

# Einführung in die Wahrscheinlichkeitstheorie und Statistik im SS2012 – Kurzschrift

Prof. Dr. C. Löh

Sommersemester 2012

---

## Inhaltsverzeichnis

-1	Literaturhinweise	2
0	Einführung	3
1	Das wahrscheinlichkeitstheoretische Modell – Wahrscheinlichkeitsräume und Zufallsvariablen	4
1.1	Wahrscheinlichkeitsräume	4
1.2	Zufallsvariablen	7
1.3	Verteilungsfunktionen	8
1.4	Exkurs: Integration auf Wahrscheinlichkeitsräumen	10
1.5	Erwartungswert und Varianz	16
1.6	Klassische Verteilungen	19
2	Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten	23
2.1	Stochastische Unabhängigkeit	23
2.2	Stochastische Unabhängigkeit und Produkte	24
2.3	Unkorreliertheit	28
2.4	Bedingte Wahrscheinlichkeiten	29
3	Gesetze der großen Zahlen und der zentrale Grenzwertsatz	32
3.1	Das schwache Gesetz der großen Zahlen	32
3.2	Null-/Eins-Gesetze	33
3.3	Das starke Gesetz der großen Zahlen	34
3.4	Der zentrale Grenzwertsatz	35
4	Einführung in die Schätz- und Testtheorie	38
4.1	Das statistische Modell	38
4.2	Schätzer	39
4.3	Alternativtestprobleme	42
4.4	Konfidenzbereiche	48

Version vom 27. Juli 2012

clara.loeh@mathematik.uni-regensburg.de

Fakultät für Mathematik, Universität Regensburg, 93040 Regensburg

## -1 Literaturhinweise

Die folgenden Listen enthalten eine kleine Auswahl an Literatur zur Wahrscheinlichkeitstheorie und Statistik.

### *Maßtheorie*

- [1] H. Bauer. *Maß- und Integrationstheorie*, De Gruyter, zweite Auflage, 1992.
- [2] J.L. Doob. *Measure Theory*, Springer, 1994.
- [3] T. Tao. *An Introduction to Measure Theory*, AMS, 2001.

### *Wahrscheinlichkeitstheorie und Statistik*

Manche dieser Bücher enthalten auch die nötigen Aspekte der Maßtheorie.

- [4] H.-O. Georgii. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*, De Gruyter, vierte Auflage, 2009.
- [5] A. Klenke. *Wahrscheinlichkeitstheorie*, Springer, zweite Auflage, 2008.
- [6] D. Meintrup, S. Schäffler. *Stochastik: Theorie und Anwendungen*, Springer, 2005.

... und viele weitere Bücher; je nach eigenen Vorlieben werden Ihnen manche Bücher besser gefallen als andere.

### *Weiterführende Literatur*

- [7] J. Havil. *Nonplussed!: Mathematical Proof of Implausible Ideas*, Princeton University Press, 2010.
- [8] D. Huff, I. Geis. *How to Lie with Statistics*. W W Norton & Co, 1993.
- [9] L. Gonick, W. Smith. *Cartoon Guide to Statistics*, Collins Reference, 1993.

## 0 Einführung

Diese Vorlesung gibt eine Einführung in die Wahrscheinlichkeitstheorie und Statistik; dabei handelt es sich um Gebiete der praktischen Mathematik, d.h. diese Gebiete bestehen jeweils aus dem Übergang von Situationen der Praxis zu einem geeigneten mathematischen Modell, einem stringenten mathematischen Anteil und der Interpretation der Aussagen über diese Modelle in der praktischen Situation.

**Caveat 0.1.** „Schlechte“ Modellbildung führt dazu, dass korrekte Resultate über das entsprechende mathematische Modell keine adäquate Aussage über das ursprüngliche Problem liefern.

- *Wahrscheinlichkeitstheorie*: Wahrscheinlichkeitstheorie ist die Lehre vom Zufall. Mathematisch basiert Wahrscheinlichkeitstheorie auf den Formalismen der Maßtheorie (jedoch mit einem etwas anderen Blickwinkel).

Typische Fragestellungen sind: Mit welcher „Wahrscheinlichkeit“ tritt ein gewisses „Ereignis“ ein? Welches von zwei gegebenen „Ereignissen“ ist „wahrscheinlicher“?

Anwendungen hat die Wahrscheinlichkeitstheorie zum Beispiel in folgenden Gebieten:

- Glücksspiel
- Finanzmathematik
- Quantenmechanik
- Reine Mathematik (probabilistische Methode, messbare Gruppentheorie)
- ...
- *Statistik*: In der Statistik werden Methoden zum Umgang mit Daten untersucht. Insbesondere wird in der *mathematischen Statistik* versucht, von Beobachtungen auf unterliegende Gesetzmäßigkeiten zu schließen. Die Grundlage dafür liefert die Wahrscheinlichkeitstheorie.

Die *deskriptive Statistik* hingegen befasst sich mit der Beschreibung und Visualisierung von Daten.

Typische Fragestellungen der mathematischen Statistik sind: Mit welcher „Sicherheit“ sind gegebene (empirische) Daten auf eine gewisse Gesetzmäßigkeit zurückzuführen? Mit welcher „Sicherheit“ kann man eine Hypothese durch gewisse Daten überprüfen?

Anwendungen hat die mathematische Statistik zum Beispiel in folgenden Gebieten:

- (experimentelle) Naturwissenschaften
- Medizin
- Vorhersagen aller Art
- ...

Wir werden uns zunächst mit dem wahrscheinlichkeitstheoretischen Modell vertraut machen und viele Beispiele von klassischen Verteilungen kennenlernen. Danach werden wir weitere wichtige Aspekte der Wahrscheinlichkeitstheorie wie stochastische Unabhängigkeit, bedingte Verteilungen und den zentralen Grenzwertsatz behandeln. Zum Schluss werden wir uns mit den Grundlagen der Schätz- und Testtheorie befassen.

# 1 Das wahrscheinlichkeitstheoretische Modell – Wahrscheinlichkeitsräume und Zufallsvariablen

Unser erstes Ziel ist es, „Zufall“ mathematisch mit Hilfe der Maßtheorie zu modellieren. Dies führt zum Begriff des Wahrscheinlichkeitsraums. Außerdem werden wir die gebräuchliche und nützliche Sprache der Zufallsvariablen einführen, Kenngrößen von Zufallsvariablen studieren und klassische Beispiele für Verteilungen betrachten.

## 1.1 Wahrscheinlichkeitsräume

Wie kann man ein Zufallsexperiment beschreiben?

- Was sind die möglichen Ergebnisse des Experiments?
- Was sind die „interessanten“ bzw. „sinnvollen“ Ereignisse des Experiments?
- Mit welcher „Wahrscheinlichkeit“ treten die Ereignisse ein?

Die Beschreibung von Zufallsexperimenten basierend auf diesen Fragen führt ganz natürlich zur Modellierung mit Hilfe von Wahrscheinlichkeitsräumen.

**Definition 1.1** (messbarer Raum,  $\sigma$ -Algebra). Ein *messbarer Raum* ist ein Paar  $(\Omega, S)$ , wobei  $\Omega$  eine Menge und  $S$  eine  $\sigma$ -Algebra auf  $\Omega$  ist, d.h.  $S \subset \text{Pot}(\Omega)$  erfüllt die folgenden Eigenschaften:

- Es ist  $\emptyset \in S$  und  $\Omega \in S$ .
- Für alle  $A \in S$  ist  $\Omega \setminus A \in S$ .
- Für alle Folgen  $(A_n)_{n \in \mathbb{N}}$  in  $S$  ist  $\bigcup_{n \in \mathbb{N}} A_n \in S$ .

Hierbei bezeichnet  $\text{Pot}(\Omega)$  die Potenzmenge von  $\Omega$ .

Wie auch bei anderen Begriffen, wie zum Beispiel  $\sigma$ -Kompaktheit oder  $\sigma$ -Additivität (s.u.), bezieht sich das „ $\sigma$ “ auf eine Abzählbarkeitsaussage (in diesem Fall die Abgeschlossenheit unter höchstens abzählbaren Vereinigungen).

**Bemerkung 1.2** (erzeugte  $\sigma$ -Algebra). Ist  $\Omega$  eine Menge und  $T \subset \text{Pot}(\Omega)$ , so gibt es eine bezüglich Inklusion kleinste  $\sigma$ -Algebra auf  $\Omega$ , die  $T$  enthält. Man nennt diese die *von  $T$  erzeugte  $\sigma$ -Algebra auf  $\Omega$*  und bezeichnet sie mit  $\sigma(T)$ .

**Definition 1.3** (Borel- $\sigma$ -Algebra). Sei  $(X, T)$  ein topologischer Raum. Die von  $T$  erzeugte  $\sigma$ -Algebra auf  $X$  heißt *Borel- $\sigma$ -Algebra auf  $X$*  und wird mit  $B(X, T)$  oder (falls die Topologie auf  $X$  aus dem Kontext klar ist) mit  $B(X)$  bezeichnet.

**Definition 1.4** (Wahrscheinlichkeitsmaß, Wahrscheinlichkeitsraum).

- Sei  $(\Omega, S)$  ein messbarer Raum. Ein *Wahrscheinlichkeitsmaß auf  $(\Omega, S)$*  ist eine Abbildung  $P: S \rightarrow [0, 1]$  mit den folgenden Eigenschaften:
  - Es ist  $P(\emptyset) = 0$  und  $P(\Omega) = 1$ .
  - $\sigma$ -Additivität: Ist  $(A_n)_{n \in \mathbb{N}}$  eine Folge paarweise disjunkter Mengen aus  $S$ , so ist

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} P(A_n).$$

- Ein *Wahrscheinlichkeitsraum* ist ein Tripel  $(\Omega, S, P)$ , wobei  $(\Omega, S)$  ein messbarer Raum und  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, S)$  ist.

**Caveat 1.5** (Wahrscheinlichkeiten von (nicht-disjunkten) Vereinigungen). Ist  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sind  $A, B \in S$  nicht disjunkt, so gilt im allgemeinen *nicht*, dass  $P(A \cup B) = P(A) + P(B)$  ist. Dies ist häufig eine Quelle von Fehlern!

Mathematisch gesehen ist die Wahrscheinlichkeitstheorie also ein Teil der Maßtheorie. Allerdings ist der Blickwinkel und damit auch die Art der betrachteten Probleme anders als in der Maßtheorie.

Die atomaren Maße beschreiben deterministische Zufallsexperimente:

**Definition 1.6** (atomares Maß). Sei  $\Omega$  eine nicht-leere Menge und sei  $\omega \in \Omega$ . Das Wahrscheinlichkeitsmaß

$$\delta_\omega: \text{Pot}(\Omega) \longrightarrow [0, 1]$$

$$A \longmapsto \begin{cases} 0 & \text{falls } \omega \notin A \\ 1 & \text{falls } \omega \in A \end{cases}$$

auf  $(\Omega, \text{Pot}(\Omega))$  heißt *atomares Maß auf  $\Omega$  konzentriert in  $\omega$* .

Wichtig für die Modellierung sind außerdem die Gleichverteilungen:

**Definition 1.7** (Gleichverteilung auf endlichen Mengen). Sei  $\Omega$  eine nicht-leere, *endliche* Menge. Das Wahrscheinlichkeitsmaß

$$\text{Pot}(\Omega) \longrightarrow [0, 1]$$

$$A \longmapsto \frac{|A|}{|\Omega|}$$

auf  $(\Omega, \text{Pot}(\Omega))$  heißt *Gleichverteilung auf  $\Omega$*  oder *Laplaceverteilung auf  $\Omega$* .

Die Definition der Gleichverteilung auf endlichen Mengen lässt sich als

$$\frac{\text{Zahl der günstigen Fälle}}{\text{Zahl der möglichen Fälle}}$$

interpretieren. Um Zähler und Nenner solcher Brüche zu bestimmen, verwendet man häufig Methoden aus der Kombinatorik.

Verwendet man nicht Mächtigkeiten, sondern das Lebesgue-Maß, so erhält man analog Gleichverteilungen auf reellen Borelmengen (ein wichtiger Spezialfall ist insbesondere die Borelmenge  $[0, 1] \subset \mathbb{R}$ ):

**Definition 1.8** (Gleichverteilung auf reellen Borelmengen). Sei  $n \in \mathbb{N}$  und sei  $A \in B(\mathbb{R}^n)$  mit  $0 < \lambda^n(A) < \infty$ ; hierbei bezeichnet  $B(\mathbb{R}^n)$  die Borel- $\sigma$ -Algebra auf  $\mathbb{R}^n$  bezüglich der Standardtopologie und  $\lambda^n$  bezeichnet das Lebesgue-Maß auf  $\mathbb{R}^n$ . Das Wahrscheinlichkeitsmaß

$$B(A) \longrightarrow [0, 1]$$

$$B \longmapsto \frac{\lambda^n(B)}{\lambda^n(A)}$$

auf  $(A, B(A))$  (wobei  $A \subset \mathbb{R}^n$  mit der Teilraumtopologie versehen wird) heißt *Gleichverteilung auf A*.

Ein weiterer wichtiger Spezialfall von Wahrscheinlichkeitsräumen ist der Fall, in dem die Masse des Wahrscheinlichkeitsmaßes auf höchstens abzählbar vielen Punkten konzentriert ist. Man beachte in diesem Zusammenhang, dass absolut konvergente Reihen umgeordnet werden können (ohne dass sich der Wert der Reihe ändert), und daher die untenstehenden Reihen wohldefiniert sind.

**Definition 1.9** (Zähldichte, diskreter Wahrscheinlichkeitsraum). Sei  $\Omega$  eine Menge, sei  $\Omega' \subset \Omega$  eine (höchstens) abzählbare Teilmenge und sei  $p: \Omega' \rightarrow [0, 1]$  mit  $\sum_{\omega \in \Omega'} p(\omega) = 1$ . Dann ist

$$P: \text{Pot}(\Omega) \rightarrow [0, 1]$$

$$A \mapsto \sum_{\omega \in A \cap \Omega'} p(\omega)$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, \text{Pot}(\Omega))$  und  $p$  ist eine *Zähldichte für P*. Wahrscheinlichkeitsmaße dieser Form heißen *diskrete Wahrscheinlichkeitsmaße* und die zugehörigen Wahrscheinlichkeitsräume heißen *diskrete Wahrscheinlichkeitsräume*.

Wir werden Zähldichten später noch in einen allgemeineren Kontext von Dichten einordnen (Abschnitt 1.4.3).

**Caveat 1.10** (Maßproblem). Im allgemeinen ist die Potenzmenge des Ergebnisraums „zu groß“ um „vernünftige“ Wahrscheinlichkeitsmaße zuzulassen (unter Annahme des Auswahlaxioms). Dies ist verwandt mit der Tatsache, dass (unter Annahme des Auswahlaxioms) das Lebesgue-Maß auf  $\mathbb{R}$  nicht auf  $\text{Pot}(\mathbb{R})$  fortgesetzt werden kann. Daher ist es unerlässlich die Stufe der  $\sigma$ -Algebren mit in den Formalismus aufzunehmen.

**Notation 1.11** (fast nie, fast sicher). In der Wahrscheinlichkeitstheorie wird die folgende Sprechweise verwendet: Sei dazu  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $A \in S$ .

- Das Ereignis  $A$  tritt *(P-)fast nie* ein, falls  $P(A) = 0$ .
- Das Ereignis  $A$  tritt *(P-)fast sicher* ein, falls  $P(A) = 1$ .

Man beachte, dass im ersten Fall im allgemeinen *nicht*  $A = \emptyset$  und im zweiten Fall im allgemeinen *nicht*  $A = \Omega$  gilt.

Der folgende Satz liefert ein nützliches Kriterium um Wahrscheinlichkeitsmaße auf einem gemeinsamen messbaren Raum zu vergleichen:

**Satz 1.12** (Eindeutigkeitssatz für Wahrscheinlichkeitsmaße). Sei  $(\Omega, S)$  ein messbarer Raum, sei  $T \subset S$  ein schnitt-stabiles Erzeugendensystem der  $\sigma$ -Algebra  $S$  und seien  $P, Q: S \rightarrow [0, 1]$  Wahrscheinlichkeitsmaße auf  $(\Omega, S)$  mit  $P|_T = Q|_T$ . Dann folgt

$$P = Q.$$

Dabei heißt  $T \subset \text{Pot}(\Omega)$  *schnitt-stabil*, wenn für alle  $A, B \in T$  auch  $A \cap B \in T$  gilt. Der Beweis des Eindeutigkeitssatzes verwendet sogenannte *Dynkin-Systeme*.

## 1.2 Zufallsvariablen

Eine mathematische Theorie („Kategorie“) besteht aus einer geeigneten Klasse von Objekten (in unserem Fall: messbare Räume bzw. Wahrscheinlichkeitsräume) und strukturerhaltenden Morphismen zwischen diesen Objekten (in unserem Fall: messbare Abbildungen bzw. sogenannte Zufallsvariablen).

Auch in der Modellierung ergibt sich der Bedarf nach solchen strukturerhaltenden Abbildungen; im Beispiel der Wahrscheinlichkeitstheorie stehen dabei insbesondere die folgenden Aspekte im Vordergrund:

- Manchmal interessiert nur ein gewisser Aspekt eines Modells und nicht das gesamte Zufallsexperiment. Man sucht also ein geeignetes Abstraktionswerkzeug in der Modellierung, das es erlaubt, Information zu filtern und neu zu kombinieren.
- Außerdem möchte man manchmal über „Variablen“ mit „zufälligen Werten“ sprechen.

Beide Aspekte können mit sogenannten Zufallsvariablen umgesetzt werden.

**Definition 1.13** (messbare Abbildung, Zufallsvariable).

- Seien  $(\Omega, S)$  und  $(\Omega', S')$  messbare Räume. Eine Abbildung  $X: \Omega \rightarrow \Omega'$  heißt (bezüglich  $S$  und  $S'$ ) *messbar*, falls Urbilder messbarer Mengen unter  $X$  messbar sind, d.h. falls

$$\forall A' \in S' \quad X^{-1}(A') \in S.$$

- Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $(\Omega', S')$  ein messbarer Raum. Eine  $(\Omega', S')$ -wertige *Zufallsvariable* auf  $(\Omega, S, P)$  ist eine (bezüglich  $S$  und  $S'$ ) messbare Abbildung  $\Omega \rightarrow \Omega'$ .  
Ist  $(\Omega', S') = (\mathbb{R}, B(\mathbb{R}))$ , so spricht man von *reellwertigen Zufallsvariablen*.

**Bemerkung 1.14** (stetige Abbildungen sind messbar). Seien  $(X, T)$  und  $(X', T')$  topologische Räume. Dann ist jede stetige Abbildung  $X \rightarrow X'$  bezüglich den Borel- $\sigma$ -Algebren auf  $X$  bzw.  $X'$  messbar.

In der Wahrscheinlichkeitstheorie und Statistik hat sich die folgende Notation eingebürgert, die vor allem den zweiten Aspekt von Zufallsvariablen verdeutlicht:

**Notation 1.15.** Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(\Omega', S')$  ein messbarer Raum und sei  $X: \Omega \rightarrow \Omega'$  eine  $(\Omega', S')$ -wertige Zufallsvariable auf  $(\Omega, S, P)$ .

- Ist  $A' \subset \Omega'$ , so schreibt man

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

- Ist  $c' \in \Omega'$ , so schreibt man

$$\{X = c'\} := \{\omega \in \Omega \mid X(\omega) = c'\} = X^{-1}(\{c'\}).$$

Analog definiert man für reellwertige Zufallsvariablen  $X$  und reelle Zahlen  $c \in \mathbb{R}$  die Mengen  $\{X \leq c\}, \dots$

**Definition 1.16** (induzierte Verteilung). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(\Omega', S')$  ein messbarer Raum und sei  $X: \Omega \rightarrow \Omega'$  eine  $(\Omega', S')$ -wertige Zufallsvariable auf  $(\Omega, S, P)$ . Dann ist

$$P_X: S' \rightarrow [0, 1] \\ A' \rightarrow P(X^{-1}(A')) = P(\{X \in A'\})$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega', S')$ , die (*Wahrscheinlichkeits-*)*Verteilung von X*.

Man ändert nun den Blickwinkel: Im Normalfall interessiert sich man sich nur für die *Verteilung* einer Zufallsvariable, nicht aber für den Wahrscheinlichkeitsraum, der den Definitionsbereich der Zufallsvariablen bildet oder die genaue Definition der Zufallsvariablen.

**Definition 1.17** (identisch verteilt). Seien  $(\Omega_1, S_1, P_1)$ ,  $(\Omega_2, S_2, P_2)$  Wahrscheinlichkeitsräume, sei  $(\Omega, S)$  ein messbarer Raum und seien  $X_1: \Omega_1 \rightarrow \Omega$  und  $X_2: \Omega_2 \rightarrow \Omega$  zwei  $(\Omega, S)$ -wertige Zufallsvariablen auf  $(\Omega_1, S_1, P_1)$  bzw.  $(\Omega_2, S_2, P_2)$ . Die Zufallsvariablen  $X_1$  und  $X_2$  heißen *identisch verteilt*, wenn sie dasselbe Wahrscheinlichkeitsmaß auf  $(\Omega, S)$  induzieren, d.h., falls  $(P_1)_{X_1} = (P_2)_{X_2}$ .

### 1.3 Verteilungsfunktionen

Wie kann man entscheiden, ob reellwertige Zufallsvariablen identisch verteilt sind? Wir werden sehen, dass sich Verteilungen von reellwertigen Zufallsvariablen (bzw. Wahrscheinlichkeitsmaße auf  $(\mathbb{R}, B(\mathbb{R}))$ ) durch ihre sogenannten Verteilungsfunktionen charakterisieren lassen. Verteilungsfunktionen werden später außerdem bei der Betrachtung von Konvergenzbegriffen eine wichtige Rollen spielen.

**Definition 1.18** (Verteilungsfunktion eines reellen Wahrscheinlichkeitsmaßes/einer reellen Zufallsvariable).

- Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, B(\mathbb{R}))$ . Die *Verteilungsfunktion von P* ist definiert als

$$F_P: \mathbb{R} \rightarrow [0, 1] \\ x \mapsto P((-\infty, x]).$$

- Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $X: \Omega \rightarrow \mathbb{R}$  eine reellwertige Zufallsvariable auf  $(\Omega, S, P)$ . Die *Verteilungsfunktion von X* ist definiert als  $F_X := F_{P_X}: \mathbb{R} \rightarrow [0, 1]$ .

**Caveat 1.19.** Im allgemeinen lassen sich Verteilungsfunktionen nicht explizit in geschlossener Form durch elementare Funktionen darstellen.

**Proposition 1.20** (Eigenschaften von Verteilungsfunktionen). *Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, B(\mathbb{R}))$ . Dann gilt:*

1. Die Verteilungsfunktion  $F_P$  von  $P$  ist monoton wachsend.



2. Die Funktion  $F_P$  ist rechtsseitig stetig, d.h. für alle  $x \in \mathbb{R}$  ist

$$F_P(x) = \lim_{[x, \infty) \ni z \rightarrow x} F_P(z).$$

3. Es gilt (und insbesondere existieren diese Grenzwerte)

$$\lim_{x \rightarrow -\infty} F_P(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F_P(x) = 1.$$

4. Für alle  $x \in \mathbb{R}$  gilt: Die Funktion  $F_P$  ist genau dann in  $x$  stetig, wenn  $P(\{x\}) = 0$  ist.

Der Beweis beruht auf der  $\sigma$ -Stetigkeit von Wahrscheinlichkeitsmaßen:

**Lemma 1.21** ( $\sigma$ -Stetigkeit von Wahrscheinlichkeitsmaßen). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum.

1. Sei  $(A_n)_{n \in \mathbb{N}}$  eine monoton wachsende Folge in  $S$  (d.h. für alle  $n \in \mathbb{N}$  ist  $A_n \subset A_{n+1}$ ). Dann gilt

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

2. Sei  $(A_n)_{n \in \mathbb{N}}$  eine monoton fallende Folge in  $S$ . Dann gilt

$$P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Man ändert nun die Perspektive und führt den folgenden Abstraktionsschritt durch: Man verwendet die ersten drei Eigenschaften von Proposition 1.20 als *Definition* für eine Klasse von Funktionen:

**Definition 1.22** (reelle Verteilungsfunktion). Eine *reelle Verteilungsfunktion* ist eine monoton wachsende, rechtsseitig stetige Funktion  $F: \mathbb{R} \rightarrow [0, 1]$  mit

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Der folgende Satz liefert die gewünschte Charakterisierung von Verteilungen reellwertiger Zufallsvariablen durch ihre Verteilungsfunktion:

**Satz 1.23** (Korrespondenzsatz für Verteilungsfunktionen).

1. Ist  $F: \mathbb{R} \rightarrow [0, 1]$  eine reelle Verteilungsfunktion, so gibt es ein Wahrscheinlichkeitsmaß  $\lambda_F$  auf  $(\mathbb{R}, B(\mathbb{R}))$  mit  $F_{\lambda_F} = F$ , das sogenannte Lebesgue-Stieltjes-Maß zu  $F$ .
2. Ist  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, B(\mathbb{R}))$ , so ist  $P = \lambda_{F_P}$ .

Die Konstruktion des Lebesgue-Stieltjes-Maßes erfolgt analog zur Konstruktion des Lebesgue-Maßes. Die Eindeutigkeitsaussage beruht auf dem Eindeutigkeitssatz für Wahrscheinlichkeitsmaße (Satz 1.12).

**Korollar 1.24.** Insbesondere sind reellwertige Zufallsvariablen genau dann identisch verteilt, wenn sie dieselbe Verteilungsfunktion besitzen.

Welche Rolle spielen Verteilungsfunktionen in den Anwendungen? Zum Beispiel betrachtet man in der Statistik unter anderem den Begriff des Medians bzw. allgemeiner der Quantile.

**Definition 1.25** (Quantil). Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, B(\mathbb{R}))$  und sei  $p \in (0, 1)$ . Dann ist

$$\inf\{x \in \mathbb{R} \mid F_P(x) \geq p\} \in \mathbb{R}$$

das  $p$ -Quantil von  $P$ . Analog definiert man Quantile für reellwertige Zufallsvariablen als die Quantile ihrer Verteilung.

**Bemerkung 1.26.** Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $X: \Omega \rightarrow \mathbb{R}$  eine reellwertige Zufallsvariable auf  $(\Omega, S, P)$ , sei  $p \in (0, 1)$  und sei  $x_p \in \mathbb{R}$  das  $p$ -Quantil von  $X$ . Dann gilt

$$P_X(X \leq x_p) = F_X(x_p) \geq p$$

und für alle  $\varepsilon \in \mathbb{R}_{>0}$  ist

$$P_X(X \leq x_p - \varepsilon) = F_X(x_p - \varepsilon) < p.$$

Ein wichtiger Spezialfall ist das  $1/2$ -Quantil, das eine zentrale Kenngröße bei der Darstellung von Daten bzw. Verteilungen ist:

**Definition 1.27** (Median). Sei  $X$  eine reellwertige Zufallsvariable. Dann heißt das  $1/2$ -Quantil von  $X$  auch *Median* von  $X$ .

## 1.4 Exkurs: Integration auf Wahrscheinlichkeitsräumen

Da viele wichtige Kenngrößen von reellwertigen Zufallsvariablen (z.B. Erwartungswert und Varianz) auf Integration beruhen, stellen wir im folgenden die Grundlagen der Integration auf Wahrscheinlichkeitsräumen zusammen. Details und Beweise finden sich in allen Büchern zu Maß- und Integrationstheorie.

Analog zur Konstruktion des Lebesgue-Integrals auf  $(\mathbb{R}^n, B(\mathbb{R}^n))$  aus dem Lebesgue-Maß auf  $(\mathbb{R}^n, B(\mathbb{R}^n))$  kann man zu jedem Maßraum einen zugehörigen Integralbegriff für geeignete reellwertige Funktionen definieren.

Wir werden diese Konstruktion für Wahrscheinlichkeitsräume skizzieren (da dies technisch ein bisschen einfacher ist als für allgemeine Maßräume) und die für die Anwendungen wichtigsten Sätze zusammenstellen.

### 1.4.1 Skizze der Konstruktion des Integrals auf Wahrscheinlichkeitsräumen

Wie im Fall des Lebesgue-Integrals definiert man zunächst eine geeignete Klasse von Treppenfunktionen und ein Integral für diese Treppenfunktionen. Dieses Integral wird dann in einem weiteren Schritt durch Grenzwertbildung auf eine geeignete Klasse messbarer Funktionen erweitert.

Im folgenden sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum.

*Treppenfunktionen.*

- Ist  $A \subset \Omega$ , so schreiben wir

$$\begin{aligned}\chi_A: \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto \begin{cases} 0 & \text{falls } \omega \notin A, \\ 1 & \text{falls } \omega \in A \end{cases}\end{aligned}$$

für die *charakteristische Funktion* zu  $A$  auf  $\Omega$ . Man beachte, dass  $\chi_A$  genau dann bezüglich  $S$  und  $B(\mathbb{R})$  messbar ist, wenn  $A \in S$  ist.

- Der *Raum der Treppenfunktionen* auf  $(\Omega, S, P)$  ist definiert als

$$T(\Omega, S, P) := \left\{ \sum_{j=1}^k a_j \cdot \chi_{A_j} \mid k \in \mathbb{N}, a_1, \dots, a_k \in \mathbb{R}, A_1, \dots, A_k \in S \right\}.$$

- Außerdem schreiben wir

$$T_{\geq 0}(\Omega, S, P) := \left\{ \sum_{j=1}^k a_j \cdot \chi_{A_j} \mid k \in \mathbb{N}, a_1, \dots, a_k \in \mathbb{R}_{\geq 0}, A_1, \dots, A_k \in S \right\}.$$

*Integration von Treppenfunktionen.* Das *Integral für Treppenfunktionen* auf  $(\Omega, S, P)$  ist definiert durch

$$\begin{aligned}\int \cdot dP: T(\Omega, S, P) &\longrightarrow \mathbb{R} \\ \sum_{j=1}^k a_j \cdot \chi_{A_j} &\longmapsto \sum_{j=1}^k a_j \cdot P(A_j);\end{aligned}$$

man kann zeigen, dass dies tatsächlich wohldefiniert ist (d.h., dass der Wert des Integrals *nicht* von der gewählten Zerlegung einer Treppenfunktion in messbare „Stufen“ abhängt).

*Fortsetzung des Integrals.*

- Sei  $X: \Omega \longrightarrow \mathbb{R}_{\geq 0}$  bezüglich  $S$  und  $B(\mathbb{R}_{\geq 0}) \subset B(\mathbb{R})$  messbar. Dann ist  $X$  *bezüglich  $P$  integrierbar*, falls es eine Folge  $(T_n)_{n \in \mathbb{N}}$  in  $T_{\geq 0}(\Omega, S, P)$  von Treppenfunktionen gibt, die punktweise monoton wachsend ist, punktweise gegen  $f$  konvergiert, und für die außerdem der Grenzwert  $\lim_{n \rightarrow \infty} \int T_n dP$  in  $\mathbb{R}$  existiert. Dann heißt

$$\int X dP := \lim_{n \rightarrow \infty} \int T_n dP \in \mathbb{R}$$

*Integral von  $X$  bezüglich  $P$* ; man kann zeigen, dass dies tatsächlich wohldefiniert ist (d.h., dass der Wert des Integrals nicht von der gewählten approximierenden Folge von Treppenfunktionen abhängt).

- Sei  $X: \Omega \longrightarrow \mathbb{R}$  bezüglich  $S$  und  $B(\mathbb{R})$  messbar. Dann ist  $X$  *bezüglich  $P$  integrierbar*, falls

$$X_+ := \max(X, 0) \quad \text{und} \quad X_- := -\min(X, 0)$$

im obigen Sinne bezüglich  $P$  integrierbar sind. In diesem Fall ist

$$\int X \, dP := \int X_+ \, dP - \int X_- \, dP \in \mathbb{R}$$

das *Integral von  $X$  bezüglich  $P$* . Es ist nicht schwer zu zeigen, dass eine messbare Funktion  $X: \Omega \rightarrow \mathbb{R}$  genau dann bezüglich  $P$  integrierbar ist, wenn  $|X|$  bezüglich  $P$  integrierbar ist.

- Außerdem verwenden wir die folgende Notation: Ist  $A \in \mathcal{S}$ , ist  $X: \Omega \rightarrow \mathbb{R}$  messbar und ist  $X \cdot \chi_A$  bezüglich  $P$  integrierbar, so schreiben wir auch

$$\int_A X \, dP := \int X \cdot \chi_A \, dP.$$

**Caveat 1.28** (Integrierbarkeit und Beschränktheit). Auf einem Wahrscheinlichkeitsraum sind alle messbaren und (fast überall) beschränkten Funktionen integrierbar. (Auf allgemeinen Maßräumen ist dies im allgemeinen *nicht* wahr!)

Umgekehrt ist *nicht* jede auf einem Wahrscheinlichkeitsraum integrierbare Funktion (fast überall) beschränkt!

Vom Blickwinkel der Funktionalanalysis erhält man das Integral  $\int \cdot \, dP$  durch Vervollständigung von  $T(\Omega, \mathcal{S}, P)$  bezüglich der durch das Integral von Treppenfunktionen gegebenen Halbnorm  $\|\cdot\|_1$  und durch stetige, lineare Fortsetzung des Integrals von Treppenfunktionen auf diese Vervollständigung.

#### 1.4.2 Wichtige Eigenschaften des Integrals

Analog zum Lebesgue-Integral auf  $(\mathbb{R}^n, B(\mathbb{R}^n))$  erhalten wir für die von Wahrscheinlichkeitsmaßen induzierten Integrale die folgenden Eigenschaften/Sätze:

- Positivität des Integrals
- Linearität des Integrals
- Satz von der monotonen Konvergenz
- Satz von der dominierten Konvergenz
- Transformationssatz
- Satz von Fubini.

Wir geben nun exakte Formulierungen dieser Sätze an:

**Satz 1.29** (Positivität des Integrals). *Sei  $(\Omega, \mathcal{S}, P)$  ein Wahrscheinlichkeitsraum und sei  $X: \Omega \rightarrow \mathbb{R}_{\geq 0}$  bezüglich  $P$  integrierbar.*

1. *Dann ist  $\int X \, dP \geq 0$ .*
2. *Es gilt genau dann  $\int X \, dP = 0$ , wenn  $P(\{X = 0\}) = 1$  ist.*

**Satz 1.30** (Linearität des Integrals). *Sei  $(\Omega, \mathcal{S}, P)$  ein Wahrscheinlichkeitsraum, seien  $X, Y: \Omega \rightarrow \mathbb{R}$  bezüglich  $P$  integrierbar und seien  $a, b \in \mathbb{R}$ . Dann ist auch*

$$\begin{aligned} a \cdot X + b \cdot Y: \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto a \cdot X(\omega) + b \cdot Y(\omega) \end{aligned}$$

bezüglich  $P$  integrierbar und es gilt

$$\int (a \cdot X + b \cdot Y) dP = a \cdot \int X dP + b \cdot \int Y dP.$$

Aus der Positivität und der Linearität des Integrals folgt die Monotonie des Integrals:

**Korollar 1.31** (Monotonie des Integrals). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $X, Y: \Omega \rightarrow \mathbb{R}$  bezüglich  $P$  integrierbar.*

1. *Ist  $X \leq Y$  (punktweise), so ist*

$$\int X dP \leq \int Y dP.$$

2. *Es gilt*

$$\left| \int X dP \right| \leq \int |X| dP.$$

Die nächsten beiden Sätze sind die zentralen Konvergenzsätze für Integrale, die hinreichende Kriterien dafür angeben, wann (punktweise) Grenzwerte und Integrale miteinander verträglich sind:

**Satz 1.32** (monotone Konvergenz). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge bezüglich  $P$  integrierbarer Funktionen  $\Omega \rightarrow \mathbb{R}$ , die punktweise monoton wachsend ist, mit der Eigenschaft, dass die Menge  $\{\int X_n dP \mid n \in \mathbb{N}\} \subset \mathbb{R}$  beschränkt ist.*

*Dann existiert eine bezüglich  $S$  und  $B(\mathbb{R})$  messbare Funktion  $X: \Omega \rightarrow \mathbb{R}$ , für die  $P$ -fast überall  $X = \lim_{n \rightarrow \infty} X_n$  gilt (punktweise); diese Funktion  $X$  ist dann bezüglich  $P$  integrierbar und es gilt*

$$\int X dP = \lim_{n \rightarrow \infty} \int X_n dP.$$

**Satz 1.33** (dominierte Konvergenz). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge bezüglich  $P$  integrierbarer Funktionen, die  $P$ -fast überall (punktweise) gegen eine (bezüglich  $S$  und  $B(\mathbb{R})$  messbare) Funktion  $X: \Omega \rightarrow \mathbb{R}$  konvergiert. Außerdem gebe es eine bezüglich  $P$  integrierbare Funktion  $Y: \Omega \rightarrow \mathbb{R}$  mit*

$$\forall n \in \mathbb{N} \quad |X_n| \leq Y.$$

*Dann ist auch  $X$  bezüglich  $P$  integrierbar und es gilt*

$$\int X dP = \lim_{n \rightarrow \infty} \int X_n dP.$$

Mithilfe des Transformationssatzes können wir Integration bezüglich induzierten Verteilungen beschreiben:

**Satz 1.34** (Transformationssatz). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $(\Omega', S')$  ein messbarer Raum. Sei  $X: \Omega \rightarrow \Omega'$  eine  $(\Omega', S')$ -wertige Zufallsvariable*

auf  $(\Omega, S, P)$  und sei  $Y: \Omega' \rightarrow \mathbb{R}$  eine bezüglich  $S'$  und  $B(\mathbb{R})$  messbare Abbildung. Dann ist  $Y$  genau dann bezüglich  $P_X$  integrierbar, wenn  $Y \circ X$  bezüglich  $P$  integrierbar ist, und in diesem Fall gilt

$$\int Y dP_X = \int Y \circ X dP.$$

Der Satz von Fubini erlaubt es, Integrale auf Produkträumen (Definition 2.9, Satz 2.13) durch iterierte Integrale auszudrücken:

**Satz 1.35** (Satz von Fubini). Seien  $(\Omega_1, S_1, P_1)$  und  $(\Omega_2, S_2, P_2)$  Wahrscheinlichkeitsräume und sei  $X: \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$  bezüglich  $S_1 \otimes S_2$  und  $B(\mathbb{R})$  messbar. Dann ist  $X$  genau dann bezüglich  $P_1 \otimes P_2$  integrierbar, wenn folgende Bedingung erfüllt ist:

- Für  $P_1$ -fast alle  $\omega_1 \in \Omega_1$  ist

$$\begin{aligned} \Omega_2 &\longrightarrow \mathbb{R} \\ \omega_2 &\longmapsto X(\omega_1, \omega_2) \end{aligned}$$

bezüglich  $P_2$  integrierbar,

– und

$$\begin{aligned} \Omega_1 &\longrightarrow \mathbb{R} \\ \omega_1 &\longmapsto \int X(\omega_1, \omega_2) dP_2(\omega_2) \end{aligned}$$

ist bezüglich  $P_1$  integrierbar.

In diesem Fall gilt

$$\int X d(P_1 \otimes P_2) = \int \int X(\omega_1, \omega_2) dP_2(\omega_2) dP_1(\omega_1).$$

(Analog gilt dies auch für die andere Reihenfolge der Faktoren.)

#### 1.4.3 Integration bezüglich Wahrscheinlichkeitsmaßen, die durch Dichten gegeben sind

Im Spezialfall der diskreten Wahrscheinlichkeitsräume hat das Integral die folgende Gestalt:

**Satz 1.36** (Integration auf diskreten Wahrscheinlichkeitsräumen). Sei  $(\Omega, \text{Pot}(\Omega), P)$  ein diskreter Wahrscheinlichkeitsraum, wobei  $P$  durch die Zähldichte  $p: \Omega' \rightarrow [0, 1]$  gegeben sei (und  $\Omega' \subset \Omega$  höchstens abzählbar ist). Eine Funktion  $X: \Omega \rightarrow \mathbb{R}$  ist genau dann bezüglich  $P$  integrierbar, wenn die Reihe  $\sum_{\omega \in \Omega'} |X(\omega)| \cdot p(\omega)$  konvergiert, und in diesem Fall ist

$$\int X dP = \sum_{\omega \in \Omega'} X(\omega) \cdot p(\omega).$$

Man beachte, dass im obigen Satz die Reihenfolge der Summation in den betrachteten Reihen keine Rolle spielt, da es sich um absolute Konvergenz bzw. Werte von absolut konvergente Reihen handelt.

Etwas allgemeiner gilt der obige Satz für Wahrscheinlichkeitsmaße, die durch sogenannte Dichten gegeben sind:

**Definition 1.37** (Wahrscheinlichkeitsdichte). Sei  $(\Omega, S, \mu)$  ein Maßraum (nicht notwendig ein Wahrscheinlichkeitsraum). Eine *Wahrscheinlichkeitsdichte* auf  $(\Omega, S, \mu)$  ist eine  $\mu$ -integrierbare Funktion  $f: \Omega \rightarrow \mathbb{R}_{\geq 0}$  mit

$$\int f d\mu = 1.$$

Die wichtigsten Fälle für uns sind, dass die Dichten auf  $\mathbb{N}$  mit dem Zählmaß (dies sind die uns bereits bekannten Zähldichten) bzw. auf  $(\mathbb{R}^n, B(\mathbb{R}^n))$  mit dem Lebesgue-Maß definiert sind.

**Proposition 1.38** (von einer Dichte induziertes Wahrscheinlichkeitsmaß). Sei  $(\Omega, S, \mu)$  ein Maßraum (nicht notwendig ein Wahrscheinlichkeitsraum) und sei  $f: \Omega \rightarrow \mathbb{R}_{\geq 0}$  eine Wahrscheinlichkeitsdichte auf  $(\Omega, S, \mu)$ . Dann ist

$$\begin{aligned} f \odot \mu: S &\longrightarrow [0, 1] \\ A &\longmapsto \int_A f d\mu = \int f \cdot \chi_A d\mu \end{aligned}$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, S)$ , das von der Wahrscheinlichkeitsdichte  $f$  bezüglich  $\mu$  induzierte Wahrscheinlichkeitsmaß auf  $(\Omega, S)$ .

Dies folgt aus der Monotonie, Linearität und den Konvergenzsätzen für allgemeine Integrale. Mit einer geeigneten maßtheoretischen Induktion erhält man für solche Wahrscheinlichkeitsmaße:

**Satz 1.39** (Integration bezüglich von Dichten induzierten Wahrscheinlichkeitsmaßen). Sei  $(\Omega, S, \mu)$  ein Maßraum, sei  $f: \Omega \rightarrow \mathbb{R}_{\geq 0}$  eine Wahrscheinlichkeitsdichte auf  $(\Omega, S, \mu)$  und sei  $P := f \odot \mu$  das induzierte Wahrscheinlichkeitsmaß auf  $(\Omega, S)$ . Dann ist eine messbare Funktion  $X: \Omega \rightarrow \mathbb{R}$  genau dann bezüglich  $P$  integrierbar, wenn  $X \cdot f$  bezüglich  $\mu$  integrierbar ist, und es gilt in diesem Fall

$$\int X dP = \int X \cdot f d\mu.$$

Der Vollständigkeit halber erwähnen wir noch den Satz von Radon-Nikodym, der charakterisiert, unter welchen Bedingungen ein Wahrscheinlichkeitsmaß von einer Dichte bezüglich eines gegebenen Maßes induziert ist:

**Satz 1.40** (Satz von Radon-Nikodym). Sei  $(\Omega, S, P)$  ein Maßraum, wobei  $\mu$  ein sogenanntes  $\sigma$ -endliches Maß ist, und sei  $P$  ein Wahrscheinlichkeitsmaß auf  $(\Omega, S)$ . Dann sind folgende Aussagen äquivalent:

1. Es gilt  $P \ll \mu$ , d.h. für alle  $A \in S$  mit  $\mu(A) = 0$  ist auch  $P(A) = 0$ .
2. Es gibt eine Wahrscheinlichkeitsdichte  $f: \Omega \rightarrow \mathbb{R}_{\geq 0}$  auf  $(\Omega, S, P)$  mit  $P = f \odot \mu$ .

## 1.5 Erwartungswert und Varianz

Wir führen nun wichtige Kenngrößen von reellwertigen Zufallsvariablen ein: Erwartungswert (eine Verallgemeinerung des Mittelwerts) und Varianz (eng verwandt mit der Standardabweichung). Vom maßtheoretischen Standpunkt aus handelt es sich hierbei um Größen, die man per Integration aus reellwertigen Zufallsvariablen erhält.

### 1.5.1 Erwartungswert

Der Erwartungswert ist eine Verallgemeinerung des arithmetischen Mittels einer endlichen Folge reeller Zahlen; die Mittelung geschieht per Integration über das entsprechende Wahrscheinlichkeitsmaß, d.h. die Funktionswerte werden gemäß dem Wahrscheinlichkeitsmaß „gewichtet“ und „aufsummiert“.

**Definition 1.41** (Erwartungswert). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $X: \Omega \rightarrow \mathbb{R}$  eine bezüglich  $P$  integrierbare reellwertige Zufallsvariable auf  $(\Omega, S, P)$ . Dann bezeichnet man

$$E(X) := \int X \, dP$$

als *Erwartungswert von  $X$* .

Mithilfe des Transformationssatzes sieht man, dass der Erwartungswert einer integrierbaren reellwertigen Zufallsvariable nur von ihrer Verteilung abhängt, und man spricht daher auch oft vom Erwartungswert dieser Verteilung (statt vom Erwartungswert der Zufallsvariablen).

Mit den Sätzen 1.36 und 1.39 können wir außerdem Integrierbarkeit und Erwartungswert reformulieren, wenn das Wahrscheinlichkeitsmaß auf dem Definitionsbereich der Zufallsvariablen durch eine Wahrscheinlichkeitsdichte gegeben ist.

Aus den grundlegenden Eigenschaften des Integrals erhalten wir die entsprechenden Eigenschaften des Erwartungswerts:

**Proposition 1.42** (grundlegende Eigenschaften des Erwartungswerts). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $X, Y: \Omega \rightarrow \mathbb{R}$  bezüglich  $P$  integrierbare reellwertige Zufallsvariablen auf  $(\Omega, S, P)$ .

1. Monotonie. Ist  $X \leq Y$ , so ist  $E(X) \leq E(Y)$ . Außerdem ist  $|E(X)| \leq E(|X|)$ .
2. Linearität. Sind  $a, b \in \mathbb{R}$ , so ist auch  $a \cdot X + b \cdot Y$  bezüglich  $P$  integrierbar und

$$E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y).$$

3. Normierung. Es gilt  $E(1) = 1$ .

Analog können wir natürlich auch alle weiteren Sätze über Integration in Sätze über Erwartungswerte übersetzen (insbesondere die Konvergenzsätze).

**Proposition 1.43** (Markov-Ungleichung). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $X: \Omega \rightarrow \mathbb{R}$  eine bezüglich  $P$  integrierbare reellwertige Zufallsvariable auf  $(\Omega, S, P)$ , und sei  $c \in \mathbb{R}_{>0}$ . Dann gilt

$$P(|X| \geq c) \leq \frac{1}{c} \cdot E(|X|).$$



Man beachte jedoch, dass die Abschätzung aus der Markov-Ungleichung im allgemeinen sehr grob ist (da nur der Erwartungswert, aber nicht die Verteilung eingeht).

Mithilfe von Erwartungswerten können wir außerdem eine wichtige Eigenschaft des Medians formulieren:

**Proposition 1.44** (der Median minimiert die absolute Abweichung). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $X: \Omega \rightarrow \mathbb{R}$  eine bezüglich  $P$  integrierbare reellwertige Zufallsvariable auf  $(\Omega, S, P)$  und sei  $m \in \mathbb{R}$  der Median von  $X$ . Dann gilt für alle  $a \in \mathbb{R}$ , dass*

$$E(|X - m|) \leq E(|X - a|).$$

### 1.5.2 Varianz

Der Erwartungswert einer reellwertigen Zufallsvariable ist eine sehr grobe Kenngröße: Zufallsvariablen können denselben Erwartungswert besitzen und trotzdem sehr unterschiedliche Verteilungen haben. Daher führt man weitere Kenngrößen ein, die messen, wie stark sich eine reelle Wahrscheinlichkeitsverteilung um ihren Erwartungswert konzentriert; dies geschieht durch Integration höherer Potenzen von Zufallsvariablen. Ein Beispiel ist die Varianz, die die quadratische Abweichung vom Erwartungswert misst.

**Bemerkung 1.45** (Quadratintegrierbarkeit impliziert Integrierbarkeit). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $X: \Omega \rightarrow \mathbb{R}$  eine reellwertige Zufallsvariable auf  $(\Omega, S, P)$ . Ist  $X$  bezüglich  $P$  quadratintegrierbar (d.h. ist  $X^2$  bezüglich  $P$  integrierbar), so ist auch  $X$  bezüglich  $P$  integrierbar. Insbesondere gilt: Ist  $X^2$  bezüglich  $P$  integrierbar, so auch  $(X - E(X))^2$ .

Man beachte jedoch, dass

- integrierbare Zufallsvariablen im allgemeinen *nicht* quadratintegrierbar sind, und dass
- auf allgemeinen Maßräumen quadratintegrierbare Funktionen im allgemeinen *nicht* integrierbar sind.

**Definition 1.46** (Varianz, Standardabweichung). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $X: \Omega \rightarrow \mathbb{R}$  eine quadratintegrierbare reellwertige Zufallsvariable auf  $(\Omega, S, P)$ .

- Die *Varianz* von  $X$  ist

$$\text{Var}(X) := E((X - E(X))^2) \in \mathbb{R}_{\geq 0}.$$

- Die *Standardabweichung* von  $X$  ist

$$\sigma(X) := \sqrt{\text{Var}(X)}.$$

Mithilfe des Transformationssatzes kann man zeigen, dass die Varianz/Standardabweichung einer quadratintegrierbaren reellwertigen Zufallsvariable nur von ihrer Verteilung abhängt, und man spricht daher auch oft von der Varianz/Standardabweichung dieser Verteilung (statt von der Varianz/Standardabweichung der Zufallsvariablen).

Manchmal ist es einfacher, die Varianz über die folgende Darstellung zu berechnen:

**Proposition 1.47** (alternative Darstellung der Varianz). *Sei  $X$  eine quadratintegrierbare reellwertige Zufallsvariable. Dann gilt*

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

Analog zum Erwartungswert lassen sich natürlich auch Quadratintegrierbarkeit und Varianz/Standardabweichung bezüglich durch Dichten gegebene Wahrscheinlichkeitsmaße mithilfe der Sätze 1.36 und 1.39 reformulieren.

**Proposition 1.48** (grundlegende Eigenschaft der Varianz). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $X: \Omega \rightarrow \mathbb{R}$  eine quadratintegrierbare reellwertige Zufallsvariable auf  $(\Omega, S, P)$ .*

1. Charakterisierung deterministischer Zufallsvariablen. *Genau dann ist  $\text{Var}(X) = 0$ , wenn  $X$  deterministisch ist (d.h.  $X$  ist  $P$ -fast überall konstant).*
2. Affine Transformation. *Für alle  $a, b \in \mathbb{R}$  ist auch  $a \cdot X + b: \Omega \rightarrow \mathbb{R}$  quadratintegrierbar und*

$$\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X).$$

3. Standardisierung. *Ist  $\text{Var}(X) \neq 0$ , so ist die reellwertige Zufallsvariable*

$$\frac{1}{\sigma(X)} \cdot (X - E(X))$$

*quadratintegrierbar und hat Erwartungswert 0 und Varianz 1.*

Der Beweis des ersten Teils beruht auf der folgenden Aussage über Integration:

**Lemma 1.49** (nicht-negative Funktionen mit verschwindendem Integral). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $X: \Omega \rightarrow \mathbb{R}$  eine integrierbare reellwertige Zufallsvariable mit  $X \geq 0$ . Dann gilt genau dann  $\int X dP = 0$ , wenn  $P(X > 0) = 0$  ist, d.h., wenn  $P$ -fast überall  $X = 0$  gilt.*

**Caveat 1.50** (Varianz von Summen). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $X, Y: \Omega \rightarrow \mathbb{R}$  quadratintegrierbare reellwertige Zufallsvariablen auf  $(\Omega, S, P)$ . Man kann zeigen, dass dann auch  $X + Y$  quadratintegrierbar ist, aber im allgemeinen lässt sich  $\text{Var}(X + Y)$  nicht durch  $\text{Var}(X)$  und  $\text{Var}(Y)$  ausdrücken – man muss zusätzlich verstehen, wie  $X$  und  $Y$  voneinander „abhängen“ (Abschnitt 2.3).

Aus der Markov-Ungleichung erhalten wir die Tschebyschev-Ungleichung:

**Proposition 1.51** (Tschebyschev-Ungleichung). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $X: \Omega \rightarrow \mathbb{R}$  eine quadratintegrierbare reellwertige Zufallsvariable auf  $(\Omega, S, P)$  und sei  $c \in \mathbb{R}_{>0}$ . Dann ist*

$$P(|X - E(X)| \geq c) \leq \frac{1}{c^2} \cdot \text{Var}(X).$$

Außerdem können wir mithilfe der Varianz die folgende Eigenschaft des Erwartungswertes formulieren:

**Proposition 1.52** (der Erwartungswert minimiert die quadratische Abweichung). *Sei  $X$  eine quadratintegrierbare reellwertige Zufallsvariable. Für alle  $a \in \mathbb{R}$  gilt dann*

$$E((X - a)^2) \geq \text{Var}(X),$$

wobei Gleichheit genau dann eintritt, wenn  $a = E(X)$  ist.

## 1.6 Klassische Verteilungen

Im folgenden stellen wir einige klassische Wahrscheinlichkeitsverteilungen und ihre Kenngrößen, sowie ihre Einsatzgebiete vor; die Begründung für letzteres folgt zum Teil erst später, wenn wir die nötigen Hilfsmittel entwickelt haben. Weitere wichtige Verteilungen werden wir in der Statistik kennenlernen.

**Caveat 1.53** (Existenz von Erwartungswert und Varianz). Im allgemeinen existieren Erwartungswert bzw. Varianz von Wahrscheinlichkeitsverteilungen auf  $(\mathbb{R}, B(\mathbb{R}))$  *nicht* unbedingt!

### 1.6.0 Deterministische Verteilung

- *Beschreibung.* Sei  $x \in \mathbb{R}$ . Die in  $x$  konzentrierte atomare Wahrscheinlichkeitsverteilung  $\delta_x$  auf  $(\mathbb{R}, B(\mathbb{R}))$  (Definition 1.6) stimmt mit der durch die Zähldichte

$$\begin{aligned} \{x\} &\longrightarrow [0, 1] \\ x &\longmapsto 1 \end{aligned}$$

auf  $(\mathbb{R}, B(\mathbb{R}))$  gegebenen Wahrscheinlichkeitsverteilung überein.

- *Erwartungswert.* Der Erwartungswert von  $\delta_x$  ist  $x$ .
- *Varianz.* Die Varianz von  $\delta_x$  ist 0 (Proposition 1.48).
- *Anwendung.* Man verwendet atomare Wahrscheinlichkeitsmaße zur Modellierung von deterministischen Prozessen.

### 1.6.1 Gleichverteilung

- *Diskrete Gleichverteilungen:*
  - *Beschreibung.* Sei  $n \in \mathbb{N}_{>0}$ . Die (diskrete) Gleichverteilung (bzw. Laplaceverteilung)  $U_{\{1, \dots, n\}}^d$  auf  $\{1, \dots, n\} \subset \mathbb{R}$  (Definition 1.7) stimmt mit dem Wahrscheinlichkeitsmaß auf  $(\{1, \dots, n\}, \text{Pot}(\{1, \dots, n\}))$  [bzw.  $(\mathbb{R}, B(\mathbb{R}))$ ] überein, das durch die Zähldichte

$$\begin{aligned} \{1, \dots, n\} &\longrightarrow [0, 1] \\ j &\longmapsto \frac{1}{n} \end{aligned}$$

gegeben ist.

- *Erwartungswert.* Der Erwartungswert von  $U_{\{1, \dots, n\}}^d$  ist  $\frac{n+1}{2}$ .
- *Varianz.* Die Varianz von  $U_{\{1, \dots, n\}}^d$  ist  $\frac{n^2-1}{12}$ .

- *Anwendung.* Anwendungen sind z.B. der faire Münz- bzw. Würfelwurf etc.
- *Stetige Gleichverteilungen:*
  - *Beschreibung.* Seien  $a, b \in \mathbb{R}$  mit  $a < b$ . Die Gleichverteilung  $U_{[a,b]}$  auf  $[a, b]$  (Definition 1.8) stimmt mit dem Wahrscheinlichkeitsmaß auf  $([a, b], B([a, b]))$  [bzw.  $(\mathbb{R}, B(\mathbb{R}))$ ] überein, das durch die  $\lambda^1$ -Wahrscheinlichkeitsdichte

$$\frac{1}{b-a} \cdot \chi_{[a,b]}$$

gegeben ist.

- *Erwartungswert.* Der Erwartungswert von  $U_{[a,b]}$  ist  $\frac{a+b}{2}$ .
- *Varianz.* Die Varianz von  $U_{[a,b]}$  ist  $\frac{(b-a)^2}{12}$ .
- *Anwendung.* Zum Beispiel kann man mithilfe dieser Gleichverteilungen durch geeignete Zufallsvariablen viele andere Verteilungen simulieren; dies wird beim sogenannten Sampling genutzt.

### 1.6.2 Bernoulli-/Binomialverteilung

- *Definition.* Sei  $n \in \mathbb{N}_{>0}$  und sei  $p \in [0, 1]$ . Die Wahrscheinlichkeitsverteilung auf  $(\{0, \dots, n\}, \text{Pot}(\{0, \dots, n\}))$  [bzw.  $(\mathbb{R}, B(\mathbb{R}))$ ], die durch die Zähldichte (binomische Formel!)

$$\begin{aligned} \{0, \dots, n\} &\longrightarrow [0, 1] \\ j &\longmapsto \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} \end{aligned}$$

gegeben ist, heißt *Binomialverteilung* zu den Parametern  $n$  und  $p$  und wird mit  $B(n, p)$  bezeichnet. Im Fall  $n = 1$  nennt man diese Verteilung auch *Bernoulliverteilung* zum Parameter  $p$ .

- *Erwartungswert.* Der Erwartungswert von  $B(n, p)$  ist  $n \cdot p$ .
- *Varianz.* Die Varianz von  $B(n, p)$  ist  $n \cdot p \cdot (1-p)$ .
- *Anwendung.* Die Bernoulliverteilung modelliert z.B. einen (unfairen) Münzwurf. Die Binomialverteilung modelliert die „Summe“ „unabhängiger“ (unfairer) Münzwürfe.

### 1.6.3 Poissonverteilung

- *Definition.* Sei  $\lambda \in \mathbb{R}_{>0}$ . Die *Poissonverteilung*  $\text{Poi}(\lambda)$  auf  $\mathbb{N} \subset \mathbb{R}$  ist die Wahrscheinlichkeitsverteilung auf  $(\mathbb{N}, \text{Pot}(\mathbb{N}))$  [bzw.  $(\mathbb{R}, B(\mathbb{R}))$ ], die durch die Zähldichte (Exponentialreihe!)

$$\begin{aligned} \mathbb{N} &\longrightarrow [0, 1] \\ k &\longmapsto e^{-\lambda} \cdot \frac{\lambda^k}{k!} \end{aligned}$$

gegeben ist.

- *Erwartungswert.* Der Erwartungswert von  $\text{Poi}(\lambda)$  ist  $\lambda$ .
- *Varianz.* Die Varianz von  $\text{Poi}(\lambda)$  ist  $\lambda$ .
- *Anwendung.* Poissonverteilungen eignen sich, um Zählvorgänge „seltener“ „unabhängiger“ Ereignisse (z.B. Anzahl von Mutationen in der DNA einer Zelle) zu modellieren. Ein erster Schritt zur Begründung ist die folgende Approximationseigenschaft.

**Satz 1.54** (Poisson-Approximation der Binomialverteilung). *Sei  $\lambda \in \mathbb{R}_{>0}$  und sei  $(p_n)_{n \in \mathbb{N}}$  eine Folge in  $[0, 1]$  mit  $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda$ . Dann gilt für alle  $k \in \mathbb{N}$ , dass*

$$\lim_{n \rightarrow \infty} B(n, p_n)(\{k\}) = \text{Poi}(\lambda)(\{k\}).$$

#### 1.6.4 Geometrische und Exponentialverteilung

- *Geometrische Verteilungen:*
  - *Definition.* Sei  $p \in (0, 1)$ . Die Wahrscheinlichkeitsverteilung auf  $(\mathbb{N}, \text{Pot}(\mathbb{N}))$  [bzw.  $(\mathbb{R}, B(\mathbb{R}))$ ], die durch die Zähldichte (geometrische Reihe!)

$$\begin{aligned} \mathbb{N} &\longrightarrow [0, 1] \\ k &\longmapsto p \cdot (1 - p)^k \end{aligned}$$

gegeben ist, heißt *geometrische Verteilung zum Parameter  $p$*  und wird mit  $G(p)$  bezeichnet.

- *Erwartungswert.* Der Erwartungswert von  $G(p)$  ist  $\frac{1-p}{p}$ .
- *Varianz.* Die Varianz von  $G(p)$  ist  $\frac{1-p}{p^2}$ .
- *Anwendung.* Die geometrische Verteilung tritt als Wartezeitverteilung bei „unabhängiger“ Wiederholung von Bernoulli-Experimenten auf (Anzahl der Fehlversuche vor dem ersten „Erfolg“).
- *Exponentialverteilungen:*
  - *Definition.* Sei  $\lambda \in \mathbb{R}_{>0}$ . Das Wahrscheinlichkeitsmaß auf  $(\mathbb{R}_{>0}, B(\mathbb{R}_{>0}))$  [bzw.  $(\mathbb{R}, B(\mathbb{R}))$ ], das durch die  $\lambda^1$ -Dichte

$$\begin{aligned} \mathbb{R}_{>0} [\text{bzw. } \mathbb{R}] &\longrightarrow \mathbb{R}_{\geq 0} \\ x &\longmapsto \lambda \cdot e^{-\lambda \cdot x} \cdot \chi_{(0, \infty)}(x) \end{aligned}$$

gegeben ist, heißt *Exponentialverteilung zum Parameter  $\lambda$*  und wird mit  $\text{Exp}(\lambda)$  bezeichnet.

- *Erwartungswert.* Der Erwartungswert von  $\text{Exp}(\lambda)$  ist  $\frac{1}{\lambda}$ .
- *Varianz.* Die Varianz von  $\text{Exp}(\lambda)$  ist  $\frac{1}{\lambda^2}$ .
- *Anwendung.* Exponentialverteilungen treten bei der Modellierung von Warteprozessen mit „kontinuierlicher“ Zeit auf bzw. bei der Modellierung von Lebensdauern (Bemerkung 2.32).

#### 1.6.5 Normalverteilung

Die Normalverteilung ist eine der wichtigsten Verteilungen in der Wahrscheinlichkeitstheorie und Statistik.

**Proposition 1.55** ((mehrdimensionale) Normalverteilung). Sei  $n \in \mathbb{N}_{>0}$ , sei  $C \in \mathbb{R}^{n \times n}$  symmetrisch und positiv definit, und sei  $a \in \mathbb{R}^n$ . Dann ist

$$f_{N(a,C)}: \mathbb{R}^n \longrightarrow \mathbb{R}_{\geq 0}$$

$$x \longmapsto \frac{1}{\sqrt{(2 \cdot \pi)^n \cdot \det(C)}} \cdot \exp\left(-\frac{1}{2} \cdot (x - a)^\top \cdot C^{-1} \cdot (x - a)\right)$$

eine Wahrscheinlichkeitsdichte auf  $(\mathbb{R}^n, B(\mathbb{R}^n), \lambda^n)$ . Man nennt das von dieser  $\lambda^n$ -Wahrscheinlichkeitsdichte auf  $(\mathbb{R}^n, B(\mathbb{R}^n))$  induzierte Wahrscheinlichkeitsmaß  $n$ -dimensionale Normalverteilung mit den Parametern  $a$  und  $C$  und bezeichnet diese mit dem Symbol  $N(a, C)$ .

Der Beweis verwendet die Choleskyzerlegung (eine Folgerung aus der Hauptachsentransformation) von symmetrischen positiv definiten Matrizen um dann mit Hilfe einer geeigneten affinen Transformation das zu berechnende Integral auf das klassische Integral

$$\int_{\mathbb{R}} e^{-\frac{1}{2} \cdot x^2} d\lambda^1(x) = \sqrt{2 \cdot \pi}$$

zurückzuführen.

**Korollar 1.56** ((zum Beweis) affine Transformationen von standardnormalverteilten Zufallsvariablen). Sei  $n \in \mathbb{N}$ , sei  $T \in \text{GL}(n, \mathbb{R})$  und sei  $a \in \mathbb{R}^n$ . Ist  $X$  eine  $(\mathbb{R}^n, B(\mathbb{R}^n))$ -wertige Zufallsvariable, die  $N(0, E_n)$ -verteilt ist, so gilt: Die  $(\mathbb{R}^n, B(\mathbb{R}^n))$ -wertige Zufallsvariable  $T \cdot X + a$  ist  $N(a, T \cdot T^\top)$ -verteilt. (Hierbei bezeichnet  $E_n$  die  $n$ -dimensionale Einheitsmatrix.)

**Proposition 1.57** (Erwartungswert und Varianz von eindimensionalen Normalverteilungen). Sei  $a \in \mathbb{R}$  und  $c \in \mathbb{R}_{>0}$ . Dann existieren Erwartungswert und Varianz von  $N(a, c)$ . Der Erwartungswert von  $N(a, c)$  ist  $a$  und die Varianz von  $N(a, c)$  ist  $c$ .

**Caveat 1.58.** Da man Stammfunktionen von

$$\mathbb{R} \longrightarrow \mathbb{R}$$

$$x \longmapsto e^{-x^2}$$

nicht geschlossen durch elementare Funktionen ausdrücken kann, kann man auch die Verteilungsfunktion  $F_{N(0,1)}$  der eindimensionalen Standardnormalverteilung nicht geschlossen durch elementare Funktionen ausdrücken (!). Analog lassen sich auch die Quantile von Normalverteilungen nicht geschlossen explizit, sondern nur näherungsweise, berechnen; daher gibt es für die Normalverteilungen Tabellen der Quantile.

Die „normalisierte“ Summe vieler „unabhängiger“ und identisch verteilter Zufallsvariablen (mit existierender Varianz) „konvergiert“ nach dem zentralen Grenzwertsatz (Satz 3.15) gegen eine Normalverteilung. Daher tritt die Normalverteilung in vielen solchen Situationen in der Modellierung auf.

**Caveat 1.59.** Oft wird jedoch in der Modellierung auch dann die Normalverteilung verwendet, wenn dies eigentlich nicht gerechtfertigt ist! Die entsprechenden Modelle haben dann also keine Aussagekraft für die ursprünglichen praktischen Probleme!

## 2 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

Wir wollen nun einen Formalismus einführen, der es uns erlaubt, darüber zu sprechen, ob zufällige Ereignisse/Prozesse voneinander im Sinne der Wahrscheinlichkeitstheorie unabhängig sind bzw. wie stark eine solche (Un)Abhängigkeit ist.

### 2.1 Stochastische Unabhängigkeit

Zwei Ereignisse  $A$  und  $B$  in einem Wahrscheinlichkeitsraum  $(\Omega, S, P)$  werden aus dem Blickwinkel der Stochastik als „unabhängig“ angesehen, wenn das Eintreten des Ereignisses  $A$  nichts an der Wahrscheinlichkeit dafür, dass das Ereignis  $B$  eintritt, ändert (und umgekehrt) – d.h., falls „ $P(A \cap B)/P(A) = P(B)$ “ ist. Man definiert daher (symmetrischer und die Division vermeidend):

**Definition 2.1** (stochastische Unabhängigkeit zweier Ereignisse). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $A, B \in S$ . Dann heißen die Ereignisse  $A$  und  $B$  *stochastisch unabhängig* (bezüglich  $P$ ), wenn

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

**Bemerkung 2.2.** Ereignisse, die fast sicher oder fast nie eintreten, sind von jedem anderen Ereignis stochastisch unabhängig.

**Caveat 2.3.** Stochastische Unabhängigkeit ist *nicht* dasselbe wie kausale Unabhängigkeit der entsprechenden Ereignisse!

Wir verallgemeinern den Begriff der stochastischen Unabhängigkeit auf größere Ereignissysteme und auf Zufallsvariablen, indem wir uns auf geeignete endliche Teilsysteme zurückziehen; dies ist formal ähnlich zur Definition linear unabhängiger Familien von Vektoren.

**Definition 2.4** (stochastisch unabhängig). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $I$  eine Menge.

- Eine Familie  $(A_i)_{i \in I}$  von Ereignissen aus  $S$  heißt *stochastisch unabhängig* (bezüglich  $P$ ), wenn folgendes gilt: Für alle endlichen Teilmengen  $J \subset I$  ist

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

- Eine Familie  $(S_i)_{i \in I}$  von Teilmengen von  $S$  heißt *stochastisch unabhängig* (bezüglich  $P$ ), wenn folgendes gilt: Alle Familien  $(A_i)_{i \in I} \in \prod_{i \in I} S_i$  sind im obigen Sinne stochastisch unabhängig.
- Eine Familie  $(X_i: (\Omega, S) \rightarrow (\Omega_i, S_i))_{i \in I}$  von Zufallsvariablen auf  $(\Omega, S, P)$  heißt *stochastisch unabhängig* (bezüglich  $P$ ), wenn die Familie  $(X_i^{-1}(S_i))_{i \in I}$  stochastisch unabhängig ist; dabei verwenden wir zu  $i \in I$  die Notation

$$X_i^{-1}(S_i) = \{X_i^{-1}(A) \mid A \in S_i\}.$$

**Caveat 2.5** (paarweise stochastische Unabhängigkeit vs. stochastische Unabhängigkeit). Wie im Fall linearer (oder algebraischer ...) Unabhängigkeit genügt paarweise stochastische Unabhängigkeit im allgemeinen *nicht* für stochastische Unabhängigkeit!

Im nächsten Abschnitt werden wir systematisch untersuchen wie man stochastisch unabhängige Familien von Ereignissen/Ereignissystemen/Zufallsvariablen konstruieren kann (was insbesondere viele Beispiele liefert). Wir geben nun noch einige grundlegende Eigenschaften von stochastisch unabhängigen Familien an:

**Proposition 2.6** (schnitt-stabile Erzeugendensysteme und stochastische Unabhängigkeit). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $I$  eine Menge und sei  $(T_i)_{i \in I}$  eine stochastisch unabhängige Familie schnitt-stabiler Teilmengen von  $S$ . Dann ist auch die Familie  $(\sigma(T_i))_{i \in I}$ , der von den  $T_i$  erzeugten  $\sigma$ -Algebren auf  $\Omega$ , stochastisch unabhängig.

Der Beweis dieser Proposition beruht auf denselben Methoden wie der Beweis des Maßeindeutigkeitssatzes (Satz 1.12); alternativ kann man diese Proposition auch auf den Maßeindeutigkeitssatz zurückführen.

Insbesondere erhält man die folgende Charakterisierung von stochastischer Unabhängigkeit für reellwertige Zufallsvariablen (analog kann man die obige Proposition auch verwenden, um stochastische Unabhängigkeit diskret verteilter reellwertige Zufallsvariablen konkreter zu charakterisieren):

**Korollar 2.7** (Charakterisierung stochastischer Unabhängigkeit reellwertiger Zufallsvariablen). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $I$  eine Menge und sei  $(X_i)_{i \in I}$  eine Familie von reellwertigen Zufallsvariablen auf einem Wahrscheinlichkeitsraum  $(\Omega, S, P)$ . Dann ist die Familie  $(X_i)_{i \in I}$  genau dann stochastisch unabhängig, wenn folgendes gilt: Für alle endlichen Teilmengen  $J \subset I$  und alle Folgen  $(a_j)_{j \in J}$  reeller Zahlen ist

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

**Proposition 2.8** (messbare Verarbeitung stochastisch unabhängiger Zufallsvariablen). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $I$  eine Menge und sei  $(X_i: (\Omega, S) \rightarrow (\Omega_i, S_i))_{i \in I}$  eine stochastisch unabhängige Familie von Zufallsvariablen auf  $(\Omega, S, P)$ . Sei außerdem  $(f_i: (\Omega_i, S_i) \rightarrow (\Omega'_i, S'_i))_{i \in I}$  eine Familien messbarer Abbildungen. Dann ist auch die Familie der Zufallsvariablen  $(f_i \circ X_i)_{i \in I}$  auf  $(\Omega, S, P)$  stochastisch unabhängig.

## 2.2 Stochastische Unabhängigkeit und Produkte

Wir geben nun eine Charakterisierung für stochastische Unabhängigkeit von Familien von Zufallsvariablen mithilfe von Produkträumen und konstruieren stochastisch unabhängige Familien von Zufallsvariablen mithilfe von Produkträumen. Außerdem lassen sich über diesen Zugang viele Eigenschaften stochastisch unabhängiger Familien von Zufallsvariablen herleiten.

Wir beginnen mit dem nötigen theoretischen Hintergrund über Produkträume. Als ersten Schritt betrachten wir die zugehörige messbare Struktur:



**Definition 2.9** (Produkt- $\sigma$ -Algebra). Sei  $I$  eine Menge und sei  $((\Omega_i, S_i))_{i \in I}$  eine Familie messbarer Räume.

- Ist  $J \subset I$  nicht-leer, so schreiben wir

$$\begin{aligned} \pi_J: \prod_{i \in I} \Omega_i &\longrightarrow \prod_{i \in J} \Omega_i \\ \omega &\longmapsto (\omega_i)_{i \in J} \end{aligned}$$

für die zugehörige Projektion.

- Ist  $J \subset I$  endlich und nicht-leer und ist  $(A_i)_{i \in J} \in \prod_{i \in J} S_i$ , so heißt das Urbild  $\pi_J^{-1}(\prod_{i \in J} A_i) \subset \prod_{i \in I} \Omega_i$  *Zylindermenge* zu  $(A_i)_{i \in J}$ .
- Die *Produkt- $\sigma$ -Algebra* von  $((\Omega_i, S_i))_{i \in I}$  auf  $\prod_{i \in I} \Omega_i$  ist definiert als

$$\bigotimes S_i := \sigma \left( \left\{ \pi_J^{-1} \left( \prod_{i \in J} A_i \right) \mid J \subset I \text{ endlich, } J \neq \emptyset, (A_i)_{i \in J} \in \prod_{i \in J} S_i \right\} \right)$$

(d.h. als die von den Zylindermengen erzeugte  $\sigma$ -Algebra auf  $\prod_{i \in I} \Omega_i$ ).

**Bemerkung 2.10** (endliche Produkte abzählbarer diskreter messbarer Räume). Sei  $I$  eine *endliche* Menge und sei  $(\Omega_i)_{i \in I}$  eine Familie von (höchstens) abzählbaren Mengen. Dann ist

$$\bigotimes_{i \in I} \text{Pot}(\Omega_i) = \text{Pot} \left( \prod_{i \in I} \Omega_i \right).$$

**Caveat 2.11.**

- Im allgemeinen sind *nicht* alle Elemente von Produkt- $\sigma$ -Algebren kartesische Produkte von messbaren Mengen aus den Faktoren.
- Im allgemeinen sind Einpunktmengen in Produkten *nicht* in der Produkt- $\sigma$ -Algebra, selbst wenn alle  $\sigma$ -Algebren auf den Faktoren diskret sind(!).

Analog zu Produkten in anderen Kategorien (z.B. in der Topologie, Algebra, ...) erfüllt die Produkt- $\sigma$ -Algebra die universelle Eigenschaft für Produkte in der Kategorie der messbaren Räume:

**Proposition 2.12** (universelle Eigenschaft der Produkt- $\sigma$ -Algebra). Sei  $I$  eine Menge und sei  $((\Omega_i, S_i))_{i \in I}$  eine Familie messbarer Räume. Sei  $(\Omega, S) := (\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} S_i)$  der zugehörige Produktraum und sei  $(\pi_i: \Omega \rightarrow \Omega_i)_{i \in I}$  die entsprechende Familie der Projektionen. Dann besitzt das Paar  $((\Omega, S), (\pi_i)_{i \in I})$  die folgenden Eigenschaften:

1. Für alle  $i \in I$  ist  $\pi_i: \Omega \rightarrow \Omega_i$  bezüglich  $S$  und  $S_i$  messbar.
2. Ist  $(\Omega', S')$  ein messbarer Raum und ist  $(f_i: (\Omega', S') \rightarrow (\Omega_i, S_i))_{i \in I}$  eine Familie messbarer Abbildungen, so gibt es genau eine bezüglich  $S'$  und  $S$  messbare Abbildung  $f: \Omega' \rightarrow \Omega$  mit der Eigenschaft, dass für alle  $i \in I$  gilt, dass  $\pi_i \circ f = f_i$  ist.

Insbesondere gilt: Ist  $((\tilde{\Omega}, \tilde{S}), (\tilde{\pi}_i: \tilde{\Omega} \rightarrow \Omega_i)_{i \in I})$  ein weiteres Paar, das die Eigenschaften 1 und 2 erfüllt, so gibt es genau einen messbaren Isomorphismus  $(\Omega, S) \rightarrow (\tilde{\Omega}, \tilde{S})$ , der mit  $(\pi_i)_{i \in I}$  und  $(\tilde{\pi}_i)_{i \in I}$  verträglich ist. Ein messbarer Isomorphismus zwischen messbaren Räumen ist dabei eine messbare Abbildung, die ein messbares Inverses besitzt.

Insbesondere zeigt die universelle Eigenschaft, dass eine Abbildung in ein Produkt messbarer Räume genau dann bezüglich der Produkt- $\sigma$ -Algebra messbar ist, wenn die Komposition mit allen Projektionen auf die einzelnen Faktoren messbar ist.

Als zweiten Schritt konstruieren wir nun geeignete Wahrscheinlichkeitsmaße auf Produkten von Wahrscheinlichkeitsräumen:

**Satz 2.13** (Satz von Andersen-Jessen und Definition des Produktmaßes). *Sei  $I$  eine Menge und sei  $((\Omega_i, S_i, P_i))_{i \in I}$  eine Familie von Wahrscheinlichkeitsräumen. Dann gibt es genau ein Wahrscheinlichkeitsmaß  $P$  auf  $(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} S_i)$  mit der folgenden Eigenschaft: Für alle endlichen nicht-leeren Teilmengen  $J \subset I$  und alle  $(A_i)_{i \in J} \in \prod_{i \in J} S_i$  gilt*

$$P\left(\pi_J^{-1}\left(\prod_{i \in J} A_i\right)\right) = \prod_{i \in J} P_i(A_i).$$

Man nennt  $P$  das Produktmaß von  $(P_i)_{i \in I}$  und verwendet die Notationen

$$\bigotimes_{i \in I} P_i := P \quad \text{und} \quad \bigotimes_{i \in I} (\Omega_i, S_i, P_i) = \left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} S_i, P\right).$$

Die Eindeutigkeitsaussage folgt aus dem Maßeindeutigkeitssatz 1.12; die Existenzaussage folgt mit einer geschickten Anwendung des Maßfortsetzungssatzes von Carathéodory.

In gewissen Spezialfällen können wir das Produktmaß auch expliziter angeben:

**Proposition 2.14** (Produktmaß endlich viele diskreter Wahrscheinlichkeitsmaße). *Sei  $I$  eine endliche Menge und zu jedem  $i \in I$  sei ein diskreter Wahrscheinlichkeitsraum  $(\Omega_i, S_i, P_i)$  gegeben, der jeweils durch eine Zähldichte  $p_i: \Omega'_i \rightarrow [0, 1]$  (wobei  $\Omega'_i \subset \Omega_i$  höchstens abzählbar ist) beschrieben ist. Dann ist das Produktmaß  $\bigotimes_{i \in I} P_i$  auf  $(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} S_i)$  durch die Zähldichte*

$$p: \prod_{i \in I} \Omega'_i \rightarrow [0, 1]$$

$$\omega \mapsto \prod_{i \in I} p_i(\omega_i)$$

gegeben.

Insbesondere zeigt diese Proposition, dass das endliche Produkt von Laplaceverteilungen wieder eine Laplaceverteilung ergibt.

**Caveat 2.15.** Unendliche Produkte von diskreten Wahrscheinlichkeitsmaßen sind im allgemeinen *nicht* durch eine Zähldichte gegeben.

**Bemerkung 2.16** (endliche Produkte und Wahrscheinlichkeitsdichten). Analog zur obigen Proposition kann man allgemeiner mithilfe der entsprechenden Version des Satzes von Fubini zeigen, dass endliche Produkte von durch Wahrscheinlichkeitsdichten (bezüglich  $\sigma$ -endlichen Maßen) gegebenen Wahrscheinlichkeitsmaßen auch wieder durch eine geeignete Wahrscheinlichkeitsdichte (nämlich das „Produkt“ der ursprünglichen Dichten) bezüglich des Produktmaßes der den Wahrscheinlichkeitsdichten auf den Faktoren unterliegenden Maße gegeben ist.

Wir wenden uns nun wieder unserem eigentlichen Ziel zu – nämlich dem Verständnis stochastisch unabhängiger Familien von Zufallsvariablen mithilfe von Produkträumen; dies sind Folgerungen des Satzes von Andersen-Jessen (Satz 2.13):

**Korollar 2.17** (stochastische Unabhängigkeit der Projektionen). *Sei  $I$  eine Menge und sei  $((\Omega_i, S_i, P_i))_{i \in I}$  eine Familie von Wahrscheinlichkeitsräumen. Dann ist die Familie  $(\pi_i: \prod_{j \in I} \Omega_j \rightarrow \Omega_i)_{i \in I}$  der Projektionen bezüglich dem Produktmaß  $\bigotimes_{i \in I} P_i$  auf  $(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} S_i)$  stochastisch unabhängig.*

Insbesondere erlaubt uns dies, jede Familie von Verteilungen durch eine stochastisch unabhängige Familie von Zufallsvariablen zu realisieren, was für die Modellierung vieler Zufallsexperimente von großer Wichtigkeit ist:

**Korollar 2.18** (Existenz stochastisch unabhängiger Zufallsvariablen mit gegebenen Verteilungen). *Sei  $I$  eine Menge und sei  $((\Omega_i, S_i, P_i))_{i \in I}$  eine Familie von Wahrscheinlichkeitsräumen. Dann existiert ein Wahrscheinlichkeitsraum  $(\Omega, S, P)$  und eine stochastisch unabhängige Familie  $(X_i: (\Omega, S) \rightarrow (\Omega_i, S_i))_{i \in I}$  von Zufallsvariablen auf  $(\Omega, S, P)$ , so dass  $P_{X_i} = P_i$  für alle  $i \in I$  gilt.*

Außerdem können wir mit diesen Methoden stochastisch unabhängige Familien von Zufallsvariablen mithilfe des Produktmaßes charakterisieren:

**Satz 2.19** (Charakterisierung stochastischer Unabhängigkeit von Zufallsvariablen). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum,  $I$  eine Menge, sei  $(X_i: (\Omega, S) \rightarrow (\Omega_i, S_i))_{i \in I}$  eine Familie von Zufallsvariablen und sei  $X: (\Omega, S) \rightarrow (\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} S_i)$  die durch die universelle Eigenschaft gegebene messbare Abbildung in den Produktraum. Dann ist die Familie  $(X_i)_{i \in I}$  genau dann bezüglich  $P$  stochastisch unabhängig, wenn*

$$P_X = \bigotimes_{i \in I} P_{X_i}$$

*ist.*

Insbesondere hängt zum Beispiel die Verteilung der Summe von stochastisch unabhängigen Zufallsvariablen nur von den Verteilungen der einzelnen Zufallsvariablen ab – etwas, das im allgemeinen (also ohne stochastische Unabhängigkeit) nicht zutrifft.

**Korollar 2.20** (Summenverteilungen stochastisch unabhängiger Zufallsvariablen). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $J$  eine endliche Menge und sei  $(X_j)_{j \in J}$  eine stochastisch unabhängige Familie reellwertiger Zufallsvariablen auf  $(\Omega, S, P)$ . Dann ist*

$$P_{\sum_{j \in J} X_j} = \left( \bigotimes_{j \in J} P_{X_j} \right)_Z,$$

*wobei*

$$\begin{aligned} Z: \prod_J \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \sum_{j \in J} x_j \end{aligned}$$

*die entsprechende reelle Summenabbildung ist.*

## 2.3 Unkorreliertheit

Ein wichtiger mit stochastischer Unabhängigkeit verwandter Begriff ist die sogenannte Unkorreliertheit bzw. der Korrelationskoeffizient. Das Ziel dabei ist, die Stärke der (linearen) Abhängigkeit zwischen zwei reellwertigen Zufallsvariablen durch eine reelle Zahl zu messen (im allgemeinen ist dies jedoch sehr grob!).

**Definition 2.21** (Kovarianz, (un)korreliert, positiv/negativ korreliert). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $X, Y$  reellwertige Zufallsvariablen auf  $(\Omega, S, P)$  mit der Eigenschaft, dass  $X, Y$  und  $X \cdot Y$  integrierbar sind.

– Dann heißt

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

*Kovarianz von  $X$  und  $Y$  (bezüglich  $P$ ).*

- Ist  $\text{Cov}(X, Y) \geq 0$ , so heißen  $X$  und  $Y$  *positiv korreliert*.
- Ist  $\text{Cov}(X, Y) = 0$ , so heißen  $X$  und  $Y$  *unkorreliert*.
- Ist  $\text{Cov}(X, Y) \leq 0$ , so heißen  $X$  und  $Y$  *negativ korreliert*.

Wir werden sehen (Proposition 2.28), dass die Kovarianz (im wesentlichen) die bezüglich der quadratischen Abweichung beste lineare Approximation an die Abhängigkeit zwischen den betrachteten Zufallsvariablen darstellt.

**Proposition 2.22** (grundlegenden Eigenschaften der Kovarianz). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $X, Y$  reellwertige Zufallsvariablen auf  $(\Omega, S, P)$ .

1. Ist  $X$  quadratintegrierbar, so ist  $\text{Cov}(X, X) = \text{Var}(X)$ .
2. Sind  $X, Y, X \cdot Y$  integrierbar, so gilt

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

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

und für alle  $a, b, c, d \in \mathbb{R}$  ist

$$\text{Cov}(a + b \cdot X, c + d \cdot Y) = b \cdot d \cdot \text{Cov}(X, Y).$$

3. Sind  $X$  und  $Y$  quadratintegrierbar, so ist  $X \cdot Y$  integrierbar (und somit auch  $X + Y$  quadratintegrierbar) und es gilt

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

sowie

$$(\text{Cov}(X, Y))^2 \leq \text{Var}(X) \cdot \text{Var}(Y).$$

Insbesondere bietet es sich daher an, die Kovarianz wie folgt zu normieren:

**Definition 2.23** (Korrelationskoeffizient). Seien  $X$  und  $Y$  quadratintegrierbare reellwertige Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum und sei  $\text{Var}(X) \neq 0$  und  $\text{Var}(Y) \neq 0$ . Der *Korrelationskoeffizient von  $X$  und  $Y$*  ist dann definiert als

$$\varrho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)} \in [-1, 1].$$

Unkorreliertheit ist eine schwächere Bedingung als stochastische Unabhängigkeit (falls die betrachteten Zufallsvariablen hinreichend integrierbar sind):

**Proposition 2.24** (stochastische Unabhängigkeit impliziert Unkorreliertheit). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $X, Y$  integrierbare reellwertige, stochastisch unabhängige Zufallsvariablen auf  $(\Omega, S, P)$ . Dann ist auch  $X \cdot Y$  integrierbar und  $X$  und  $Y$  sind unkorreliert.*

**Korollar 2.25** (Varianz von Summen unabhängiger Zufallsvariablen). *Seien  $X, Y$  stochastisch unabhängige quadratintegrierbare reellwertige Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum. Dann ist*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

**Caveat 2.26.** Unkorrelierte Zufallsvariablen sind jedoch im allgemeinen *nicht* stochastisch unabhängig!

**Caveat 2.27.** In der Statistik betrachtet man unter anderem Testverfahren, die überprüfen, ob gewisse Phänomene korreliert sind oder nicht. Dabei ist jedoch immer zu berücksichtigen, dass Korreliertheit höchstens ein Indiz für Kausalität sein kann, aber im allgemeinen *keinen* kausalen Zusammenhang impliziert!

Die Hauptmotivation hinter der Definition der Kovarianz bzw. des Korrelationskoeffizienten ist der folgende Sachverhalt:

**Proposition 2.28** (Kleinste-Quadrate-Regressionsgerade). *Seien  $X$  und  $Y$  quadratintegrierbare reellwertige Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum und es gelte  $\text{Var}(X) \neq 0$  und  $\text{Var}(Y) \neq 0$ . Für alle  $a, b \in \mathbb{R}$  gilt*

$$E((Y - a \cdot X - b)^2) \geq (1 - \varrho(X, Y)^2) \cdot \text{Var}(Y),$$

und Gleichheit gilt genau dann, wenn

$$a = \frac{\sigma(Y)}{\sigma(X)} \cdot \varrho(X, Y) \quad \text{und} \quad b = E(Y) - \frac{\sigma(Y)}{\sigma(X)} \cdot \varrho(X, Y) \cdot E(X)$$

ist.

Eine ganz zentrale Rolle werden stochastische Unabhängigkeit bzw. Unkorreliertheit bei den Konvergenzsätzen (Gesetze der großen Zahlen, zentraler Grenzwertsatz) spielen (Abschnitt 3).

## 2.4 Bedingte Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten erlauben es, Vorwissen über das Eintreten eines Ereignisses und die damit verbundene Neubewertung der Wahrscheinlichkeit des Eintretens anderer Ereignisse zu modellieren.

**Definition 2.29** (bedingte Wahrscheinlichkeit). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und seien  $A, B \in S$  mit  $P(B) > 0$ . Dann heißt

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

bedingte Wahrscheinlichkeit von  $A$  unter  $B$  (bezüglich  $P$ ).

**Proposition 2.30** (grundlegende Eigenschaften bedingter Wahrscheinlichkeiten). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum.

1. Ist  $B \in S$  mit  $P(B) > 0$ , so ist

$$\begin{aligned} P(\cdot | B) &: S \longrightarrow [0, 1] \\ A &\longmapsto P(A | B) \end{aligned}$$

ein Wahrscheinlichkeitsmaß auf  $(\Omega, S)$ .

2. Multiplikationsformel/Pfadregel. Sei  $n \in \mathbb{N}$ , seien  $A_1, \dots, A_n \in S$  und es gelte  $P(A_1 \cap \dots \cap A_{n-1}) > 0$ . Dann ist

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

3. Formel von der totalen Wahrscheinlichkeit/Fallunterscheidungsregel. Sei  $I$  eine höchstens abzählbare Menge und sei  $(A_i)_{i \in I}$  eine Familie von paarweise disjunkten Elementen von  $S$  mit  $\bigcup_{i \in I} A_i = \Omega$  und außerdem gelte  $P(A_i) > 0$  für alle  $i \in I$ . Dann gilt für alle  $A \in S$ , dass

$$P(A) = \sum_{i \in I} P(A_i) \cdot P(A | A_i).$$

4. Formel von Bayes. Sei  $I$  eine höchstens abzählbare Menge und sei  $(A_i)_{i \in I}$  eine Familie von paarweise disjunkten Elementen von  $S$  mit  $\bigcup_{i \in I} A_i = \Omega$  und außerdem gelte  $P(A_i) > 0$  für alle  $i \in I$ . Seien  $A, B \in S$  mit  $P(A) > 0$  und  $P(B) > 0$ . Dann gilt

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} = \frac{P(B | A) \cdot P(A)}{\sum_{i \in I} P(A_i) \cdot P(B | A_i)}.$$

Wichtige Anwendungsbeispiele für bedingte Wahrscheinlichkeiten sind zum Beispiel diagnostische Tests in der Medizin, das Ziegenproblem und die Gedächtnislosigkeit der Exponentialverteilung.

**Satz 2.31** (Gedächtnislosigkeit der Exponentialverteilung). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $X$  eine reellwertige Zufallsvariable auf  $(\Omega, S, P)$  mit folgenden Eigenschaften: Es ist  $P(X > 0) = 1$  und für alle  $t \in \mathbb{R}_{>0}$  gilt  $P(X > t) > 0$ . Dann ist  $X$  genau dann exponentialverteilt, wenn  $X$  im folgenden Sinne gedächtnislos ist: Für alle  $s, t \in \mathbb{R}_{>0}$  gilt

$$P(X > t + s | X > t) = P(X > s).$$

**Bemerkung 2.32** (Modellierung von Wartezeiten durch Exponentialverteilungen). Exponentialverteilungen haben also die folgende Eigenschaft: Sind  $s, t \in \mathbb{R}_{>0}$  und haben wir bereits die Zeit  $t$  auf das entsprechende Ereignis gewartet, so ist die Wahrscheinlichkeit, dass wir noch mindestens  $s$  warten müssen, genauso groß wie wenn wir noch gar nicht gewartet hätten. Und umgekehrt ist jede Verteilung mit dieser Eigenschaft bereits eine Exponentialverteilung. Daher eignen sich Exponentialverteilungen zur Modellierung gewisser Wartezeiten.

Unser Konzept von bedingter Wahrscheinlichkeit setzt immer voraus, dass die betrachtete Bedingung *positive* Wahrscheinlichkeit besitzt. Mit etwas aufwendigeren Konzepten (bedingte Erwartungswerte, bedingte Verteilungen) kann man unseren Begriff geeignet verallgemeinern. Dies ist vor allem dann wichtig, wenn man sich systematisch mit sogenannten stochastischen Prozessen (z.B. dem Poisson-Prozess oder der Brownschen Bewegung) beschäftigen möchte.

### 3 Gesetze der großen Zahlen und der zentrale Grenzwertsatz

Wir werden uns in diesem Kapitel mit dem folgenden Typ von Fragestellungen beschäftigen: Sei  $(X_n)_{n \in \mathbb{N}_{>0}}$  eine hinreichend unabhängige Folge von hinreichend integrierbaren reellwertigen Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum. Was passiert dann mit den normalisierten Summen  $\frac{1}{n} \cdot \sum_{k=1}^n X_k$  für „große“  $n$ ? „Konvergieren“ diese normalisierten Summen? Wenn ja, wogegen?

Insbesondere werden wir die folgenden Resultate, die Antworten auf diese Fragen geben, studieren:

- das schwache Gesetz der großen Zahlen
- das starke Gesetz der großen Zahlen
- der zentrale Grenzwertsatz.

Diese Resultate bilden insbesondere auch das Fundament der mathematischen Statistik.

#### 3.1 Das schwache Gesetz der großen Zahlen

Wir wollen nun den folgenden Eindruck wahrscheinlichkeitstheoretisch untermauern: Werfen wir (unabhängig) „oft“ eine Münze, so erwarten wir, dass die relative Häufigkeit

$$\frac{\text{Anzahl der Würfe, in denen „Kopf“ gefallen ist}}{\text{Gesamtanzahl der Würfe}}$$

von „Kopf“ die Wahrscheinlichkeit dafür, dass „Kopf“ fällt „approximiert“. Bzw. umgekehrt erlaubt dies eine Interpretation „abstrakter“ Wahrscheinlichkeiten als relative Häufigkeiten.

Wir beginnen mit einem vergleichsweise schwachen Konvergenzbegriff:

**Definition 3.1** (Konvergenz in Wahrscheinlichkeit/stochastische Konvergenz). Seien  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reellwertiger Zufallsvariablen auf  $(\Omega, S, P)$ . Dann *konvergiert*  $(X_n)_{n \in \mathbb{N}}$  *stochastisch* (bzw. *in Wahrscheinlichkeit*) *bezüglich*  $P$  gegen die reellwertige Zufallsvariable  $X$  auf  $(\Omega, S, P)$ , falls: Für alle  $\varepsilon \in \mathbb{R}_{>0}$  ist

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

In diesem Fall schreiben wir auch kurz  $X_n \xrightarrow[n \rightarrow \infty]{\text{stoch}} X$  oder  $X_n \xrightarrow[n \rightarrow \infty]{P} X$ .

**Satz 3.2** (schwaches Gesetz der großen Zahlen). Sei  $(X_n)_{n \in \mathbb{N}_{>0}}$  eine Folge quadratintegrierbarer reellwertiger paarweise unkorrelierter reellwertiger Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, S, P)$  mit

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \cdot \sum_{k=1}^n \text{Var}(X_k) = 0.$$

Dann erfüllt  $(X_n)_{n \in \mathbb{N}_{>0}}$  das schwache Gesetz der großen Zahlen, d.h.:

$$\frac{1}{n} \cdot \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow[n \rightarrow \infty]{\text{stoch}} 0.$$



**Korollar 3.3** (schwaches Gesetz der großen Zahlen für identisch verteilte Zufallsvariablen). Sei  $(X_n)_{n \in \mathbb{N}_{>0}}$  eine Folge quadratintegrierbarer reellwertiger paarweise unkorrelierter identisch verteilter Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum. Sei  $m := E(X_1)$  und zu  $n \in \mathbb{N}$  sei

$$S_n := \frac{1}{n} \cdot \sum_{k=1}^n X_k.$$

Dann gilt  $S_n \xrightarrow[n \rightarrow \infty]{stoch} m$ .

Der Mittelwert liegt bei häufiger hinreichend unabhängiger Wiederholung also mit hoher Wahrscheinlichkeit nahe am Erwartungswert.

Es gibt viele Varianten des schwachen Gesetzes der großen Zahlen; z.B. kann man unter stärkeren Unabhängigkeitsbedingungen die Integrierbarkeitsbedingungen etwas abschwächen.

Eine Anwendung des schwachen Gesetzes der großen Zahlen in der theoretischen Mathematik ist, dass es erlaubt, einen probabilistischen Beweis des Weierstraßschen Approximationssatzes zu geben (mithilfe von Bernsteinpolynomen).

### 3.2 Null-/Eins-Gesetze

Unser nächstes Ziel ist es, die Konvergenzaussagen aus dem schwachen Gesetz der großen Zahlen zu verbessern. Als Vorbereitung betrachten wir sogenannte Null-/Eins-Gesetze (die wir dann als technisches Hilfsmittel benötigen werden).

**Proposition 3.4** (Lemma von Borel-Cantelli). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(A_n)_{n \in \mathbb{N}}$  eine Folge in  $S$  und sei

$$A := \limsup_{n \rightarrow \infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{k \in \mathbb{N}_{\geq n}} A_k.$$

Dann gilt:

1. Falls die Reihe  $\sum_{n=0}^{\infty} P(A_n)$  konvergiert, ist  $P(A) = 0$ .
2. Falls die Reihe  $\sum_{n=0}^{\infty} P(A_n)$  nicht konvergiert und die Familie  $(A_n)_{n \in \mathbb{N}}$  stochastisch unabhängig ist, ist  $P(A) = 1$ .

Allgemeiner besagt das Kolmogorowsche Null-/Eins-Gesetz, dass „terminale“ Ereignisse (d.h. Ereignisse, die „nicht von endlichen Anfangsstücken abhängen“) immer nur fast sicher oder fast nie eintreten können.

**Definition 3.5** (terminale  $\sigma$ -Algebra, terminale Ereignisse). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $(X_n: (\Omega, S) \rightarrow (\Omega_n, S_n))_{n \in \mathbb{N}}$  eine Folge von Zufallsvariablen auf  $(\Omega, S, P)$ .

- Ist  $I \subset \mathbb{N}$  mit  $I \neq \emptyset$ , so schreiben wir

$$\begin{aligned} X_I: (\Omega, S) &\rightarrow \left( \prod_{i \in I} \Omega_i, \bigotimes_{i \in I} S_i \right) \\ \omega &\mapsto (X_i(\omega))_{i \in I} \end{aligned}$$

- für die von  $(X_i)_{i \in I}$  induzierte Zufallsvariable in den Produktraum.
- Die *terminale  $\sigma$ -Algebra* zu  $(X_n)_{n \in \mathbb{N}}$  ist definiert als

$$\sigma_\infty((X_n)_{n \in \mathbb{N}}) := \bigcap_{n \in \mathbb{N}} X_{\mathbb{N}_{\geq n}}^{-1} \left( \bigotimes_{k \in \mathbb{N}_{\geq n}} S_k \right).$$

- Elemente der terminalen  $\sigma$ -Algebra  $\sigma_\infty((X_n)_{n \in \mathbb{N}})$  heißen *terminale Ereignisse* bezüglich  $(X_n)_{n \in \mathbb{N}}$ .

**Satz 3.6** (Null-/Eins-Gesetz von Kolmogorov). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(X_n: (\Omega, S) \rightarrow (\Omega_n, S_n))_{n \in \mathbb{N}}$  eine stochastisch unabhängige Folge von Zufallsvariablen auf  $(\Omega, S, P)$  und sei  $A$  ein bezüglich  $(X_n)_{n \in \mathbb{N}}$  terminales Ereignis. Dann gilt*

$$P(A) = 0 \quad \text{oder} \quad P(A) = 1.$$

Der Beweis beruht darauf, dass man zeigt, dass terminale Ereignisse zu sich selbst stochastisch unabhängig sind, indem man die Koordinatenbereiche geschickt aufteilt und Proposition 2.6 anwendet.

### 3.3 Das starke Gesetz der großen Zahlen

Wir möchten nun das schwache Gesetz der großen Zahlen zu einem entsprechenden Satz mit einer stärkeren Konvergenzaussage verbessern.

**Definition 3.7** (fast sichere Konvergenz). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reellwertiger Zufallsvariablen auf  $(\Omega, S, P)$ . Die Folge  $(X_n)_{n \in \mathbb{N}}$  *konvergiert  $P$ -fast sicher* gegen eine reellwertige Zufallsvariable  $X$  auf  $(\Omega, S, P)$ , wenn

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

In diesem Fall schreibt man  $X_n \xrightarrow[n \rightarrow \infty]{P\text{-f.s.}} X$ .

**Proposition 3.8** (fast sichere Konvergenz impliziert stochastische Konvergenz). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reellwertiger Zufallsvariablen auf  $(\Omega, S, P)$  und sei  $X$  eine reellwertige Zufallsvariable auf  $(\Omega, S, P)$ . Gilt  $X_n \xrightarrow[n \rightarrow \infty]{P\text{-f.s.}} X$ , so folgt auch  $X_n \xrightarrow[n \rightarrow \infty]{\text{stoch}} X$ .*

Die Umkehrung gilt jedoch im allgemeinen *nicht*, d.h. nicht jede stochastisch konvergente Folge von Zufallsvariablen ist auch fast sicher konvergent.

**Satz 3.9** (starkes Gesetz der großen Zahlen). *Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum und sei  $(X_n)_{n \in \mathbb{N}_{>0}}$  eine Folge quadratintegrierbarer paarweise unkorrelierter reellwertiger Zufallsvariablen auf  $(\Omega, S, P)$ . Die Folge  $(\text{Var}(X_n))_{n \in \mathbb{N}_{>0}}$  sei beschränkt. Dann gilt*

$$\frac{1}{n} \cdot \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow[n \rightarrow \infty]{P\text{-f.s.}} 0.$$

Der Beweis dieser Version des starken Gesetzes der großen Zahlen beruht auf dem Lemma von Borel-Cantelli (Proposition 3.4) und der Betrachtung einer geeigneten Teilfolge.

Ähnlich wie im Fall des schwachen Gesetzes der großen Zahlen erhalten wir außerdem den folgenden Spezialfall (d.h., dass der „Mittelwert“ bei häufiger Wiederholung fast sicher punktweise gegen den Erwartungswert konvergiert):

**Korollar 3.10** (starkes Gesetz der großen Zahlen für identisch verteilte Zufallsvariablen). *Sei  $(X_n)_{n \in \mathbb{N}_{>0}}$  eine Folge quadratintegrierbarer reellwertiger paarweise unkorrelierter identisch verteilter Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, S, P)$ . Sei  $m := E(X_1)$  und zu  $n \in \mathbb{N}$  sei*

$$S_n := \frac{1}{n} \cdot \sum_{k=1}^n X_k.$$

*Dann gilt  $S_n \xrightarrow[n \rightarrow \infty]{P-f.s.} m$ .*

**Bemerkung 3.11.**

- Das starke Gesetz der großen Zahlen gilt auch mit den etwas schwächeren Voraussetzungen an die Varianzen wie im schwachen Gesetz der großen Zahlen. Der Beweis ist jedoch deutlich aufwendiger.
- Ebenso wie beim schwachen Gesetz der großen Zahlen kann man bei stärkeren Unabhängigkeitsvoraussetzungen die Integrierbarkeitsvoraussetzungen etwas abschwächen.

### 3.4 Der zentrale Grenzwertsatz

Wir wollen nun genauer verstehen, wie stark sich gewisse Mittelwerte von identisch verteilten Zufallsvariablen im Grenzwert um den Erwartungswert konzentrieren. Wir müssen daher Mittelwerte betrachten, bei denen nicht nur die Erwartungswerte, sondern auch die Varianzen normiert sind. Man betrachtet daher zu einer hinreichend unabhängigen Folge  $(X_n)_{n \in \mathbb{N}_{>0}}$  hinreichend integrierbarer identisch verteilter reellwertiger Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum die Folge

$$\left( \frac{1}{\sqrt{n}} \cdot \sum_{k=1}^n \frac{1}{\sqrt{\text{Var}(X_k)}} \cdot (X_k - E(X_k)) \right)_{n \in \mathbb{N}_{>0}}.$$

Es stellt sich heraus, dass in diesem Kontext der folgende Konvergenzbegriff und die Standardnormalverteilung eine zentrale Rolle spielen.

**Definition 3.12** (Verteilungskonvergenz). Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reellwertiger Zufallsvariablen und sei  $X$  eine reellwertige Zufallsvariable. Dann *konvergiert*  $(X_n)_{n \in \mathbb{N}}$  *in Verteilung gegen*  $X$ , wenn für alle  $x \in \mathbb{R}$ , in denen  $F_X$  stetig ist, gilt, dass

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

In diesem Fall schreiben wir  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ .

Man beachte dabei, dass Verteilungskonvergenz nur von den induzierten Verteilungen abhängt (da die Verteilungsfunktionen nur von den induzierten Verteilungen abhängen). Man schreibt daher gegebenenfalls auch Ausdrücke wie „ $X_n \xrightarrow[n \rightarrow \infty]{d} N(0,1)$ “.

Verteilungskonvergenz lässt sich alternativ auch als eine sogenannte schwache Konvergenz im Sinne der Funktionalanalysis charakterisieren:

**Proposition 3.13** (Charakterisierungen von Verteilungskonvergenz). *Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reellwertiger Zufallsvariablen und sei  $X$  eine reellwertige Zufallsvariable. Dann sind folgende Aussagen äquivalent:*

1. *Es gilt  $X_n \xrightarrow{d} X$ .*
2. *Für alle stetigen und beschränkte Funktionen  $f: \mathbb{R} \rightarrow \mathbb{R}$  gilt*

$$\lim_{n \rightarrow \infty} E(f \circ X_n) = E(f \circ X).$$

3. *Für alle  $f \in \tilde{C}^2(\mathbb{R}, \mathbb{R})$  gilt*

$$\lim_{n \rightarrow \infty} E(f \circ X_n) = E(f \circ X),$$

wobei  $\tilde{C}^2(\mathbb{R}, \mathbb{R})$  die Menge aller Funktionen  $f \in C^2(\mathbb{R}, \mathbb{R})$  bezeichne, die beschränkt sind und für die sowohl  $f'$  als auch  $f''$  beschränkt und gleichmäßig stetig sind.

**Bemerkung 3.14** (Vergleich mit anderen Konvergenzarten). Sei  $(\Omega, S, P)$  ein Wahrscheinlichkeitsraum, sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reellwertiger Zufallsvariablen auf  $(\Omega, S, P)$  und sei  $X$  eine reellwertige Zufallsvariable auf  $(\Omega, S, P)$ . Dann gilt:

- Gilt  $X_n \xrightarrow[n \rightarrow \infty]{P\text{-f.s.}} X$ , so folgt  $X_n \xrightarrow[n \rightarrow \infty]{\text{stoch}} X$ . Die Umkehrung gilt im allgemeinen *nicht*.
- Gilt  $X_n \xrightarrow[n \rightarrow \infty]{\text{stoch}} X$ , so folgt  $X_n \xrightarrow[n \rightarrow \infty]{d} X$ . Die Umkehrung gilt im allgemeinen *nicht*.

Betrachtet man eine stochastisch unabhängige Folge  $(X_n)_{n \in \mathbb{N}_{>0}}$  von  $B(1, 1/2)$ -verteilten Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum und die zugehörigen Histogramme (d.h. flächentreue Darstellungen der entsprechenden Zähldichten) der Folge

$$\left( \frac{1}{\sqrt{n}} \cdot \sum_{k=1}^n \frac{1}{\sqrt{\text{Var}(X_k)}} \cdot (X_k - E(X_k)) \right)_{n \in \mathbb{N}_{>0}},$$

so stellt man fest, dass sich diese Histogramme der  $\lambda^1$ -Dichte  $f_{N(0,1)}$  der Standardnormalverteilung  $N(0,1)$  annähern. Da Histogramme flächentreu sind, entspricht dies einer Aussage über Verteilungskonvergenz.

Dieses Beispiel ist eine Instanz eines ganz allgemeinen und für die Wahrscheinlichkeitstheorie und Statistik zentralen Phänomens:

**Satz 3.15** (Zentraler Grenzwertsatz). *Sei  $(X_n)_{n \in \mathbb{N}_{>0}}$  eine stochastisch unabhängige Folge von identisch verteilten quadratintegrierbaren reellwertigen Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum und es gelte  $\text{Var}(X_1) > 0$ . Dann gilt*

$$\frac{1}{\sqrt{n}} \cdot \sum_{k=1}^n \frac{1}{\sqrt{\text{Var}(X_k)}} \cdot (X_k - E(X_k)) \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

**Caveat 3.16** (Normalapproximation). In Anwendungen werden standardisierte Summen wie im zentralen Grenzwertsatz häufig durch Normalverteilungen ersetzt (!). Es ist allerdings zu berücksichtigen, dass der obige Satz *keine* Aussage über die Konvergenzgeschwindigkeit und *keine* Fehlerabschätzung liefert. So eine Ersetzung kann also fahrlässig sein.

Die folgende Eigenschaft der Standardnormalverteilung ist ein entscheidender Baustein des Beweises:

**Proposition 3.17** (Stabilität der Standardnormalverteilung). *Sei  $(X_n)_{n \in \mathbb{N}_{>0}}$  eine stochastisch unabhängige Folge standardnormalverteilter Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, S, P)$ . Dann gilt für alle  $n \in \mathbb{N}_{>0}$ , dass*

$$P_{\frac{1}{\sqrt{n}} \cdot \sum_{k=1}^n \frac{1}{\sqrt{\text{Var}(X_k)}} (X_k - E(X_k))} = N(0, 1).$$

*Insbesondere erfüllt die Folge  $(X_n)_{n \in \mathbb{N}_{>0}}$  somit die Aussage des zentralen Grenzwertsatzes.*

Umgekehrt folgt aus dem zentralen Grenzwertsatz, dass die Standardnormalverteilung die einzige reelle Wahrscheinlichkeitsverteilung mit existierender, positiver Varianz ist, die diese Stabilitätseigenschaft besitzt.

Ein Beweis des zentralen Grenzwertsatzes vergleicht das Konvergenzverhalten der gegebenen Folge mit einer weiteren Folge mithilfe der Charakterisierung von Verteilungskonvergenz über Testfunktionen in  $\tilde{C}^2(\mathbb{R}, \mathbb{R})$  (Proposition 3.13). Die entsprechenden Differenzen der Erwartungswerte werden dann durch geeignete Teleskopsummen und Taylorapproximationen abgeschätzt. Als letzten Schritt wendet man diese Erkenntnisse dann auf eine Vergleichsfolge an, deren Folgenglieder standardnormalverteilt sind, und macht von der Stabilität der Standardnormalverteilung Gebrauch.

**Bemerkung 3.18.** Ein alternativer, konzeptionellerer, Beweis beruht auf Fouriertransformation (in der Wahrscheinlichkeitstheorie auch oft mit „charakteristischen Funktionen“ bezeichnet).

## 4 Einführung in die Schätz- und Testtheorie

Die Grundfragestellung der mathematischen Statistik ist, wie man mithilfe von Beobachtungen/Stichproben/Messungen auf zugrundeliegende Gesetzmäßigkeiten schließen kann bzw. mit welcher „Sicherheit“ man solche Schlüsse ziehen kann.

Als unterliegende mathematische Theorie/Sprache verwenden wir die Wahrscheinlichkeitstheorie.

Im folgenden werden wir die folgenden Aspekte betrachten:

- statistische Modellbildung und grundlegende Fragestellungen
- (Punkt-)Schätzprobleme
- Alternativtestprobleme/Hypothesentestprobleme
- Konfidenzintervalle.

### 4.1 Das statistische Modell

Das grundlegende Objekt in der Statistik ist das sogenannte statistische Modell; es besteht aus:

- einem Stichprobenraum (modelliert durch einen messbaren Raum), und
- einer Familie von in Frage kommenden Gesetzmäßigkeiten (modelliert durch eine Familie von Wahrscheinlichkeitsmaßen auf dem Stichprobenraum).

**Definition 4.1** (statistisches Modell). Ein *statistisches Modell* ist ein Tripel

$$(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$$

bestehend aus einem messbaren Raum  $(\Omega, S)$  und einer nicht-leeren Familie  $(P_\vartheta)_{\vartheta \in \Theta}$  von Wahrscheinlichkeitsmaßen auf  $(\Omega, S)$ .

Ist  $\vartheta \in \Theta$ , so schreiben wir zur besseren Unterscheidbarkeit im folgenden  $E_{P_\vartheta}$  bzw.  $\text{Var}_{P_\vartheta}$  für den Erwartungswert bzw. die Varianz bezüglich  $P_\vartheta$ .

Das Grundziel der Statistik lässt sich dann also folgendermaßen formulieren: Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell. Zu gegebenem/beobachtetem  $\omega \in \Omega$  möchte man dann einen Parameter  $\vartheta \in \Theta$  (oder einen Bereich in  $\Theta$ ) finden, der „möglichst gut zu  $\omega$  passt.“

Oft möchte man mehrere, unabhängige, Stichproben nehmen; dies wird durch entsprechende Produkte modelliert:

**Definition 4.2** (Produkt eines statistischen Modells). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell und sei  $I$  eine nicht-leere Menge. Dann ist das  *$I$ -fache Produkt* von  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  das statistische Modell

$$(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})^{\otimes I} := \left( \prod_I \Omega, S^{\otimes I}, (P_\vartheta^{\otimes I})_{\vartheta \in \Theta} \right),$$

wobei wir  $S^{\otimes I} := \bigotimes_I S$  und  $P_\vartheta^{\otimes I} := \bigotimes_I P_\vartheta$  für alle  $\vartheta \in \Theta$  schreiben.

Außerdem ist es häufig nützlich, sich auf folgende spezielle Situation zu beschränken:

**Definition 4.3** ((diskretes) statistisches Standardmodell).

- Ein statistisches Modell  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ist ein *statistisches Standardmodell*, wenn es ein  $\sigma$ -endliches Maß  $\mu$  auf  $(\Omega, S)$  gibt, so dass es für jedes  $\vartheta \in \Theta$  eine  $\mu$ -Wahrscheinlichkeitsdichte  $f_\vartheta$  auf  $(\Omega, S)$  mit  $P_\vartheta = f_\vartheta \odot \mu$  gibt.
- Ein statistisches Standardmodell ist ein *diskretes statistisches Standardmodell*, wenn der unterliegende messbare Raum diskret ist und die betrachteten Wahrscheinlichkeitsmaße durch Zähldichten gegeben sind.

## 4.2 Schätzer

Wir studieren nun Schätzprobleme und geeignete/gewünschte Eigenschaften von Schätzfunktionen:

**Definition 4.4** ((Punkt-)Schätzer). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell, sei  $(\Omega', S')$  ein messbarer Raum und sei  $\tau: \Theta \rightarrow \Omega'$  eine Abbildung. Ein *(Punkt-)Schätzer* für  $\tau$  ist eine messbare Abbildung vom Typ  $(\Omega, S) \rightarrow (\Omega', S')$ .

Man beachte, dass diese Definition eines Schätzers sehr schwach ist; man wird im Normalfall noch zusätzliche Bedingungen an Schätzer stellen:

- Der Schätzer sollte nicht systematisch fehlerhaft sein (dies wird zum Beispiel durch Erwartungstreue beschrieben).
- Man sollte eine ganze Folge von Schätzern konstruieren, die in einem geeigneten Sinne gegen die zu schätzende Funktion „konvergiert“ (dies wird zum Beispiel durch Konsistenz beschrieben).

**Definition 4.5** (erwartungstreuer Schätzer, konsistente Folge von Schätzern). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell und sei  $\tau: \Theta \rightarrow \mathbb{R}$  eine Funktion.

- Ein Schätzer  $T: (\Omega, S) \rightarrow (\mathbb{R}, B(\mathbb{R}))$  für  $\tau$  heißt *erwartungstreu*, falls: Für alle  $\vartheta \in \Theta$  ist  $T$  bezüglich  $P_\vartheta$  integrierbar und

$$E_{P_\vartheta}(T) = \tau(\vartheta).$$

- Eine Folge  $(T_n: (\Omega, S) \rightarrow (\mathbb{R}, B(\mathbb{R})))_{n \in \mathbb{N}}$  von Schätzern für  $\tau$  heißt *konsistent*, falls: Für alle  $\vartheta \in \Theta$  konvergiert  $(T_n)_{n \in \mathbb{N}}$  bezüglich  $P_\vartheta$  stochastisch gegen  $\tau(\vartheta)$ .

Wie kann man relativ systematisch „vernünftige“ Schätzer bzw. Schätzfolgen finden? Eine viel verwendete Methode sind die sogenannten Maximum-Likelihood-Schätzer. Die Grundidee dabei ist, dass davon ausgegangen wird, dass tatsächlich beobachtete Stichproben zu sehr „plausiblen“ (ausgedrückt durch eine Maximierung der entsprechenden Dichten an der gegebenen Stichprobe) Parametern gehören sollten.

**Definition 4.6** (Maximum-Likelihood-Schätzer). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Standardmodell, sei  $\mu$  ein zugehöriges  $\sigma$ -endliches Maß auf  $(\Omega, S)$  und sei  $(f_\vartheta: \Omega \rightarrow \mathbb{R}_{\geq 0})_{\vartheta \in \Theta}$  eine passende Familie von  $\mu$ -Wahrscheinlichkeitsdichten.

- Die Funktion

$$\begin{aligned} L: \Omega \times \Theta &\rightarrow \mathbb{R}_{\geq 0} \\ (\omega, \vartheta) &\mapsto f_\vartheta(\omega) \end{aligned}$$

ist dann eine/die zugehörige *Likelihoodfunktion*.

- Sei  $S_\Theta$  eine  $\sigma$ -Algebra auf  $\Theta$ . Ein *Maximum-Likelihood-Schätzer* für  $\text{id}_\Theta$  für dieses Modell (mit den oben gegebenen Zusatzdaten) ist eine messbare Abbildung  $T: (\Omega, S) \rightarrow (\Theta, S_\Theta)$  mit der folgenden Eigenschaft: Für alle  $\omega \in \Omega$  ist

$$L(\omega, T(\omega)) = \max_{\vartheta \in \Theta} L(\omega, \vartheta).$$

- Ist  $(\Omega', S')$  ein weiterer messbarer Raum und  $\tau: \Theta \rightarrow \Omega'$  eine Abbildung, so heißt ein Schätzer  $T': \Omega \rightarrow \Omega'$  für  $\tau$  ein *Maximum-Likelihood-Schätzer* für  $\tau$ , falls es einen Maximum-Likelihood-Schätzer  $T$  für  $\text{id}_\Theta$  gibt mit

$$T' = \tau \circ T.$$

Man beachte den Unterschied zwischen *probability* (d.h. Wahrscheinlichkeiten auf dem Stichprobenraum) und *likelihood* (d.h. der Plausibilität von Parametern).

**Bemerkung 4.7** (log-Likelihood). Um Maximum-Likelihood-Schätzer bestimmen, sind also Maximierungsprobleme zu lösen. Es bietet sich dabei oft an, statt Likelihoodfunktionen  $L$  die monoton transformierten Funktionen (sogenannte *log-Likelihoodfunktionen*  $\ln \circ L$  zu betrachten (da viele Dichtefunktionen aus Produkten und Exponentialfunktionen bestehen).

In vielen klassischen Beispielen liefern Maximum-Likelihood-Schätzer intuitiv erscheinende Schätzer. Man beachte jedoch, dass das Konzept des Maximum-Likelihood-Schätzers im allgemeinen *nicht* mit dem Gütekriterium der Erwartungstreue vergleichbar bzw. vereinbar ist; es handelt sich um eine vollständig andere Art der Bewertung der Güte von Schätzern.

**Caveat 4.8.** Maximum-Likelihood-Schätzer sind im allgemeinen *nicht* erwartungstreu!

Unter gewissen zusätzlichen Voraussetzungen liefern Maximum-Likelihood-Schätzer aber immerhin konsistente Schätzfolgen; wir geben hier nur eine einfache Variante (ohne Beweis):

**Satz 4.9** (Konsistenz von Maximum-Likelihood-Schätzern). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Standardmodell und sei  $L: \Omega \times \Theta \rightarrow \mathbb{R}_{\geq 0}$  eine Likelihoodfunktion für dieses Modell. Es gelte außerdem:

- Es ist  $\Theta \subset \mathbb{R}$  ein offenes Intervall und für alle  $\vartheta, \vartheta' \in \Theta$  mit  $\vartheta \neq \vartheta'$  ist  $P_\vartheta \neq P_{\vartheta'}$ .
- Für alle  $n \in \mathbb{N}_{>0}$  gebe es einen Maximum-Likelihood-Schätzer  $T_n: \Omega^n \rightarrow \mathbb{R}$  für  $\text{id}_\Theta$  auf  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})^{\otimes n}$  (bezüglich  $\mu^{\otimes n}$  und der Produkt-Likelihoodfunktion

$$L^{\otimes n}: \Omega^n \times \Theta \rightarrow \mathbb{R}_{\geq 0}$$

$$(\omega, \vartheta) \mapsto \prod_{j=1}^n L(\omega_j, \vartheta)$$

sowie der Borel- $\sigma$ -Algebra auf  $\Theta \subset \mathbb{R}$ ) mit folgender Eigenschaft: Für alle  $\omega \in \Omega^n$  ist  $L^{\otimes n}(\omega, \cdot)|_{\Theta < T_n(\omega)}$  monoton wachsend und  $L^{\otimes n}(\omega, \cdot)|_{\Theta > T_n(\omega)}$  monoton fallend.



Dann ist  $(T_n \circ \pi_{\{1, \dots, n\}})_{n \in \mathbb{N}_{>0}}$  eine konsistente Schätzfolge für  $\text{id}_\Theta$  auf  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta}^{\otimes \mathbb{N}_{>0}})$ .

[Der Beweis beruht auf elementaren Abschätzungen und einer geeigneten Version des schwachen Gesetzes der großen Zahlen.]

**Bemerkung 4.10.** Da konsistente Schätzfolgen unter geeigneten zusätzlichen Annahmen „asymptotisch erwartungstreu“ sind, sind Maximum-Likelihood-Schätzer trotz der im allgemeinen mangelnden Erwartungstreue in vielen Fällen eine vernünftige Wahl.

Neben Erwartungstreue und Konsistenz gibt es weitere Gütekriterien für Schätzer, die helfen können zu spezifizieren, was ein „guter“ oder gar „optimaler“ Schätzer ist. Zum Beispiel ist es erstrebenswert, die mittlere quadratische Abweichung des Schätzers von der zu schätzenden Funktion zu minimieren. Im Spezialfall der Klasse der erwartungstreuen Schätzer lässt sich dies wie folgt als Minimierung der Streuung formulieren:

**Definition 4.11** (gleichmäßig bester Schätzer). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell und sei  $\tau: \Theta \rightarrow \mathbb{R}$  eine Abbildung. Ein Schätzer  $T: (\Omega, S) \rightarrow (\mathbb{R}, B(\mathbb{R}))$  für  $\tau$  heißt *gleichmäßig bester (erwartungstreuer) Schätzer für  $\tau$* , wenn  $T$  ein erwartungstreuer Schätzer für  $\tau$  ist,  $T$  für alle  $\vartheta \in \Theta$  bezüglich  $P_\vartheta$  quadratintegrierbar ist, und folgendes gilt: Ist  $\tilde{T}: (\Omega, S) \rightarrow (\mathbb{R}, B(\mathbb{R}))$  ein erwartungstreuer Schätzer für  $\tau$  und ist  $\tilde{T}$  für alle  $\vartheta \in \Theta$  bezüglich  $P_\vartheta$  quadratintegrierbar, so ist

$$\text{Var}_{P_\vartheta}(T) \leq \text{Var}_{P_\vartheta}(\tilde{T}).$$

**Caveat 4.12.** Im allgemeinen existieren natürlich *nicht* unbedingt gleichmäßig beste Schätzer; selbst wenn sie existieren, ist es oft schwierig, solche optimalen Schätzer zu finden oder zu überprüfen, ob ein gegebener Schätzer ein gleichmäßig bester Schätzer ist.

**Bemerkung 4.13** (Eindeutigkeit gleichmäßig bester Schätzer). Gleichmäßig beste Schätzer sind im wesentlichen eindeutig (d.h. sie stimmen fast sicher bezüglich jedem Wahrscheinlichkeitsmaß aus der zu dem entsprechenden statistischen Modell gehörenden Familie überein).

Wir stellen im folgenden eine spezielle Klasse von statistischen Modellen vor, in denen gewisse gleichmäßig beste Schätzer existieren und konkret angegeben werden können:

**Definition 4.14** (Exponentialfamilie). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell, sei  $\mu$  ein  $\sigma$ -endliches Maß auf  $(\Omega, S)$  und sei  $T: (\Omega, S) \rightarrow (\mathbb{R}, B(\mathbb{R}))$  messbar. Dann ist  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  eine *Exponentialfamilie bezüglich  $\mu$  und  $T$* , falls folgendes gilt: Es gibt Abbildungen  $a, b: \Theta \rightarrow \mathbb{R}$  und eine messbare Abbildung  $h: (\Omega, S) \rightarrow (\mathbb{R}_{\geq 0}, B(\mathbb{R}_{\geq 0}))$ , so dass für jedes  $\vartheta \in \Theta$  die Funktion

$$\begin{aligned} \Omega &\longrightarrow \mathbb{R}_{\geq 0} \\ \omega &\longmapsto h(\omega) \cdot \exp(a(\vartheta) \cdot T(\omega) - b(\vartheta)) \end{aligned}$$

eine  $\mu$ -Wahrscheinlichkeitsdichte von  $P_\vartheta$  ist.

Der Begriff der Exponentialfamilie wird nicht in allen Quellen einheitlich verwendet – manchmal werden noch zusätzliche Regularitätsbedingungen an die auftretenden Funktionen gestellt.

Viele klassische Familien von Wahrscheinlichkeitsmaßen sind Exponentialfamilien.

**Bemerkung 4.15** (Produkte von Exponentialfamilien). Ist  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  eine Exponentialfamilie bezüglich eines  $\sigma$ -endlichen Maßes  $\mu$  auf  $(\Omega, S)$  und einer messbaren Abbildung  $T: (\Omega, S) \rightarrow (\mathbb{R}, B(\mathbb{R}))$ , und ist  $n \in \mathbb{N}_{>0}$ , so ist  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})^{\otimes n}$  eine Exponentialfamilie bezüglich  $\mu^{\otimes n}$  auf  $(\Omega, S)^{\otimes n}$  und

$$\begin{aligned}\Omega^n &\longrightarrow \mathbb{R} \\ \omega &\longmapsto \frac{1}{n} \cdot \sum_{k=1}^n T(\omega_k).\end{aligned}$$

Für Exponentialfamilien kann man unter geeigneten Regularitätsvoraussetzungen gleichmäßig beste Schätzer beschreiben (ohne Beweis):

**Satz 4.16** (gleichmäßig beste Schätzer für Exponentialfamilien). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  eine Exponentialfamilie bezüglich eines  $\sigma$ -endlichen Maßes  $\mu$  auf  $(\Omega, S)$  und einer messbaren Abbildung  $T: (\Omega, S) \rightarrow (\mathbb{R}, B(\mathbb{R}))$ . Dabei seien folgende Regularitätsbedingungen erfüllt:

- Für alle  $\vartheta \in \Theta$  sei  $T$  bezüglich  $P_\vartheta$  quadratintegrierbar und es sei  $T$  nicht  $P_\vartheta$ -fast sicher konstant.
- Die Menge  $\{a(\vartheta) \mid \vartheta \in \Theta\} \subset \mathbb{R}$  enthalte ein offenes nicht-leeres Intervall; hierbei bezeichne  $a$  eine entsprechende Funktion wie in der Definition von Exponentialfamilien (Definition 4.14).

Dann ist  $T$  ein gleichmäßig bester erwartungstreuer Schätzer für die Abbildung

$$\begin{aligned}\Theta &\longrightarrow \mathbb{R} \\ \vartheta &\longmapsto E_{P_\vartheta}(T).\end{aligned}$$

[Der Beweis dieser allgemeinen Fassung beruht auf dem Satz von Lehmann und Scheffé über gleichmäßig beste Schätzer beim Vorliegen von „Suffizienz“ und „Vollständigkeit“. Unter zusätzlichen Regularitätsbedingungen kann auch ein Beweis gegeben werden, der auf elementaren Abschätzungen aus der eindimensionalen reellen Analysis basiert.]

### 4.3 Alternativtestprobleme

Im folgenden untersuchen wir sogenannte Alternativtestprobleme, d.h. das Testen von Hypothesen (z.B. „Hat ein gegebenes Medikament starke Nebenwirkungen?“). Wir beginnen mit der Formalisierung:

**Definition 4.17** (Grundbegriffe für Alternativtestprobleme). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell und sei eine disjunkte Zerlegung  $\Theta = \Theta_0 \cup \Theta_1$  von  $\Theta$  in die Nullhypothese  $\Theta_0$  und die Alternative  $\Theta_1$  gegeben.

- Ein Test von  $\Theta_0$  gegen  $\Theta_1$  (auf dem gegebenen Modell) ist eine messbare Abbildung  $T: (\Omega, \mathcal{S}) \rightarrow ([0, 1], \mathcal{B}([0, 1]))$ . Ist  $T(\Omega) \subset \{0, 1\}$ , so heißt  $T$  *nicht-randomisiert*; ist  $T(\Omega) \not\subset \{0, 1\}$ , so heißt  $T$  *randomisiert*.
- Sei  $\alpha \in [0, 1]$ . Ein Test  $T$  von  $\Theta_0$  gegen  $\Theta_1$  ist ein *Test zum Irrtumsniveau  $\alpha$* , falls

$$\sup_{\vartheta \in \Theta_0} E_{P_\vartheta}(T) \leq \alpha.$$

- Sei  $\alpha \in [0, 1]$ . Ein Test  $T: \Omega \rightarrow [0, 1]$  von  $\Theta_0$  gegen  $\Theta_1$  ist ein *gleichmäßig bester Test für  $\Theta_0$  gegen  $\Theta_1$  zum Irrtumsniveau  $\alpha$* , falls  $T$  ein Test zum Irrtumsniveau  $\alpha$  ist und für jeden Test  $\tilde{T}: \Omega \rightarrow [0, 1]$  von  $\Theta_0$  gegen  $\Theta_1$  zum Irrtumsniveau  $\alpha$  gilt, dass

$$\forall \vartheta \in \Theta_1 \quad 1 - E_{P_\vartheta}(T) \leq 1 - E_{P_\vartheta}(\tilde{T}).$$

- Sei  $\alpha \in [0, 1]$ . Ein Test  $T: \Omega \rightarrow [0, 1]$  von  $\Theta_0$  gegen  $\Theta_1$  zum Irrtumsniveau  $\alpha$  heißt *unverfälscht zum Niveau  $\alpha$* , falls

$$\sup_{\vartheta \in \Theta_0} E_{P_\vartheta}(T) \leq \alpha \leq \inf_{\vartheta \in \Theta_1} E_{P_\vartheta}(T).$$

**Bemerkung 4.18** (Randomisierung). Ist  $T: \Omega \rightarrow [0, 1]$  in der Situation der obigen Definition ein Test für  $\Theta_0$  gegen  $\Theta_1$  und ist  $\omega \in \Omega$ , so interpretiert man den Wert  $T(\omega) \in [0, 1]$  wie folgt:

- Falls  $T(\omega) = 0$  ist, entscheidet man sich bei der Beobachtung  $\omega$  für die Nullhypothese  $\Theta_0$ .
- Falls  $T(\omega) = 1$  ist, entscheidet man sich bei der Beobachtung  $\omega$  für die Alternative  $\Theta_1$ .
- Falls  $T(\omega) \in (0, 1)$  ist, führt man ein zusätzliches  $B(1, T(\omega))$ -Zufallsexperiment durch und entscheidet sich mit Wahrscheinlichkeit  $T(\omega)$  für die Alternative bzw. mit Wahrscheinlichkeit  $1 - T(\omega)$  für die Nullhypothese.

Das Konzept der Randomisierung ist essentiell, um in vielen klassischen Beispielen die Existenz gleichmäßig bester Tests zu gewährleisten.

**Caveat 4.19** (Fehler erster/zweiter Art). Die Begriffsbildung der Alternativtestprobleme ist *nicht* symmetrisch in der Nullhypothese und der Alternative. Durch das Irrtumsniveau werden Tests nur gegen Fehler erster Art (es liegt die Nullhypothese vor, aber man entscheidet sich für die Alternative) abgesichert, nicht gegen Fehler zweiter Art (es liegt die Alternative vor, aber man entscheidet sich für die Alternative). Dies ist bei der Modellierung praktischer Probleme zu berücksichtigen.

**Caveat 4.20.** In der Praxis müssen statistisches Modell, Nullhypothese/Alternative und das Irrtumsniveau *vor* der eigentlichen Durchführung des Experiments festgelegt sein! Eine nachträgliche Anpassung macht die Tests wertlos.

Wir beginnen mit dem Spezialfall, dass Nullhypothese und Alternative jeweils einelementige Mengen sind, sogenannte Alternativtestprobleme mit einfacher Nullhypothese und einfacher Alternative:

**Definition 4.21** (Neyman-Pearson-Test bei einfacher Nullhypothese und einfacher Alternative). Sei  $(\Omega, S, (P_0, P_1))$  ein statistisches (Standard)Modell (bezüglich dem  $\sigma$ -endlichen Maß  $\mu$  auf  $(\Omega, S)$  und den  $\mu$ -Wahrscheinlichkeitsdichten  $f_0$  bzw.  $f_1$  von  $P_0$  bzw.  $P_1$ ; nach dem Satz von Radon-Nikodym können wir statistische Modelle mit endlicher Parametermenge immer als Standardmodelle auffassen). Sei  $c \in \mathbb{R}_{\geq 0}$ . Ein Test  $T: \Omega \rightarrow [0, 1]$  von  $\{0\}$  gegen  $\{1\}$  auf diesem Modell ist ein *Neyman-Pearson-Test* zum Schwellenwert  $c$ , falls für alle  $\omega \in \Omega$  gilt, dass:

- Ist  $f_1(\omega) < c \cdot f_0(\omega)$ , so ist  $T(\omega) = 0$ .
- Ist  $f_1(\omega) > c \cdot f_0(\omega)$ , so ist  $T(\omega) = 1$ .

**Satz 4.22** (Neyman-Pearson-Lemma bei einfacher Nullhypothese und einfacher Alternative). Sei  $(\Omega, S, (P_0, P_1))$  ein statistisches (Standard)Modell (bezüglich dem  $\sigma$ -endlichen Maß  $\mu$  auf  $(\Omega, S)$  und den  $\mu$ -Wahrscheinlichkeitsdichten  $f_0$  bzw.  $f_1$  von  $P_0$  bzw.  $P_1$ ) und sei  $\alpha \in (0, 1)$ . Dann gilt:

1. Es gibt einen Neyman-Pearson-Test  $T$  für  $\{0\}$  gegen  $\{1\}$  auf diesem Modell mit  $E_{P_0}(T) = \alpha$ .
2. Jeder Neyman-Pearson-Test  $T$  für  $\{0\}$  gegen  $\{1\}$  auf diesem Modell mit  $E_{P_0}(T) = \alpha$  ist ein gleichmäßig bester Test von  $\{0\}$  gegen  $\{1\}$  auf diesem Modell zum Irrtumsniveau  $\alpha$ .
3. Ist  $T$  ein gleichmäßig bester Test für  $\{0\}$  gegen  $\{1\}$  auf diesem Modell zum Irrtumsniveau  $\alpha$ , so stimmt  $T$  bereits  $\mu$ -fast überall mit einem Neyman-Pearson-Test überein.

**Bemerkung 4.23.** Der Beweis des Neyman-Pearson-Lemmas ist in dem Sinne konstruktiv, dass er eine Anleitung gibt, wie man solche gleichmäßig besten Tests über Quantile finden kann; Quantile können jedoch im allgemeinen *nicht* explizit bestimmt werden – daher verwendet man numerische Approximationen der Quantile bzw. Quantil-Tabellen.

Das Neyman-Pearson-Lemma besitzt für gewisse Klassen von Verteilungsfamilien Verallgemeinerungen auf einseitige bzw. zweiseitige Tests mit etwas komplizierterer Nullhypothese bzw. Alternative. Wir geben im folgenden nur eine Auswahl an prominenten Beispielen solcher Tests, die in der Praxis viel verwendet werden.

Als Vorbereitung benötigen wir zwei weitere wichtige Familien von Wahrscheinlichkeitsverteilungen:

**Proposition 4.24** ( $\chi^2$ -Verteilungen). Sei  $n \in \mathbb{N}_{>0}$ . Dann ist

$$\begin{aligned} \mathbb{R} &\longrightarrow \mathbb{R}_{\geq 0} \\ x &\longmapsto \chi_{(0, \infty)}(x) \cdot \frac{1}{\Gamma(\frac{n}{2}) \cdot 2^{n/2}} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}} \end{aligned}$$

eine  $\lambda^1$ -Wahrscheinlichkeitsdichte auf  $(\mathbb{R}, B(\mathbb{R}))$ . Die zugehörige Wahrscheinlichkeitsverteilung heißt  $\chi^2$ -Verteilung mit  $n$  Freiheitsgraden und wird mit  $\chi_n^2$  bezeichnet.

Hierbei ist  $\Gamma$  die *Gamma-Funktion* (eine Fortsetzung der Fakultätsfunktion):

$$\begin{aligned}\Gamma: \mathbb{R}_{>0} &\longrightarrow \mathbb{R} \\ r &\longmapsto \int_{\mathbb{R}_{>0}} x^{r-1} \cdot e^{-x} d\lambda^1(x).\end{aligned}$$

**Bemerkung 4.25.** Ist  $n \in \mathbb{N}_{>0}$  und ist  $(X_1, \dots, X_n)$  eine stochastisch unabhängige Familie von  $N(0, 1)$ -verteilten reellwertigen Zufallsvariablen (auf einem gemeinsamen Wahrscheinlichkeitsraum), so hat  $\sum_{k=1}^n X_k^2$  die Verteilung  $\chi_n^2$ .

**Proposition 4.26** (*t-Verteilungen*). Sei  $n \in \mathbb{N}_{>0}$ . Dann ist

$$\begin{aligned}\mathbb{R} &\longrightarrow \mathbb{R}_{\geq 0} \\ x &\longmapsto \frac{1}{\sqrt{n} \cdot B(\frac{1}{2}, \frac{n}{2})} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n \cdot \pi} \cdot \Gamma(\frac{n}{2})} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}\end{aligned}$$

eine  $\lambda^1$ -Wahrscheinlichkeitsdichte auf  $(\mathbb{R}, B(\mathbb{R}))$ . Die zugehörige Wahrscheinlichkeitsverteilung auf  $(\mathbb{R}, B(\mathbb{R}))$  heißt *t-Verteilung* mit  $n$  Freiheitsgraden und wird mit  $t_n$  bezeichnet.

Hierbei ist  $B$  die *Beta-Funktion*:

$$\begin{aligned}B: \mathbb{R}_{>0} \times \mathbb{R}_{>0} &\longrightarrow \mathbb{R} \\ (a, b) &\longmapsto \int_{[0,1]} x^{a-1} \cdot (1-x)^{b-1} d\lambda^1(x).\end{aligned}$$

**Bemerkung 4.27.** Ist  $n \in \mathbb{N}_{>0}$  und ist  $(X_0, \dots, X_n)$  eine stochastisch unabhängige Familie von  $N(0, 1)$ -verteilten Zufallsvariablen (auf einem gemeinsamen Wahrscheinlichkeitsraum), so hat

$$\frac{X_0}{\sqrt{\frac{1}{n} \cdot \sum_{k=1}^n X_k^2}}$$

die Verteilung  $t_n$ .

Wichtige Tests für Normalverteilungsfamilien auf Produktmodellen sind:

- *Einseitiger Gauß-Test für den Erwartungswert bei bekannter Varianz*: „Ist der Erwartungswert einer Normalverteilung (bei bekannter Varianz) höchstens so groß wie ein vorgegebener Wert?“  
[Hierfür werden die Quantile der Standardnormalverteilung benötigt.]
- *Zweiseitiger Gauß-Test für den Erwartungswert bei bekannter Varianz*: „Stimmt der Erwartungswert einer Normalverteilung (bei bekannter Varianz) mit einem vorgegebenen Wert überein?“  
[Hierfür werden die Quantile der Standardnormalverteilung benötigt. Die Optimierung ist in diesem Fall nur in der Klasse der unverfälschten Tests möglich.]
- *Einseitiger t-Test für den Erwartungswert*: „Ist der Erwartungswert einer Normalverteilung höchstens so groß wie ein vorgegebener Wert?“  
[Hierfür werden die Quantile der t-Verteilungen benötigt. Die Optimierung ist in diesem Fall nur in der Klasse der unverfälschten Tests möglich.]

- *Zweiseitiger  $t$ -Test für den Erwartungswert:* „Stimmt der Erwartungswert einer Normalverteilung mit einem vorgegebenen Wert überein?“  
[Hierfür werden die Quantile der  $t$ -Verteilungen benötigt. Die Optimierung ist in diesem Fall nur in der Klasse der unverfälschten Tests möglich.]
- *Einseitiger  $\chi^2$ -Test für die Varianz bei bekanntem Erwartungswert:* „Ist die Varianz einer Normalverteilung (bei bekanntem Erwartungswert) höchstens so groß wie ein vorgegebener Wert?“  
[Hierfür werden die Quantile der  $\chi^2$ -Verteilungen benötigt.]
- *Linksseitiger  $\chi^2$ -Test für die Varianz:* „Ist die Varianz einer Normalverteilung höchstens so groß wie ein vorgegebener Wert?“  
[Hierfür werden die Quantile der  $\chi^2$ -Verteilungen benötigt. Für den entsprechenden rechtsseitigen Test beachte man, dass die Optimierung in diesem Fall nur in der Klasse der unverfälschten Tests möglich ist.]
- ...  
[Es gibt noch viele weitere solche speziellen Tests, die aber im Prinzip alle nach demselben Schema aufgebaut sind.]

Wir beschreiben diese Tests nun etwas genauer (ohne Beweis); diese Resultate basieren alle auf entsprechenden Varianten des Neyman-Pearson-Lemmas:

**Notation 4.28** (Stichprobenmittel/modifizierte Stichprobenvarianz).

- Zu  $n \in \mathbb{N}_{>0}$  bezeichne

$$s_n: \mathbb{R}^n \longrightarrow \mathbb{R}$$

$$\omega \longmapsto \frac{1}{n} \cdot \sum_{k=1}^n \omega_k$$

das *Stichprobenmittel*.

- Zu  $n \in \mathbb{N}_{>1}$  bezeichne

$$v_n^*: \mathbb{R}^n \longrightarrow \mathbb{R}$$

$$\omega \longmapsto \frac{1}{n-1} \cdot \sum_{k=1}^n (\omega_k - s_n(\omega))^2$$

die *modifizierte Stichprobenvarianz*.

**Satz 4.29** (einseitiger Gauß-Test bei bekannter Varianz). *Sei  $n \in \mathbb{N}_{>0}$ , sei  $a_0 \in \mathbb{R}$  und sei  $c \in \mathbb{R}_{>0}$ . Wir betrachten das statistische Modell*

$$(\mathbb{R}, B(\mathbb{R}), (N(a, c))_{a \in \mathbb{R}})^{\otimes n}.$$

*Sei außerdem  $\alpha \in (0, 1)$  und sei  $q$  das  $(1 - \alpha)$ -Quantil von  $N(0, 1)$ . Dann ist*

$$\mathbb{R}^n \longrightarrow [0, 1]$$

$$\omega \longmapsto \begin{cases} 0 & \text{falls } \sqrt{n/c} \cdot (s_n(\omega) - a_0) \leq q \\ 1 & \text{sonst} \end{cases}$$

ein gleichmäßig bester Test von  $(-\infty, a_0]$  gegen  $(a_0, \infty)$  auf diesem Modell zum Irrtumsniveau  $\alpha$ .

**Satz 4.30** (zweiseitiger Gauß-Test bei bekannter Varianz). Sei  $n \in \mathbb{N}_{>0}$ , sei  $a_0 \in \mathbb{R}$  und sei  $c \in \mathbb{R}_{>0}$ . Wir betrachten das statistische Modell

$$(\mathbb{R}, B(\mathbb{R}), (N(a, c))_{a \in \mathbb{R}})^{\otimes n}.$$

Sei außerdem  $\alpha \in (0, 1)$  und sei  $q$  das  $(1 - \alpha/2)$ -Quantil von  $N(0, 1)$ . Dann ist

$$\begin{aligned} \mathbb{R}^n &\longrightarrow [0, 1] \\ \omega &\longmapsto \begin{cases} 0 & \text{falls } \sqrt{n/c} \cdot |s_n(\omega) - a_0| \leq q \\ 1 & \text{sonst} \end{cases} \end{aligned}$$

ein gleichmäßig bester unverfälschter Test von  $\{a_0\}$  gegen  $\mathbb{R} \setminus \{a_0\}$  auf diesem Modell zum Irrtumsniveau  $\alpha$  in der Klasse der unverfälschten Tests.

**Satz 4.31** (einseitiger  $t$ -Test für den Erwartungswert). Sei  $n \in \mathbb{N}_{>1}$  und sei  $a_0 \in \mathbb{R}$ . Wir betrachten das statistische Modell

$$(\mathbb{R}, B(\mathbb{R}), (N(a, c))_{(a, c) \in \mathbb{R} \times \mathbb{R}_{>0}})^{\otimes n}.$$

Sei außerdem  $\alpha \in (0, 1)$  und sei  $q$  das  $(1 - \alpha)$ -Quantil von  $t_{n-1}$ . Dann ist

$$\begin{aligned} \mathbb{R}^n &\longrightarrow [0, 1] \\ \omega &\longmapsto \begin{cases} 0 & \text{falls } \sqrt{n/v_n^*(\omega)} \cdot (s_n(\omega) - a_0) \leq q \\ 1 & \text{sonst} \end{cases} \end{aligned}$$

ein gleichmäßig bester unverfälschter Test von  $(-\infty, a_0]$  gegen  $(a_0, \infty)$  auf diesem Modell zum Irrtumsniveau  $\alpha$  in der Klasse der unverfälschten Tests.

**Satz 4.32** (zweiseitiger  $t$ -Test für den Erwartungswert). Sei  $n \in \mathbb{N}_{>1}$  und sei  $a_0 \in \mathbb{R}$ . Wir betrachten das statistische Modell

$$(\mathbb{R}, B(\mathbb{R}), (N(a, c))_{(a, c) \in \mathbb{R} \times \mathbb{R}_{>0}})^{\otimes n}.$$

Sei außerdem  $\alpha \in (0, 1)$  und sei  $q$  das  $(1 - \alpha/2)$ -Quantil von  $t_{n-1}$ . Dann ist

$$\begin{aligned} \mathbb{R}^n &\longrightarrow [0, 1] \\ \omega &\longmapsto \begin{cases} 0 & \text{falls } \sqrt{n/v_n^*(\omega)} \cdot |s_n(\omega) - a_0| \leq q \\ 1 & \text{sonst} \end{cases} \end{aligned}$$

ein gleichmäßig bester unverfälschter Test von  $\{a_0\}$  gegen  $\mathbb{R} \setminus \{a_0\}$  auf diesem Modell zum Irrtumsniveau  $\alpha$  in der Klasse der unverfälschten Tests.

**Satz 4.33** (einseitiger  $\chi^2$ -Test für die Varianz (bei bekanntem Erwartungswert)). Sei  $n \in \mathbb{N}_{>0}$ , sei  $a \in \mathbb{R}$  und sei  $c_0 \in \mathbb{R}_{>0}$ . Wir betrachten das statistische Modell

$$(\mathbb{R}, B(\mathbb{R}), (N(a, c))_{c \in \mathbb{R}_{>0}})^{\otimes n}.$$

Sei außerdem  $\alpha \in (0, 1)$  und sei  $q$  das  $(1 - \alpha)$ -Quantil von  $\chi_n^2$ . Dann ist

$$\begin{aligned} \mathbb{R}^n &\longrightarrow [0, 1] \\ \omega &\longmapsto \begin{cases} 0 & \text{falls } \frac{1}{c_0} \cdot \sum_{k=1}^n (\omega_k - a)^2 \leq q \\ 1 & \text{sonst} \end{cases} \end{aligned}$$

ein gleichmäßig bester Test von  $(0, c_0]$  gegen  $(c_0, \infty)$  auf diesem Modell zum Irrtumsniveau  $\alpha$ .

**Satz 4.34** (linksseitiger  $\chi^2$ -Test für die Varianz). Sei  $n \in \mathbb{N}_{>0}$  und sei  $c_0 \in \mathbb{R}_{>0}$ . Wir betrachten das statistische Modell

$$(\mathbb{R}, B(\mathbb{R}), (N(a, c))_{(a, c) \in \mathbb{R} \times \mathbb{R}_{>0}})^{\otimes n}.$$

Sei außerdem  $\alpha \in (0, 1)$  und sei  $q$  das  $(1 - \alpha)$ -Quantil von  $\chi_{n-1}^2$ . Dann ist

$$\begin{aligned} \mathbb{R}^n &\longrightarrow [0, 1] \\ \omega &\longmapsto \begin{cases} 0 & \text{falls } \frac{1}{c_0} \cdot \sum_{k=1}^n (\omega_k - s_n(\omega))^2 \leq q \\ 1 & \text{sonst} \end{cases} \end{aligned}$$

ein gleichmäßig bester Test von  $(0, c_0]$  gegen  $(c_0, \infty)$  auf diesem Modell zum Irrtumsniveau  $\alpha$ .

#### 4.4 Konfidenzbereiche

Ein weiteres wichtiges Ziel in der Statistik ist es, unbekannte Größen mit einer zusätzlichen „Sicherheit“ zu schätzen. Dies wird durch Konfidenzbereiche formalisiert:

**Definition 4.35** (Konfidenzbereich). Sei  $(\Omega, S, (P_\vartheta)_{\vartheta \in \Theta})$  ein statistisches Modell, sei  $\Omega'$  eine Menge, sei  $\tau: \Theta \longrightarrow \Omega'$  eine Abbildung und sei  $\alpha \in [0, 1]$ . Ein *Konfidenzbereich* für  $\tau$  zum Irrtumsniveau  $\alpha$  auf diesem Modell ist eine Abbildung  $C: \Omega \longrightarrow \text{Pot}(\Omega')$  mit folgenden Eigenschaften:

- Für alle  $\omega' \in \Omega'$  ist

$$\{\omega \in \Omega \mid \omega' \in C(\omega)\} \in S.$$

- Und es gilt

$$\inf_{\vartheta \in \Theta} P_\vartheta(\{\omega \in \Omega \mid \tau(\vartheta) \in C(\omega)\}) \geq 1 - \alpha.$$

Das Ziel ist es dann, bei „niedrigem“ Irrtumsniveau möglichst „kleine“ Konfidenzbereiche zu finden. Die Definition legt nahe, dieses Problem (wie im Falle von Alternativtestproblemen) auf entsprechende Quantilprobleme zurückzuführen; tatsächlich lassen



sich sogar die meisten Konfidenzbereichprobleme in Alternativtestprobleme übersetzen.

Zum Beispiel erhält man so in Analogie zu den  $t$ -Tests für den Erwartungswert (ohne Beweis):

**Satz 4.36** (Konfidenzintervall für den Erwartungswert im Gauß-Modell). *Sei  $n \in \mathbb{N}_{>1}$ , sei  $\alpha \in (0, 1)$  und sei  $q \in \mathbb{R}$  das  $(1 - \alpha/2)$ -Quantil von  $t_{n-1}$ . Dann ist*

$$C: \mathbb{R}^n \longrightarrow \text{Pot}(\mathbb{R}^n)$$

$$\omega \longmapsto \left( s_n(\omega) - q \cdot \sqrt{v_n^*(\omega)/n}, s_n(\omega) + q \cdot \sqrt{v_n^*(\omega)/n} \right)$$

ein Konfidenzbereich zum Irrtumsniveau  $\alpha$  für die Abbildung

$$\mathbb{R} \times \mathbb{R}_{>0} \longrightarrow \mathbb{R}$$

$$(a, c) \longmapsto a$$

auf dem Modell  $(\mathbb{R}, B(\mathbb{R}), (N(a, c))_{(a, c) \in \mathbb{R} \times \mathbb{R}_{>0}})^{\otimes n}$ . Hierbei bezeichnen  $s_n$  bzw.  $v_n^*$  das Stichprobenmittel bzw. die modifizierte Stichprobenvarianz (Notation 4.28).