uniformen Verteilung  $\mathcal{U}$  ist. Der Raum  $([0, 1), \mathcal{U})$  ist aber kein diskreter  $W$ -Raum.

## 2.3 Konvergenz gegen geomtr. Vertng.

**Anmerkung.** Wir interessieren uns für konvergente Folgen diskreter W-Maße, die alle auf dieselbe abzählbare Menge konzentriert sind. Dann ist der Grenzwert wieder ein diskretes W-Maß. In diesem Abschnitt betrachten wir ein relativ einfaches und anschauliches Beispiel, in dem der Grenzwert die geometrische Verteilung ist. In den folgenden Abschnitten untersuchen wir dann die Konvergenz der Binomialverteilung bei wachsendem  $n$  und zeigen:

Bei geeigneter Parameterwahl konvergiert die Binomialverteilung  $\mathcal{B}_{n,p}$  einerseits gegen die

→ Poisson-Verteilung auf  $\mathbb{N}$ , also wieder eine diskrete Verteilung.

und bei anderer Wahl der Parameter gegen die

→ Normalverteilung auf  $\mathbb{R}$ , also eine Verteilung mit stetiger Dichte.

### 2.3.1 Bsp. (Konvergenz gegen die geom. V.)

Wir betrachten zunächst Zufallsexperimente mit einem endlichen Ergebnisraum. Aus einer Urne mit  $K$  schwarzen Kugeln und  $N - K$  roten Kugeln wird eine geordnete Stichprobe  $\omega := (\omega_1, \dots, \omega_n)$  vom Umfang  $N$  gezogen. Es werden also alle Kugeln gezogen. Der Raum  $\Omega_{\text{ord}}$  der geordneten Stichproben sei mit Laplace-Wahrscheinlichkeit versehen. Es gibt  $\binom{N}{K}$  mögliche Plätze für die schwarzen Kugeln, die roten kommen auf die restlichen Plätze. Also ist:

$$|\Omega_{\text{ord}}| = \binom{N}{K}.$$

Die Zufallsvariable  $X : \Omega_{\text{ord}} \rightarrow \mathbb{N}$  gibt an, wieviele rote Kugeln gezogen werden, bevor die erste schwarze kommt:

$$X : \omega \mapsto \max\{i \mid 0 \leq i \leq N - K, \omega_1, \dots, \omega_i \text{ rot}\}$$

Wenn die erste schwarze Kugel im  $k + 1$ -ten Zug erscheint, sind bereits  $k$  rote Kugeln und eine schwarze gezogen worden. Es verbleiben also noch  $\binom{N-k-1}{K-1}$  mögliche Plätze für die verbleibenden  $K - 1$  schwarzen Kugeln. Also ist

$$P_{N,K}\{X = k\} = \frac{\binom{N-k-1}{K-1}}{\binom{N}{K}} \quad \text{für } k = 0, \dots, N - K.$$

Wir betrachten den Raum  $\mathbb{N}_0$  mit dem diskreten W-Maß  $P_{N,K}$ . Es gilt also

$$\begin{aligned} P_{N,K}\{X = k\} &= \frac{(N - k - 1)!}{(K - 1)!(N - K - k)!} \frac{K!(N - K)!}{N!} \\ &= \frac{K}{N - k} \prod_{i=0}^{k-1} \frac{N - K - i}{N - i} \end{aligned} \quad (2.3.1)$$

für  $k = 0, \dots, N - K$  und 0 sonst.

Man kann die Gleichung (2.3.1) auch direkt interpretieren, als die Wahrscheinlichkeit, zunächst  $k$  rote Kugeln zu ziehen und dann eine schwarze.

**Grenzübergang:** Wie verhält sich die Verteilung  $P_{N,K}$  wenn man immer mehr Kugeln in der Urne hat?

Man führe einen Grenzübergang  $N, K \rightarrow \infty$  derart aus, daß der Anteil der schwarzen Kugeln  $\frac{K}{N} \rightarrow p \in (0, 1)$  konvergiert. D.h. man wähle eine Folge  $(K_N)_N$  so daß

$$\lim_{N \rightarrow \infty} \frac{K_N}{N} = p \quad \text{mit } 0 < p < 1.$$

Aus der Gleichung (2.3.1) erhält man in diesem Fall

$$\lim_{N \rightarrow \infty} P_{N,K_N}\{k\} = p(1 - p)^k.$$

Der Grenzwert ist die geometrische Wahrscheinlichkeit zum Parameter  $p$ :

$$\lim_{N \rightarrow \infty} P_{N,K} = \mathcal{G}_p \quad (2.3.2)$$

Das Ergebnis ist anschaulich zu verstehen. Bei festem  $k$  wähle man  $N$  und  $K_N \approx pN$  sehr groß. Dann ändert sich beim Ziehen von  $k$  Kugeln aus der Urne der Restbestand praktisch nicht und folglich sind die Wahrscheinlichkeiten für die Farbe der nächsten Kugeln annähernd konstant  $p$  bzw.  $1 - p$ . Die Wahrscheinlichkeit, erst  $k$  rote und dann eine schwarze Kugel zu ziehen ist also  $\approx (1 - p)^k p$ .

**Anmerkung. (Bose-Einstein-Statistik)** Man kann das W-Maß in dem obigen Beispiel 2.3.1 auch als die Verteilung der Anzahl der Objekte in einer Zelle bei einer Belegung mit nichtunterscheidbaren Objekten interpretieren (→ Bezeichnung 1.2.5).

$R$  rote Kugeln werden durch einen Zufallsmechanismus auf  $K + 1$  Zellen mit den Nummern  $0, 1, \dots, K$  verteilt. Als Zufallsmechanismus verwende man ein Urne mit  $K$  schwarzen Kugeln und  $R = N - K$  roten Kugeln, aus der man alle Kugeln der Reihe nach entnimmt. Man hat also eine geordnete Stichprobe von roten und schwarzen Kugeln. Die roten Kugeln werden dann nach folgender Regel auf die  $K + 1$  Zellen verteilt:

Begint die Stichprobe mit einer roten Kugel, so kommt diese und die folgenden roten Kugeln solange in die Zelle mit der Nummer 0, bis eine schwarze Kugel gezogen wird, andernfalls bleibt die Zelle mit der Nummer 0 leer. Die schwarzen Kugeln bilden nun die Trennwände zwischen den folgenden Zellen mit den Nummern  $1, \dots, K - 1$ . Die roten Kugeln zwischen der  $i$ -ten und der  $i + 1$ -ten schwarzen Kugel kommen in die Zelle mit der Nummer  $i$ . Die roten Kugeln, die auf die  $K$ -te schwarze Kugel folgen, kommen in die letzte Zelle mit der Nummer  $K$ .

Ein Beispiel sagt mehr als tausend Worte: Es sei  $K = 5, R = 6$ . Die Stichprobe

$$r, r, s, s, r, r, r, s, s, r, s$$

ergibt die Belegungszahlen  $2, 0, 3, 0, 1, 0$ .

Da alle  $\binom{R+K}{R}$  Anordnungen von  $R$  roten und  $K$  schwarzen Kugeln gleichwahrscheinlich sind, bewirkt dieser Zufallsmechanismus, daß alle so erzeugten Belegungen der Zellen gleichwahrscheinlich sind. Die Zufallsvariable  $X$  mit der durch die Gleichung (2.3.1) gegebene Verteilung gibt die Anzahl der roten Kugeln in der Zelle mit der Nummer 0 an.

$$P_{R,K}\{X = k\} := \frac{K}{R + K - k} \prod_{i=0}^{k-1} \frac{R - i}{R + K - i} \quad \text{für } k = 0, \dots, R.$$

Für  $R, K \rightarrow \infty$  mit  $\frac{R}{K} \rightarrow \lambda$  ist

$$\lim_{R,K \rightarrow \infty} P_{R,K}\{k\} = \frac{1}{\lambda + 1} \left( \frac{\lambda}{1 + \lambda} \right)^k = \mathcal{G}_p\{k\}, \quad (2.3.3)$$

wobei  $p = \frac{\lambda}{1 + \lambda}$  und  $\lambda$  der Erwartungswert von  $\mathcal{G}_p$  ist.

Aus Symmetriegründen hat die Anzahl der Kugeln in jeder anderen Zelle die gleiche Verteilung wie  $X$  und ist also approximativ geometrisch verteilt.

$\binom{R+K}{R}$  ist die Anzahl Belegungen von  $K + 1$  Zellen mit  $R$  nicht unterscheidbaren Objekten (→ Folgerung 1.2.6). In der Physik heißt diese Laplace-Wahrscheinlichkeit auf den Belegungen die **Bose-Einstein Statistik**. Man findet sie in der Quantenmechanik bei der Beschreibung der sogenannten Bose-Teilchen. Elementarteilchen von diesem Typ sind **ununterscheidbar**, d.h. der Austausch zweier Teilchen ändert nichts am Zustand des Systems.

Es können mehrere Bose-Teilchen in der selben Zelle sein oder physikalisch ausgedrückt, denselben Zustand haben. Ein anderer Typ sind die Fermi-Dirac-Teilchen, die ebenfalls ununterscheidbar sind, von denen aber immer höchstens eines in einer Zelle sein kann. Die Gleichverteilung auf den möglichen Belegungen impliziert also, dass die Zahl der Teilchen in der gleichen Zelle annähernd geometrisch verteilt ist.

**Anmerkung.** Im allgemeinen hat man bei Belegungen mit unterscheidbaren Objekten nicht die Gleichverteilung (Bose-Einstein-Statistik), sondern die Multinomialverteilung. Die Anzahl der Objekte in einer Zelle ist dann die Randverteilung, also binomialverteilt ( $\rightarrow$  Beispiel 1.8.5).

Wir werden analog zum Vorgehen im Beispiel 2.3.1 für Zufallsvariable  $X_n$ , die  $\mathcal{B}_{n,p}$  verteilt sind, den Grenzwert der Verteilung für  $n \rightarrow \infty$  unter der Nebenbedingung  $E(X_n) \rightarrow \lambda$  untersuchen. Dabei stoßen wir die auf die  $\rightarrow$  Poissonverteilung auf  $\mathbb{N}$ .

## 2.4 Poisson-Verteilung

**Anmerkung.** Die Exponentialfunktion  $e^x$  gilt

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \quad \text{für } x \in \mathbb{R}. \quad (2.4.1)$$

$e^x$  hat die Potenzreihenentwicklung

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad \text{für } x \in \mathbb{R}. \quad (2.4.2)$$

Für  $\lambda > 0$  ist  $e^{-\lambda} \frac{\lambda^n}{n!} > 0$  und

$$\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} = 1. \quad (2.4.3)$$

Man kann also mit diesen Werten ein diskretes W-Maß auf  $\mathbb{N}_0$  bilden.

### 2.4.1 Bez. (Poisson-Verteilung)

1. Für  $\lambda > 0$  heißt das Wahrscheinlichkeitsmaß  $\mathcal{P}_\lambda$  auf  $\mathbb{N}_0$  mit

$$\mathcal{P}_\lambda\{k\} = e^{-\lambda} \frac{\lambda^k}{k!}$$

die Poisson-Verteilung zum Parameter  $\lambda$

2. Eine Zufallsvariable  $X$  heißt Poisson-verteilt, wenn sie ihre Werte in  $\mathbb{N}_0$  annimmt und

$$P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}$$

ist.

### 2.4.2 Festst. (Eigenschaften: Poisson-Verteilung)

Eine Poisson-verteilte Zufallsvariable  $X$  hat den Erwartungswert

$$E(X) = \lambda$$

und die Varianz

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

$\lambda$  ist also zugleich der Erwartungswert und die Varianz ein  $\mathcal{P}_\lambda$ -verteilter Zufallsvariablen.

**Beweis.**

$$E(X) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^\lambda = \lambda.$$

Aus Feststellung 1.18.3 (ii) folgt

$$\text{Var}(X) = E(X(X-1)) + E(X) - (E(X))^2.$$

$$\begin{aligned} E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda} e^\lambda = \lambda^2 \end{aligned}$$

und somit

$$\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

### 2.4.3 Satz (Konvergenz gegen Poisson-Verteilung)

Für eine Folge von Binomialverteilungen  $\mathcal{B}_{n,p_n}$ , ( $n \in \mathbb{N}$ ), deren Mittelwerte  $\mu_n := np_n$  konvergieren:

$$\lim_{n \rightarrow \infty} \mu_n =: \lambda \quad (2.4.4)$$

gilt

$$\lim_{n \rightarrow \infty} \mathcal{B}_{n,p_n}\{k\} = e^{-\lambda} \frac{\lambda^k}{k!}. \quad (2.4.5)$$

Unter der Nebenbedingung (2.4.4) konvergiert die Binomialverteilung gegen die Poissonverteilung.

**Beweis.** Es ist

$$\begin{aligned} \mathcal{B}_{n,p_n}\{k\} &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} n(n-1)\cdots(n-k+1) p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \mu_n^k \left(1 - \frac{\mu_n}{n}\right)^{n-k} \\ &= \frac{1}{k!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \mu_n^k \left(1 - \frac{\mu_n}{n}\right)^{-k} \left(1 - \frac{\mu_n}{n}\right)^n \end{aligned}$$

Letzteres konvergiert für  $n \rightarrow \infty$  gegen

$$\mathcal{B}_{n,p_n}\{k\} \rightarrow \frac{1}{k!} e^{-\lambda}.$$

Dabei haben wir das folgende Bemerkung 2.4.4 benutzt:

**2.4.4 Bem.** Man kann die Gleichung (2.4.1) noch etwas verschärfen: Für  $0 \leq \mu_n \rightarrow \lambda$  gilt

$$\left(1 - \frac{\mu_n}{n}\right)^n \rightarrow e^{-\lambda}$$

Dies sieht man folgendermaßen ein: Da der natürliche Logarithmus  $\log x$  die Ableitung  $\frac{d}{dx} \log x = 1/x$  hat, folgt aus dem Mittelwertsatz der Differentialrechnung

$$\log\left(1 - \frac{\mu_n}{n}\right) - \log 1 = -\frac{\mu_n}{n} \frac{1}{1 - \xi_n}$$

mit  $0 < \xi_n < \mu_n/n$ . Also gilt  $\xi_n \rightarrow 0$  für  $n \rightarrow \infty$ . Da  $\log 1 = 0$  ist folgt hieraus

$$\begin{aligned} \left(1 - \frac{\mu_n}{n}\right)^n &= \exp\left(n \log\left(1 - \frac{\mu_n}{n}\right)\right) \\ &= \exp\left(n \left(-\frac{\mu_n}{n} \frac{1}{1 - \xi_n}\right)\right) \\ &\rightarrow \exp(-\lambda). \end{aligned}$$

**Anmerkung.** Die Approximation der Binomialverteilung  $\mathcal{B}_{n,p}$  durch die Poisson-Verteilung  $\mathcal{P}_\lambda$  mit  $\lambda = np$  findet ihre Anwendung für große Werte von  $n$  und sehr kleine Werte von  $p$ . Das ist aus folgenden Gründen angebracht:

- Die Binomialkoeffizienten  $\binom{n}{k}$  werden schnell sehr groß und die Werte der Binomialverteilung lassen sich für große  $n$  und kleine  $p$  nicht gut berechnen. Die Poisson-Verteilung kann gut rekursiv berechnen:

$$\mathcal{P}_\lambda(0) = e^{-\lambda}, \mathcal{P}_\lambda(k) = \frac{\lambda}{k} \mathcal{P}_\lambda(k-1)$$

- Manchmal sind  $n$  und  $p$  nicht genau bekannt, man weiß nur, daß  $n$  groß und  $p$  klein ist. Wenn man den Erwartungswert  $\lambda = np$  schätzen schätzen, kann man die Verteilung trotzdem gut approximieren.

### 2.4.5 Bsp. (Verteilung der Schadensmeldungen)

Die Anzahl der Schadensmeldungen in einem Zeitraum  $(0, t]$ ,  $t > 0$ , sei proportional zur Länge des Intervalls  $(0, t]$ , also gleich  $\alpha t$  mit einem empirisch zu ermittelnden Faktor  $\alpha$ . Die Schadensmeldungen treffen zufällig ein, z.B. bei einer Versicherung. Wie ist die Verteilung  $X$  der Schadensmeldungen in einem festen Zeitraum  $(0, t]$  anzusetzen?

Klar ist,  $X$  nimmt seine Werte in  $\mathbb{N}_0$  an. Wir zerlegen das Intervall  $(0, t]$  in  $n$  gleichlange Teilintervalle. Wenn  $n$  groß ist und somit die Teilintervalle kurz sind, ist anzunehmen, daß in jedem Teilintervall höchstens ein Schadenfall eintritt. Weiterhin nehmen wir an, daß das Auftreten eines Schadens in einem Teilintervall nicht davon abhängt, ob und in welchen anderen Teilintervallen Schäden auftreten. D.h., man kann so tun, als ob die Schadensfälle in den einzelnen gleichlangen Teilintervallen  $n$ -fach Bernoulli-verteilt mit dem Parameter

$p = \alpha t/n$  sind Die Zufallsvariable  $X$  ist die Summe der „Erfolge“. und folglich annähernd binomialverteilt mit den Parametern  $n$  und  $p \approx \alpha t/n$ .

Diese Überlegung kann man für alle hinreichen großen  $n \in \mathbb{N}$  anstellen. Das liefert den Ansatz für die Verteilung von  $X$ :

$$\begin{aligned} P\{X = k\} &= \lim_{n \rightarrow \infty} \mathcal{B}_{n,\alpha t/n}\{k\} \\ &= \mathcal{P}_{\alpha t}\{k\} = e^{-\alpha t} \frac{(\alpha t)^k}{k!} \quad \text{für } k \in \mathbb{N}_0. \end{aligned} \tag{2.4.6}$$

für  $k \in \mathbb{N}_0$ .

**Anmerkung.** Wir halten fest, die Poisson-Verteilung  $\mathcal{P}_\lambda$  ist ein Modell für die Anzahl rein zufällig auftretender Zeitpunkte in einem Zeitintervall. Man denke an die Zeitpunkte, in denen im Beispiel 2.4.5 ein Schaden eintritt. Andere typische Beispiele sind die in einem Callcenter eingehenden Anrufe, die über einen Mail-Server weiterzuleitenden Emails, die Anzahl der Kunden an einem Schalter, die Anzahl der Zeitpunkte, in denen ein Atom einer radioaktiven Substanz zerfällt. Das Intervall kann auch eine andere Dimension haben, z.B. die Anzahl der Fahrzeuge auf einem Straßenabschnitt.

**Anmerkung. (zum Beispiel 2.4.5 Schadensmeldungen)** In der obigen Diskussion haben wir darauf verzichtet, einen formalen W-Raum  $(\Omega, P)$  für  $X : (\Omega, P) \rightarrow \mathbb{N}_0$  anzugeben. Wir gehen mal davon aus, daß einen solchen Raum gibt und wollen die Herleitung der Gleichung (2.4.6) etwas präzisieren:

1. Man zerlegt das Intervall  $(0, t]$  in  $n$ -Teile und bezeichnet mit  $X_{n,j}$ , ( $j = 1, \dots, n$ ) die Anzahl der Schadensmeldungen im  $j$ -ten Teilintervall. Dann ist

$$X = X_{n,1} + \dots + X_{n,n}.$$

Wir haben vorausgesetzt:

$X_{n,1}, \dots, X_{n,n}$  **unabhängig und identisch verteilt.** (\*)

2. Dann haben wir  $X_{n,j}$  durch die Indikatorvariable  $\mathbb{1}_{\{X_{n,j} \geq 1\}}$  ersetzt und  $X$  durch die Summe

$$S_n := \sum_{j=1}^n \mathbb{1}_{\{X_{n,j} \geq 1\}}.$$

angenähert. Da die Ereignisse  $\{X_{n,j} > 1\}$  unabhängig sind und die gleiche Wahrscheinlichkeit  $p_n := P\{X_{n,1} \geq 1\}$  besitzen, ist  $S_n$  eine  $\mathcal{B}_{n,p_n}$  verteilte Zufallsvariable. Aus

$$\mathbb{1}_{\{X_{n,1} \geq 1\}} \leq X_{n,1}$$

folgt

$$p_n = E(\mathbb{1}_{\{X_{n,1} \geq 1\}}) \leq E(X_{n,1}) = \frac{\alpha t}{n}.$$

3. Wir haben dann mit „munteren Worten“  $p_n \approx \alpha t/n$  gesetzt, d.h. wir setzen voraus, daß

$$\lim_{n \rightarrow \infty} np_n = \lim_{n \rightarrow \infty} E(S_n) \stackrel{!}{=} E(X) = \alpha t \tag{**}$$

Hiermit folgt aus Satz 2.4.3

$$\lim_{n \rightarrow \infty} P\{S_n = k\} = \lim_{n \rightarrow \infty} \mathcal{B}_{n,p_n}\{k\} = \mathcal{P}_{\alpha t}. \tag{2.4.7}$$

Wir wollen nun schließen, daß  $X$  ebenfalls Poisson-verteilt mit dem Parameter  $\alpha t$  ist.

4. Dazu zeigen wir die folgende Aussage:

Aus (\*\*) folgt

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{j=1}^n \{X_{n,j} \geq 2\}\right) = 0. \tag{2.4.8}$$

D.h. bei immer feiner werdender Zerlegung wird das Auftreten von mehr als einem Schaden immer unwahrscheinlicher. Dies sieht man folgendermaßen ein:

Wir benutzen in der folgenden Gleichung die  $\sigma$ -Additivität des nicht genauer angegebenen Wahrscheinlichkeitsmaßes  $P$ . Im Falle diskreter W-Räume hatten wir die  $\sigma$ -Additivität in Satz 2.2.3 gezeigt.

$$\begin{aligned} E(S_n) &= \sum_{j=1}^n P\{\mathbb{1}_{\{X_{n,j} \geq 1\}}\} = \sum_{j=1}^n \sum_{k=1}^{\infty} P\{X_{n,j} = k\} \\ &\leq \sum_{j=1}^n \sum_{k=1}^{\infty} kP\{X_{n,j} = k\} \stackrel{!}{=} \sum_{j=1}^n E(X_{n,j}) = E(X). \end{aligned}$$

Aus (\*\*\*) folgt somit

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n \sum_{k=2}^{\infty} (k-1)P\{X_{n,j} = k\} = 0.$$

Aus der  $\sigma$ -Additivität von  $P$  folgt nun

$$\begin{aligned} P\left(\bigcup_{j=1}^n \{X_{n,j} \geq 2\}\right) &\leq \sum_{j=1}^n \sum_{k=2}^{\infty} P\{X_{n,j} = k\} \\ &\leq \sum_{j=1}^n \sum_{k=2}^{\infty} (k-1)P\{X_{n,j} = k\} \rightarrow 0 \end{aligned}$$

und somit die Gleichung (2.4.8).

5. Aus den Gleichungen (2.4.7) und (2.4.8) erhalten wir nun die Verteilung von  $X$ . Wir zerlegen dazu die Menge  $\{X = k\}$  in die folgenden Fälle:

$$\begin{aligned} P\{X = k\} &= P(\{X = k\} \cap \{X = S_n\}) + P(\{X = k\} \cap \{X \neq S_n\}) \\ &= P\{S_n = k\} \cap \{X = S_n\} + P(\{X = k\} \cap \{X \neq S_n\}) \\ &= P\{S_n = k\} - P(\{S_n = k\} \cap \{X \neq S_n\}) \\ &\quad + P(\{X = k\} \cap \{X \neq S_n\}) \end{aligned}$$

Da

$$\{X \neq S_n\} \subset \bigcup_{j=1}^n \{X_{n,j} \geq 2\}$$

ist, folgt aus (2.4.8)

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\{S_n = k\} \cap \{X \neq S_n\}) &= 0, \\ \lim_{n \rightarrow \infty} P(\{X = k\} \cap \{X \neq S_n\}) &= 0. \end{aligned}$$

$$\text{Also ist } P\{X = k\} = \lim_{n \rightarrow \infty} P\{S_n = k\} = \mathcal{P}_{\text{ot}}\{k\}.$$

**Anmerkung.** Bei der Herleitung der Poisson-Verteilung in der obigen Anmerkung haben wir vorausgesetzt, daß sich  $X$  als Summe von  $n$  unabhängigen, identisch verteilten Zufallsvariablen schreiben läßt. Wir wollen zeigen, daß dies für eine Poisson-verteilte Zufallsvariable für jedes  $n$  einzeln möglich ist, indem wir  $X$  durch eine passende Zufallsvariable  $\tilde{X}$  mit der gleichen Verteilung ersetzen. Eigentlich benötigen wir einen W-Raum, auf dem dies für alle  $n$  zugleich möglich ist.

In der Übung wird gezeigt, daß für unabhängige Zufallsvariable  $X$  und  $Y$ , die Poisson-verteilt sind mit den Parametern  $\lambda$  bzw  $\mu$ , die Summe  $X + Y$  wieder Poisson-verteilt ist mit dem Parameter  $\lambda + \mu$ .

Es sei  $X$  Poisson-verteilt mit dem Parameter  $\lambda$ . Man wähle eine Zufallsvariable  $Y$  die Poisson-verteilt mit dem Parameter  $\lambda/n$  ist. Z.B. die identische Abbildung von  $(\Omega, P) := (\mathbb{N}_0, \mathcal{P}_{\lambda/n})$ .

Man bilde nun das  $n$ -fache Produkt  $(\Omega^n, P^{\otimes n})$  mit den Projektionen  $\text{pr}_\nu : \Omega^n \rightarrow \Omega$  und setze

$$X_\nu = Y \circ \text{pr}_\nu \quad \text{für } \nu = 1, \dots, n.$$

Die  $X_\nu$  sind unabhängig und  $\mathcal{P}_{\lambda/n}$  verteilt. Also ist

$$\tilde{X} := \sum_{\nu=1}^n X_\nu$$

$\mathcal{P}_\lambda$  verteilt.

Da  $X$  und  $\tilde{X}$  dieselbe Verteilung haben, kann man  $X$  durch  $\tilde{X}$  ersetzen. Die letztere Zufallsvariable ist die Summe von  $n$  identisch verteilten unabhängigen Zufallsvariablen.

## 2.5 Allgemeine W-Räume

**Anmerkung.** Wir geben die formale Definition eines W-Raumes. Die diskreten W-Räume ordnen sich hier als wichtiges Beispiel ein. Wir werden keine tieferliegenden Ergebnisse über allgemeine W-Räume herleiten, sondern benötigen die Definition im wesentlichen, um in den folgenden Beispielen die übliche Terminologie verwenden zu können.

Es erweist sich als zweckmäßig und vielfach aus Gründen der Logik als notwendig, daß man bei überabzählbaren Ereignisräumen, wie z.B. bei den reellen Zahlen  $\mathbb{R}$ , als Ereignisse nicht mehr alle Teilmengen des Grundraumes zuläßt, sondern sich auf eine kleineres Mengensystem beschränkt, daß alle „interessierenden“ Ereignisse enthält. Im Fall der reellen Zahlen wird man z. B. die Wahrscheinlichkeit von Intervallen berechnen wollen.

Wie wir bei der Untersuchung diskreter W-Räume bereits gesehen haben, ist das wesentliche theoretische und rechnerische Hilfsmittel die Approximation durch einfachere Modelle. Der Übergang zu Grenzwerten vereinfacht einerseits die Theorie aber auch die praktische Berechnung. Der Grenzwert ist der ideale genaue Wert, mit dem man jede Approximation vergleichen kann. Die Kenntnis des Grenzwertes ermöglicht es oft andere, bessere Näherungen zu finden, als die, mit denen man zunächst die Existenz des Grenzwertes und seine Eigenschaften gefunden hat.

Um Grenzwerte bilden zu können, müssen wir für die Ereignisse abzählbare Mengenoperationen (Vereinigung, Durchschnitt, Komplement) zulassen. Im Falle der reellen Zahlen kommt man also nicht umhin, auch all die Teilmengen zu betrachten, die sich durch „beliebig wiederholte“ abzählbare Mengenoperationen aus Intervallen bilden lassen. Man nennt dies das System der *Borel-Mengen*  $\mathcal{B}(\mathbb{R})$  auf  $\mathbb{R}$ . Es wird sehr aufwendig und ist auch nicht nötig, dieses Mengensystem durch Konstruktionsvorschriften zu beschreiben. Man charakterisiert die Borel-Mengen  $\mathcal{B}(\mathbb{R})$  als das kleinste Mengensystem, das alle Intervalle enthält und aus dem abzählbaren Mengenoperationen nicht hinausführen.

Um diese Gedanken zu präzisieren gibt man eine axiomatische Charakterisierung und gelangt so zum Begriff der  $\sigma$ -Algebra:

### 2.5.1 Def. ( $\sigma$ -Algebra)

Es sei  $\Omega$  eine Menge. eine Teilmenge  $\mathcal{A} \subset 2^\Omega$  heißt ein Mengensystem auf  $\Omega$ .

Ein Mengensystem  $\mathcal{A}$  heißt eine  $\sigma$ -Algebra, wenn folgendes gilt:

1.  $\Omega \in \mathcal{A}$
2. Mit  $A \in \mathcal{A}$  ist auch das komplementäre Menge  $A^c \in \mathcal{A}$ .
3. Ist  $(A_n)_{n \in \mathbb{N}}$  in  $\mathcal{A}$ , so ist auch

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$$

In der W-Theorie nennt das Paar  $(\Omega, \mathcal{A})$  einen **Ereignisraum** und die Elemente von  $\mathcal{A}$  Ereignisse. In der Maßtheorie sagt man **Meßraum** oder **meßbarer Raum** dazu.

#### Anmerkung.

1. Man beachte, es nicht gefordert wird, daß eine  $\sigma$ -Algebra für beliebige Familien  $(A_i)_{i \in I}$  die Vereinigungen  $\bigcup_{i \in I} A_i$  enthält. Diese Eigenschaft ist i.a. auch falsch.

Ein einfaches Beispiel ist die – sonst nicht weiter interessante –  $\sigma$ -Algebra

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

Es sind alle einpunktigen Mengen  $\{a\} \in \mathcal{A}$ . Jede Menge ist Vereinigung ihrer Punkte. Da  $\mathbb{R}$  überabzählbar ist, ist  $\mathcal{A} \neq 2^\mathbb{R}$ , z.B. ist das Intervall  $[0, \infty) \notin \mathcal{A}$ .

2. In den Axiomen einer  $\sigma$ -Algebra steht auch nicht, daß die einpunktigen Mengen zu  $\mathcal{A}$  gehören. Das wird aber häufig der Fall sein.

### 2.5.2 Festst. (Rechenregeln: $\sigma$ -Algebren)

Es sei  $\mathcal{A}$  eine  $\sigma$ -Algebra auf der Menge  $\Omega$ .

(i) Es ist  $\emptyset = \Omega^c \in \mathcal{A}$ .

(ii) Sind  $A_1, \dots, A_n \in \mathcal{A}$ , so ergänze man sie durch  $\emptyset = A_{n+1} = A_{n+2} = \dots$  zu einer Folge. Somit ist die endliche Vereinigung  $A_1 \cup \dots \cup A_n = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ .

(iii) Es sei  $(A_n)_{n \in \mathbb{N}}$  in  $\mathcal{A}$ . Nach der de Morganschen Regel gilt:

$$\bigcap_{n=1}^{\infty} A_n = \left( \bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{A}.$$

(iv) Analog zum Punkt (ii) ist mit  $A_1, \dots, A_n \in \mathcal{A}$  auch der Durchschnitt  $A_1 \cap \dots \cap A_n \in \mathcal{A}$ .

(v) Ist  $(A_n)_{n \in \mathbb{N}}$  in  $\mathcal{A}$  eine fallende Folge, d.h.

$$A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots,$$

dann kann man  $A_k$  als disjunkte Vereinigung schreiben:

$$A_k = \left( \bigcap_{n=1}^{\infty} A_n \right) \cup \bigcup_{n=k}^{\infty} (A_n \setminus A_{n+1}),$$

Das System  $\bigcap_{j=1}^{\infty} A_j$  und  $(A_n \setminus A_{n+1})_n$  besteht aus paarweise disjunkten Mengen. Ist  $(A_n)_n$  eine beliebige Folge in  $\mathcal{A}$ , dann kann man die Vereinigung der  $A_n$  als disjunkte Vereinigung schreiben:

$$\bigcup_{n=k}^{\infty} A_n = \bigcup_{n=1}^{\infty} \left( A_n \setminus \bigcup_{i=1}^{n-1} A_i \right)$$

Also

$$\bigcup_{n=k}^{\infty} A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \dots$$

### 2.5.3 Bsp. ( $\sigma$ -Algebren)

1. Für eine beliebige Menge ist die Potenzmenge  $2^\Omega$  eine  $\sigma$ -Algebra auf  $\Omega$ .  $2^\Omega$  ist offensichtlich die **größte**  $\sigma$ -Algebra auf  $\Omega$ . Sie ist i.a. viel zu groß.

Die **kleinste**  $\sigma$ -Algebra auf  $\Omega$  ist  $\{\Omega, \emptyset\}$ .

2. Der Durchschnitt einer beliebigen Familie von  $\sigma$ -Algebren auf  $\Omega$  ist eine  $\sigma$ -Algebra.

3. Wenn  $\Omega = \{\omega_n \mid n \in \mathbb{N}\}$  höchstens abzählbar unendlich ist, so ist  $2^\Omega$  die kleinste  $\sigma$ -Algebra auf  $\Omega$ , die alle einelementigen Teilmengen (Punkte) von  $\Omega$  enthält. Aus diesem Grunde rechnet man bei diskreten W-Räumen immer mit der vollen Potenzmenge.

4. Zu  $\mathcal{M} \subset 2^\Omega$  gibt es eine **kleinste**  $\mathcal{M}$  **umfassende**  $\sigma$ -Algebra, die man mit  $\sigma(\mathcal{M})$  bezeichnet.

**Beweis.** Die Potenzmenge  $2^\Omega$  ist eine  $\sigma$ -Algebra, die  $\mathcal{M}$  enthält. Nun bilde man den Durchschnitt aller  $\sigma$ -Algebren, die  $\mathcal{M}$  enthalten. Dieser Durchschnitt ist nicht leer, ist eine  $\sigma$ -Algebra, und enthält  $\mathcal{M}$ .

5. Ist insbesondere  $\Omega = \mathbb{R}$ , so bezeichnet man mit  $\mathcal{B}(\mathbb{R})$  die kleinste  $\sigma$ -Algebra auf  $\mathbb{R}$ , die alle offenen Mengen enthält.  $\mathcal{B}(\mathbb{R})$  heißt die **Borel-Algebra** von  $\mathbb{R}$ .

Jedes abgeschlossene oder halbabgeschlossene Intervall ist Durchschnitt von abzählbar vielen offenen Intervallen:

$$[a, b] = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, b + \frac{1}{n}),$$

$$[a, b) = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, b).$$

Weiterhin ist jede offene Menge in  $\mathbb{R}$  die Vereinigung von höchstens abzählbar vielen offenen Intervallen. Daher ist  $\mathcal{B}(\mathbb{R})$  auch die kleinste  $\sigma$ -Algebra die alle beschränkten Intervalle enthält.

$\mathcal{B}(\mathbb{R})$  ist auch die kleinste  $\sigma$ -Algebra, die alle Intervalle der Form  $(-\infty, a]$  enthält.

Weiterhin ist jede einpunktige Menge  $\{a\} = [a, a] \in \mathcal{B}(\mathbb{R})$  und somit auch jede abzählbare Menge in  $\mathcal{B}(\mathbb{R})$ .

Wenn man von dem Ereignisraum  $\mathbb{R}$  spricht, ist immer  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  gemeint.

6. Für ein Intervall  $I \subset \mathbb{R}$  sei

$$\mathcal{B}(I) := \{A \mid A \in \mathcal{B}(\mathbb{R}) \text{ und } A \subset I\}.$$

$\mathcal{B}(I)$  ist die kleinste  $\sigma$ -Algebra auf  $I$ , die alle Teilintervalle von  $I$  enthält.

7. Die Borel-Algebra  $\mathcal{B}(\mathbb{R}^d)$  ist die kleinste  $\sigma$ -Algebra, die alle offenen Teilmengen des  $\mathbb{R}^d$  enthält. Wie im Falle  $n = 1$  zeigt man, daß  $\mathcal{B}(\mathbb{R}^n)$  die kleinste  $\sigma$ -Algebra ist, die alle offenen oder alle halboffenen oder alle abgeschlossenen achsenparallelen Quader enthält.

Weiterhin ist  $\mathcal{B}(\mathbb{R}^d)$  die kleinste  $\sigma$ -Algebra, die alle Mengen der Form

$$(-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_n]$$

enthält.

**2.5.4 Bem. (Bild und Urbild einer  $\sigma$ -Algebra)**

1. Es seien  $(\Omega, \mathcal{A})$  ein Maßraum,  $\tilde{\Omega}$  eine Menge und  $X : \Omega \rightarrow \tilde{\Omega}$  eine Abbildung. Dann ist

$$\tilde{\mathcal{A}} := \{B \in 2^{\tilde{\Omega}} \mid X^{-1}(B) \in \mathcal{A}\}$$

eine  $\sigma$ -Algebra.

Man beachte, das das Mengensystem  $\{X(A) \mid A \in \mathcal{A}\}$  i.a. keine  $\sigma$ -Algebra ist, selbst wenn  $X$  surjektiv ist, da i.a.  $X(A)^c \neq X(A^c)$  ist.

2. Es seien  $\Omega$  eine Menge,  $(\tilde{\Omega}, \tilde{\mathcal{A}})$  eine Maßraum und  $X : \Omega \rightarrow \tilde{\Omega}$  eine Abbildung. Dann ist

$$X^{-1}(\tilde{\mathcal{A}}) := \{A \in 2^{\Omega} \mid A = X^{-1}(B) \text{ für ein } B \in \tilde{\mathcal{A}}\}$$

eine  $\sigma$ -Algebra.

**Anmerkung.** Wir definieren nun den allgemeinste Version eines W-Raumes.

**2.5.5 Def. (W-Maß auf einer  $\sigma$ -Algebra)**

Es sei  $(\Omega, \mathcal{A})$  ein Maßraum. Eine Abbildung  $P : \mathcal{A} \rightarrow [0, 1]$  heißt ein W-Maß, wenn die Axiome von Kolmogorov gelten:

1.  $P(\Omega) = 1$ .

2. Für eine Folge  $(A_n)_{n \in \mathbb{N}}$  paarweise disjunkter Ereignisse in  $\mathcal{A}$  gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad (\sigma\text{-additiv})$$

Das Tripel  $(\Omega, \mathcal{A}, P)$  heißt ein **W-Raum**.

**2.5.6 Satz (Rechenregeln: W-Maße)**

Es sei  $(\Omega, \mathcal{A}, P)$  ein W-Raum ( $\rightarrow$  Definition 2.5.5). Dann gilt

(i) Für zwei disjunkte Mengen  $A, B \in \mathcal{A}$  ist

$$P(A \cup B) = P(A) + P(B) \quad (\text{additiv})$$

Es gelten also alle Rechenregeln aus Feststellung 1.1.6 über endliche Mengenoperationen.

(ii) Für eine beliebige Folge  $(A_n)_n$  in  $\mathcal{A}$  gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) \quad (\sigma\text{-subadditiv})$$

(iii) Für eine monoton fallende Folge  $(A_n)_{n \in \mathbb{N}}$  in  $\mathcal{A}$  gilt

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n). \quad (\text{von oben stetig})$$

(iv) Für eine monoton wachsende Folge  $(A_n)_{n \in \mathbb{N}}$  in  $\mathcal{A}$  gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n). \quad (\text{von unten stetig})$$

**2.5.7 Bsp. (Uniforme Verteilung)**

1. Man zeigt in allen gängigen Lehrbüchern zur Analysis III, daß es genau eine W-Maß  $\mathcal{U}$  auf den Borel-Mengen  $\mathcal{B}([0, 1])$  gibt, so das

$$\mathcal{U}[a, b] := b - a \quad \text{für } 0 \leq a \leq b < 1.$$

gilt. Aus Stetigkeit von oben bzw. unten folgt dann

$$\mathcal{U}(a, b) = \mathcal{U}(a, b) = \mathcal{U}(a, b) = \mathcal{U}[a, b]$$

Man nennt diese W-Maß die uniforme Verteilung auf  $[0, 1]$ . In den Lehrbüchern zur Analysis heißt diese W-Maß das **Lebesgue-Maß** und wird mit  $\lambda$  bezeichnet.

2. Das Lebesgue-Maß  $\lambda$  existiert auch auf jedem anderen echten, beschränkten Intervall  $I \subset \mathbb{R}$ . Um die uniforme Verteilung auf  $I$  zu erhalten, muß man  $\lambda$  normieren. Man setzt

$$\mathcal{U}_I(A) := \frac{\lambda(A)}{\lambda(I)} \quad \text{für } A \in \mathcal{B}(I).$$

Ist  $I = [a, b]$ , so gilt also

$$\mathcal{U}_I([c, d]) = \frac{d - c}{b - a} \quad \text{für } a \leq c < d \leq b.$$

## 2.6 Zufallsvariable und ihre Verteilung

### 2.6.1 Def. (meßbare Abb, Zufallsvariable)

1. Es seien  $(\Omega, \mathcal{A})$  und  $(\tilde{\Omega}, \tilde{\mathcal{A}})$  Meßräume. Eine Abbildung  $X : \Omega \rightarrow \tilde{\Omega}$  heißt **meßbar**, wenn

$$X^{-1}(B) \in \mathcal{A} \quad \text{für } B \in \tilde{\mathcal{A}}.$$

2. Ist  $(\Omega, \mathcal{A}, P)$  ein W-Raum und  $(\tilde{\Omega}, \tilde{\mathcal{A}})$  ein Meßraum, dann heißt eine meßbare Abbildung  $X : \Omega \rightarrow \tilde{\Omega}$  eine **Zufallsvariable**.

Wir schreiben hierfür kurz  $X : (\Omega, \mathcal{A}, P) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$ .

**Anmerkung.** Für jede Zufallsvariable  $X : (\Omega, \mathcal{A}, P) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$  definiert jedes  $B \in \tilde{\mathcal{A}}$  ein Ereignis  $\{X \in B\} = X^{-1}(B) \in \mathcal{A}$ , dessen Wahrscheinlichkeit  $P\{X \in B\}$  man bilden kann. Wie im endlichen und diskreten Fall ( $\rightarrow$  Satz 1.6.3) heißt

$$P_X : B \mapsto P\{X \in B\} \quad \text{für } B \in \tilde{\mathcal{A}}$$

die Bildverteilung von  $X$ .

### 2.6.2 Festst. (Regeln: Zufallsvariable)

(i) Es seien  $(\Omega, \mathcal{A}, P)$  ein W-Raum,  $(\tilde{\Omega}, \tilde{\mathcal{A}})$  und  $(\hat{\Omega}, \hat{\mathcal{A}})$  Meßräume. Ist  $X : \Omega \rightarrow \tilde{\Omega}$  eine Zufallsvariable und  $\varphi : \tilde{\Omega} \rightarrow \hat{\Omega}$  meßbar, dann ist  $\varphi \circ X$  wieder Meßbar, also eine Zufallsvariable.

(ii) Es seien  $(\Omega, \mathcal{A}, P)$  ein W-Raum,  $(\tilde{\Omega}, \tilde{\mathcal{A}})$  ein Meßraum und Ist  $\mathcal{D}$  ein Erzeugendensystem von  $\tilde{\mathcal{A}}$ , d.h.,  $\sigma(\mathcal{D}) = \tilde{\mathcal{A}}$ .

Eine Abbildung  $X : \Omega \rightarrow \tilde{\Omega}$  ist bereits eine Zufallsvariable, wenn

$$X^{-1}(D) \in \mathcal{A} \quad \text{für alle } D \in \mathcal{D} \text{ gilt.}$$

**Beweis.** (i) Nach Voraussetzung ist für alle  $C \in \hat{\mathcal{A}}$

$$(\varphi \circ X)^{-1}(C) = X^{-1}(\varphi^{-1}(C)) \in \mathcal{A}.$$

(ii) Das Mengensystem

$$\hat{\mathcal{A}} := \{A \in 2^{\tilde{\Omega}} \mid X^{-1}(A) \in \mathcal{A}\}$$

ist eine  $\sigma$ -Algebra, die  $\mathcal{D}$  umfaßt. Also ist  $\tilde{\mathcal{A}} = \sigma(\mathcal{D}) \subset \hat{\mathcal{A}}$ .

### 2.6.3 Satz (Verteilung einer Zufallsvariablen)

Es seien  $(\Omega, \mathcal{A}, P)$  ein W-Raum,  $(\tilde{\Omega}, \tilde{\mathcal{A}})$  ein Meßraum und  $X : \Omega \rightarrow \tilde{\Omega}$  eine Zufallsvariable. Dann ist die Abbildung

$$P_X : \tilde{\mathcal{A}} \rightarrow [0, 1],$$

$$P_X : B \mapsto P(X^{-1}(B)) \quad \text{für } B \in \tilde{\mathcal{A}}$$

ein W-Maß auf  $\tilde{\mathcal{A}}$ .

$P_X$  heißt die Verteilung von  $X$  oder das Bildmaß von  $P$  unter  $X$ .  $(\tilde{\Omega}, \tilde{\mathcal{A}}, P_X)$  heißt der von  $X$  induzierte W-Raum.

**Beweis.** Wir weisen die Eigenschaften aus Definition 2.5.5 nach.  $P_X$  ist eine Abbildung von  $\tilde{\mathcal{A}}$  in  $[0, 1]$  mit  $P_X(\tilde{\Omega}) = 1$ . Wir zeigen die  $\sigma$ -Additivität: Für eine Folge  $(B_n)_n$  paarweise disjunkter Mengen in  $\tilde{\mathcal{A}}$  gilt

$$\begin{aligned} P_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= P\left(X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) \\ &= P\left(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\right) \\ &= \sum_{n=1}^{\infty} P(X^{-1}(B_n)) = \sum_{n=1}^{\infty} P_X(B_n). \end{aligned}$$

Man beachte, wenn  $B_1, B_2$  disjunkt sind, so sind auch die Urbilder  $X^{-1}(B_1)$  und  $X^{-1}(B_2)$  disjunkt.

**Anmerkung.** Die Feststellung 1.6.4 gilt sinngemäß weiter:

### 2.6.4 Bem. (Funktionen von Zufallsvariablen)

Gegeben seien ein W-Raum  $(\Omega, \mathcal{A}, P)$ , Meßräume  $(\tilde{\Omega}, \tilde{\mathcal{A}})$   $(\hat{\Omega}, \hat{\mathcal{A}})$  und Zufallsvariable

$$X : \Omega \rightarrow \tilde{\Omega}, \quad Y : \Omega \rightarrow \hat{\Omega}$$

und eine meßbare Abbildung  $Z : \tilde{\Omega} \rightarrow \hat{\Omega}$  so, daß

$$X = Z \circ Y$$

ist. Man bilde den W-Raum  $(\tilde{\Omega}, P_Y)$ . Dann haben die Zufallsvariablen

$$X : (\Omega, P) \rightarrow \hat{\Omega} \quad \text{und} \quad Z : (\tilde{\Omega}, P_Y) \rightarrow \hat{\Omega}$$

die gleiche Verteilung:  $(\hat{\Omega}, P_X) = (\hat{\Omega}, P_Z)$  oder kurz

$$(P_Y)_Z = P_{Z \circ Y}. \quad (2.6.1)$$

## 2.7 Eindeutigkeit eines W-Maßes

**Anmerkung. (Problem: W-Funktion zu  $\mathcal{U}$  identisch 0)**

1. Für die Uniforme Verteilung auf  $[0, 1]$  hat jede einpunktige Menge das Maß Null:

$$\mathcal{U}\{a\} = \lim_{n \rightarrow \infty} \mathcal{U}\left[\left[a - \frac{1}{n}, a + \frac{1}{n}\right]\right] = 0.$$

Bei endlichen und diskreten W-Räumen  $(\Omega, P)$  ist die W-Funktion  $p : \omega \mapsto P\{\omega\}$  ein nützliches Hilfsmittel ( $\rightarrow$  Satz 1.1.7). Im Falle der uniformen Verteilung ist die W-Funktion identisch 0 und nützt nichts bei der Berechnung von  $\mathcal{U}$ .

2. Da es i.a. nicht einfach ist, die Werte  $P(A)$  eines W-Maßes  $P$  für alle  $A \in \mathcal{A}$  zu berechnen, sucht man nach einem möglichst einfachen Mengensystem  $\mathcal{D} \subset \mathcal{A}$ , derart daß  $P$  durch seine Werte auf  $\mathcal{D}$  eindeutig festgelegt ist. Das obige Beispiel zeigt, daß die einpunktigen Mengen dazu i.a. nicht geeignet sind.

Der folgende Satz aus der Maßtheorie, den wir ohne Beweis angeben, sagt, wie ein solches Mengensystem  $\mathcal{D}$  beschaffen sein muß. Für einen Beweis siehe z.B. [2, Anhang M.4].

3. In den Anwendungen rechnet man die meiste Zeit mit der Wahrscheinlichkeit von Ereignissen aus  $\mathcal{D}$ . Aber ab und zu ist es doch wichtig, daß man nicht auf diese kleine Mengensystem beschränkt ist, sondern sich gedanklich in der von  $\mathcal{D}$  erzeugten  $\sigma$ -Algebra frei bewegen kann.

### 2.7.1 Satz (Dynkin: Eindeutigkeit v. W-Maßen)

Es seien  $(\Omega, \mathcal{A})$  ein Maßraum und  $P, Q$  zwei W-Maße auf  $\mathcal{A}$ . Ist  $\mathcal{D} \subset \mathcal{A}$  ein  $\cap$ -stabiles Mengensystem und stimmen  $P$  und  $Q$  auf  $\mathcal{D}$  überein, so stimmen sie auch auf der erzeugten  $\sigma$ -Algebra  $\sigma(\mathcal{D})$  überein. Wenn also  $\sigma(\mathcal{D}) = \mathcal{A}$  ist, so folgt  $P = Q$ .

Dabei bedeutet,  $\mathcal{D}$  ist  $\cap$ -stabil:

$$D \cap E \in \mathcal{D} \quad \text{für alle } D, E \in \mathcal{D}.$$

### 2.7.2 Bsp. (zum Satz von Dynkin)

Das Mengensystem

$$\mathcal{D} := \{(-\infty, x_1] \times \cdots \times (-\infty, x_n] \mid x \in \mathbb{R}^d\}$$

ist  $\cap$ -stabil und erzeugt die Borel-Algebra  $\mathcal{B}(\mathbb{R}^d)$ . Satz 2.7.1 führt zu der Definition 2.8.1 der Verteilungsfunktion  $F(x) := P((-\infty, x])$  eines W-Maßes auf  $\mathbb{R}$ .

**Bsp.** Sei  $\Omega = \{1, 2, 3, 4\}$  und  $\mathcal{D} = \{\{1, 2\}, \{2, 3\}\}$ . Da

$$\begin{aligned} \{1\} &= \{1, 2\} \setminus \{2, 3\}, \\ \{2\} &= \{1, 2\} \cap \{2, 3\}, \\ \{3\} &= \{2, 3\} \setminus \{1, 2\}, \\ \{4\} &= \Omega \setminus (\{1, 2\} \cup \{2, 3\}) \end{aligned}$$

gilt, ist die von  $\mathcal{D}$  erzeugte  $\sigma$ -Algebra gleich der Potenzmenge  $2^\Omega$ .

$\mathcal{D}$  ist nicht  $\cap$ -stabil. Hier sind zwei verschieden W-Maße auf  $\Omega$ , die auf  $\mathcal{D}$  übereinstimmen:

$$\begin{aligned} P &\text{ Laplace-W. auf } \Omega, \\ Q &:= P(\cdot | \{2, 4\}). \end{aligned}$$

Dieses Beispiel zeigt, daß im Satz 2.7.1 die Voraussetzung, das erzeugende System  $\mathcal{D}$  ist  $\cap$ -stabil, wirklich benötigt wird.

## 2.8 Verteilungsfunktion

**Anmerkung.** Das Beispiel 2.7.2 führt zu der folgenden Definition:

### 2.8.1 Def. (Verteilungsfunktion)

Ein W-Maß  $P$  auf  $\mathcal{B}(\mathbb{R}^d)$  ist eindeutig durch seine Verteilungsfunktion

$$\begin{aligned} F : \mathbb{R}^d &\rightarrow [0, 1], \\ F : x &\mapsto P((-\infty, x_1] \times \cdots \times (-\infty, x_n]) \end{aligned}$$

festgelegt.

Manchmal nennt man  $F$  auch kurz die Verteilung von  $P$ . Die Bezeichnung „Verteilung“ wird aber ebenfalls für das Bildmaß einer Zufallsvariablen benutzt. Bei einer Zufallsvariablen  $X : (\Omega, P) \rightarrow \mathbb{R}^d$  muß man sorgfältig zwischen dem Bildmaß  $P_X$  und der Verteilungsfunktion  $F_X$  von  $P_X$  unterscheiden! Zur Unterscheidung wird für  $F_X$  auch die Bezeichnung *kumulative Verteilung* verwendet.

### 2.8.2 Festst. (Verteilungsfkt. rechtsseitig stetig)

Es sei  $P$  ein W-Maß auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mit Verteilungsfunktion  $F$ . Dann gilt

(i) Für jede monoton fallende Folge  $(x_n)_n$ , mit  $\lim_{n \rightarrow \infty} x_n = b$  gilt

$$\lim_{n \rightarrow \infty} F(x_n) = P\left(\bigcap_{n=1}^{\infty} (-\infty, x_n]\right) = P((-\infty, b]) = F(b).$$

$F$  ist also rechtsseitig stetig.

(ii) Für jede monoton fallende Folge  $(x_n)_n$ , mit  $\lim_{n \rightarrow \infty} x_n = -\infty$  gilt

$$\lim_{n \rightarrow \infty} F(x_n) = P\left(\bigcap_{n=1}^{\infty} (-\infty, x_n]\right) = P(\emptyset) = 0.$$

Es ist also  $\lim_{x \rightarrow -\infty} F(x) = 0$ .

(iii) Für jede monoton wachsende Folge  $(x_n)_n$ , mit  $\lim_{n \rightarrow \infty} x_n = \infty$  gilt

$$\lim_{n \rightarrow \infty} F(x_n) = P\left(\bigcup_{n=1}^{\infty} (-\infty, x_n]\right) = P((-\infty, \infty)) = 1.$$

Es ist also  $\lim_{x \rightarrow \infty} F(x) = 1$ .

**Anmerkung.** In der Maßtheorie zeigt man die Umkehrung zu Feststellung 2.8.2.

### 2.8.3 Satz (Maß zu einer Verteilungsfkt.)

Sei  $F : \mathbb{R} \rightarrow [0, 1]$  eine monoton wachsende, rechtsseitig stetige Funktion mit

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Dann gibt es genau eine W-Maß  $P$  auf der Borel-Algebra  $\mathcal{B}(\mathbb{R})$ , so daß  $F$  die Verteilungsfunktion von  $P$  ist:

$$P((-\infty, x]) = F(x) \quad \text{für } x \in \mathbb{R}.$$

**Anmerkung. (Zum Bew. von Satz 2.8.3)** 1. Wir konstruieren das gesuchte Maß  $P$  als die Verteilung einer Zufallsvariablen

$$X : (0, 1) \rightarrow \mathbb{R},$$

die man die **Quantilfunktion** von  $F$  nennt.

Dabei trägt  $(0, 1)$  die uniforme Verteilung  $\mathcal{U}$  ( $\rightarrow$  Beispiel 2.5.7). und  $P$  ist das Bildmaß  $\mathcal{U}_X$ . Die Abbildung  $X$  ist eine Art „Umkehrfunktion“ zu  $F$ . Die genaue Konstruktion, die noch weitere Anwendungen findet, formulieren wir in Feststellung 2.8.4.

2. Wenn  $F : \mathbb{R} \rightarrow (0, 1)$  stetig und streng monoton wachsend ist, und

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

gilt, so folgt aus dem Zwischenwertsatz, daß das Bild  $F(\mathbb{R}) = (0, 1)$  ist.  $F$  ist also bijektiv und hat eine stetige Umkehrfunktion  $X : (0, 1) \rightarrow \mathbb{R}$ , die die gewünschte Gleichung (2.8.2) erfüllt.

3. Wenn  $F$  nur monoton wachsend ist oder Sprungstellen hat, definiert man eine verallgemeinerte Umkehrfunktion, die sogenannte Quantilfunktion:

### 2.8.4 Festst. (Quantilfunktion)

Sei  $F : \mathbb{R} \rightarrow [0, 1]$  eine monoton wachsende, rechtsseitig stetige Funktion mit

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

(i) Die Quantilfunktion (Quantil-Transformation)  $X$  zu  $F$  wird definiert durch

$$X(t) := \min\{x \in \mathbb{R} \mid F(x) \geq t\} \quad \text{für } t \in (0, 1). \quad (2.8.1)$$

Man beachte, da  $F$  rechtsseitig stetig ist, wird das Minimum in Gleichung (2.8.1) wirklich angenommen. Man zeichne ein Bild, das zeigt

- was in einer Sprungstelle von  $F$  passiert,
- was passiert, wenn  $F$  auf einem Teilintervall konstant ist

und wie  $X$  jeweils aussieht.

(ii) Nach Konstruktion der Quantilfunktion  $X$  gilt für  $t \in (0, 1)$ ,  $x \in \mathbb{R}$

$$X(t) \leq x \quad \Leftrightarrow \quad t \leq F(x) \\ X^{-1}((-\infty, x]) = (0, F(x)] \in \mathcal{B}((0, 1)), \quad (2.8.2)$$

Aus Gleichung (2.8.2) folgt, daß  $X : ((0, 1), \mathcal{B}((0, 1))) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  meßbar ist.

**Anmerkung.** Aus Gleichung (2.8.2) folgt für das Bildmaß der Quantilfunktion  $X$ :

### 2.8.5 Folg. (Verteilung der Quantilfunktion)

Die Quantilfunktion  $X$  ist eine reelle Zufallsvariable auf dem  $W$ -Raum  $(0, 1)$  mit der uniformen Verteilung  $\mathcal{U}$  und es gilt

$$\mathcal{U}(X^{-1}(-\infty, x]) = \mathcal{U}((0, F(x)]) = F(x).$$

D.h.  $F$  ist die Verteilungsfunktion ihrer Quantilfunktion  $X$ .

**Anmerkung.** Da jedes  $W$ -Maß  $P$  auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  durch seine Verteilungsfunktion  $F$  eindeutig bestimmt ist ( $\rightarrow$  Definition 2.8.1), kann man die Folgerung 2.8.5 auch so aussprechen:

### 2.8.6 Folg.

Jedes  $W$ -Maß  $P$  auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  ist Bildmaß einer reellen Zufallsvariablen auf dem Wahrscheinlichkeitsraum  $((0, 1), \mathcal{B}(0, 1), \mathcal{U}_{(0,1)})$ .

**Anmerkung.** Da viele Programmiersprachen eine Funktion (random) bieten, die eine uniform auf  $(0, 1)$  verteilte Zufallsvariable nachbilden, ermöglicht die Konstruktion in Folgerung 2.8.6 die Simulation beliebiger reellwertiger Zufallsprozesse.

**Anmerkung. (Quartile und Median)** Mit der Quantilfunktion bestimmt man die sogenannten Quantile eines  $W$ -Maßes  $P$ , bzw einer Verteilungsfunktion  $F$ , bzw einer Zufallsvariablen  $X$ . Eine Zahl  $q \in \mathbb{R}$  heißt ein  $\alpha$ -Quantil,  $0 < \alpha < 1$ , von  $P$ , wenn

$$P((-\infty, q]) \geq \alpha \quad \text{und} \quad P([q, \infty)) \geq 1 - \alpha.$$

Ein  $\frac{1}{2}$ -Quantil  $q_{1/2}$  heißt ein **Median**. Ein Median ist ein Punkt in  $\mathbb{R}$ , an dem die Verteilungsfunktion das Niveau  $1/2$  gerade übersteigt oder überspringt. Im Fall einer Sprungstelle zählt dieser Punkt sowohl bei der linken wie bei der rechten Hälfte mit. Wenn die Verteilungsfunktion in diesem Punkt nicht strikt monoton wächst, ist der Median nicht eindeutig bestimmt.

Ist  $P$  die Verteilung einer reellen Zufallsvariablen  $X$ , so besagt „ $q$  ist ein  $\alpha$ -Quantil von  $X$ “, daß die Wahrscheinlichkeit für Beobachtungen von  $X$ ,

die  $\leq q$  sind, mindestens  $\alpha$  ist,

die  $\geq q$  sind, mindestens  $(1 - \alpha)$  ist.

Für eine reelle Zufallsvariable  $X$  auf einem Laplaceraum teilt das  $1/4$ -Quantil  $q_{1/4}$  grob gesprochen die Anzahl der Beobachtungen von  $X$  in einer langen Beobachtungsreihe im Verhältnis  $1/4$  zu  $3/4$  auf.

## 2.9 Dichtefunktion, stetige Verteilungen

### Anmerkung. (Lebesgue-Maß und Lebesgue-Integral)

1. Auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  existiert eindeutig ein Maß  $\lambda$  so, daß das Maß  $\lambda(I)$  eines jeden beschränkten Intervalls  $I$  gleich seiner Länge ist. Dieses Maß  $\lambda$  heißt das **Lebesgue-Maß** auf den Borel-Mengen  $\mathcal{B}(\mathbb{R})$ .

Definiert wird das Lebesgue-Maß auf  $\mathcal{B}(\mathbb{R})$  wie folgt: Ist  $G \subset \mathbb{R}$  offen, so ist  $G$  Vereinigung von abzählbar vielen offenen Intervallen  $G = \bigcup_{n=1}^{\infty} (a_n, b_n)$ , wobei  $-\infty \leq a_n < b_n \leq \infty$  gilt. Man setzt dann

$$\lambda(G) = \begin{cases} \sum_{n=1}^{\infty} (b_n - a_n) & \text{wenn alle } a_n, b_n \in \mathbb{R}, \\ \infty & \text{sonst.} \end{cases}$$

Für eine beliebige Menge  $A \in \mathcal{B}(\mathbb{R})$  betrachtet man alle offenen Mengen  $G$ , die  $A$  umfassen und setzt dann

$$\lambda(A) = \inf_{A \subset G} \lambda(G).$$

Nun kann man zeigen, daß so ein eindeutiges, *wohldefiniertes* Maß  $\lambda$  auf  $\mathcal{B}(\mathbb{R})$  erklärt wird, und, daß für Intervalle  $I$  das Lebesgue-Maß  $\lambda(I)$  die Länge des Intervalls ist.

2. Analog gibt es auf dem  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  eindeutig das  **$d$ -dimensionale Lebesgue-Maß**  $\lambda^d$ , welches für Quader und andere elementare Körper den elementaren Inhalt angibt.

3. Da  $\lambda(\mathbb{R}) = \infty$  ist, kann man durch Normierung aus  $\mathbb{R}$  keinen W-Raum machen. Um weitere Beispiele zu gewinnen, gehen wir folgendermaßen vor: Wenn  $f : \mathbb{R} \rightarrow [0, \infty)$  eine *integrierbare*, nichtnegative Funktion mit

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

ist, so setze man für Intervalle  $(a, b]$ ,  $-\infty \leq a < b < \infty$ ,

$$P_f((a, b]) = \int_a^b f(x) dx. \quad (**)$$

Dies erzeugt dann ein eindeutig bestimmtes W-Maß  $P_f$  auf den Borel-Mengen  $\mathcal{B}(\mathbb{R})$ . Man nennt  $f$  die **Dichtefunktion** dieses Maßes  $P_f$ .

4. Wir müssen noch ein paar Worte darüber verlieren, was *integrierbare* Funktion bedeutet und was das Integral in der Gleichung (\*) bzw. (\*\*) eigentlich ist. Für eine tragkräftige Theorie braucht man hier das **Lebesgue-Integral**, Dieses ist eine Erweiterung des Riemann-Integrals mit vielen schönen und praktischen Eigenschaften, die man bei dem enger gefaßten Riemann-Integral noch nicht findet.

In vielen praktischen Fällen ist die Funktion  $f$  stetig oder stückweise stetig, so daß man die Integrale (\*) und (\*\*) mit Hilfe einer Stammfunktion  $F$  bilden kann:

$$P_f((a, b]) = \int_a^b f(x) dx = F(b) - F(a) \quad (**')$$

$$P_f(\mathbb{R}) = \int_{-\infty}^{\infty} f(x) dx = \lim_{x \rightarrow \infty} F(x) - \lim_{y \rightarrow -\infty} F(y) \quad (*')$$

Aber auch in diesem einfachen Fall ist das Lebesgue-Integral, das natürlich den selben Wert hat, ein nützlicher Helfer.

5. Wir geben die allgemeine Definition einer Dichtefunktion, beschränken uns in den Beispielen dann aber auf den stückweise stetigen Fall.

### 2.9.1 Def. (Dichtefunktion, stetiges W-Maß)

Eine integrierbare, nicht-negative Funktion  $f : \mathbb{R} \rightarrow [0, \infty)$  heißt **Dichtefunktion eines Wahrscheinlichkeitsmaßes**  $P$  auf  $\mathcal{B}(\mathbb{R})$ , wenn für alle  $a, b$  mit  $-\infty \leq a < b < \infty$  gilt

$$P((a, b]) = \int_a^b f(x) dx. \quad (2.9.1)$$

Man nennt  $f$  auch **Wahrscheinlichkeitsdichte** oder kurz **Dichte** von  $P$ .

Ein Wahrscheinlichkeitsmaß auf  $\mathcal{B}(\mathbb{R})$  mit Dichtefunktion heißt ein **stetiges W-Maß**.

### Anmerkung.

1. Wir benötigen im folgenden nur den Fall, daß  $f$  stetig oder stückweise ist.

2. Wenn ein W-Maß  $P$  auf  $\mathcal{B}(\mathbb{R})$  eine Dichtefunktion  $f$  hat, so ist diese nicht ganz eindeutig bestimmt. So kann man  $f$  in endlich vielen Punkten abändern, ohne den Wert des Integrals (2.9.1) zu verändern.

Man kann zeigen, sind  $f_1$  und  $f_2$  Dichtefunktionen von  $P$ , so ist die Menge  $\{f_1 \neq f_2\}$  eine Lebesgue-Nullmenge:

$$\lambda\{f_1 \neq f_2\} = 0.$$

Dann ist auch  $P\{f_1 \neq f_2\} = 0$ .

### 2.9.2 Festst. (Verteilungs- und Dichtefunktion)

Es sei  $P$  ein W-Maß auf  $\mathbb{R}$  mit Verteilungsfunktion  $F$ .

1.  $f$  ist genau dann die Dichtefunktion von  $P$ , wenn

$$F(x) = \int_{-\infty}^x f(\xi) d\xi \quad \text{für } x \in \mathbb{R} \quad (2.9.2)$$

gilt.

2. Man kann den Begriff der Ableitung einer Funktion dahingehend verallgemeinern, daß man die Gleichung (2.9.2) als  $F' = f$  schreiben kann. Wir wollen diese Verallgemeinerung nicht weiter verfolgen. Aus dem Hauptsatz der Differential und Integralrechnung folgt:

Ist die Verteilungsfunktion  $F$  bis auf endlich viele Ausnahmepunkte stetig differenzierbar und gilt  $F' = f$ , so ist  $f$  die Dichtefunktion von  $P$ .

### 2.9.3 Bsp. (Dichtfunkt. der uniform. Verteilung.)

1. Ein W-Maße  $P$  auf einem Intervall  $I$  kann man als W-Maße auf ganz  $\mathbb{R}$  auffassen. Man setze

$$P(A) := P(A \cap I) \quad \text{für } A \in \mathcal{B}(\mathbb{R}).$$

Das so fortgesetzte W-Maß  $P$  ist auf  $I$  konzentriert und die bedingte Wahrscheinlichkeit  $P(\cdot|I) = P$ .

2. Faßt man die uniforme Verteilung  $\mathcal{U}_{[0,1]}$  im obigen Sinne als W-Maß auf  $\mathbb{R}$  auf, so hat  $\mathcal{U}_{[0,1]}$  die Verteilungsfunktion

$$F(x) := \begin{cases} 0 & \text{für } x < 0, \\ x & \text{für } 0 \leq x \leq 1, \\ 1 & \text{für } x > 1. \end{cases}$$

und die Dichtefunktion

$$f(x) := \begin{cases} 0 & \text{für } x < 0, \\ 1 & \text{für } 0 \leq x \leq 1, \\ 0 & \text{für } x > 1. \end{cases}$$

Da es auf einzelne Werte der Dichtefunktion nicht ankommt, kann man auch  $f$  in den Punkten 0 und 1 auf den Wert 0 setzen. D.h.  $f = \mathbb{1}_{[0,1]}$  oder  $f = \mathbb{1}_{(0,1)}$  usw.

### 2.9.4 Bem. (Dichtefunktion auf $\mathbb{R}^d$ )

Es sei  $P$  ein W-Maß auf  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  mit Verteilungsfunktion

$$F(x_1, \dots, x_d) = P((-\infty, x_1] \times \dots \times (-\infty, x_d])$$

für  $x = (x_i)_{i=1}^d \in \mathbb{R}^d$ . Eine integrierbare, nichtnegative Funktion  $f : \mathbb{R}^d \rightarrow [0, \infty)$  ist die Dichtefunktion des W-Maßes  $P$ , wenn für alle  $x = (x_i)_{i=1}^d \in \mathbb{R}^d$  gilt

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(\xi_1, \dots, \xi_d) d\xi_1 \cdots d\xi_d.$$

Dann gilt für alle Quader  $A = (a_1, b_1] \times \cdots \times (a_d, b_d]$

$$P(A) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(\xi_1, \dots, \xi_d) d\xi_1 \cdots d\xi_d.$$

## 2.10 Exponentialverteilung

### 2.10.1 Bez. (Exponentialverteilung)

Die Exponentialverteilung ist das kontinuierliche Analogon zur geometrischen Verteilung ( $\rightarrow$  Beispiel 2.1.3).

(i) Die Exponentialverteilung mit dem Parameter  $\lambda > 0$  hat die Dichtefunktion

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0 \end{cases} \\ = \lambda e^{-\lambda x} \mathbb{1}_{(0, \infty)}(x).$$

und die Verteilungsfunktion

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0 \end{cases} \\ = (1 - e^{-\lambda x}) \mathbb{1}_{(0, \infty)}(x).$$

Für die Quantilfunktion erhält man

$$X(t) := -\frac{\log(1-t)}{\lambda} \quad \text{für } t \in (0, 1).$$

(ii) Die Bedeutung des Parameters  $\lambda > 0$  erklärt sich aus dem  $\rightarrow$  Erwartungswert und der  $\rightarrow$  Varianz einer exponentialverteilten Zufallsvariablen  $X$ :

$$E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda},$$

$$\text{Var}(X) = E(X^2) - (EX)^2 = \frac{1}{\lambda^2}.$$

**Anmerkung.** In den folgenden beiden Bemerkungen diskutieren wir Zufallsexperimente, in denen die Exponentialverteilung typischerweise auftritt.

### 2.10.2 Bem. (geom. Vert. $\rightarrow$ exp. Vert.)

Um den Zusammenhang zur geometrischen Verteilung ( $\rightarrow$  Beispiel 2.1.3) herzustellen, sei  $(X_n)_n$  eine Folge geometrisch verteilter Zufallsvariablen mit den Parametern  $p_n$ . Wir setzen voraus, daß  $\lim_{n \rightarrow \infty} np_n = \lambda$  ist. Der Grenzwert der Erwartungswerte der Variablen  $X_n/n$  ist dann ( $\rightarrow$  Feststellung 2.2.5)

$$E\left(\frac{X_n}{n}\right) = \frac{1}{np_n} - \frac{1}{n} \rightarrow \frac{1}{\lambda}.$$

Für  $x > 0$  betrachten wir die Wahrscheinlichkeit der Menge

$$\left\{ \frac{X_n}{n} \leq x \right\} = \{X_n \leq nx\}$$

Da die geometrische Verteilung auf  $\mathbb{N}_0$  konzentriert ist, haben  $\{X_n \leq nx\}$  und  $\{X_n \leq \lfloor nx \rfloor\}$  die gleiche Wahrscheinlichkeit, wobei  $\lfloor nx \rfloor$  die größte ganze Zahl  $\leq nx$  ist. Für die Verteilungsfunktion von  $X_n/n$  folgt nun

$$\begin{aligned} P_{X_n/n}((-\infty, x]) &= P_{X_n}((-\infty, \lfloor nx \rfloor]) \\ &= \sum_{k=0}^{\lfloor nx \rfloor} (1-p_n)^k p_n \\ &= 1 - (1-p_n)^{\lfloor nx \rfloor + 1} \\ &= 1 - \left(1 - \frac{np_n}{n}\right)^{\lfloor nx \rfloor + 1}. \end{aligned}$$

Für  $n \rightarrow \infty$  konvergiert dies gegen  $1 - e^{-\lambda x}$  ( $\rightarrow$  Bemerkung 2.4.4) Somit ist  $X_n/n$  für große  $n$  annähernd exponentialverteilt mit dem Parameter  $\lambda = \lim_{n \rightarrow \infty} np_n$ .

**Anmerkung. (Interpretation der Rechnung in Bem. 2.10.2)** Die geometrische Verteilung  $\mathcal{G}_p\{k\}$ ,  $k \in \mathbb{N}_0$ , gibt die Wahrscheinlichkeit an, daß in einem oft genug wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit  $p$  der erste Erfolg nach  $k$  Mißerfolgen eintritt. Man stelle sich vor, daß die einzelnen Bernoulli-Experimente in einem festen Takt mit Taktzeit 1 durchgeführt werden. Verkürzt man die Taktzeit auf  $1/n$ ,  $n \in \mathbb{N}$ , so sei die Erfolgswahrscheinlichkeit  $p_n$ . Die Zufallsvariable  $X_n$  gebe an, wieviele Takte der Länge  $1/n$  vergangen sind, bis der erste Erfolg eintritt. Die Zufallsvariable  $X_n/n$  gibt grob gesagt die Zeit an, die verstrichen ist, bis der erste Erfolg eintritt.

Die getroffene Voraussetzung  $np_n \rightarrow \lambda$  besagt, daß der Erwartungswert für die Zeit bis zum ersten Erfolg gegen  $1/\lambda$  konvergiert. Die Rechnung zeigt, daß dann die Verteilungsfunktion von  $X_n/n$  gegen die Verteilungsfunktion der Exponentialverteilung zum Parameter  $\lambda$  konvergiert.

### 2.10.3 Bem.

Wir erinnern an das Beispiel 2.4.5 der Anzahl der Schadensmeldungen, die bei einer Versicherung nach dem Zeitpunkt 0 bis einschließlich zum Zeitpunkt  $t > 0$  eingehen. Wenn  $\lambda > 0$  die durchschnittliche Zahl der Schadensmeldungen in einem Zeitraum der Länge 1 ist, so ist die Anzahl der Schadensmeldungen im Zeitraum  $(0, t]$  Poisson-verteilt mit dem Parameter  $\lambda t$ .

Wir fragen nun nach der Wahrscheinlichkeit  $P((0, t])$ , daß im Zeitraum  $(0, t]$ , ( $t > 0$ ), mindestens ein Schaden gemeldet wurde:

$$P((0, t]) = \sum_{k=1}^{\infty} \mathcal{P}_{\lambda t}\{k\} = 1 - \mathcal{P}_{\lambda t}\{0\} = 1 - e^{-\lambda t}.$$

Da wir erst die Schäden nach dem Zeitpunkt 0 registrieren, ist  $P((-\infty, 0]) = 0$ . Die Wahrscheinlichkeit für mindestens einen Schadensfall bis zum Zeitpunkt  $t$  ist also exponentialverteilt mit Parameter  $\lambda$ :

$$P((-\infty, t]) = \begin{cases} 1 - e^{-\lambda t} & (t > 0), \\ 0 & (t \leq 0). \end{cases}$$

**Anmerkung.** Wir halten fest, bei einem Experiment mit rein zufällig auftretenden Zeitpunkten, ist die Anzahl der Zeitpunkte im Zeitintervall  $(0, t]$  Poisson-verteilt mit Parameter  $\lambda t$  und die Wartezeit auf den ersten Zeitpunkt exponentialverteilt mit Parameter  $\lambda$ . Dabei ist  $\lambda$  die durchschnittliche Anzahl der Zeitpunkte pro Zeiteinheit und  $1/\lambda$  die durchschnittliche Wartezeit.

Eine exakte Untersuchung findet man in den Lehrbüchern zur Wahrscheinlichkeitstheorie unter dem Stichwort **Poisson-Prozess**.

### 2.11 Normalverteilung

**Anmerkung. (Formeln zur Normalverteilung)** Die Funktion

$$\varphi : x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{für } x \in \mathbb{R} \quad (2.11.1)$$

ist positiv und symmetrisch um den Nullpunkt. Die Funktion  $\varphi$  fällt sehr schnell, so daß für alle Potenzen  $k \in \mathbb{N}_0$

$$\lim_{x \rightarrow \pm\infty} x^k \varphi(x) = 0 \quad (2.11.2)$$

ist. Man kann zeigen, daß diese Funktion keine Stammfunktion innerhalb der elementaren Funktionen ( $x^y, \log x, e^x, \sin x, \cos x, \dots$ ) hat. Man bezeichnet die Stammfunktion mit

$$\Phi : x \mapsto \int_{-\infty}^x e^{-\xi^2/2} d\xi \quad \text{für } x \in \mathbb{R}. \quad (2.11.3)$$

und nennt sie die **Gaußsche Fehlerfunktion** oder wegen der Form ihres Graphens auch die **Gaußsche Glockenfunktion**. Dabei ist die Integrationskonstante so gewählt, daß

$$\lim_{x \rightarrow -\infty} \Phi(x) = 0 \quad (2.11.4)$$

ist. Der Faktor  $1/\sqrt{2\pi}$  in der Definition von  $\varphi$  ist so gewählt, daß

$$\lim_{x \rightarrow \infty} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\xi^2/2} d\xi = 1 \quad (2.11.5)$$

wird.  $\Phi$  ist stetig differenzierbar mit Ableitung  $\Phi' = \varphi$ . Da  $\varphi > 0$  ist, ist

$$\Phi \text{ strikt monoton wachsend.} \quad (2.11.6)$$

Die Gleichung (2.11.4), (2.11.4) und (2.11.6) besagen, daß  $\Phi$  Verteilungsfunktion eines W-Maßes auf  $\mathbb{R}$  ist ( $\rightarrow$  Satz 2.8.3).

Da  $\varphi$  symmetrisch um den Nullpunkt ist, ist der Mittelwert dieses W-Maßes

$$\int_{-\infty}^{\infty} x\varphi(x) dx = 0. \quad (2.11.7)$$

Es ist  $\frac{d}{dx}\varphi(x) = -x\varphi(x)$ . Durch partielle Integration erhält man

$$\int_{-\infty}^{\infty} x^2\varphi(x) dx = -\left\{x \cdot \varphi(x)\right\}_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \varphi(x) dx = 1 \quad (2.11.8)$$

und somit ist die Varianz = 1.

Einige Werte von  $\Phi$

$x$	0	0,67	1,00	1,28	1,64	1,96	2,33	3,008
$\Phi(x)$	0,5	0,75	0,84	0,90	0,95	0,975	0,99	0,999

(2.11.9)

Die Werte für  $x < 0$  erhält man aus der Symmetrie  $\Phi(-x) = 1 - \Phi(x)$ .

#### 2.11.1 Bez. (Standard-Normalverteilung)

Das Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mit der Dichtefunktion

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{für } x \in \mathbb{R}$$

und der Verteilungsfunktion

$$\Phi : x \mapsto \int_{-\infty}^x e^{-\xi^2/2} d\xi \quad \text{für } x \in \mathbb{R}.$$

heißt die Standard-Normalverteilung auf  $\mathbb{R}$ .

Der Erwartungswert der Standard-Normalverteilung ist 0 und die Varianz ist 1. Daher wird die Standard-Normalverteilung mit  $\mathcal{N}_{0,1}$  bezeichnet.

**Anmerkung.** Durch Substitution

$$x = \frac{\xi - \mu}{\sigma} \quad \text{und} \quad dx = \frac{1}{\sigma} d\xi$$

folgt aus den Gleichungen (2.11.5) – (2.11.8)

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\sigma} \varphi\left(\frac{\xi - \mu}{\sigma}\right) d\xi &= 1, \\ \int_{-\infty}^{\infty} \xi \frac{1}{\sigma} \varphi\left(\frac{\xi - \mu}{\sigma}\right) d\xi &= \mu \\ \int_{-\infty}^{\infty} \xi^2 \frac{1}{\sigma} \varphi\left(\frac{\xi - \mu}{\sigma}\right) d\xi &= \sigma^2 + \mu^2. \end{aligned}$$

Also ist  $x \frac{1}{\sigma} \mapsto \varphi\left(\frac{\xi - \mu}{\sigma}\right)$  die Dichtefunktion eines W-Maßes mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ .

Für die zugehörige Verteilungsfunktion erhält man mit der Substitution  $\xi = \sigma t + \mu$

$$\int_{-\infty}^x \frac{1}{\sigma} \varphi\left(\frac{\xi - \mu}{\sigma}\right) d\xi = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(t) dt = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (2.11.10)$$

Durch Verschiebung und Skalierung erhält man aus der Standard-Normalverteilung eine ganze Familie von W-Maßen.

#### 2.11.2 Bez. (Normalverteilungen $\mathcal{N}_{\mu, \sigma^2}$ )

Das W-Maß auf  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mit der Dichtefunktion

$$\frac{1}{\sigma} \varphi\left(\frac{\xi - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

heißt die Normalverteilung mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ . Es wird mit  $\mathcal{N}_{\mu, \sigma^2}$  bezeichnet. Die zugehörige Verteilungsfunktion ist

$$\int_{-\infty}^x \frac{1}{\sigma} \varphi\left(\frac{\xi - \mu}{\sigma}\right) d\xi = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

wobei  $\Phi$  die Verteilungsfunktion von  $\mathcal{N}_{0,1}$  ist.

#### 2.11.3 Festst. (affine Transform. normalvertl. ZV)

(i) Es sei  $X$  eine Zufallsvariable mit der Verteilung  $\mathcal{N}_{\mu, \sigma^2}$ . Für  $\alpha, \beta \in \mathbb{R}, \alpha \neq 0$  hat dann  $Y := \alpha X + \beta$  die Verteilung  $\mathcal{N}_{\alpha\mu + \beta, (\alpha\sigma)^2}$ .

(ii) Es seien  $\alpha, \beta \in \mathbb{R}, \alpha > 0$ . Dann gilt für  $x \in \mathbb{R}$

$$N_{\mu, \sigma^2} \left\{-\infty, \frac{x - \beta}{\alpha}\right\} = N_{\alpha\mu + \beta, (\alpha\sigma)^2} \{-\infty, x\}.$$

(iii) Wenn  $X$  normalverteilt mit den Parametern  $\mu$  und  $\sigma^2$  ist, dann ist die zentrierte normalisierte Zufallsvariable  $X^* = (X - \mu)/\sigma$  standard-normalverteilt ( $\rightarrow$  Bezeichnung 2.12.1)

(iv) Für  $\mu, \sigma \in \mathbb{R}, \sigma > 0$

$$N_{0,1} \{-\infty, \frac{x - \mu}{\sigma}\} = N_{\mu, \sigma^2} \{-\infty, x\}.$$

**Beweis.** (i) Klar ist, daß

$$E(\alpha X + \beta) = \alpha E(X) + \beta,$$

$$\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$$

ist. Zu zeigen ist, daß  $Y$  normalverteilt ist. Es sei zunächst  $\alpha > 0$

Da

$$\{Y \leq y\} = \{\alpha X + \beta \leq y\} = \left\{X \leq \frac{y - \beta}{\alpha}\right\}$$

hat die Verteilungsfunktion  $F_Y$  die Gestalt

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{\frac{y-\beta}{\alpha}} \frac{1}{\sigma} \varphi\left(\frac{\xi - \mu}{\sigma}\right) d\xi \\ &= \int_{-\infty}^y \frac{1}{\alpha\sigma} \varphi\left(\frac{\eta - (\alpha\mu + \beta)}{\alpha\sigma}\right) d\eta \end{aligned}$$

wobei  $\xi = \frac{\eta - \beta}{\alpha}$  substituiert wurde. Die Dichtefunktion von  $F_Y$  ist also

$$y \mapsto \frac{1}{\alpha\sigma} \varphi\left(\frac{y - (\alpha\mu + \beta)}{\alpha\sigma}\right)$$

und folglich ist  $Y$  normalverteilt mit Erwartungswert  $\alpha\mu + \beta$  und Varianz  $(\alpha\sigma)^2$ .

Die Aussagen (ii), (iii) und (iv) folgen nun aus (i)

## 2.12 Grenzwertsatz von Moivre-Laplace

### 2.12.1 Bez. (normierte zentrierte ZV)

Um Zufallsvariable besser vergleichen zu können verschiebt und skaliert man ihre Werte so, daß der Erwartungswert zu 0 und die Varianz zu 1 wird:

**zentrieren:** Zu einer Zufallsvariablen  $X$  mit Erwartungswert  $\mu$  bilde man  $Y = X - \mu$ . Dann ist  $E(Y) = 0$ .

**normieren:** Zu einer Zufallsvariablen  $Y$  mit Varianz  $\sigma^2$  bilde man  $Z = Y/\sigma$ . Dann ist  $\text{Var}(Z) = 1$ .

**normieren-zentrieren** Zu einer Zufallsvariablen  $X$  mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$  bilde man

$$X^* := \frac{X - \mu}{\sigma}.$$

$X^*$  ist dann zentriert und normiert:

$$E(X^*) = 0, \quad \text{Var}(X^*) = 1.$$

### 2.12.2 Satz (Moivre-Laplace)

Es sei  $(X_k)_k$  eine Folge binomialverteilter Zufallsvariablen mit dem jeweiligen Parameter  $n_k \in \mathbb{N}$ ,  $p_k \in (0, 1)$ . Wenn

$$\lim_{k \rightarrow \infty} \text{Var}(X_k) = \infty$$

ist, dann konvergieren die Verteilungsfunktionen  $F_{X_k^*}$  der normierten und zentrierten Zufallsvariablen

$$X_k^* := \frac{X_k - E(X_k)}{\sqrt{\text{Var}(X_k)}} = \frac{X_k - n_k p_k}{\sqrt{n_k p_k (1 - p_k)}}$$

gegen die Standard-Normalverteilung. Die Konvergenz ist gleichmäßig auf  $\mathbb{R}$ :

$$\lim_{k \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_{X_k^*}(x) - \Phi(x)| = 0.$$

### 2.12.3 Folg. (Moivre-Laplace)

Es seien  $0 < p < 1$  und  $(S_n)_n$  eine Folge  $\mathcal{B}_{n,p}$ -verteilter Zufallsvariablen. Dann gilt

$$\lim_{n \rightarrow \infty} \sup_{a < b} \left| P\{a < S_n^* \leq b\} - \Phi(b) - \Phi(a) \right| = 0.$$

**Anmerkung. (zum Satz von Moivre-Laplace)** 1. Der Satz von Moivre-Laplace ist ein Spezialfall des zentralen Grenzwertsatzes.

2. Man benötigt nur die Voraussetzung, daß die  $X_k$  binomialverteilt sind und ihre Varianzen

$$\text{Var}(X_k) = n_k p_k (1 - p_k) \rightarrow \infty \quad \text{für } k \rightarrow \infty.$$

Dann konvergieren auch die Erwartungswerte

$$E(X_k) = n_k p_k \geq n_k p_k (1 - p_k) \rightarrow \infty$$

und die Anzahlen  $n_k \rightarrow \infty$ . Die  $p_k$  müssen nur  $0 < p_k < 1$  erfüllen, ansonsten ist ihr Verhalten für  $k \rightarrow \infty$  unwichtig.

3. Man beachte, daß die Verteilungen  $F_{X_k^*}$  konvergieren. Es wird nichts über die Konvergenz der Zufallsvariablen  $X_k^*$  gesagt. Letzteres macht auch keinen Sinn, denn die  $X_k$  haben zwar alle ihre Werte in  $\mathbb{N}_0 \subset \mathbb{R}$ , sie können aber ganz unterschiedliche Definitionsbereiche haben:  $X_k : (\Omega_k, P_k) \rightarrow \mathbb{N}_0$ .

4. Der Satz sagt etwas über die Konvergenz der Bildmaße  $(P_k)_{X_k^*}$  aus. Die Zufallsvariablen sind eigentlich überflüssig, machen aber die Formulierung anschaulicher.

Man kann den Satz von Moivre-Laplace auch als Grenzwertsatz für Folgen  $(\mathcal{B}_{n_k, p_k})_k$  von Binomialverteilungen formulieren. Bei dem Beweis eines vorbereitenden Hilfssatzes, des lokalen Grenzwertsatzes, werden wir diesen Standpunkt auch einnehmen.

### 2.12.4 Folg. (Approximation der BinomialV.)

Es sei  $0 < p < 1$ . Dann gilt für  $n \rightarrow \infty$

$$\sup_{t \in \mathbb{R}} \left| \mathcal{B}_{n,p}((-\infty, t]) - \mathcal{N}_{np, np(1-p)}((-\infty, t]) \right| \rightarrow 0.$$

Also ist für  $k, l \in \mathbb{N}_0$

$$\mathcal{B}_{n,p}((-\infty, k]) \approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right)$$

und

$$\mathcal{B}_{n,p}\{k, \dots, l\} = \Phi\left(\frac{l - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right).$$

Die Approximation wird noch etwas besser, wenn man die sogenannte *Stetigkeitskorrektur* vornimmt:

$$\mathcal{B}_{n,p}((-\infty, k]) \approx \Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

und

$$\mathcal{B}_{n,p}\{k, \dots, l\} = \Phi\left(\frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

Mit heutigen Computeralgebraprogrammen kann man  $\mathcal{B}_{n,p}$  auch für große  $n$  recht genau berechnen. Für formelmäßige Rechnungen ist aber die Näherung durch die Gaußsche Fehlerfunktion überlegen.

**Anmerkung.** Das folgende Bsp. ist dem Lehrbuch [1] von Krenzel entnommen:

### 2.12.5 Bsp. (Bestimmung: Stichprobenumfang)

Wir wollen den Prozentsatz der Wähler einer Partei  $A$  schätzen. Werden  $n$  Wähler befragt und sind darunter  $S_n$  Wähler der Partei  $A$ , so sei  $S_n/n$  der Schätzer für die Wahrscheinlichkeit  $p$ , daß ein zufällig ausgewählter Wähler für die Partei  $A$  stimmt.

Wie groß muß  $n$  sein, damit die Wahrscheinlichkeit eines Irrtums in der Schätzung von  $p$  von mehr als 1% kleiner als 0,05 ist?.

Es soll also gelten

$$P\left\{-0,01 \leq \frac{S_n}{n} - p \leq 0,01\right\} \approx 0,95.$$

Wir nehmen an, daß die Befragungen unabhängig sind und somit die  $S_n$   $\mathcal{B}_{n,p}$  verteilt sind.

Mit  $\mu_n = np$  und  $\sigma_n = \sqrt{np(1-p)}$  ergibt sich mit Folgerung 2.12.3

$$\begin{aligned} 0,95 &\approx P\left\{-\frac{0,01n}{\sigma_n} \leq S_n^* \leq \frac{0,01n}{\sigma_n}\right\} \\ &\approx \Phi\left(\frac{0,01n}{\sigma_n}\right) - \Phi\left(-\frac{0,01n}{\sigma_n}\right) \\ &= 2\Phi\left(\frac{0,01n}{\sigma_n}\right) - 1. \end{aligned}$$

Also  $\Phi(0,01n/\sigma_n) \approx 0,975$ .

Aus der Tabelle (2.11.9) entnimmt man  $\Phi^{-1}(0,975) = 1,96$ . Also muß

$$\frac{0,01}{\sqrt{np(1-p)}} \approx 1,96 \quad \Leftrightarrow \quad n \approx p(1-p)10.000 \cdot 1,96^2$$

sein.

Nun ist  $p$  leider unbekannt. Da aber  $\max_{0 < p < 1} p(1-p) = 0,25$  ist kommt man in jedem Fall mit

$$n = 0,25 \cdot 10.000 \cdot 1,96^2 = 9600$$

Befragungen aus. Hat man von vorneherein die Information, daß  $p \leq 0,1$  ist, so ist  $\max_{0 < p < 0,1} p(1-p) = 0,1 \cdot 0,9 = 0,09$ . Man kommt nun mit

$$n = 0,09 \cdot 10.000 \cdot 1,96^2 = 3450$$

Befragungen aus.

## Literatur

### Lehrbücher

- [1] Krengel, U. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 6. Auflage 2002, Vieweg, ISBN 3-528-57259-0
- [2] Pfanzagl, J. *Elementare Wahrscheinlichkeitsrechnung*, 2. Auflage 1991, de Gruyter, ISBN 3-11-013384-9
- [3] Henze, N. *Stochastik für Einsteiger*, 4. Auflage, 2003, Vieweg, ISBN 3-528-36894-2

### Weitere Literatur

- [4] Dehling, H. und B. Haupt *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 2003, Springer, ISBN 3-540-43384-8
- [5] Dümbgen, L. *Stochastik für Informatiker*, 2003, Springer, ISBN 3-540-00061-5
- [6] Georgii, H.-O. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 2002, de Gruyter, ISBN 3-11-017235-6
- [7] Hinderer, K. *Grundbegriffe der Wahrscheinlichkeitstheorie*, 1972, Springer
- [8] Krickeberg, H. und H. Ziezold, *Stochastische Methoden*, 4. Auflage, 1994, Springer, ISBN 3-540-57792-0