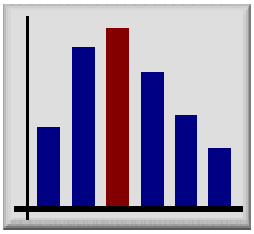

# Stochastik I


Lehrstuhl für Mathematische Statistik

Universität Würzburg

Prof. Dr. Michael Falk

# Inhaltsverzeichnis

# 1 Das Kolmogoroffsche Axiomensystem

[A. N. Kolmogoroff (1933)] Seit Euklid werden bei einem rein geometrischen Aufbau der Geometrie die Grundbegriffe „Punkt“ und „Gerade“ nicht explizit definiert, sondern axiomatisch eingeführt.

1. Man vermittelt zunächst bewusst eine vage Vorstellung von dem, was mit den Grundbegriffen gemeint ist, um die Theorie später anwenden zu können („Ein Punkt ist, was keinen Teil hat“, „Eine Gerade ist eine Linie, die gleich liegt mit den Punkten auf ihr selbst“ (Euklid)). Die vage Vorstellung wird dann im Verlauf der Beschäftigung mit der Theorie zwangsläufig immer präziser.

2. Man beschreibt mittels „Axiomen“, welche Beziehungen zwischen den Grundbegriffen bestehen.

Analog gehen wir nun bei der Axiomatisierung der Stochastik vor. Im ersten Axiom fordern wir die Existenz von Wahrscheinlichkeiten.

**Axiom 1’** Ist $\Omega$ die Menge der möglichen Ergebnisse eines Experimentes (d.h. genau ein $\omega \in \Omega$ tritt bei der Durchführung des Experimentes ein), so ist jeder Teilmenge $A \subset \Omega$ eine reelle Zahl $P(A) \geq 0$ zugeordnet, Wahrscheinlichkeit von $A$ genannt, die den Grad der Sicherheit angibt, mit dem $A$ eintritt.

**Axiom 2** $P(\Omega) = 1$.

**Axiom 3** ($\sigma$–Additivität von $P$) Für eine Folge $A_1, A_2, \ldots$ paarweise disjunkter Teilmengen von $\Omega$ gilt: $P(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$.

Das System 1’,2,3 ist zu einschränkend, wie der folgende Satz zeigt.

**Satz (G. Vitali 1905)** *Es existiert kein $P$ zu $\Omega = [0,1)$, welches die Axiome 1’,2 und 3 erfüllt und zusätzlich translationsinvariant ist, d.h. $P(A_c) = P(A)$ für $A_c := \{a + c\,(mod\ 1) : a \in A\}$, $c \geq 0$.*

Beweis: Siehe Übungen. $\square$

Axiom 1’ wird nun abgeschwächt, indem $P$ nicht mehr auf der gesamten Potenzmenge $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ definiert wird.

**Axiom 1** Ist $\Omega$ die Menge der möglichen Ergebnisse eines Experimentes, so ist einigen (nicht notwendig allen) Teilmengen von $\Omega$, Ereignisse genannt, eine reelle Zahl $P(A) \geq 0$ zugeordnet, Wahrscheinlichkeit von A genannt, die den Grad der Sicherheit angibt, mit dem $A$ eintritt.

- $\Omega$ ist ein Ereignis.
- Das Komplement $A^c = \Omega \backslash A$ eines Ereignisses $A$ ist ein Ereignis.
- Der Durchschnitt von zwei Ereignissen ist ein Ereignis.

- Die Vereinigung von abzählbar vielen disjunkten Ereignissen ist ein Ereignis.

**Definition 1.1** $\Omega$ sei eine nichtleere Menge. Dann heißt $\mathcal{A} \subset \mathcal{P}(\Omega)$ (= Potenzmenge von $\Omega$) $\sigma$–Algebra über $\Omega$:$\Leftrightarrow$

1. $\Omega \in \mathcal{A}$,
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
3. $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$
4. $A_i \in \mathcal{A},\ i \in \mathbb{N},\ A_i \cap A_j = \emptyset$ für $i \neq j \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.

**Definition 1.2** $(\Omega, \mathcal{A})$ heißt messbarer Raum $:\Leftrightarrow$ $\mathcal{A}$ ist $\sigma$–Algebra über nichtleerer Menge $\Omega$.

**Definition 1.3** $(\Omega, \mathcal{A})$ sei messbarer Raum. Eine Funktion $P : \mathcal{A} \to \mathbb{R}_+ = [0, \infty)$, die die Axiome 2 und 3 erfüllt, heißt Wahrscheinlichkeitsmaß. Das Tripel $(\Omega, \mathcal{A}, P)$ heißt in diesem Fall Wahrscheinlichkeitsraum.

# 2 Erste Folgerungen aus dem Axiomensystem

**Satz 2.1** $(\Omega, \mathcal{A})$ *messbarer Raum,* $A_n \in \mathcal{A}$, $n \in \mathbb{N}$. *Dann gilt:*

*(i)* $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$,

*(ii)* $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$,

*(iii)*
$$\begin{aligned}\limsup_{n \in \mathbb{N}} A_n &:= \{\omega \in \Omega : \omega \textit{ liegt in } \infty \textit{ vielen } A_n\} \\ &= \bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n \in \mathcal{A},\end{aligned}$$

*(iv)*
$$\begin{aligned}\liminf_{n \in \mathbb{N}} A_n &:= \{\omega \in \Omega : \omega \textit{ liegt in fast allen } A_n\} \\ &= \bigcup_{m=1}^{\infty} \bigcap_{n \geq m} A_n \in \mathcal{A},\end{aligned}$$

BEWEIS:

(i) Setze $B_1 := A_1$, $B_n := A_n \backslash (A_1 \cup \ldots \cup A_{n-1}) = A_n \cap A_1^c \cap \ldots \cap A_{n-1}^c \in \mathcal{A}$. $B_n$, $n \in \mathbb{N}$, sind paarweise disjunkt mit $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n \in \mathcal{A}$.

(ii)

$$\bigcap_{n\in\mathbb{N}} A_n = \Big(\Big(\bigcap_{n\in\mathbb{N}} A_n\Big)^c\Big)^c = \Big(\bigcup_{n\in\mathbb{N}} A_n^c\Big)^c \in \mathcal{A}.$$

(iii) und (iv) folgen unmittelbar aus (i), (ii). $\square$

**Korollar 2.2** $\Omega \neq \emptyset$, $\mathcal{A} \subset \mathcal{P}(\Omega)$. *Dann:* $\mathcal{A}$ *ist* $\sigma$*–Algebra* $\Leftrightarrow$

*(i)* $\Omega \in \mathcal{A}$,

*(ii)* $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,

*(iii)* $A_n \in \mathcal{A}$, $n \in \mathbb{N} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

**Satz 2.3** $(\Omega, \mathcal{A}, P)$ *sei Wahrscheinlichkeitsraum. Dann gilt:*

*(i)* $P(\emptyset) = 0$,

*(ii)* $P(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^n P(A_i)$, *falls* $A_1, \ldots, A_n$ *paarweise disjunkt,*

*(iii)* $0 \leq P(A) \leq 1$ *stets,*

*(iv)* $A \subset B$ $(\in \mathcal{A}) \Rightarrow P(A) \leq P(B)$ *(Monotonie von* $P$*),*

*(v)* $P(A^c) = 1 - P(A)$.

Beweis:

(i)

$$\begin{aligned} \emptyset &= \emptyset \cup \emptyset \cup \ldots \\ &\Rightarrow P(\emptyset) = P(\emptyset) + P(\emptyset) + \ldots \\ &\Rightarrow P(\emptyset) = 0. \end{aligned}$$

(ii) Wegen $P(\emptyset) = 0$ gilt:

$$\begin{aligned} P(A_1 \cup \ldots \cup A_n) &= P(A_1 \cup \ldots \cup A_n \cup \emptyset \cup \ldots) \\ &= P(A_1) + \ldots + P(A_n) + 0 + \ldots \end{aligned}$$

(v)

$$\begin{aligned} \Omega = A \cup A^c &\Rightarrow 1 = P(\Omega) = P(A) + P(A^c) \\ &\Rightarrow P(A^c) = 1 - P(A). \end{aligned}$$

(iii) Folgt unmittelbar aus (v):

$$0 \leq P(A) = 1 - \underbrace{P(A^c)}_{\geq 0} \leq 1.$$

(iv)

$$\begin{aligned} B &= A \cup (B \backslash A) = A \cup (B \cap A^c) \\ \Rightarrow_{(ii)} P(B) &= P(A) + P(B \backslash A) \geq P(A). \end{aligned}$$

□

Im folgenden sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $A_1, \ldots, A_n \in \mathcal{A}$. $|M|$ bezeichnet die Anzahl der Elemente einer Menge $M$ (Mächtigkeit von $M$).

**Satz 2.4 (Allgemeiner Additionssatz)**

$$\begin{aligned} P(A_1 \cup \ldots \cup A_n) &= \sum_{\emptyset \neq T \subset \{1,\ldots,n\}} (-1)^{|T|-1} P\left(\bigcap_{i \in T} A_i\right) \\ &= \sum_{k=1}^{n} (-1)^{k-1} S_k \end{aligned}$$

*mit*

$$S_k := \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} P\Big(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\Big).$$

Beispiel: Im Fall $n = 2$ ergibt sich

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

Im Fall $n = 3$ ergibt sich

$$\begin{aligned} & P(A_1 \cup A_2 \cup A_3) \\ = \; & P(A_1) + P(A_2) + P(A_3) \\ & \quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ & \quad + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Beweis: Mittels vollständiger Induktion; "+" bedeutet Vereinigung disjunkter Mengen.
Der Fall $n = 2$:

$$\begin{aligned} & A_1 \cup A_2 = A_1 + (A_2 \backslash A_1), \\ & A_2 = (A_2 \cap A_1) + (A_2 \backslash A_1) \\ \Rightarrow \; & P(A_1 \cup A_2) = P(A_1) + P(A_2 \backslash A_1), \\ & P(A_2) = P(A_2 \cap A_1) + P(A_2 \backslash A_1) \\ \Rightarrow \; & P(A_1 \cup A_2) - P(A_1) = P(A_2) - P(A_2 \cap A_1) \\ \Rightarrow \; & \text{Behauptung für den Fall } n = 2. \end{aligned}$$

Der Induktionsschritt $n \to n+1$:

$$\begin{aligned}
&P((A_1 \cup \ldots \cup A_n) \cup A_{n+1}) \\
&= \quad P((A_1 \cup \ldots \cup A_n)) + P(A_{n+1}) \\
&\qquad -P\left((A_1 \cap A_{n+1}) \cup (A_2 \cap A_{n+1}) \cup \ldots \right. \\
&\qquad\qquad \left. \cup (A_n \cap A_{n+1})\right) \\
&= \sum_{\emptyset \neq T \subset \{1,\ldots,n\}} (-1)^{|T|-1} P\left(\bigcap_{i \in T} A_i\right) + P(A_{n+1}) \\
&\quad - \sum_{\emptyset \neq T \subset \{1,\ldots,n\}} (-1)^{|T|-1} P\left(\bigcap_{i \in T} A_i \cap A_{n+1}\right) \\
&= \sum_{\emptyset \neq T \subset \{1,\ldots,n+1\},\, n+1 \notin T} (-1)^{|T|-1} P\left(\bigcap_{i \in T} A_i\right) + P(A_{n+1}) \\
&\quad + \sum_{T \subset \{1,\ldots,n+1\},\, n+1 \in T,\, T \cap \{1,\ldots,n\} \neq \emptyset} (-1)^{|T|-1} P\left(\bigcap_{i \in T} A_i\right) \\
&= \sum_{\emptyset \neq T \subset \{1,\ldots,n+1\}} (-1)^{|T|-1} P\left(\bigcap_{i \in T} A_i\right).
\end{aligned}$$

□

**Satz 2.5** *Sei $B_k$ das Ereignis, dass genau $k$ der Ereignisse $A_1, \ldots, A_n$ eintreten, d.h. $\omega \in B_k \Leftrightarrow \omega \in A_i$ für genau $k$ der Indizes $i = 1, \ldots, n$. Dann gilt:*

$$\begin{aligned}
P(B_k) &= \sum_{U \subset \{1,\ldots,n\},\, |U| \geq k} \binom{|U|}{k} (-1)^{|U|-k} P\left(\bigcap_{i \in U} A_i\right) \\
&= \sum_{m=k}^{n} \binom{m}{k} (-1)^{m-k} S_m,
\end{aligned}$$

*$S_m$ wie in Satz 2.4, $S_0 := 1$.*

**Bemerkung** $B_0 = (A_1 \cup \ldots \cup A_n)^c \Rightarrow P(B_0) = 1 - P\left(\cup_{1 \leq i \leq n} A_i\right) =_{2.4} \sum_{m=0}^{n} (-1)^m S_m$.

Beweis:

$$B_k = \sum_{S \subset \{1,\ldots,n\},\, |S|=k} \left(\left(\bigcap_{i \in S} A_i\right) \cap \left(\bigcap_{i \in S^c} A_i^c\right)\right),$$

(disjunkte Zerlegung von $B_k$). Es folgt:

$$
\begin{aligned}
& P(B_k) \\
&= \sum_{S\subset\{1,\dots,n\},\,|S|=k} P\left(\left(\bigcap_{i\in S} A_i\right)\cap\left(\bigcap_{i\in S^c} A_i^c\right)\right) \\
&= \sum_{S\subset\{1,\dots,n\},\,|S|=k} \left\{1-P\left(\left(\bigcap_{i\in S} A_i\right)^c \cup \left(\bigcup_{i\in S^c} A_i\right)\right)\right\} \\
&=_{2.4} \sum_{S\subset\{1,\dots,n\},\,|S|=k} \left\{1-\left[P\left(\left(\bigcap_{i\in S} A_i\right)^c\right)\right.\right. \\
&\qquad + \sum_{\emptyset\neq T\subset S^c} (-1)^{|T|-1} P\left(\bigcap_{i\in T} A_i\right) \\
&\qquad \left.\left. -P\left(\bigcup_{i\in S^c}\left(\left(\bigcap_{j\in S} A_j\right)^c \cap A_i\right)\right)\right]\right\} \\
&= \sum_{S\subset\{1,\dots,n\},\,|S|=k} \left\{P\left(\bigcap_{i\in S} A_i\right)\right. \\
&\qquad -\left[\sum_{\emptyset\neq T\subset S^c} (-1)^{|T|-1} P\left(\underbrace{\bigcap_{i\in T} A_i}_{=C}\right)\right. \\
&\qquad \left.\left. -\sum_{\emptyset\neq T\subset S^c} (-1)^{|T|-1} P\left(\underbrace{\left(\bigcap_{j\in S} A_j\right)^c}_{=D^c} \cap \underbrace{\bigcap_{i\in T} A_i}_{=C}\right)\right]\right\};
\end{aligned}
$$

wegen $P(C)-P(D^c\cap C)=P(C\cap D)$ folgt

$$
\begin{aligned}
&= \sum_{S\subset\{1,\dots,n\},\,|S|=k} \left\{P\left(\bigcap_{i\in S} A_i\right)\right. \\
&\qquad \left. -\sum_{\emptyset\neq T\subset S^c} (-1)^{|T|-1} P\left(\underbrace{\left(\bigcap_{j\in S} A_j\right)\cap\left(\bigcap_{i\in T} A_i\right)}_{=\bigcap_{i\in S\cup T} A_i}\right)\right\} \\
&= \sum_{S\subset\{1,\dots,n\},\,|S|=k}\ \sum_{T\subset S^c} (-1)^{|T|} P\left(\bigcap_{i\in S\cup T} A_i\right) \\
&= \sum_{S\subset\{1,\dots,n\},\,|S|=k}\ \sum_{U\supset S,\,U\subset\{1,\dots,n\}} (-1)^{|U|-k} P\left(\bigcap_{i\in U} A_i\right).
\end{aligned}
$$

Der Summand $(-1)^{|U|-k}P\left(\bigcap_{i\in U} A_i\right)$ tritt hierbei so oft auf, wie es $k$–elementige Teilmengen $S$ von $U$ gibt, also $\binom{|U|}{k}$–mal. Hieraus folgt der erste Teil der Behauptung sowie

$$= \sum_{m=k}^{n} \binom{m}{k} (-1)^{m-k} \underbrace{\sum_{U \subset \{1,\ldots,n\},\, |U|=m} P\left(\bigcap_{i \in U} A_i\right)}_{=S_m}.$$

□

**Satz 2.6** *Sei $C_k$ das Ereignis, dass mindestens $k$ der Ereignisse $A_1, \ldots, A_n$ eintreten. Dann gilt:*

$$P(C_k) = \sum_{m=k}^{n} \binom{m-1}{k-1} (-1)^{m-k} S_m.$$

Beweis:

$$\begin{aligned} P(C_k) &= \sum_{j=k}^{n} P(B_j) \\ &=_{2.5} \sum_{j=k}^{n} \sum_{m=j}^{n} \binom{m}{j} (-1)^{m-j} S_m \\ &= \sum_{m=k}^{n} \left( \sum_{j=k}^{m} \binom{m}{j} (-1)^{m-j} \right) S_m. \end{aligned}$$

Für die innere Summe folgt aus der Beziehung $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$:

$$\begin{aligned} &\binom{m}{m} - \binom{m}{m-1} + \binom{m}{m-2} - \ldots \\ &\quad +(-1)^{m-k}\binom{m}{k} \\ &= \underbrace{\binom{m-1}{m}}_{=0} + \underbrace{\binom{m-1}{m-1} - \binom{m-1}{m-1}}_{=0} \\ &\quad \underbrace{-\binom{m-1}{m-2} + \ldots + (-1)^{m-k}\binom{m-1}{k}}_{=0} \\ &\quad +(-1)^{m-k}\binom{m-1}{k-1} \\ &= (-1)^{m-k}\binom{m-1}{k-1}. \end{aligned}$$

□

**Bemerkung** Der Allgemeine Additionssatz 2.4 ist in 2.6 enthalten ($k = 1$).

**Bemerkung** Zur Geschichte der Stochastik: Briefwechsel (1654) zwischen P. Fermat und B. Pascal (u.a. wg. Chevalier de Méré); inzwischen stürmische Entwicklung (A.N. Kolmogoroff (1933)$\longrightarrow \ldots$)

# 3 Grundlagen der Kombinatorik

**Definition 3.1** Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ heißt Laplace–Experiment $:\Leftrightarrow |\Omega| < \infty$ und alle einelementigen Teilmengen von $\Omega$ sind Ereignisse mit der gleichen Wahrscheinlichkeit.

**Satz 3.2** $(\Omega, \mathcal{A}, P)$ *Laplace–Experiment, $A \subset \Omega$. Dann gilt:*

$$\begin{aligned} P(A) &= \frac{|A|}{|\Omega|} \\ &= \frac{\textit{Anzahl der für } A \textit{ günstigen Ausgänge}}{\textit{Anzahl aller möglichen Ausgänge}}. \end{aligned}$$

**Satz 3.3 (Additionsprinzip der Kombinatorik)** *Für disjunkte endliche Mengen $A_1$, $A_2$ gilt:*

$$|A_1 + A_2| = |A_1| + |A_2|.$$

**Korollar** *Für disjunkte endliche Mengen $A_1, \ldots, A_k$ gilt:*

$$|A_1 \cup \ldots \cup A_k| = |A_1| + \ldots + |A_k|.$$

**Satz 3.4 (Multiplikationssatz der Kombinatorik)** *$A_1$ sei eine Menge der Mächtigkeit $n_1 \in \mathbb{Z}_+ = \mathbb{N} \cup \{0\}$, $B_2$ eine beliebige Menge und $n_2 \in \mathbb{Z}_+$. Jedem $a_1 \in A_1$ sei genau eine $n_2$–elementige Teilmenge $B(a_1) \subset B_2$ zugeordnet, und es sei*

$$A_2 := \{(a_1, a_2) : a_1 \in A_1, a_2 \in B(a_1)\}.$$

*Dann gilt: $|A_2| = n_1 n_2$.*

Beweis: Folgt aus 3.3. □

**Korollar 3.5** *$|A_1| = n_1 \in \mathbb{Z}_+$, $B_1, \ldots, B_n$ seien beliebige Mengen und $n_2, \ldots, n_k \in \mathbb{Z}_+$. Für $i = 1, \ldots, k-1$ sei jedem $i$–Tupel $(a_1, \ldots, a_i) \in A_i$ eine $n_{i+1}$–elementige Teilmenge $B(a_1, \ldots, a_i) \subset B_{i+1}$ zugeordnet, und es sei*

$$\begin{aligned} A_{i+1} \;:=\; & \{(a_1, \ldots, a_i, a_{i+1}) : (a_1, \ldots, a_i) \in A_i, \\ & \quad a_{i+1} \in B(a_1, \ldots, a_i)\}, \end{aligned}$$

*(Definition durch Induktion (Rekursion)). Dann gilt:*

$$|A_k| = n_1 n_2 \ldots n_k.$$

**Korollar** $|A_1 \times A_2 \times \ldots \times A_k| = n_1 n_2 \cdots n_k$, *falls* $|A_i| = n_i$, $i = 1, \ldots, k$.

Obiges Korollar ergibt speziell für $A_i = A$, $i = 1, \ldots, k$: $|\underbrace{A \times \ldots \times A}_{k \text{ mal}}| = |A^k| = |A|^k$.

Die Menge $A^k$ aller $k$–Tupel von Elementen aus $A$ heißt <u>geordnete Probe</u> zu $A$ vom Umfang $k$ <u>mit Wiederholung</u>.

**Satz 3.6** *Es gibt $n^k$ geordnete Proben zu einer $n$–elementigen Menge vom Umfang $k$ mit Wiederholung.*

**Beispiel** $A$, $B$ endliche Mengen, $B^A :=$ Menge aller Abbildungen von $A$ nach $B$. Dann: $\left|B^A\right| = |B|^{|A|}$.

**Korollar 3.7** *Eine $n$–elementige Teilmenge besitzt $2^n$ verschiedene Teilmengen.*

BEWEIS: $A$ sei eine $n$–elementige Menge, dann: $|\{0,1\}^A| = 2^{|A|}$; |Menge aller Abbildungen von $A \to \{0,1\}| = |\mathcal{P}(A)|$. □

Ein $k$–Tupel $(a_1, \ldots, a_k) \in A^k$ mit $a_i \neq a_j$ für $j \neq i$ heißt <u>geordnete Probe</u> aus $A$ vom Umfang $k$ <u>ohne Wiederholung</u>.

**Satz 3.8** *Zu einer $n$–elementigen Menge gibt es $(n)_k := n(n-1)\cdots(n-k+1)$ geordnete Proben vom Umfang $k \geq 1$ ohne Wiederholung.*

BEWEIS: Für eine geordnete Probe $(a_1, \ldots, a_k)$ vom Umfang $k$ ohne Wiederholung gilt: $a_1 \in A$, $a_2 \in A\backslash\{a_1\}$, $a_3 \in A\backslash\{a_1, a_2\}, \ldots, a_k \in A\backslash\{a_1, \ldots, a_{k-1}\}$. Aus dem Multiplikationsprinzip, genauer 3.5, folgt nun die Behauptung. □

Speziell für $k = n$ erhalten wir

**Satz 3.9** *$n$ verschiedene Elemente können auf $(n)_n = n!$ verschiedene Arten angeordnet werden, d.h. es existieren $n!$ Permutationen einer $n$–elementigen Menge.*

Eine <u>ungeordnete Probe</u> vom Umfang $k$ <u>mit</u> bzw. <u>ohne</u> Wiederholung erhalten wir, indem wir geordnete Proben, die sich nur in der Reihenfolge unterscheiden, identifizieren. Die ungeordneten Proben vom Umfang $k$ <u>ohne</u> Wiederholung sind demnach einfach die $k$–elementigen Teilmengen von $A$.

**Satz 3.10** *Eine $n$–elementige Menge besitzt*

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

*verschiedene $k$–elementige Teilmengen.*

Beweis: Eine geordnete Probe vom Umfang $k$ ohne Wiederholung besteht aus einer $k$–elementigen Teilmenge und einer Anordnung. Es gibt $k!$ verschiedene Möglichkeiten der Anordnung (3.9), also (3.8):

$$(n)_k = \text{Anzahl der } k\text{–elementigen Teilmengen} \times k!.$$

Hieraus folgt die Behauptung. □

**Korollar 3.11** *(i) Es gibt $n^k$ Möglichkeiten, $k$ unterscheidbare Kugeln auf $n$ unterscheidbare Urnen zu verteilen.*

*(ii) Es gibt $(n)_k$ Möglichkeiten, $k$ unterscheidbare Kugeln so auf $n$ unterscheidbare Urnen zu verteilen, dass keine Urne mehr als eine Kugel enthält.*

*(iii) Es gibt $\binom{n}{k}$ Möglichkeiten, $k$ ununterscheidbare Kugeln so auf $n$ unterscheidbare Urnen zu verteilen, dass keine Urne mehr als eine Kugel enthält.*

**Satz 3.12** *Es gibt*

$$\binom{k}{k_1, k_2, \ldots, k_n} := \frac{k!}{k_1!k_2!\cdots k_n!}$$

*Möglichkeiten, $k$ unterscheidbare Kugeln so auf $n$ unterscheidbare Urnen zu verteilen, dass genau $k_i$ Kugeln in die Urne Nummer $i$ kommen ($k_i \geq 0$, $i = 1, \ldots, n$; $k_1 + k_2 + \ldots + k_n = k$).*

Beweis: Es gibt

$$\binom{k}{k_1} \text{Möglichkeiten der } k_1 \text{ Kugeln für Urne 1}$$

$$\binom{k-k_1}{k_2} \text{Möglichkeiten der } k_2 \text{ Kugeln für Urne 2}$$

$$\vdots \quad \vdots \quad \vdots$$

$$\binom{k-k_1-\ldots-k_{n-2}}{k_{n-1}} \quad \text{Möglichkeiten der } k_{n-1} \text{ Kugeln für Urne } n-1.$$

Ausmultiplikation liefert nun:

$$
\begin{aligned}
&\text{Möglichkeiten insgesamt}\\
&= \binom{k}{k_1}\binom{k-k_1}{k_2}\cdots\binom{k-k_1-\ldots-k_{n-2}}{k_{n-1}}\\
&= \frac{k!}{k_1!(k-k_1)!}\times\frac{(k-k_1)!}{k_2!(k-k_1-k_2)!}\times\ldots\\
&\qquad\times\frac{(k-k_1-\ldots-k_{n-2})!}{k_{n-1}!(k-k_1-\ldots-k_{n-1})!}\\
&= \frac{k!}{k_1!\cdots k_n!}.
\end{aligned}
$$

$\square$

**Bemerkung** Die Größen $\binom{k}{k_1,\ldots,k_n}$ heißen Polynomialkoeffizienten. Wegen $\binom{n}{k}=\binom{n}{k,n-k}$ verallgemeinern sie die Binomialkoeffizienten $\binom{n}{k}$.

**Korollar 3.13**

$$
\begin{aligned}
&(a_1+\ldots+a_n)^k\\
&= \sum_{k_1\geq 0,\ldots,k_n\geq 0,\, k_1+\ldots+k_n=k}\binom{k}{k_1,\ldots,k_n}a_1^{k_1}a_2^{k_2}\cdots a_n^{k_n}.
\end{aligned}
$$

**Korollar 3.14** *(i)*

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

*(ii)*

$$\sum_{k=0}^{r}\binom{n}{k}\binom{m}{r-k}=\binom{m+n}{r}.$$

*(iii)*

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

Beweis:

(i) Zerlegung der Potenzmenge einer $n$–elementigen Menge gemäß Mächtigkeit der Teilmenge; 3.7 $\Rightarrow$ Behauptung.

(ii) $\binom{m+n}{r}$ = Anzahl der $r$–elementigen Teilmengen von $\{1,\ldots,n,n+1,\ldots,m+n\}$. Die Anzahl der Möglichkeiten, hierbei $k$ Elemente aus $\{1,\ldots,n\}$ und somit $r-k$ aus $\{n+1,\ldots,n+m\}$ auszuwählen, ist $\binom{n}{k}\binom{m}{r-k}$.

(iii) Folgt mit $m=r=n$ aus (ii) wegen $\binom{n}{k}^2=\binom{n}{k}\binom{n}{n-k}$.

$\square$

# 4 Vermischte Aufgaben

**Aufgabe 4.1** *Aus einer Schulklasse mit 20 Schülern wird eine Woche lang (5 Tage) jeden Morgen ein Schüler zufällig ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass mindestens 1 Schüler mehrmals ausgewählt wird?*

LÖSUNG: Laplace–Experiment mit $\Omega = \{1, \ldots, 20\}^5$, $|\Omega| = 20^5$; ungünstige Fälle: alle geordneten Proben vom Umfang 5 ohne Wiederholung, d.h. $(20)_5$. Also:

$$\begin{aligned} &\text{gesuchte Wahrscheinlichkeit} \\ &= \frac{20^5 - (20)_5}{20^5} \\ &= 1 - \frac{20 \times 19 \times \ldots \times 16}{20^5} = 0,4186. \end{aligned}$$

□

**Aufgabe 4.2** *Sack mit $N$ Nüssen, darunter $S$ schlechte Nüsse. Gezogen wird eine Stichprobe vom Umfang $n$. Wie groß ist die Wahrscheinlichkeit $p(s)$, dass in der Stichprobe genau $s$ schlechte Nüsse sind, $s = 0, 1, \ldots, n$?*

LÖSUNG: $\{1, \ldots, S\} \doteq$ Menge der schlechten Nüsse von $\{1, \ldots, N\}$. Laplace–Experiment mit $\Omega = \{A \subset \{1, \ldots, N\} : |A| = n\}$, $|\Omega| = \binom{N}{n}$. Dann:

$$\begin{aligned} &p(s) \\ &= \frac{|A \in \Omega \text{ mit } |A \cap \{1, \ldots, S\}| = s|}{\binom{N}{n}} \\ &= \frac{\binom{S}{s}\binom{N-S}{n-s}}{\binom{N}{n}} \\ &=: H_{N,S,n}(s). \end{aligned}$$

$H_{N,S,n}$ heißt Hypergeometrische Verteilung zu den Parametern $N$, $S$, $n$ (Qualitätskontrolle).

□

**Aufgabe 4.3** *Skatspiel: 32 Karten, 3 Spieler, je 10 Karten; „Skat“ mit 2 Karten. Es gibt vier Buben.*

(i) *Spieler A habe 2 Buben. Wie groß ist die Wahrscheinlichkeit, dass die Spieler B und C jeweils 1 Buben besitzen?*

LÖSUNG: Es gibt $\binom{22}{10,10,2}$ mögliche Verteilungen der 22 Karten, die A nicht besitzt, auf B,C und den Skat. Diese sind gleich wahrscheinlich.

Die Anzahl der günstigen Möglichkeiten beträgt $\binom{20}{9,9,2} \times \binom{2}{1,1,0}$. Also beträgt die gesuchte Wahrscheinlichkeit

$$\frac{\binom{20}{9,9,2} \times \binom{2}{1,1,0}}{\binom{22}{10,10,2}} = \frac{100}{231}.$$

□

(ii) *Gesucht: Wahrscheinlichkeit, dass einer der beiden Spieler beide Buben besitzt.*

LÖSUNG:

$$\frac{2 \times \binom{20}{8,10,2} \times \binom{2}{2,0,0}}{\binom{22}{10,10,2}} = \frac{90}{231}.$$

□

(iii) *Gesucht: Wahrscheinlichkeit, dass beide Buben im Skat liegen.*

LÖSUNG:

$$\frac{\binom{20}{10,10,0} \times \binom{2}{0,0,2}}{\binom{22}{10,10,2}} = \frac{1}{231}.$$

□

(iv) *Gesucht: Wahrscheinlichkeit, dass genau 1 Bube im Skat liegt:*

LÖSUNG:

$$\frac{2 \times \binom{20}{9,10,1} \times \binom{2}{1,0,1}}{\binom{22}{10,10,2}} = \frac{40}{231}.$$

□

(v) *Bilderschecks in Warenprodukten: $k$ Warenpackungen (Cornflakes). In jeder Packung ist genau 1 von $n$ möglichen Sammelmarken (etwa $n =$ 11 Fußballspieler).*

*Annahme: Laplace–Experiment, es gibt $n^k$ Möglichkeiten der Verteilung. Gesucht: Wahrscheinlichkeit $p_m$, dass wenigstens $m$ Sammelmarken fehlen.*

LÖSUNG: $\Omega =$ Menge aller möglichen Verteilungen von $k$ unterscheidbaren Kugeln ($\doteq$ Packungen) auf $n$ unterscheidbare Urnen ($\doteq$ Sammelmarken). $|\Omega| = n^k$.

$A_i :=$ Menge aller Verteilungen, bei denen die $i$–te Urne leer ist.

Für $1 \leq i_1 < \ldots < i_r \leq n$ gilt:

$$P\Big(A_{i_1} \cap \ldots \cap A_{i_r}\Big) = \frac{(n-r)^k}{n^k}.$$

*Es folgt mit der Bezeichnung von 2.4*

$$\begin{aligned} S_r &= \sum_{1 \leq i_1 < \ldots < i_r \leq n} P\Big(A_{i_1} \cap \ldots \cap A_{i_r}\Big) \\ &= \binom{n}{r} \frac{(n-r)^k}{n^k} \end{aligned}$$

*und damit aus 2.6*

$$\begin{aligned} p_m &= P(C_m) \\ &= \sum_{r=m}^{n} (-1)^{r-m} \binom{r-1}{m-1} \binom{n}{r} \frac{(n-r)^k}{n^k}. \end{aligned}$$

□

# 5 Bedingte Wahrscheinlichkeiten

$(\Omega, \mathcal{A}, P)$ sei Wahrscheinlichkeitsraum, $B \in \mathcal{A}$ mit $P(B) > 0$.
Es sei bekannt, dass das Ereignis $B$ eingetreten ist.
Neues Experiment: Ergebnismenge $\Omega' = B$.
Heuristisch: Wahrscheinlichkeit, dass $A$ eintritt, wenn bereits bekannt ist, dass $B$ eingetreten ist, ist $P(A \cap B)/P(B)$.

**Definition 5.1** $(\Omega, \mathcal{A}, P)$ sei Wahrscheinlichkeitsraum, $A \in \mathcal{A}$ und $B \in \mathcal{A}$ mit $P(B) > 0$. Dann heißt

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

bedingte Wahrscheinlichkeit von $A$ unter (der Bedingung) $B$.

**Beispiel 5.2** Für die beiden ersten Kinder einer Familie seien die 4 Geschlechtskombinationen $J-J$, $M-M$, $J-M$ und $M-J$ gleich wahrscheinlich. Von einer Familie sei bekannt, dass wenigstens eines der Kinder ein Junge ist. Wie groß ist die Wahrscheinlichkeit, dass diese Familie sogar zwei Jungen hat?

LÖSUNG: $A_1 :\doteq$ 1. Kind ist ein Junge, $A_2 :\doteq$ 2. Kind ist ein Junge. Damit:

$$\begin{aligned} & P(A_1 \cap A_2 | A_1 \cup A_2) \\ &= \frac{P((A_1 \cap A_2) \bigcap (A_1 \cup A_2))}{P(A_1 \cup A_2)} \\ &= \frac{P(A_1 \cap A_2)}{P(A_1 \cup A_2)} \\ &= \frac{P(A_1 \cap A_2)}{P(A_1) + P(A_2) - P(A_1 \cap A_2)} \\ &= \frac{\frac{1}{4}}{\frac{1}{2} + \frac{1}{2} - \frac{1}{4}} \\ &= \frac{1}{3}. \end{aligned}$$

□

**Satz 5.3** *$(\Omega, \mathcal{A}, P)$ sei Wahrscheinlichkeitsraum, $B \in \mathcal{A}$ mit $P(B) > 0$. Die bedingte Wahrscheinlichkeit $P_B : \mathcal{A} \to \mathbb{R}_+$ ist eine Wahrscheinlichkeit, d.h. $P_B$ erfüllt die Axiome 1,2,3.*

Beweis: Trivial, Axiome nachprüfen. □

**Satz 5.4** *$A_1, \ldots, A_n$ Ereignisse mit $P(A_1 \cap \ldots \cap A_{n-1}) > 0$. Dann:*

$$\begin{aligned} & P(A_1 \cap \ldots \cap A_n) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \times \ldots \\ &\quad \times P(A_n|A_1 \cap \ldots \cap A_{n-1}). \end{aligned}$$

Beweis: $\underline{n = 2:}$ $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$;
$\underline{n \to n+1:}$

$$\begin{aligned} & P(A_1 \cap \ldots \cap A_{n+1}) \\ &= P(A_1 \cap \ldots \cap A_n)P(A_{n+1}|A_1 \cap \ldots \cap A_n) \\ &=_{\text{Ind. V.}} P(A_1)P(A_2|A_1) \times \cdots \\ &\quad \times P(A_{n+1}|A_1 \cap \ldots \cap A_n). \end{aligned}$$

□

**Beispiel 5.5** 16 weiße, 16 schwarze Schachfiguren liegen im Kasten. 3 Figuren werden zufällig ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass alle 3 Figuren schwarz sind?

Lösung: $A_i$ sei das Ereignis, dass die $i$–te Figur schwarz ist. Dann:

$$\begin{aligned} & P(A_1 \cap A_2 \cap A_3) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\ &= \frac{16}{32} \times \frac{15}{31} \times \frac{14}{30} \qquad \left( = \frac{\binom{16}{3}}{\binom{32}{3}} \right). \end{aligned}$$

□

**Satz 5.6 (Totale Wahrscheinlichkeit)** *$(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum, $B_1, \ldots, B_n$ seien disjunkte Ereignisse mit $\sum_{i=1}^n B_i = \Omega$, $P(B_i) > 0$, $i = 1, \ldots, n$. Dann:*

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i), \qquad A \in \mathcal{A}.$$

BEWEIS:

$$\begin{aligned} P(A) &= P(\Omega \cap A) \\ &= P\left(\sum_{i=1}^{n}(B_i \cap A)\right) \\ &= \sum_{i=1}^{n} P(B_i \cap A) \\ &= \sum_{i=1}^{n} P(B_i)P(A|B_i) \end{aligned}$$

□

**Beispiel 5.7 (Zweistufiges Experiment)** In Urne 1 liegen 2 weiße und 8 schwarze Kugeln, in Urne 2 liegen 4 weiße und 6 schwarze Kugeln.
Zunächst wird gewürfelt. Bei einer 5 oder 6 erfolgt eine Ziehung aus Urne 1, bei einer 1-4 wird aus Urne 2 gezogen.
Wie groß ist die Wahrscheinlichkeit, eine weiße Kugel zu ziehen?

LÖSUNG: $A \doteq$ eine weiße Kugel wird gezogen,
$B_1 \doteq 5$ oder 6 beim Würfeln,
$B_2 \doteq 1$–4 beim Würfeln. Dann:

$$\begin{aligned} P(A) &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) \\ &= \frac{1}{3} \times \frac{1}{5} + \frac{2}{3} \times \frac{2}{5} = \frac{1}{3}. \end{aligned}$$

□

**Satz 5.8 (Bayessche Formel)** *Zusätzlich zu den Voraussetzungen von Satz 5.6 gelte $P(A) > 0$. Dann gilt:*

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}.$$

BEWEIS:

$$\begin{aligned} P(B_i|A) &= \frac{P(A \cap B_i)}{P(A)} \\ &= \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}. \end{aligned}$$

□

**Beispiel 5.9 (Fortsetzung von Beispiel 5.7)** $A \doteq$ Ziehen einer weißen Kugel, $P(A) = 1/3$.
Wie groß ist die Wahrscheinlichkeit von $B_i$, $i = 1, 2$, wenn bekannt ist, dass eine weiße Kugel gezogen wurde, d.h. wie groß ist die Wahrscheinlichkeit, dass eine weisse Kugel aus Urne $i$ stammt?

LÖSUNG:

$$\begin{aligned} P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)} \\ &= \frac{1/3 \times 1/5}{1/3 \times 1/5 + 2/3 \times 2/5} = \frac{1}{5} \\ &\Rightarrow P(B_2|A) = \frac{4}{5}. \end{aligned}$$

$P(B_i|A)$ heißt a posteriori Wahrscheinlichkeit von $B_i$, $P(B_i)$ heißt a priori Wahrscheinlichkeit von $B_i$. □

**Beispiel 5.10 (Überprüfung, ob radikal)** $R :\doteq$ Kandidat ist radikal, $B :\doteq$ Kandidat wird für radikal erklärt.
Eine Überprüfung ergebe mit der Wahrscheinlichkeit 0,95 ein richtiges Ergebnis, d.h.

$$P(B|R) = 0,95;\ P(B^c|R^c) = 0,95.$$

Es sei $P(R) = 0,005$.
Wie groß ist die Wahrscheinlichkeit, das ein als radikal erklärter Kandidat tatsächlich radikal ist?

LÖSUNG: Gesucht:

$$\begin{aligned} P(R|B) &= \frac{P(R \cap B)}{P(B)} \\ &= \frac{P(R)P(B|R)}{P(R)P(B|R) + P(R^c)P(B|R^c)} \\ &= \frac{5/1000 \times 95/100}{5/1000 \times 95/100 + 995/1000 \times 5/100} \\ &= \frac{95}{1090}\ (!). \end{aligned}$$

□

**Beispiel 5.11 (Brustkrebs-Screening durch Mammographie)** [1]
Die Wahrscheinlichkeit, dass eine 50jährige Frau Brustkrebs hat, ist bei etwa 0,8% anzusiedeln.

[1]Aus: Christian Hesse (2010). *Warum Mathematik glücklich macht.* C.H. Beck, München, S. 199ff.

Die Wahrscheinlichkeit, dass das Mammogramm einer Patientin positiv ist, wenn sie Brustkrebs hat, liegt bei etwa 90% (sog. Sensitivität des Untersuchungsverfahrens).
Die Wahrscheinlichkeit, dass ein durchgeführtes Mammogramm positiv ist, wenn die Patientin keinen Brustkrebs hat, liegt bei etwa 7% (sog. Falsch-Positiv-Rate).
Angenommen, eine 50-jährige Frau unterzieht sich einer Mammographie und der Befund ist positiv. Wie wahrscheinlich ist es, dass die Frau tatsächlich Brustkrebs hat?

LÖSUNG: $B :\doteq$ Brustkrebserkrankung, $M :\doteq$ Mammogramm ist positiv. Dann:

$$P(B) = \frac{8}{1000}, \quad P(M \mid B) = \frac{90}{100}, \quad P(M \mid B^c) = \frac{7}{100}.$$

Gesucht:

$$\begin{aligned} P(B \mid M) &= \frac{P(B)P(M \mid B)}{P(B)P(\mid B) + P(B^c)P(M \mid B^c)} \\ &= \frac{\frac{8}{1000}\frac{90}{100}}{\frac{8}{1000}\frac{90}{100} + \frac{992}{1000}\frac{7}{100}} \\ &= \frac{720}{7664} \\ &\approx \frac{9}{100} \text{ (!)}. \end{aligned}$$

□

**Beispiel 5.12 (Laplacescher Folgesatz)** In einer Urne liegen $N$ Kugeln, $W$ weiße und $N - W$ schwarze. $W$ sei unbekannt; alle $N + 1$ möglichen Mischungsverhältnisse besitzen die gleiche Wahrscheinlichkeit $1/(N + 1)$.
Es werden nacheinander $n + 1$ Kugeln ohne Zurücklegen gezogen.
Wie groß ist die Wahrscheinlichkeit, dass die $n + 1$-te Kugel weiß ist, wenn die ersten $n$ Kugeln weiß gewesen sind?

LÖSUNG: $A_i :\doteq$ nur weiße Kugeln unter den ersten $i$ Ziehungen.
Offenbar gilt $A_1 \supset A_2 \supset \ldots$ und gesucht ist

$$P(A_{n+1}|A_n) = \frac{P(A_{n+1} \cap A_n)}{P(A_n)} = \frac{P(A_{n+1})}{P(A_n)}.$$

Wir erhalten:

$$
\begin{aligned}
P(A_n) &= \sum_{w=0}^{N} P\{W=w\}P(A_n|\{W=w\}) \\
&= \sum_{w=0}^{N} \frac{1}{N+1} \times \frac{\binom{w}{n}}{\binom{N}{n}} \\
&= \frac{1}{(N+1)\binom{N}{n}} \sum_{w=n}^{N} \binom{w}{n} \\
&= \frac{1}{N+1} \times \frac{\binom{N+1}{n+1}}{\binom{N}{n}} \\
&= \frac{1}{n+1}.
\end{aligned}
$$

Behauptung:

$$\sum_{w=n}^{N} \binom{w}{n} = \binom{N+1}{n+1}.$$

Denn: $\binom{N+1}{n+1}$ ist die Anzahl der Möglichkeiten, aus der Menge $\{1,2,\ldots,N+1\}$ eine $n+1$–elementige Teilmenge auszuwählen. Dabei gibt es $\binom{w}{n}$ Möglichkeiten, die Auswahl so zu treffen, dass $w+1$ das größte der ausgewählten Elemente ist, $w = n,\ldots,N$.

Insgesamt erhalten wir somit:

$$P(A_{n+1}|A_n) = \frac{n+1}{n+2}$$

unabhängig von $N$! □

# 6 Unabhängigkeit

Gegeben ist ein Würfel, $A := \{2,4,6\}$, $B := \{5,6\}$, $P(A) = 1/2$.

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{1/6}{2/6} = \frac{1}{2} = P(A),$$

d.h. die zusätzliche Information des Eintretens von $B$ hat in diesem speziellen Fall keinen Einfluss auf die Wahrscheinlichkeit des Eintretens von $A$.

**Definition 6.1** $A, B$ Ereignisse mit $P(B) > 0$; dann:

$$A \text{ unabhängig von } B :\Leftrightarrow P(A) = P(A|B).$$

**Satz 6.2** *$A, B$ Ereignisse mit $P(A) > 0$ und $P(B) > 0$; dann:*

$$\begin{aligned} & A \textit{ unabhängig von } B \\ \Leftrightarrow \quad & B \textit{ unabhängig von } A \\ \Leftrightarrow \quad & P(A \cap B) = P(A)P(B). \end{aligned}$$

BEWEIS:

$$\begin{aligned} & A \text{ unabhängig von } B \\ \Leftrightarrow \quad & P(A) = P(A \cap B)/P(B) \\ \Leftrightarrow \quad & P(B) = P(A \cap B)/P(A) \\ \Leftrightarrow \quad & P(A \cap B) = P(A)P(B). \end{aligned}$$

□

**Definition 6.3** $A, B$ beliebige Ereignisse, dann:

$$A, B \text{ unabhängig } :\Leftrightarrow P(A \cap B) = P(A)P(B).$$

**Satz 6.4** *$A, B$ unabhängig, dann gilt:*

- *$A^c, B$ sind unabhängig,*
- *$A, B^c$ sind unabhängig,*
- *$A^c, B^c$ sind unabhängig.*

BEWEIS:

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

□

**Satz 6.5** *$A, B$ seien unabhängig; $A, C$ seien unabhängig; $B \cap C = \emptyset$. Dann gilt:*

$$A, B \cup C \textit{ sind unabhängig.}$$

Beweis:

$$
\begin{aligned}
P(A\cap(B\cup C)) &= P((A\cap B)+(A\cap C))\\
&= P(A\cap B)+P(A\cap C)\\
&= P(A)P(B)+P(A)P(C)\\
&= P(A)P(B\cup C).
\end{aligned}
$$

□

**Bemerkung** $A, B, C$ paarweise unabhängig $\not\Rightarrow P(A\cap B\cap C) = P(A)P(B)P(C)$. Gegenbeispiele: Siehe Übungen.

**Definition 6.6** $\Omega \neq \emptyset$; $\mathcal{A} \subset \mathcal{P}(\Omega)$ ist Algebra über $\Omega :\Leftrightarrow$

(i) $\Omega \in \mathcal{A}$,

(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,

(iii) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

**Definition 6.7** $\emptyset \neq \mathcal{S} \subset \mathcal{P}(\Omega)$. Dann:

$$
\begin{aligned}
\alpha(\mathcal{S}) &:= \bigcap_{\mathcal{P}(\Omega)\supset\mathcal{D}\supset\mathcal{S},\,\mathcal{D}\text{ Algebra}} \mathcal{D}\\
&=: \text{kleinste Algebra, die } \mathcal{S} \text{ enthält,}\\
\sigma(\mathcal{S}) &:= \bigcap_{\mathcal{P}(\Omega)\supset\mathcal{D}\supset\mathcal{S},\,\mathcal{D}\,\sigma\text{–Algebra}} \mathcal{D}\\
&=: \text{kleinste } \sigma\text{–Algebra, die } \mathcal{S} \text{ enthält.}
\end{aligned}
$$

Beachte: Der beliebige Durchschnitt von ($\sigma$–) Algebren ist wieder eine ($\sigma$–) Algebra; $\mathcal{P}(\Omega)$ ist eine ($\sigma$–) Algebra mit $\mathcal{P}(\Omega) \neq \emptyset$.

**Satz 6.8** $\emptyset \neq \mathcal{S} \subset \mathcal{P}(\Omega)$. *Dann: $\alpha(\mathcal{S})$ = Menge aller endlichen, disjunkten Vereinigungen von endlichen Durchschnitten von Mengen aus $\mathcal{S}$ oder deren Komplemente, d.h.*

$$
\begin{aligned}
\mathcal{S}_1 &:= \{S \subset \Omega : S \in \mathcal{S} \text{ oder } S^c \in \mathcal{S}\},\\
\mathcal{S}_2 &:= \{S_1 \cap \ldots \cap S_n : n \in \mathbb{N}, S_i \in \mathcal{S}_1, i = 1, \ldots, n\}\\
\mathcal{S}_3 &:= \{T_1 \cup \ldots \cup T_n : T_j \in \mathcal{S}_2, j = 1, \ldots, n,\\
&\qquad\qquad \text{paarweise disjunkt}, n \in \mathbb{N}\}\\
&\Rightarrow \alpha(\mathcal{S}) = \mathcal{S}_3.
\end{aligned}
$$

Beweis:

1. $\mathcal{S}_3 \subset \alpha(\mathcal{S})$ (trivial, da eine Algebra $\cap$–stabil, $\cup$–stabil und Komplement–stabil ist).

2. $\mathcal{S}_3 \neq \emptyset$, da $\mathcal{S}_3 \supset \mathcal{S}_2 \supset \mathcal{S}_1 \supset \mathcal{S} \neq \emptyset$.

3. $T \in \mathcal{S}_2 \Rightarrow T^c \in \mathcal{S}_3$.
   (Denn: $T = S_1 \cap \ldots \cap S_n \in \mathcal{S}_2$ mit $S_i \in \mathcal{S}_1 \Rightarrow$

$$\begin{aligned} T^c &= S_1^c \cup \ldots \cup S_n^c \\ &= S_1^c \cup (S_2^c \backslash S_1^c) \cup S_3^c \backslash (S_1^c \cup S_2^c) \cup \ldots \\ &\qquad \cup S_n^c \backslash (S_1^c \cup \ldots \cup S_{n-1}^c) \\ &= S_1^c + (S_2^c \cap S_1) + (S_3^c \cap S_1 \cap S_2) + \ldots \\ &\qquad + (S_n^c \cap S_1 \cap \ldots \cap S_{n-1}) \in \mathcal{S}_3. \end{aligned}$$

4. $T_1, T_2 \in \mathcal{S}_2 \Rightarrow T_1 \cap T_2 \in \mathcal{S}_2$ (trivial).

5. $U_1, U_2 \in \mathcal{S}_3 \Rightarrow U_1 = T_{11} + \ldots + T_{1n_1}$ mit disjunkten $T_{1j} \in \mathcal{S}_2$, $U_2 = T_{21} + \ldots + T_{2n_2}$ mit disjunkten $T_{2j} \in \mathcal{S}_2$

$$\Rightarrow U_1 \cap U_2 = \underbrace{\bigcup_{i,j} ( \underbrace{T_{1i} \cap T_{2j}}_{\in \mathcal{S}_2 \text{ wegen } 4.} )}_{\text{disjunkte Vereinigung}} \in \mathcal{S}_3.$$

6. $U \in \mathcal{S}_3 \Rightarrow U = T_1 + \ldots + T_n$, $T_i \in \mathcal{S}_2, i = 1, \ldots, n$, disjunkt $\Rightarrow U^c = T_1^c \cap \ldots \cap T_n^c \in \mathcal{S}_3$ wegen 3. und 5. ($T_j^c \in \mathcal{S}_3$ wegen 3.).

Wegen 2., 5. und 6. ist $\mathcal{S}_3$ eine Algebra. Hieraus und aus 1. folgt, dass $\mathcal{S}_3 = \alpha(S)$.

$\square$

**Definition 6.9** $A_1, \ldots, A_n$ seien beliebige Ereignisse, dann:

$$\begin{aligned} &A_1, \ldots, A_n \text{ (global) unabhängig} \\ :\Leftrightarrow \quad &A_i, B \text{ unabhängig für } 1 \leq i \leq n \\ &\qquad\qquad \text{und } B \in \alpha(\{A_j : j \neq i\}). \end{aligned}$$

**Satz 6.10** $A_1, \ldots, A_n$ *sind unabhängig* $\Leftrightarrow$

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}) \tag{1}$$

*für* $2 \leq k \leq n$, $1 \leq i_1 < \ldots < i_k \leq n$.

BEWEIS:

„⇒“

$$
\begin{aligned}
& A_{i_2} \cap \ldots \cap A_{i_k} \in \alpha(\{A_j : j \neq i_1\}) \\
& \quad \Rightarrow \quad A_{i_1}, A_{i_2} \cap \ldots \cap A_{i_k} \text{ unabhängig} \\
& \quad \Rightarrow \quad P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2} \cap \ldots \cap A_{i_k}) \\
& \qquad\qquad\qquad\qquad \text{u.s.w. (Induktion)}
\end{aligned}
$$

„⇐“ Es genügt zu zeigen: $A_1$ ist unabhängig von jedem Ereignis aus $\alpha(\{A_2, \ldots, A_n\})$.
In (1) beliebige der $A_{i_j}$ durch $A_{i_j}^c$ ersetzbar (z.B. $P(A_{i_1} \cap A_{i_2}^c \cap \ldots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}^c) \cdots P(A_{i_k})$, siehe 6.4) $\Rightarrow$

$$
\begin{aligned}
& P(A_1 \cap A_{i_2}^{(c)} \cap \ldots \cap A_{i_k}^{(c)}) \\
& \quad = \quad P(A_1) P(A_{i_2}^{(c)}) \cdots P(A_{i_k}^{(c)})
\end{aligned}
$$

mit $A_{i_j}^{(c)} = A_{i_j}$ oder $A_{i_j}^c$,
d.h. $A_1$ unabhängig von allen Ereignissen aus $\mathcal{S}_2(A_2, \ldots, A_n)$
$\Rightarrow_{6.5}$ $A_1$ unabhängig von allen Ereignissen aus $\mathcal{S}_3(A_2, \ldots, A_n) =_{6.8} \alpha(\{A_2, \ldots, A_n\})$.

$\square$

**Definition 6.11** $A_i, i \in I$, beliebige Ereignisse, dann:

$$
\begin{aligned}
& A_i, i \in I, \text{ unabhängig} \\
& \quad :\Leftrightarrow \quad A_i, G \text{ unabhängig für beliebiges} \\
& \qquad\qquad\qquad i \in I \text{ und } G \in \alpha(\{A_j : j \in I, j \neq i\}).
\end{aligned}
$$

**Bemerkung 6.12** (i)

$$
\begin{aligned}
& A_i, i \in I, \text{ unabhängig} \\
& \quad \Leftrightarrow \quad A_i, i \in I_0 \text{ unabhängig} \\
& \qquad\qquad\qquad \text{für alle endlichen Teilmengen } I_0 \text{ von } I.
\end{aligned}
$$

(ii) $A_1, A_2, \ldots$ sei eine Folge von Ereignissen, dann:

$$
\begin{aligned}
& A_1, A_2, \ldots \text{ unabhängig} \\
& \quad \Leftrightarrow \quad A_1, \ldots, A_n \text{ unabhängig für alle } n \in \mathbb{N}.
\end{aligned}
$$

Beweis:

(i)

„⇒“ trivial.

„$\Leftarrow$" Nach Satz 6.8 existiert zu jedem $G \in \alpha(\{A_j : j \neq i\})$ ein $I_0 \subset I$ mit $|I_0| < \infty$ und $G \in \alpha(\{A_j : j \in I_0\})$.

$\square$

**Definition 6.13** (i) $\mathcal{G}_i, i \in I$, unabhängige Algebren von Ereignissen

:$\Leftrightarrow$ $\mathcal{G}_i$ ist Ereignis–Algebra, $i \in I$, und für jedes $i \in I$ ist jedes $G \in \mathcal{G}_i$ unabhängig von allen $H \in \alpha\left(\bigcup_{j\neq i} \mathcal{G}_j\right)$.

(ii) $\mathcal{A}_i, i \in I$, unabhängige $\sigma$–Algebren von Ereignissen

:$\Leftrightarrow$ $\mathcal{A}_i$ ist $\sigma$–Algebra von Ereignissen, $i \in I$, und für jedes $i \in I$ ist jedes $A \in \mathcal{A}_i$ unabhängig von allen $B \in \sigma\left(\bigcup_{j\neq i} \mathcal{A}_j\right)$.

**Satz 6.14** $A_i, i \in I$ *unabhängig* $\Rightarrow \alpha(\{A_i\})$, $i \in I$, *unabhängige Algebren.*

Beweis:

$$\alpha(\{A_i\}) = \{A_i, A_i^c, \emptyset, \Omega\},$$

d.h.

$$\alpha(\{A_j : j \in I, j \neq i\}) = \alpha\left(\bigcup_{j\neq i} \alpha(\{A_j\})\right).$$

$\emptyset$ und $\Omega$ sind von allen Ereignissen unabhängig. $\square$

**Satz 6.15 (Borel–Cantelli Lemma)** $A_1, A_2, \ldots$ *seien Ereignisse;*

$$\begin{aligned} A &:= \{\omega \in \Omega : \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N}\} \\ &= \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m. \end{aligned}$$

*Dann gilt:*

*(i)* $P(A) = 0$, *falls* $\sum_{n\in\mathbb{N}} P(A_n) < \infty$.

*(ii)* $P(A) = 1$, *falls* $\sum_{n\in\mathbb{N}} P(A_n) = \infty$ *und* $A_1, A_2, \ldots$ *unabhängig.*

Beweis: Es gilt (siehe Übungen):

„(i)"

$$\begin{aligned} P(A) &= P\Big(\bigcap_{n\in\mathbb{N}} \underbrace{\bigcup_{m\geq n} A_m}_{\text{absteigende Folge}}\Big) \\ &= \lim_{n\in\mathbb{N}} P\left(\bigcup_{m\geq n} A_m\right), \end{aligned}$$

wobei

$$\begin{aligned} P\left(\bigcup_{m\geq n} A_m\right) &= \lim_{k\in\mathbb{N}} P\left(\bigcup_{m=n}^{k} A_m\right) \\ &\leq_{\text{s. Üb.}} \sum_{m\geq n} P(A_m) \to_{n\in\mathbb{N}} 0 \end{aligned}$$

als Rest einer konvergenten Reihe.

„(ii)“

$$\begin{aligned} P(A^c) &= P\Big(\bigcup_{n\in\mathbb{N}} \underbrace{\bigcap_{m\geq n} A_m^c}_{\text{aufsteigende Folge}}\Big) \\ &= \lim_{n\in\mathbb{N}} P\Big(\bigcap_{m\geq n} A_m^c\Big) \\ &= \lim_{n\in\mathbb{N}} P\Big(\bigcap_{N\geq n} \underbrace{\bigcap_{n\leq m\leq N} A_m^c}_{\text{absteigende Folge}}\Big) \\ &= \lim_{n\in\mathbb{N}} \Big(\lim_{N\in\mathbb{N}} P\big(\bigcap_{n\leq m\leq N} \underbrace{A_m^c}_{\text{unabhängig}}\big)\Big) \\ &= \lim_{n\in\mathbb{N}} \Big(\lim_{N\in\mathbb{N}} \underbrace{\prod_{n\leq m\leq N} (\underbrace{1-P(A_m)}_{\leq \exp(-P(A_m))})}_{\leq \exp\left(-\sum_{m=n}^{N} P(A_m)\right) \to_{N\in\mathbb{N}} 0}\Big), \end{aligned}$$

denn $1-x \leq \exp(-x)$ wegen Taylor–Entwicklung:

$$\exp(-x) = 1 - x + \underbrace{\exp(-\vartheta x)\frac{x^2}{2}}_{\geq 0} \geq 1 - x.$$

□

**Satz 6.16 (Fortsetzungssatz)** *$\mathcal{G}$ sei Algebra über $\Omega$, $Q$ ein $\sigma$–additiver, normierter Inhalt auf $\mathcal{G}$ (d.h. $Q : \mathcal{G} \to [0,\infty)$ mit $Q(\Omega) = 1$ und $Q(\bigcup_{n\in\mathbb{N}} G_n) = \sum_{n\in\mathbb{N}} Q(G_n)$ für disjunkte $G_n \in \mathcal{G}$, $n \in \mathbb{N}$ mit $\bigcup_{n\in\mathbb{N}} G_n \in \mathcal{G}$).*
*Dann gilt: Es existiert genau ein Wahrscheinlichkeitsmaß $P$ auf $\mathcal{A} := \sigma(\mathcal{G})$ mit $P/\mathcal{G} = Q/\mathcal{G}$.*

Beweis: Siehe Maßtheorie. (Etwa Satz 4.9 im Maßtheorie-Skript (\ ~`falk`\`downloads`\).

□

**Satz 6.17** *$\mathcal{G}_1, \mathcal{G}_2$ seien unabhängige Ereignis–Algebren. Dann sind $\mathcal{A}_1 := \sigma(\mathcal{G}_1)$, $\mathcal{A}_2 := \sigma(\mathcal{G}_2)$ unabhängige $\sigma$–Algebren.*

Beweis: Zu zeigen:

$$\begin{aligned} &\forall A_1 \in \mathcal{A}_1, \forall A_2 \in \mathcal{A}_2 : \\ &\qquad P(A_1 \cap A_2) = P(A_1)P(A_2). \end{aligned}$$

Sei $G_1 \in \mathcal{G}_1$ gegeben.

**1. Fall:** $P(G_1) = 0$. Dann gilt:

$$\begin{aligned} 0 &= P(\underbrace{G_1 \cap A_2}_{\subset G_1}) \\ &= P(G_1)P(A_2) = 0 \end{aligned}$$

für alle $A \in \mathcal{A}_2$.

**2. Fall:** $P(G_1) > 0$. Setze

$$Q(A_2) := \frac{P(G_1 \cap A_2)}{P(G_1)} \text{ für } A_2 \in \mathcal{A}_2.$$

Dann gilt

1. $$\forall G_2 \in \mathcal{G}_2 : Q(G_2) = P(G_2)$$

   wegen der Unabhängigkeit von $\mathcal{G}_1, \mathcal{G}_2$.
2. $Q$ ist Wahrscheinlichkeitsmaß auf $\mathcal{A}_2$; denn:
   (a) $Q(\Omega) = 1$.
   (b) $A_2^1, A_2^2, \ldots$ sei Folge disjunkter Mengen aus $\mathcal{A}_2$, dann:

$$\begin{aligned} Q\left(\sum_{n\in\mathbb{N}} A_2^n\right) &= \frac{P\left(G_1 \cap \sum_{n\in\mathbb{N}} A_2^n\right)}{P(G_1)} \\ &= \frac{P\left(\sum_{n\in\mathbb{N}}(G_1 \cap A_2^n)\right)}{P(G_1)} \\ &= \frac{\sum_{n\in\mathbb{N}} P\left(G_1 \cap A_2^n\right)}{P(G_1)} \\ &= \sum_{n\in\mathbb{N}} Q\left(A_2^n\right). \end{aligned}$$

Aus 1. und 2. folgt mittels des Fortsetzungssatzes 6.16, dass $Q(A_2) = P(A_2)$ für alle $A_2 \in \mathcal{A}_2$, d.h.

$$P(G_1 \cap A_2) = P(G_1)P(A_2) \text{ für alle } A_2 \in \mathcal{A}_2.$$

Analog schließt man: $G_1 \in \mathcal{G}_1$ durch $A_1 \in \mathcal{A}_1$ ersetzbar. $\square$

**Satz 6.18** *$\mathcal{G}_i$, $i \in I$, seien beliebige Algebren von Ereignissen zum Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann sind äquivalent:*

*(i)* *$\mathcal{G}_i$, $i \in I$, sind unabhängig.*

*(ii)* *$\mathcal{A}_i := \sigma(\mathcal{G}_i)$, $i \in I$, sind unabhängig.*

*(iii)* *Für je endlich viele verschiedene $i_1, \ldots, i_k \in I$ und $G_{i_1} \in \mathcal{G}_{i_1}, \ldots, G_{i_k} \in \mathcal{G}_{i_k}$ gilt:*

$$P\left(G_{i_1} \cap \ldots \cap G_{i_k}\right) = P\left(G_{i_1}\right) \cdots P\left(G_{i_k}\right).$$

Beweis:

**„(ii)⇒(i)"** Trivial.

**„(i)⇒(iii)"** $G_{i_1}$ ist unabhängig von allen Ereignissen aus $\alpha\Big(\bigcup_{j \neq i_1} \mathcal{G}_j\Big)$, speziell von $G_{i_2} \cap \ldots \cap G_{i_k}$, d.h.

$$\begin{aligned} &P\left(G_{i_1} \cap \left(G_{i_2} \cap \ldots \cap G_{i_k}\right)\right) \\ =& \quad \ldots = P\left(G_{i_1}\right) \cdots P\left(G_{i_k}\right) \end{aligned}$$

mittels Induktion.

**„(iii)⇒(ii)"** Zu zeigen ist: $A_{i_0} \in \mathcal{A}_{i_0}$ ist unabhängig von allen $A \in \sigma\Big(\bigcup_{j \neq i_0} \mathcal{A}_j\Big)$.

Sei $G_{i_0} \in \mathcal{G}_{i_0}$ und $H \in \alpha\Big(\bigcup_{j \neq i_0} \mathcal{G}_j\Big)$

$$\begin{aligned} \Rightarrow_{6.8} \quad & H \in \alpha(\{G_{i_1}, \ldots, G_{i_r}\}) \text{ für geeignete} \\ & \qquad G_{i_1} \in \mathcal{G}_{i_1}, \ldots, G_{i_r} \in \mathcal{G}_{i_r} \\ \Rightarrow_{6.10} \quad & G_{i_0} \text{ und } H \text{ sind unabhängig, d.h.} \\ & \qquad \mathcal{G}_{i_0} \text{ und } \alpha\left(\bigcup_{j \neq i_0} \mathcal{G}_j\right) \text{ sind unabhängig} \\ \Rightarrow_{6.17} \quad & \mathcal{A}_{i_0} = \sigma(\mathcal{G}_{i_0}), \sigma\left(\alpha\left(\bigcup_{j \neq i_0} \mathcal{G}_j\right)\right) \text{ unabhängig }. \end{aligned}$$

Die Behauptung folgt nun aus den Gleichungen:

$$\begin{aligned} \sigma\left(\alpha\left(\bigcup_{j \neq i_0} \mathcal{G}_j\right)\right) \Bigg( &= \sigma\left(\bigcup_{j \neq i_0} \mathcal{G}_j\right) \Bigg) \\ &= \sigma\left(\bigcup_{j \neq i_0} \underbrace{\sigma(\mathcal{G}_j)}_{=\mathcal{A}_j}\right). \end{aligned}$$

Zum zweiten „=":

„⊂": Trivial.

„$\supset$“: Für $i \neq i_0$ gilt:

$$
\begin{aligned}
& \bigcup_{j \neq i_0} \mathcal{G}_j \supset \mathcal{G}_i \\
\Rightarrow \quad & \sigma\left(\bigcup_{j \neq i_0} \mathcal{G}_j\right) \supset \sigma(\mathcal{G}_i) \\
\Rightarrow \quad & \sigma\left(\bigcup_{j \neq i_0} \mathcal{G}_j\right) \supset \bigcup_{i \neq i_0} \sigma(\mathcal{G}_i) \\
\Rightarrow \quad & \sigma\left(\bigcup_{j \neq i_0} \mathcal{G}_j\right) \supset \sigma\left(\bigcup_{i \neq i_0} \underbrace{\sigma(\mathcal{G}_i)}_{=\mathcal{A}_i}\right)
\end{aligned}
$$

□

Gegeben seien zwei Zufallsexperimente, die durchgeführt werden, ohne dass sie sich gegenseitig beeinflussen. Gesucht ist ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ zur mathematischen Beschreibung des Zufallsexperimentes, welches darin besteht, dass die beiden Experimente ohne wechselseitige Beeinflussung — also unabhängig — durchgeführt werden.
Die einzelnen Experimente werden durch $(\Omega_1, \mathcal{A}_1, P_1)$, $(\Omega_2, \mathcal{A}_2, P_2)$ beschrieben. Nahe liegend:

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

Jedes $A_1 \in \mathcal{A}_1$ kann identifiziert werden mit $A_1 \times \Omega_2$,
jedes $A_2 \in \mathcal{A}_2$ kann identifiziert werden mit $\Omega_1 \times A_2$,
d.h. Forderung:

$$A_1 \times \Omega_2,\ \Omega_1 \times A_2 \text{ sind Ereignisse}$$

$$\Rightarrow (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) = A_1 \times A_2 \text{ Ereignis.}$$

Daher:

$$\mathcal{A} := \sigma\left(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}\right).$$

Ferner soll die Forderung

$$P_1(A_1) = P(A_1 \times \Omega_2),\ P_2(A_2) = P(\Omega_1 \times A_2)$$

erfüllt sein.
Zur Unabhängigkeit: $A_1 \times \Omega_2$ und $\Omega_1 \times A_2$ sollen stets unabhängig sein für $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, d.h. es soll gelten

$$
\begin{aligned}
P(A_1 \times A_2) &= P((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) \\
&= P(A_1 \times \Omega_2) P(\Omega_1 \times A_2) \\
&= P_1(A_1) P_2(A_2).
\end{aligned}
$$

**Satz 6.19** *$(\Omega_1, \mathcal{A}_1, P_1), \ldots, (\Omega_n, \mathcal{A}_n, P_n)$ seien Wahrscheinlichkeitsräume. Setze*

$$\begin{aligned}\Omega &:= \Omega_1 \times \ldots \times \Omega_n \\ &:= \{(\omega_1, \ldots, \omega_n) : \omega_i \in \Omega_i, 1 \leq i \leq n\}\end{aligned}$$

*und*

$$\mathcal{A} := \sigma(\{A_1 \times \ldots \times A_n : A_i \in \mathcal{A}_i, 1 \leq i \leq n\}.$$

*Dann gilt: Es existiert genau ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ mit*

$$P(A_1 \times \ldots \times A_n) = P_1(A_1) \cdots P_n(A_n).$$

Beweisskizze: Definiere $P$ auf Mengensystem $\mathcal{Z} := \{A_1 \times \ldots \times A_n : A_i \in \mathcal{A}_i, 1 \leq i \leq n\}$ durch

$$P(A_1 \times \ldots \times A_n) := \prod_{i=1}^{n} P(A_i).$$

Dann Fortsetzung von $P$ (zu einem (eindeutig bestimmten) $\sigma$–additiven Inhalt) auf $\alpha(\mathcal{Z})$. Die Behauptung folgt dann aus dem Fortsetzungssatz 6.16.

□

**Definition 6.20** $(\Omega, \mathcal{A}, P)$ ist unabhängiges Produkt der Wahrscheinlichkeitsräume $(\Omega_1, \mathcal{A}_1, P_1)$, ..., $(\Omega_n, \mathcal{A}_n, P_n)$ :⇔ $(\Omega, \mathcal{A}, P)$ wird definiert gemäß 6.19.

Schreibweise: $\Omega = \Omega_1 \times \ldots \times \Omega_n$, $\mathcal{A} = \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$, $P = P_1 \times \ldots \times P_n$. Im Fall

$$(\Omega_1, \mathcal{A}_1, P_1) = \ldots = (\Omega_n, \mathcal{A}_n, P_n) = (\Omega, \mathcal{A}, P),$$

d.h. unabhängige $n$–fache Wiederholung von $(\Omega, \mathcal{A}, P)$, schreiben wir kurz $(\Omega^n, \mathcal{A}^n, P^n)$.

# 7 Zufallsvariablen

$(\Omega, \mathcal{A}, P)$ zufälliges Experiment, $f : \Omega \to \Omega'$ eine Abbildung. Es sei $\omega$ ein Ergebnis; häufig interessiert weniger der exakte Ausgang $\omega$ sondern nur der Wert $f(\omega)$. Beispielsweise interessiert beim Schuss auf eine Zielscheibe weniger die genaue Lage des Einschusses sondern der Abstand zum Mittelpunkt. Daher wird man vor allem Ereignisse der Gestalt

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

betrachten, wobei $A'$ die im Bildraum von $f$ interessierenden Ereignisse durchläuft.

**Satz 7.1** *$(\Omega, \mathcal{A}, P)$ sei Wahrscheinlichkeitsraum und $f : \Omega \to \Omega'$ eine Abbildung. Setze*

$$\mathcal{A}' := \{A' \subset \Omega' : f^{-1}(A') \in \mathcal{A}\}$$

*und*

$$P'(A') := P(f^{-1}(A')) \text{ für alle } A' \in \mathcal{A}'.$$

*Dann ist $(\Omega', \mathcal{A}', P')$ ein Wahrscheinlichkeitsraum.*

Beweis:

1. $f^{-1}(\Omega') = \Omega \in \mathcal{A}$, d.h. $\Omega' \in \mathcal{A}'$.

2. Sei $A' \in \mathcal{A}'$

$$\begin{aligned} &\Rightarrow \; f^{-1}(A') =: A \in \mathcal{A} \\ &\Rightarrow \; f^{-1}(\Omega' \backslash A') = A^c \in \mathcal{A} \\ &\Rightarrow \; A'^c = \Omega' \backslash A' \in \mathcal{A}'. \end{aligned}$$

3. Sei $A'_n \in \mathcal{A}'$, $n \in \mathbb{N}$

$$\begin{aligned} &\Rightarrow \; f^{-1}(A'_n) \in \mathcal{A}, \; n \in \mathbb{N} \\ &\Rightarrow \; f^{-1}\left(\bigcup_{n \in \mathbb{N}} A'_n\right) = \bigcup_{n \in \mathbb{N}} f^{-1}(A'_n) \in \mathcal{A} \\ &\Rightarrow \; \bigcup_{n \in \mathbb{N}} A'_n \in \mathcal{A}'. \end{aligned}$$

Also ist $\mathcal{A}'$ eine $\sigma$–Algebra. Ferner ist $P'$ ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathcal{A}')$:

1. $P'(\Omega') = P(f^{-1}(\Omega')) = P(\Omega) = 1$.

2. Für paarweise disjunkte $A'_n$, $n \in \mathbb{N}$, gilt:

$$\begin{aligned} P'\left(\sum_{n \in \mathbb{N}} A'_n\right) &= P\left(f^{-1}\left(\sum_{n \in \mathbb{N}} A'_n\right)\right) \\ &= P\left(\sum_{n \in \mathbb{N}} f^{-1}\left(A'_n\right)\right) \\ &= \sum_{n \in \mathbb{N}} P\left(f^{-1}\left(A'_n\right)\right) \\ &= \sum_{n \in \mathbb{N}} P'\left(A'_n\right). \end{aligned}$$

□

**Bemerkung 7.2** $\mathcal{A}'$ heißt finale $\sigma$–Algebra bezüglich $f$; $P'$ heißt das durch $P$ und $f$ auf $\mathcal{A}'$ induzierte Wahrscheinlichkeitsmaß oder Bildmaß von $P$ unter $f$.

**Definition 7.3** $(\Omega, \mathcal{A}, P)$ sei ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{A}')$ ein messbarer Raum, dann: $f : \Omega \to \Omega'$ heißt $\underline{\mathcal{A}, \mathcal{A}'\text{–messbar}}$ oder <u>Zufallsgröße</u>, falls $f^{-1}(\mathcal{A}') \subset \mathcal{A}$, d.h. falls für alle $A' \in \mathcal{A}'$ gilt:

$$f^{-1}(A') \in \mathcal{A},$$

(d.h. $\mathcal{A}'$ ist sub–$\sigma$–Algebra der finalen $\sigma$–Algebra.) Schreibweise:

$$f : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}').$$

**Satz 7.4** $f : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$. *Dann:*

$$f^{-1}(\mathcal{A}') := \{f^{-1}(A') : A' \in \mathcal{A}'\} \text{ ist sub–}\sigma\text{–Algebra von } \mathcal{A}.$$

*(= <u>zu $f$ gehörende Vergröberung von $\mathcal{A}$</u>, <u>durch $f$ bestimmte Ereignisse von $\mathcal{A}$</u>).*

Beweis:

1. $\Omega = f^{-1}(\Omega')$.
2. $A \in f^{-1}(\mathcal{A}')$

$$\begin{aligned} &\Rightarrow \quad \exists A' \in \mathcal{A}' : A = f^{-1}(A') \\ &\Rightarrow \quad A^c = f^{-1}(A'^c) \in f^{-1}(\mathcal{A}'). \end{aligned}$$

3. $A_n \in f^{-1}(\mathcal{A}')$, $n \in \mathbb{N}$

$$\begin{aligned} &\Rightarrow \quad \exists A'_n \in \mathcal{A}' : A_n = f^{-1}(A'_n), \; n \in \mathbb{N} \\ &\Rightarrow \quad \bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} f^{-1}(A'_n) \\ &\qquad\qquad\qquad = f^{-1}\left(\bigcup_{n \in \mathbb{N}} A'_n\right) \in f^{-1}(\mathcal{A}'). \end{aligned}$$

$\square$

**Satz 7.5** $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ *messbare Räume,* $\mathcal{S} \subset \mathcal{P}(\Omega')$ *mit* $\sigma(\mathcal{S}) = \mathcal{A}'$. *Dann gilt:*

$$\begin{aligned} &f : \Omega \to \Omega' \text{ ist } \mathcal{A}, \mathcal{A}'\text{–messbar} \\ &\quad \Leftrightarrow \quad f^{-1}(S) \in \mathcal{A} \text{ für alle } S \in \mathcal{S}. \end{aligned}$$

Beweis:

„$\Rightarrow$“ Trivial.

„⇐“ Setze
$$\mathcal{F} := \{F \in \mathcal{P}(\Omega') : f^{-1}(F) \in \mathcal{A}\}.$$
$\mathcal{F}$ ist (die finale) $\sigma$–Algebra mit $\mathcal{F} \supset \mathcal{S} \Rightarrow$
$$\mathcal{F} \supset \sigma(\mathcal{S}) = \mathcal{A}'.$$

□

**Satz 7.6** *$(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$, $(\Omega'', \mathcal{A}'')$ messbare Räume, $f : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$,$g : (\Omega', \mathcal{A}') \to (\Omega'', \mathcal{A}'')$. Dann gilt:*
$$g \circ f : \Omega \to \Omega'' \text{ ist } \mathcal{A}, \mathcal{A}''\text{–messbar.}$$

Beweis: Sei $A'' \in \mathcal{A}''$, dann:
$$(g \circ f)^{-1}(A'') = f^{-1}(\underbrace{g^{-1}(A'')}_{\in \mathcal{A}'}) \in \mathcal{A}.$$

□

**Definition 7.7** $\mathcal{I}_n$ sei die Menge aller $n$–dimensionalen Intervalle im $\mathbb{R}^n$.
$$\mathbb{B}_n := \sigma(\mathcal{I}_n) \ \subset \mathcal{P}(\mathbb{R}^n)$$
ist das System der $n$–dimensionalen Borelmengen bzw. die Borel–$\sigma$–Algebra des $\mathbb{R}^n$.

**Bemerkung 7.8** Setze
$$\mathcal{I}_n^0 := \{(\mathbf{a}, \mathbf{b}] : \{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i, i = 1 \ldots, n\} \text{ für } \mathbf{a} = (a_1, \ldots, a_n), \mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n\}.$$
dann gilt:
$$\mathbb{B}_n = \sigma(\mathcal{I}_n^0).$$
Denn z.B.:
$$[\mathbf{a}, \mathbf{b}] = \bigcap_{m \in \mathbb{N}} \left(\mathbf{a} - \frac{1}{m}, \mathbf{b}\right]$$
mit $\mathbf{a} - 1/m = (a_1 - 1/m, \ldots, a_n - 1/m)$;
$$(\mathbf{a}, \mathbf{b}) = \bigcup_{m \in \mathbb{N}} \left(\mathbf{a}, \mathbf{b} - \frac{1}{m}\right]$$
u.s.w.
$$\begin{aligned}
&\Rightarrow \ \mathbb{B}_n \supset \sigma(\mathcal{I}_n^0) \supset \mathcal{I}_n \\
&\Rightarrow \ \mathbb{B}_n \supset \sigma(\mathcal{I}_n^0) \supset \sigma(\mathcal{I}_n) = \mathbb{B}_n \\
&\Rightarrow \ \sigma(\mathcal{I}_n^0) = \mathbb{B}_n.
\end{aligned}$$

**Satz 7.9** $\mathbb{B}_n$ *enthält alle (bezüglich der euklidischen Topologie des* $\mathbb{R}^n$*) offenen und abgeschlossenen Mengen des* $\mathbb{R}^n$*.*

Beweis: $\mathcal{I}_n^{\mathbb{Q}}$ := Menge aller $n$–dimensionalen Intervalle in $\mathcal{I}_n$ mit rationalen Endpunkten. $I_n^{\mathbb{Q}}$ ist abzählbar (denn $\mathbb{Q}^{2n}$ ist abzählbar).

$$\begin{aligned}
&G \subset \mathbb{R}^n \text{ sei offen} \\
&\quad\Rightarrow\quad G = \bigcup_{I \subset G,\, I \in \mathcal{I}_n^{\mathbb{Q}}} I \in \mathbb{B}_n \text{ (als abzählb. Vereinigung);} \\
&F \subset \mathbb{R}^n \text{ abgeschlossen} \\
&\quad\Rightarrow\quad F^c \text{ offen und damit in } \mathbb{B}_n \\
&\quad\Rightarrow\quad F \in \mathbb{B}_n.
\end{aligned}$$

□

**Satz 7.10** $\mathbb{B}_n = \underbrace{\mathbb{B} \otimes \cdots \otimes \mathbb{B}}_{n\text{–}mal} = \mathbb{B}^n$.

Beweis: $n = 2$; $\mathbb{B} \otimes \mathbb{B} = \sigma(\{B_1 \times B_2 : B_1, B_2 \in \mathbb{B}\}$.

1. $\mathbb{B}_2 \subset \mathbb{B} \otimes \mathbb{B}$, da $\mathcal{I}_2 \subset \mathbb{B} \otimes \mathbb{B}$.
2. Zu zeigen: $\mathbb{B} \otimes \mathbb{B} \subset \mathbb{B}_2$.
   Es genügt zu zeigen, dass $B_1 \times B_2 \in \mathbb{B}_2$, falls $B_1, B_2 \in \mathbb{B}$.
   (a) Setze für $I \in \mathcal{I}$
   $$\mathbb{B}_I := \{B \subset \mathbb{R} : B \times I \in \mathbb{B}_2\};$$
   $\mathbb{B}_I$ ist eine $\sigma$–Algebra(!) mit $\mathcal{I} \subset \mathbb{B}_I$, d.h. $\mathbb{B} \subset \mathbb{B}_I$.
   Also gilt $B_1 \times I \in \mathbb{B}_2$, falls $B_1 \in \mathbb{B}$, $I \in \mathcal{I}$.
   (b) Setze für $B \in \mathbb{B}$
   $$\mathbb{B}_B := \{A \subset \mathbb{R} : B \times A \in \mathbb{B}_2\};$$
   $\mathbb{B}_B$ ist eine $\sigma$–Algebra(!) mit $\mathcal{I} \subset \mathbb{B}_B$ nach (a), d.h. $\mathbb{B} \subset \mathbb{B}_B$.
   Also gilt $B_1 \times B_2 \in \mathbb{B}_2$, falls $B_1, B_2 \in \mathbb{B}$.
   Analog schließt man damit von $n$ auf $n + 1$.

□

**Definition 7.11** $(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum;

$$\begin{aligned}
f : \Omega \to \mathbb{R} \text{ ist } &\underline{\text{Zufallsvariable}} \\
:\Leftrightarrow \quad f \text{ ist } &\mathcal{A}, \mathbb{B}\text{–messbar.}
\end{aligned}$$

**Satz 7.12** $(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum, $f : \Omega \to \mathbb{R}$ beliebig. Dann sind äquivalent:*

1. *$f$ ist Zufallsvariable, d.h. $f^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathbb{B}$.*
2. *$\{f \leq y\} := \{\omega \in \Omega : f(\omega) \leq y\} \in \mathcal{A}$ für alle $y \in \mathbb{R}$.*
3. *$\{f < y\} := \{\omega \in \Omega : f(\omega) < y\} \in \mathcal{A}$ für alle $y \in \mathbb{R}$.*
4. *$\{f \in I\} := \{\omega \in \Omega : f(\omega) \in I\} \in \mathcal{A}$ für alle Intervalle $I$ in $\mathbb{R}$.*
5. *$\{f \in G\} := \{\omega \in \Omega : f(\omega) \in G\} \in \mathcal{A}$ für alle $G \in \mathcal{G} :=$ Menge der offenen Teilmengen von $\mathbb{R}$.*
6. *$\{f \in F\} := \{\omega \in \Omega : f(\omega) \in F\} \in \mathcal{A}$ für alle $F \in \mathcal{F} :=$ Menge der abgeschlossenen Teilmengen von $\mathbb{R}$.*

Beweis: Folgt aus 7.5 (s. 7.8 und 7.9), da

$$\{(-\infty, y] : y \in \mathbb{R}\}, \{(-\infty, y) : y \in \mathbb{R}\}, \mathcal{I}_1, \mathcal{G}, \mathcal{F}$$

Erzeuger der $\sigma$–Algebra $\mathbb{B}$ sind. $\square$

**Beispiel 7.13** Beispiele für Zufallsvariablen:

1. 
$$\begin{aligned} & f \in \{0,1\}^\Omega \text{ ist Zufallsvariable} \\ \Leftrightarrow \quad & \forall y \in \mathbb{R} : \{\omega \in \Omega : f(\omega) \leq y\} \in \mathcal{A} \end{aligned}$$

   wobei

$$\begin{aligned} & \{\omega \in \Omega : f(\omega) \leq y\} \\ = & \begin{cases} \Omega, & \text{falls } y \geq 1, \\ \emptyset, & \text{falls } y < 0, \\ \{\omega \in \Omega : f(\omega) = 0\}, & \text{falls } 0 \leq y < 1. \end{cases} \in \mathcal{A} \\ \Leftrightarrow \quad & f = 1_A \text{ für ein geeignetes } A \in \mathcal{A}. \end{aligned}$$

   Dabei ist

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

   die Indikatorfunktion der Menge $A$.

2. Falls $\mathcal{A} = \{\emptyset, \Omega\}$, so sind nur konstante Funktionen Zufallsvariablen.
3. Falls $\mathcal{A} = \mathcal{P}(\Omega)$, so sind alle $f : \Omega \to \mathbb{R}$ Zufallsvariablen.

**Definition 7.14** $(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum; dann:

$$\begin{aligned} & f : \Omega \to \mathbb{R}^n \text{ ist } n\text{–dimensionaler Zufallsvektor} \\ :\Leftrightarrow \quad & f \text{ ist } \mathcal{A}, \mathbb{B}_n\text{–messbar.} \end{aligned}$$

**Satz 7.15**

$$f = (f_1, \dots, f_n) : \Omega \to \mathbb{R}^n \text{ ist } n\text{–dim. Zufallsvektor}$$
$$\Leftrightarrow \quad f_i : \Omega \to \mathbb{R} \text{ ist Zufallsvariable, } i = 1, \dots, n.$$

Beweis:

„$\Rightarrow$“ Sei $B \in \mathbb{B}$, dann:

$$f_i^{-1}(B) = f^{-1}(\underbrace{\mathbb{R} \times \cdots \times \overbrace{B}^{i\text{–te Stelle}} \times \cdots \times \mathbb{R}}_{\in \mathbb{B}_n}) \in \mathcal{A}.$$

„$\Leftarrow$“ 1.

$$f^{-1}(B_1 \times \cdots \times B_n) = \bigcap_{1 \le i \le n} \underbrace{f_i^{-1}(B_i)}_{\in \mathcal{A}} \in \mathcal{A},$$

falls $B_i \in \mathbb{B}$, $i = 1, \dots, n$.

2. $\{B \subset \mathbb{R}^n : f^{-1}(B) \in \mathcal{A}\}$ ist (finale) $\sigma$–Algebra (7.1).

Aus 1. und 2. folgt, dass

$$\{B \subset \mathbb{R}^n : f^{-1}(B) \in \mathcal{A}\} \supset \mathbb{B} \otimes \cdots \otimes \mathbb{B} =_{7.10} \mathbb{B}_n,$$

d.h. $f$ ist Zufallsvektor.

□

**Definition 7.16**

$$g : \mathbb{R}^n \to \mathbb{R} \text{ ist } \underline{\text{Bairesche Funktion}}$$
$$:\Leftrightarrow \quad g \text{ ist } \mathbb{B}_n, \mathbb{B}\text{–messbar.}$$

**Satz 7.17** *$f_i : \Omega \to \mathbb{R}$ sei Zufallsvariable, $i = 1, \dots, n$, $g : \mathbb{R}^n \to \mathbb{R}$ Bairesche Funktion. Dann ist $g \circ (f_1, \dots, f_n)$ Zufallsvariable.*

Beweis: Folgt aus 7.15 und 7.6. □

**Satz 7.18** *$g : \mathbb{R}^n \to \mathbb{R}$ stetig $\Rightarrow$ $g$ ist Bairesche Funktion.*

Beweis: $g$ stetig $\Rightarrow$ $\forall$ offenen Teilmengen $O$ von $\mathbb{R}$: $g^{-1}(O)$ ist offen im $\mathbb{R}^n$, d.h. $g^{-1}(O) \in \mathbb{B}_n$. Nach Satz 7.5 gilt damit $g^{-1}(B) \in \mathbb{B}_n$ für alle $B \in \mathbb{B}$, da die offenen Mengen $\mathbb{B}$ erzeugen (7.9). □

**Satz 7.19** *$f_1, f_2$ seien Zufallsvariablen, $a \in \mathbb{R}$. Dann gilt:*

$$\begin{aligned} &af_1, \\ &f_1 + f_2, \\ &f_1 f_2, \\ &f_1/f_2 \quad \textit{(falls } f_2(\omega) \neq 0 \textit{ für alle } \omega \in \Omega\textit{)}, \\ &f_1 \vee f_2 := \max(f_1, f_2), \\ &f_1 \wedge f_2 := \min(f_1, f_2) \end{aligned}$$

*sind wieder Zufallsvariablen.*

Beweis: $\mathbb{R} \ni x \mapsto ax$, $\mathbb{R}^2 \ni (x_1, x_2) \mapsto x_1 + x_2 \in \mathbb{R}$ u.s.w. sind stetige Abbildungen von $\mathbb{R} \to \mathbb{R}$ bzw. $\mathbb{R} \to \mathbb{R}^2$, d.h. Satz 7.17 und 7.18 anwendbar.

□

**Bemerkung** Der Raum der Zufallsvariablen über $(\Omega, \mathcal{A}, P)$ ist ein linearer Raum.

**Satz 7.20** *$f_1, f_2, \ldots$ seien Zufallsvariablen.*

1. *$f_n(\omega)$, $n \in \mathbb{N}$, sei für jedes $\omega \in \Omega$ nach oben beschränkt*

$$\Rightarrow \sup_{n \in \mathbb{N}} f_n \textit{ ist Zufallsvariable.}$$

*$((\sup_{n \in \mathbb{N}} f_n)(\omega) = \sup_{n \in \mathbb{N}}(f_n(\omega)))$.*

2. *$f_n(\omega)$, $n \in \mathbb{N}$, sei für jedes $\omega \in \Omega$ nach unten beschränkt*

$$\Rightarrow \inf_{n \in \mathbb{N}} f_n \textit{ ist Zufallsvariable.}$$

3. *$f_n(\omega)$, $n \in \mathbb{N}$, sei für jedes $\omega \in \Omega$ beschränkt*

$$\Rightarrow \liminf_{n \in \mathbb{N}} f_n, \limsup_{n \in \mathbb{N}} f_n \textit{ sind Zufallsvariablen.}$$

4. *$f_n(\omega)$, $n \in \mathbb{N}$, sei für jedes $\omega \in \Omega$ konvergent*

$$\Rightarrow \lim_{n \in \mathbb{N}} f_n \textit{ ist Zufallsvariable.}$$

Beweis:

1. $\forall y \in \mathbb{R} : \{\omega \in \Omega : \sup_{n \in \mathbb{N}} f_n(\omega) \leq y\} = \bigcap_{n \in \mathbb{N}} \{\omega \in \Omega : f_n(\omega) \leq y\} \in \mathcal{A}$.
2. $\{\omega \in \Omega : \inf_{n \in \mathbb{N}} f_n(\omega) < y\} = \bigcup_{n \in \mathbb{N}} \{\omega \in \Omega : f_n(\omega) < y\} \in \mathcal{A}$.
3. $\limsup_{n \in \mathbb{N}} f_n(\omega) = \inf_{m \in \mathbb{N}} \left(\sup_{n \geq m} f_n(\omega)\right)$, $\liminf_{n \in \mathbb{N}} f_n(\omega) = \sup_{m \in \mathbb{N}} (\inf_{n \geq m} f_n(\omega))$, d.h. die Behauptung folgt aus 1. und 2.

4. $\lim_{n\in\mathbb{N}} f_n = \limsup_{n\in\mathbb{N}} f_n = \liminf_{n\in\mathbb{N}} f_n$, d.h. 3. anwendbar.

□

**Definition 7.21** $f_i, i \in I$, sei eine Familie zufälliger Größen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, wobei $f_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$, $i \in I$.
Die Familie $f_i$, $i \in I$, heißt unabhängig :⇔ die Familie $\mathcal{A}(f_i) = f_i^{-1}(\mathcal{A}_i)$, $i \in I$, der zugehörigen Vergröberungen ist unabhängig.

**Bemerkung 7.22** $f_i$, $i \in I$, unabhängig ⇔ $f_i$, $i \in I_0$, unabhängig für alle endlichen Teilmengen $I_0$ von $I$ (s. 6.18).

**Satz 7.23**

$$\begin{aligned} & f_i : (\Omega, \mathcal{A}) \to (\Omega_i', \mathcal{A}_i'),\ i = 1, \dots, n \ \textit{unabhängig} \\ \Leftrightarrow\ & P(\{f_1 \in A_1', \dots, f_n \in A_n'\} \\ & \qquad = P(\{f_1 \in A_1'\}) \cdots P(\{f_n \in A_n'\}) \\ & \qquad\qquad \textit{für beliebige } A_i' \in \mathcal{A}_i',\ i = 1, \dots, n. \end{aligned}$$

Beweis:

$$\begin{aligned} & f_1, \dots, f_n \text{ unabhängig} \\ \Leftrightarrow\ & \mathcal{A}(f_1), \dots, \mathcal{A}(f_n) \text{ unabhängig} \\ \Leftrightarrow_{6.18}\ & \text{für beliebige } A_i' \in \mathcal{A}_i',\ i = 1, \dots, n, \text{ sind die} \\ & \quad \text{Ereignisse } \{f_1 \in A_1'\}, \dots, \{f_n \in A_n'\} \\ & \quad \text{unabhängig.} \end{aligned}$$

□

**Satz 7.24** *$f_i : (\Omega, \mathcal{A}) \to (\Omega_i', \mathcal{A}_i')$, $i \in I$, unabhängig, $g_i : (\Omega_i', \mathcal{A}_i') \to (\Omega_i'', \mathcal{A}_i'')$, $i \in I$.*
*Dann sind $g_i \circ f_i : (\Omega, \mathcal{A}) \to (\Omega_i'', \mathcal{A}_i'')$, $i \in I$, unabhängig.*

Beweis:

1. Nach Satz 7.6 ist $g_i \circ f_i$ $\mathcal{A}, \mathcal{A}_i$–messbar, $i \in I$.
2. Für endliches $I_0 \subset I$ und $A_i'' \in \mathcal{A}_i''$, $i \in I_0$, gilt:

$$\begin{aligned} & P\{g_i \circ f_i \in A_i'', i \in I_0\} \\ =\ & P\{f_i \in \underbrace{g_i^{-1}(A_i'')}_{\in \mathcal{A}_i'}, i \in I_0\} \\ =\ & \prod_{i\in I_0} P\{f_i \in g_i^{-1}(A_i'')\} \quad (\text{da } f_i \text{ unabhängig}) \\ =\ & \prod_{i\in I_0} P\{g_i \circ f_i \in A_i''\}. \end{aligned}$$

□

**Definition 7.25** $f_1, f_2, \ldots$ seien Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$. $C \in \mathcal{A}$ heißt terminales Ereignis bzgl. $f_1, f_2, \ldots :\Leftrightarrow$

$$C \in \sigma\left(\bigcup_{m \geq n} \mathcal{A}(f_m)\right) \text{ für alle } n \in \mathbb{N}.$$

$\bigcap_{n\in\mathbb{N}} \sigma\left(\bigcup_{m\geq n} \mathcal{A}(f_m)\right)$ heißt die zu $f_1, f_2, \ldots$ gehörende terminale $\sigma$–Algebra.

**Beispiel 7.26**

1. $\{\omega \in \Omega : f_n(\omega) > 0 \text{ unendlich oft}\}$,
2. $\{\omega \in \Omega : f_n(\omega), n \in \mathbb{N}, \text{ ist konvergent}\}$,
3. $\left\{\omega \in \Omega : \frac{1}{n}\sum_{i=1}^n f_n(\omega), n \in \mathbb{N}, \text{ ist konvergent}\right\}$

sind terminale Ereignisse.

**Satz 7.27 (Kolmogoroffsches 0–1–Gesetz)** *$C$ sei terminales Ereignis zur Folge unabhängiger Zufallsvariablen $f_1, f_2, \ldots \Rightarrow P(C) = 0$ oder $1$;*

**Bemerkung 7.28** Vergleiche Borel–Cantelli Lemma: $A_1, A_2, \ldots$ unabhängige Ereignisse $\Rightarrow 1_{A_1}, 1_{A_2}, \ldots$ unabhängige Zufallsvariablen;

$$\limsup_{n\in\mathbb{N}} A_n = \left\{\omega \in \Omega : \sum_{n\in\mathbb{N}} 1_{A_n}(\omega) = \infty\right\}$$

ist terminales Ereignis zu $1_{A_1}, 1_{A_2}, \ldots$ mit

$$P\left(\limsup_{n\in\mathbb{N}} A_n\right) = \begin{cases} 0, & \text{falls } \sum_{n\in\mathbb{N}} P(A_n) < \infty, \\ 1, & \text{falls } \sum_{n\in\mathbb{N}} P(A_n) = \infty. \end{cases}$$

Beweis: Annahme: $P(C) > 0$. Wir zeigen: $P(C) = 1$.
Für $A \in \mathcal{A}(f_n, n \in \mathbb{N}) := \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{A}(f_n)\right)$ definieren wir

$$P^*(A) := P(A|C) = \frac{P(A \cap C)}{P(C)}$$

$\Rightarrow P^*$ ist Wahrscheinlichkeitsmaß mit

$$P^*(A) = P(A) \text{ für alle } A \in \mathcal{A}_0 := \bigcup_{n\in\mathbb{N}} \mathcal{A}(f_m : m \leq n),$$

mit

$$\mathcal{A}(f_m : m \leq n) := \sigma\left(\bigcup_{m \leq n} \mathcal{A}(f_m)\right),$$

denn für $A \in \mathcal{A}_0$ gilt: $A, C$ sind unabhängig.
$\mathcal{A}_0$ ist eine Algebra(!) mit $\mathcal{A}(f_n, n \in \mathbb{N}) = \sigma(\mathcal{A}_0)$.
Fortsetzungssatz 6.16 $\Rightarrow P^* = P$ auf $\mathcal{A}(f_n, n \in \mathbb{N})$, d.h.

$$\begin{aligned}
P^*(A) = \frac{P(A \cap C)}{P(C)} &= P(A) \\
&\quad \text{für alle } A \in \mathcal{A}(f_n, n \in \mathbb{N}) \\
\Rightarrow \quad P(A \cap C) &= P(A)P(C) \\
&\quad \text{insbesondere für } A = C \quad (C \in \mathcal{A}(f_n, n \in \mathbb{N})) \\
\Rightarrow \quad P(C) &= P(C)^2 \\
\Rightarrow \quad P(C) &= 1.
\end{aligned}$$

□

# 8 Integrationstheorie

Es sei $f$ eine Zufallsvariable, die nur die Werte $a_1, \ldots, a_m$ annimmt, $p_i := P\{f = a_i\}$, $i = 1, \ldots, m$.
Bei $N$–facher unabhängiger Wiederholung des Experimentes seien $\alpha_1, \ldots, \alpha_N$ $(\in \{a_1, \ldots, a_m\})$ die beobachteten Werte von $f$. Die Erfahrung zeigt, dass sich das arithmetische Mittel

$$\frac{1}{N} \sum_{i=1}^{N} \alpha_i$$

offenbar gegen einen gewissen Wert stabilisiert. Was ist das für eine Zahl?
Dazu anschaulich: $H_i$ (Häufigkeit) bezeichne die Anzahl des Eintretens von $a_i$ unter den $N$ Durchführungen des Experimentes, $i = 1, \ldots, m$. Nach der Erfahrung wird gelten:

$$\frac{H_i}{N} \approx p_i, \text{ d.h. } H_i \approx p_i N.$$

Damit:

$$\begin{aligned}
\frac{\alpha_1 + \cdots + \alpha_N}{N} &= \frac{H_1 a_1 + \cdots + H_m a_m}{N} \\
&= \sum_{i=1}^{m} \frac{H_i}{N} a_i \\
&\approx \sum_{i=1}^{m} a_i p_i.
\end{aligned}$$

Dies wird der zu erwartende, d.h. der Erwartungswert des arithmetischen Mittels sein. Den Begriff des Erwartungswertes werden wir in diesem Kapitel untersuchen.

**Definition 8.1**

1. $e$ <u>einfache Funktion</u> $:\Leftrightarrow$
$$e = \sum_{i=1}^{m} \alpha_i 1_{A_i}$$
mit geeignetem $m \in \mathbb{N}$, $\alpha_i \geq 0$, $A_i \in \mathcal{A}$, $i = 1, \dots, m$.

2. $\mathcal{E} :=$ Menge der einfachen Funktionen.

**Bemerkung 8.2**

1. $\mathcal{E}$ = Menge aller nicht–negativen Zufallsvariablen über $(\Omega, \mathcal{A}, P)$, die nur endlich viele Werte annehmen.

2. $e_1, e_2 \in \mathcal{E}, \alpha \in \mathbb{R}_+$
$\Rightarrow \alpha e_1, e_1 + e_2, e_1 e_2, e_1 \vee e_2, e_1 \wedge e_2 \in \mathcal{E}$.

3. $e \in \mathcal{E} \Rightarrow e = \sum_{1 \leq i \leq m} \alpha_i 1_{A_i}$
mit disjunkten $A_i \in \mathcal{A}$ und $\bigcup_{1 \leq i \leq m} A_i = \Omega$.

**Satz 8.3** *Es gelte*
$$e = \sum_{1 \leq i \leq m} \alpha_i 1_{A_i} = \sum_{1 \leq j \leq n} \beta_j 1_{B_j} \in \mathcal{E}$$
*mit disjunkten* $A_i \in \mathcal{A}$, $i = 1, \dots, m$, *und disjunkten* $B_j \in \mathcal{A}$, $j = 1, \dots, n$. *Dann folgt:*
$$\sum_{1 \leq i \leq m} \alpha_i P(A_i) = \sum_{1 \leq j \leq n} \beta_j P(B_j).$$

Beweis: O.B.d.A. annehmbar, dass $\bigcup_{1 \leq i \leq m} A_i = \bigcup_{1 \leq j \leq m} B_j = \Omega$.
$$1_{A_i} = \sum_{1 \leq j \leq n} 1_{A_i \cap B_j};\ 1_{B_j} = \sum_{1 \leq i \leq n} 1_{A_i \cap B_j}$$
$$\begin{aligned}
\Rightarrow e &= \sum_{1 \leq i \leq m} \alpha_i 1_{A_i} \\
&= \sum_{1 \leq i \leq m; 1 \leq j \leq n} \alpha_i 1_{A_i \cap B_j}; \\
e &= \sum_{1 \leq j \leq n; 1 \leq i \leq m} \beta_j 1_{B_j \cap A_i}
\end{aligned}$$
$$\Rightarrow \forall i, j \text{ mit } A_i \cap B_j \neq \emptyset : \alpha_i = \beta_j$$
$$\begin{aligned}
\Rightarrow \sum_{1 \leq i \leq m} \alpha_i P(A_i) & \\
&= \sum_{1 \leq i \leq m; 1 \leq j \leq n} \alpha_i P(A_i \cap B_j) \\
&= \sum_{1 \leq i \leq m; 1 \leq j \leq n} \beta_j P(B_j \cap A_i) \\
&= \sum_{1 \leq j \leq n} \beta_j P(B_j).
\end{aligned}$$

$\square$

**Definition 8.4** Für $e = \sum_{1\le i\le m} \alpha_i 1_{A_i} \in \mathcal{E}$ mit disjunkten $A_i$ und $\alpha_i \ge 0$ definieren wir

$$\int e\,dP := \int_\Omega e\,dP := \sum_{1\le i\le m} \alpha_i P(A_i),$$

das <u>Integral</u> von $e$ über $\Omega$.

**Satz 8.5** *Für $e, e_1, e_2 \in \mathcal{E}$ gilt:*

*1.*
$$\int \alpha e\,dP = \alpha \int e\,dP, \quad \alpha \ge 0,$$

*2.*
$$\int e_1 + e_2\,dP = \int e_1\,dP + \int e_2\,dP,$$

*3.* $e = \sum_{1\le i\le m} \alpha_i 1_{A_i}$ *mit* $\alpha_i \ge 0$ *und beliebigen* $A_i$*, d.h.* <u>*nicht notwendig disjunkten*</u> $A_i \Rightarrow$
$$\int e\,dP = \sum_{1\le i\le m} \alpha_i P(A_i),$$

*4.*
$$e_1 \le e_2 \Rightarrow \int e_1\,dP \le \int e_2\,dP.$$

Beweis:

**„1.“** $e = \sum_{1\le i\le m} \alpha_i 1_{A_i}$ mit disjunkten $A_i \Rightarrow \alpha e = \sum_{1\le i\le m} (\alpha\alpha_i) 1_{A_i} \Rightarrow$ Behauptung.

**„2.“** $e_1 = \sum_i \alpha_i 1_{A_i}$ mit disjunkten $A_i$, $e_2 = \sum_j \beta_j 1_{B_j}$ mit disjunkten $B_j$. O.b.d.A. gelte $\bigcup_i A_i = \bigcup_j B_j = \Omega$;

$$\begin{aligned}
e_1 &= \sum_{i,j} \alpha_i 1_{A_i\cap B_j},\\
e_2 &= \sum_{i,j} \beta_j 1_{A_i\cap B_j}\\
\Rightarrow e_1 + e_2 &= \sum_{i,j} (\alpha_i + \beta_j) 1_{A_i\cap B_j}\\
&\qquad \text{mit } A_i \cap B_j, (i,j), \text{ disjunkt}
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \int e_1 + e_2\,dP\\
&=_{\text{Def.}} \sum_{i,j} (\alpha_i + \beta_j) P(A_i \cap B_j)\\
&= \sum_{i,j} \alpha_i P(A_i \cap B_j) + \sum_{i,j} \beta_j P(A_i \cap B_j)\\
&= \int e_1\,dP + \int e_2\,dP.
\end{aligned}$$

**„3.“** Folgt aus 1. und 2.:

$$\int \sum_i \alpha_i 1_{A_i}\, dP = \sum_i \alpha_i \int 1_{A_i}\, dP.$$

**„4.“** Nach Beweisteil 2. besitzen $e_1$ und $e_2$ die Darstellungen

$$e_1 = \sum_{1\le i\le m} \alpha_i 1_{A_i},\ e_2 = \sum_{1\le i\le m} \beta_i 1_{A_i}$$

mit disjunkten $A_i \neq \emptyset$

$$\begin{aligned} \Rightarrow \alpha_i \le \beta_i, \qquad & 1 \le i \le m, \\ \Rightarrow \int e_1\, dP \quad &= \quad \sum_i \alpha_i P(A_i) \\ &\le \quad \sum_i \beta_i P(A_i) \\ &= \quad \int e_2\, dP. \end{aligned}$$

□

**Satz 8.6** *Zu jeder Zufallsvariablen $f \ge 0$ existieren $e_n \in \mathcal{E}$, $n \in \mathbb{N}$, mit*

$$e_1 \le e_2 \le \dots \ \textit{und}\ f = \lim_{n\in\mathbb{N}} e_n.$$

Beweis: Setze

$$e_n := \sum_{0\le i<n2^n} \frac{i}{2^n} 1_{A_{n,i}},$$

mit $A_{n,i} := \{i/2^n \le f < (i+1)/2^n\}$. □

**Satz 8.7** *$e, e_n \in \mathcal{E}$, $n \in \mathbb{N}$, mit $e \le \lim_{n\in\mathbb{N}} e_n$, $e_1 \le e_2 \le \dots$ Dann folgt:*

$$\int e\, dP \le \lim_{n\in\mathbb{N}} \int e_n\, dP.$$

Beweis: Sei $e = \sum_{1\le i\le m} \alpha_i 1_{A_i}$.
Wähle $a \in [0,1)$ und setze $K_n := \{ae \le e_n\}$, $n \in \mathbb{N}$.
$K_n \in \mathcal{A}$, da $ae - e_n$ $\mathcal{A}, \mathbb{B}$–messbar,
$e 1_{K_n} = \sum_{1\le i\le m} \alpha_i 1_{A_i \cap K_n} \in \mathcal{E}$,
$K_n \uparrow \Omega$, also $A_i \cap K_n \uparrow A_i$ und damit

$\lim_{n\in\mathbb{N}} P(A_i \cap K_n) = P(A_i)$, $i = 1, \dots, m$, $\Rightarrow$

$$\begin{aligned}
a \int e \, dP &= a \sum_{1\le i\le m} \alpha_i P(A_i) \\
&= a \lim_{n\in\mathbb{N}} \sum_{1\le i\le m} \alpha_i P(A_i \cap K_n) \\
&= \lim_{n\in\mathbb{N}} \int \underbrace{ae1_{K_n}}_{\le e_n} \, dP \\
&\le \lim_{n\in\mathbb{N}} \int e_n \, dP.
\end{aligned}$$

Für $a \uparrow 1$ folgt nun die Behauptung. □

**Korollar 8.8** *$e_1 \le e_2 \le \dots$, $e'_1 \le e'_2 \le \dots \in \mathcal{E}$ mit $\lim_{n\in\mathbb{N}} e_n = \lim_{n\in\mathbb{N}} e'_n$ $(\le \infty)$. Dann gilt:*

$$\lim_{n\in\mathbb{N}} \int e_n \, dP = \lim_{n\in\mathbb{N}} \int e'_n \, dP.$$

**Definition 8.9** *Für eine Zufallsvariable $f \ge 0$ setzen wir*

$$\int f \, dP := \lim_{n\in\mathbb{N}} \int e_n \, dP$$

*(Integral von $f$ über $(\Omega, \mathcal{A}, P)$), falls*

$$f = \lim_{n\in\mathbb{N}} e_n, \quad e_1 \le e_2 \le \dots \in \mathcal{E}.$$

**Satz 8.10** *Für nicht negative Zufallsvariablen $f, f_1, f_2$ gilt:*

1. $$\int \alpha f \, dP = \alpha \int f \, dP, \quad \alpha \ge 0,$$

2. $$\int f_1 + f_2 \, dP = \int f_1 \, dP + \int f_2 \, dP,$$

3. $$f_1 \le f_2 \Rightarrow \int f_1 \, dP \le \int f_2 \, dP.$$

Beweis: Nach Satz 8.6 existieren Folgen einfacher Funktionen $e_1 \le e_2 \le \dots$, $e'_1 \le e'_2 \le \dots$, $\tilde{e}_1 \le \tilde{e}_2 \dots$ mit $\lim_{n\in\mathbb{N}} e_n = f$, $\lim_{n\in\mathbb{N}} e'_n = f_1$, $\lim_{n\in\mathbb{N}} \tilde{e}_n = f_2$.

**„1.“** $\alpha e_n \in \mathcal{E}$, $n \in \mathbb{N}$, $\alpha e_1 \leq \alpha e_2 \leq \dots$, $\lim_{n\in\mathbb{N}} \alpha e_n = \alpha f$. Damit:

$$\begin{aligned}
\int \alpha f \, dP &=_{\text{Def.}} \lim_{n\in\mathbb{N}} \int \alpha e_n \, dP \\
&=_{8.5} \lim_{n\in\mathbb{N}} \alpha \int e_n \, dP \\
&= \alpha \lim_{n\in\mathbb{N}} \int e_n \, dP \\
&=_{\text{Def.}} \alpha \int f \, dP.
\end{aligned}$$

**„2.“** $e'_n + \tilde{e}_n \in \mathcal{E}, n \in \mathbb{N}$, $e'_1 + \tilde{e}_1 \leq e'_2 + \tilde{e}_2 \leq \cdots$, $\lim_{n\in\mathbb{N}}(e'_n + \tilde{e}_n) = f_1 + f_2$. Damit:

$$\begin{aligned}
\int f_1 + f_2 \, dP &=_{\text{Def.}} \lim_{n\in\mathbb{N}} \int e'_n + \tilde{e}_n \, dP \\
&=_{8.5} \lim_{n\in\mathbb{N}} \left( \int e'_n \, dP + \int \tilde{e}_n \, dP \right) \\
&= \int f_1 \, dP + \int f_2 \, dP.
\end{aligned}$$

**„3.“**

$$\begin{aligned}
e'_n &\leq \lim_{n\in\mathbb{N}} \tilde{e}_n \quad (= f_2) \\
\Rightarrow_{8.7} \int e'_n \, dP &\leq \lim_{n\in\mathbb{N}} \int \tilde{e}_n \, dP \\
&=_{\text{Def.}} \int f_2 \, dP \\
\Rightarrow_{\text{Def.}} \int f_1 \, dP &= \lim_{n\in\mathbb{N}} \int e'_n \, dP \\
&\leq \int f_2 \, dP.
\end{aligned}$$

□

**Satz 8.11** *$f, f_1, f_2, \dots$ seien nicht negative Zufallsvariablen mit $f_1 \leq f_2 \leq \cdots$ und $\lim_{n\in\mathbb{N}} f_n = f$. Dann gilt:*

$$\int f \, dP = \lim_{n\in\mathbb{N}} \int f_n \, dP.$$

Beweis: Zu $f_n$ existiert eine monoton wachsende Folge $e_{n,m}$, $m \in \mathbb{N}$, in $\mathcal{E}$ mit

$$\lim_{m\in\mathbb{N}} e_{n,m} = f_n,$$

also:

$$\begin{array}{ccccccc}
e_{1,1} & \le & e_{1,2} & \le & \cdots & \uparrow f_1 & \\
e_{2,1} & \le & e_{2,2} & \le & \cdots & \uparrow f_2 & \\
\vdots & & \vdots & & & \vdots & \\
e_{n,1} & \le & e_{n,2} & \le & \cdots & \uparrow f_n & \\
\vdots & & \vdots & & & & \vdots \\
 & & & & & & \uparrow f.
\end{array}$$

Setze $e_n := \max(e_{1,n}, \ldots, e_{n,n})$. Dann gilt:

1. $$e_n \in \mathcal{E} \quad (8.2,3.),\ e_1 \le e_2 \le \cdots$$

2. $$e_n \le \max(f_1, \ldots, f_n) = f_n \le f.$$

3. $$\begin{array}{rcll}
\forall m \le n : e_n & \ge & e_{m,n} & \\
\Rightarrow \lim\limits_{n\in\mathbb{N}} e_n & \ge & \lim\limits_{n\in\mathbb{N}} e_{m,n} & = f_m \\
\Rightarrow \lim\limits_{n\in\mathbb{N}} e_n & \ge & f & \\
\Rightarrow_{2.} \lim\limits_{n\in\mathbb{N}} e_n & = & f & \\
\Rightarrow \displaystyle\int f\,dP & =_{\text{Def.}} & \lim\limits_{n\in\mathbb{N}} \displaystyle\int e_n\,dP & \\
 & \le_{2.} & \lim\limits_{n\in\mathbb{N}} \displaystyle\int f_n\,dP; & \\
\displaystyle\int f\,dP & \ge_{8.10} & \displaystyle\int f_n\,dP & \\
\Rightarrow \displaystyle\int f\,dP & = & \lim\limits_{n\in\mathbb{N}} \displaystyle\int f_n\,dP. &
\end{array}$$

□

**Definition 8.12** $f$ sei beliebige Zufallsvariable.

1. $f^+ := \max(f, 0)$ ist der Positivteil von $f$,
   $f^- := \max(-f, 0)$ ist der Negativteil von $f$; $f = f^+ - f^-$.
2. $f$ ist integrierbar $:\Leftrightarrow \int f^+\,dP < \infty$ und $\int f^-\,dP < \infty$.
3. $\mathcal{L} :=$ Menge aller integrierbaren Funktionen (auf $(\Omega, \mathcal{A}, P)$).
4. $f$ ist quasiintegrierbar $:\Leftrightarrow \int f^+\,dP < \infty$ oder $\int f^-\,dP < \infty$.
5. $\int f\,dP := \int f^+\,dP - \int f^-\,dP$, falls $f$ quasiintegrierbar ist.

**Satz 8.13** *$f, f_1, f_2$ beliebige Zufallsvariablen. Dann:*

1. *$f \in \mathcal{L}$, $\alpha \in \mathbb{R} \Rightarrow \alpha f \in \mathcal{L}$ und $\int \alpha f \, dP = \alpha \int f \, dP$.*

2. *$f_1, f_2 \in \mathcal{L} \Rightarrow f_1 + f_2 \in \mathcal{L}$ und $\int f_1 + f_2 \, dP = \int f_1 \, dP + \int f_2 \, dP$.*

3. *$f_1 \leq f_2$, $f_1, f_2 \in \mathcal{L} \Rightarrow \int f_1 \, dP \leq \int f_2 \, dP$.*

4. *$f \in \mathcal{L} \Leftrightarrow |f| \in \mathcal{L}$ und es gilt in diesem Fall $|\int f \, dP| \leq \int |f| \, dP$.*

5. *$g$ $\mathcal{A}, \mathbb{B}$–messbar mit $f_1 \leq g \leq f_2$, $f_1, f_2 \in \mathcal{L} \Rightarrow g \in \mathcal{L}$.*

Beweis: Mittels 8.10. □

**Satz 8.14 (v. d. monotonen Konvergenz)**

1. *$f_n \in \mathcal{L}$, $n \in \mathbb{N}$, $f_n \uparrow f < \infty \Rightarrow f$ ist quasiintegrierbar und $\int f_n \, dP \uparrow \int f \, dP$.*

2. *$f_n \in \mathcal{L}$, $n \in \mathbb{N}$, $f_n \downarrow f > -\infty \Rightarrow f$ ist quasiintegrierbar und $\int f_n \, dP \downarrow \int f \, dP$.*

Beweis: Genügt 1. zu beweisen (Übergang zu $-f_n, -f$). O.E. sei $f_n \geq 0$ (sonst Übergang zu $f'_n := f_n - f_1$). Dann folgt die Behauptung aber aus 8.11. □

**Satz 8.15 (Lemma von Fatou)**

1. *$f_n \in \mathcal{L}$, $n \in \mathbb{N}$, $f_n \leq h$, $h \in \mathcal{L}$, $\limsup_{n\in\mathbb{N}} f_n > -\infty \Rightarrow \limsup_{n\in\mathbb{N}} f_n$ ist quasiintegrierbar und*

$$\int \limsup_{n\in\mathbb{N}} f_n \, dP \geq \limsup_{n\in\mathbb{N}} \int f_n \, dP.$$

2. *$f_n \in \mathcal{L}$, $n \in \mathbb{N}$, $f_n \geq g$, $g \in \mathcal{L}$, $\liminf_{n\in\mathbb{N}} f_n < \infty \Rightarrow \liminf_{n\in\mathbb{N}} f_n$ ist quasiintegrierbar und*

$$\int \liminf_{n\in\mathbb{N}} f_n \, dP \leq \liminf_{n\in\mathbb{N}} \int f_n \, dP.$$

Beweis: Genügt 1. zu beweisen (Übergang zu $-f_n$). Es gilt:

$$\limsup_{n\in\mathbb{N}} f_n = \lim_{n\in\mathbb{N}} \left( \sup_{m\geq n} f_m \right).$$

$\sup_{m\geq n} f_m \in \mathcal{L}$, da $f_n \leq \sup_{m\geq n} f_m \leq h$ (8.13);

$\sup_{m\geq n} f_m \downarrow \limsup_{n\in\mathbb{N}} f_n \Rightarrow_{8.14} \limsup_{n\in\mathbb{N}} f_n$ ist quasiintegrierbar und

$$\begin{aligned}
\int \sup_{m\geq n} f_m \, dP \quad &\downarrow \quad \int \limsup_{n\in\mathbb{N}} f_n \, dP \\
\Rightarrow \limsup_{n\in\mathbb{N}} \int f_n \, dP \quad &= \quad \lim_{n\in\mathbb{N}} \left( \sup_{m\geq n} \int f_m \, dP \right) \\
&\leq \quad \lim_{n\in\mathbb{N}} \left( \int \sup_{m\geq n} f_m \, dP \right) \\
&= \quad \int \limsup_{n\in\mathbb{N}} f_n \, dP.
\end{aligned}$$

□

**Satz 8.16 (v. d. dominierten Konvergenz)** *$f_n$, $n \in \mathbb{N}$, Zufallsvariablen mit $\lim_{n\in\mathbb{N}} f_n = f$, $|f_n| \leq g$, $g \in \mathcal{L}$. Dann gilt:*

$$f_n, f \in \mathcal{L} \text{ und } \lim_{n\in\mathbb{N}} \int f_n \, dP = \int f \, dP.$$

Beweis: 8.13$\Rightarrow$ $f_n, f \in \mathcal{L}$. Das Lemma von Fatou liefert:

$$\begin{aligned}
\limsup_{n\in\mathbb{N}} \int f_n \, dP \quad &\leq \quad \int \limsup_{n\in\mathbb{N}} f_n \, dP \\
&= \quad \int f \, dP \\
&= \quad \int \liminf_{n\in\mathbb{N}} f_n \, dP \\
&\leq \quad \liminf_{n\in\mathbb{N}} \int f_n \, dP.
\end{aligned}$$

□

**Satz 8.17 (Transformationssatz für Integrale)** *$(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum, $(\Omega', \mathcal{A}')$ messbarer Raum, $T : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$.*
*$P' := P * T$ sei das durch $P$ und $T$ auf $\mathcal{A}'$ induzierte Wahrscheinlichkeitsmaß (Bildmaß), d.h.*

$$P'(A') = P(T^{-1}(A')), \quad A' \in \mathcal{A}',$$

*s. Satz 7.1.*
*$f' : \Omega' \to \mathbb{R}$ sei Zufallsvariable. Dann:*

$$f' \in \mathcal{L}(\Omega', \mathcal{A}', P') \Leftrightarrow f' \circ T \in \mathcal{L}(\Omega, \mathcal{A}, P)$$

*und in diesem Fall gilt:*

$$\int_{\Omega'} f' \, dP' = \int_{\Omega} f' \circ T \, dP.$$

Beweis:

1. Sei $e' \in \mathcal{E}(\Omega', \mathcal{A}')$, d.h. $e' = \sum_{1 \le i \le m} \alpha_i 1_{A_i'}$ mit $\alpha_i \ge 0$, $A_i' \in \mathcal{A}'$,

$$\begin{aligned} \Rightarrow e := e' \circ T &= \sum_{1 \le i \le m} \alpha_i \left(1_{A_i'} \circ T\right) \\ &= \sum_{1 \le i \le m} \alpha_i 1_{A_i} \in \mathcal{E}(\Omega, \mathcal{A}) \end{aligned}$$

mit $A_i := T^{-1}(A_i')$ und

$$\begin{aligned} \int_{\Omega'} e' \, dP' &= \sum_{1 \le i \le m} \alpha_i P'(A_i') \\ &= \sum_{1 \le i \le m} \alpha_i P(T^{-1}(A_i')) \\ &= \sum_{1 \le i \le m} \alpha_i P(A_i) \\ &= \int_{\Omega} e \, dP. \end{aligned}$$

2. Sei $f' : \Omega' \to \mathbb{R}_+$ Zufallsvariable. Dann existieren $e_n' \in \mathcal{E}(\Omega', \mathcal{A}')$ mit $e_n' \uparrow f' \Rightarrow e_n := e_n' \circ T \uparrow f' \circ T$, $e_n \in \mathcal{E}(\Omega, \mathcal{A})$.
Somit gilt nach 1.:

$$\begin{aligned} \int_{\Omega'} f' \, dP' &=_{\text{Def.}} \lim_{n \in \mathbb{N}} \int_{\Omega'} e_n' \, dP' \\ &=_{1.} \lim_{n \in \mathbb{N}} \int_{\Omega} e_n \, dP \\ &=_{\text{Def.}} \int_{\Omega} f' \circ T \, dP. \end{aligned}$$

3. Für beliebiges $f$ mittels Zerlegung $f = f^+ - f^-$.

□

**Definition 8.18**

1. Es sei $A \in \mathcal{A}$, $f$ Zufallsvariable und $f 1_A$ quasiintegrierbar. Dann:

$$\int_A f \, dP := \int_A f(\omega) \, P(d\omega) := \int f 1_A \, dP.$$

2. $f$ sei quasiintegrierbar. Die Abbildung

$$\mathcal{A} \ni A \mapsto \int_A f \, dP$$

heißt <u>unbestimmtes Integral</u> von $f$.

**Bemerkung** Es seien $A$, $B \in \mathcal{A}$, $A \cap B = \emptyset$ und $f$ quasiintegrierbar. Dann gilt $\int_{A\cup B} f\,dP = \int_A f\,dP = \int_B f\,dP$.

**Satz 8.19** *$f \geq 0$ sei Zufallsvariable mit $\int f\,dP = 1$. Dann:*

$$Q : \mathcal{A} \to \mathbb{R} \textit{ mit } Q(A) := \int_A f\,dP$$

*ist ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$.*

Beweis:

1. $Q(A) \geq 0$ offensichtlich,
2. $Q(\Omega) = 1$ trivial.
3. $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, seien paarweise disjunkt. Dann:

$$\begin{aligned}
Q\left(\bigcup_{n\in\mathbb{N}} A_n\right) & \\
&= \int_{\bigcup_{n\in\mathbb{N}} A_n} f\,dP \\
&=_{\text{Def.}} \int_\Omega f 1_{\bigcup_{n\in\mathbb{N}} A_n}\,dP \\
&= \int_\Omega \sum_{n\in\mathbb{N}} f 1_{A_n}\,dP \\
&= \int_\Omega \lim_{n\in\mathbb{N}} \left(\sum_{i=1}^n f 1_{A_i}\right) dP \\
&=_{\text{mon. Konv.}} \lim_{n\in\mathbb{N}} \int_\Omega \sum_{i=1}^n f 1_{A_i}\,dP \\
&= \lim_{n\in\mathbb{N}} \sum_{i=1}^n \int_\Omega f 1_{A_i}\,dP \\
&=_{\text{Def.}} \lim_{n\in\mathbb{N}} \sum_{i=1}^n \int_{A_i} f\,dP \\
&= \lim_{n\in\mathbb{N}} \sum_{i=1}^n Q(A_i) \\
&= \sum_{n\in\mathbb{N}} Q(A_n).
\end{aligned}$$

□

**Definition 8.20** $f \geq 0$ sei Zufallsvariable mit $\int f\,dP = 1$. $Q$ sei definiert wie in 8.19. Dann heißt $f$ Dichte (genauer: eine Dichte) von $Q$ bezüglich $P$. Symbolisch:

$$Q = fP,\ f = \frac{dQ}{dP},\ dQ = f\,dP.$$

**Satz 8.21** *$f_1$ sei $P$–Dichte von $Q$, $f_2 \geq 0$ sei Zufallsvariable. Dann:*

$$f_2 \text{ ist } P\text{–Dichte von } Q \Leftrightarrow P(\{f_1 \neq f_2\}) = 0.$$

**Lemma 8.22** *$f$ sei Zufallsvariable, $A \in \mathcal{A}$ mit $P(A) = 0 \Rightarrow \int_A |f| \, dP = 0$.*

Beweis: Es gilt $|f1_A| = |f|1_A$. Es existieren $e_n \in \mathcal{E}$, $n \in \mathbb{N}$, mit $e_n \uparrow |f|$

$$\Rightarrow e_n 1_A \in \mathcal{E} \text{ mit } e_n 1_A \uparrow |f|1_A;$$

$$\begin{aligned}
\int |f|1_A \, dP \;&=_{\text{Def.}}\; \lim_{n\in\mathbb{N}} \int e_n 1_A \, dP \\
&= \lim_{n\in\mathbb{N}} \sum_{1\leq i\leq m} \alpha_i \underbrace{P(A_i \cap A)}_{=0} = 0,
\end{aligned}$$

wobei $e_n = \sum_{1\leq i\leq m} \alpha_i 1_{A_i}$, $1_{A_i} 1_A = 1_{A_i \cap A}$,

$$\Rightarrow_{8.13} |\int_A f \, dP| \leq \int_A |f| \, dP = 0.$$

$\square$

Beweis:[von 8.21]

„$\Leftarrow$“ Sei $A \in \mathcal{A}$;

$$\begin{aligned}
Q(A) &= \int f_1 1_A \, dP \\
&= \int f_1 1_{A\cap\{f_1=f_2\}} \, dP + \int f_1 1_{A\cap\{f_1\neq f_2\}} \, dP \\
&=_{8.22} \int f_1 1_{A\cap\{f_1=f_2\}} \, dP \\
&= \int f_2 1_{A\cap\{f_1=f_2\}} \, dP \\
&= \int f_2 1_{A\cap\{f_1=f_2\}} \, dP + \int f_2 1_{A\cap\{f_1\neq f_2\}} \, dP \\
&= \int f_2 1_A \, dP.
\end{aligned}$$

„$\Rightarrow$“ Aus den Voraussetzungen folgt:

$$\begin{aligned}
\forall A \in \mathcal{A} : \int_A f_1 \, dP &= \int_A f_2 \, dP \\
\Rightarrow \forall A \in \mathcal{A} : \quad & \int_A f_1 \, dP - \int_A f_2 \, dP \\
&= \int_A f_1 - f_2 \, dP \;=\; 0.
\end{aligned}$$

Speziell für

$$A_n^+ := \{f_1 - f_2 > 1/n\},\ A_n^- := \{f_1 - f_2 < -1/n\}$$

gilt also:

$$\begin{aligned} 0 &= \int_{A_n^+} f_1 - f_2 \, dP \\ &\geq_{8.13} \int_{A_n^+} \frac{1}{n} \, dP \\ &= \frac{1}{n} P(A_n^+), \end{aligned}$$

d.h. $P(A_n^+) = 0$;

$$\begin{aligned} 0 &= \int_{A_n^-} f_1 - f_2 \, dP \\ &\leq_{8.13} \int_{A_n^-} -\frac{1}{n} \, dP \\ &= -\frac{1}{n} P(A_n^-), \end{aligned}$$

d.h. $P(A_n^-) = 0$. Es folgt:

$$\begin{aligned} P\{f_1 \neq f_2\} &= P\left(\bigcup_{n\in\mathbb{N}} (A_n^+ \cup A_n^-)\right) \\ &\leq \sum_{n\in\mathbb{N}} P(A_n^+ \cup A_n^-) \ = \ 0. \end{aligned}$$

□

**Satz 8.23** *$f$ sei $P$–Dichte von $Q$, $g$ sei Zufallsvariable. Dann gilt: $g$ ist $Q$–integrierbar $\Leftrightarrow$ $gf$ ist $P$–integrierbar und in diesem Fall gilt:*

$$\int g \, dQ = \int gf \, dP.$$

Beweis:

1. Es sei $e = \sum_{1\leq i\leq m} \alpha_i 1_{A_i} \in \mathcal{E}$

$$\begin{aligned} \Rightarrow \int e \, dQ &= \sum_{i=1}^m \alpha_i Q(A_i) \\ &= \sum_{i=1}^m \alpha_i \int_{A_i} f \, dP \\ &= \sum_{i=1}^m \alpha_i \int f 1_{A_i} \, dP \\ &= \int f \sum_{i=1}^m \alpha_i 1_{A_i} \, dP \\ &= \int fe \, dP. \end{aligned}$$

2. $g \geq 0$ sei Zufallsvariable $\Rightarrow g = \lim_{n\in\mathbb{N}} e_n$ mit geeigneten $e_1 \leq e_2 \leq \ldots \in \mathcal{E}$

$$\begin{aligned}\Rightarrow \int g\,dQ \quad &=_{\text{Def.}} \quad \lim_{n\in\mathbb{N}} \int e_n\,dQ \\ &=_{1.} \quad \lim_{n\in\mathbb{N}} \int e_n f\,dP \\ &=_{8.14} \quad \int \lim_{n\in\mathbb{N}} (e_n f)\,dP \\ &= \quad \int fg\,dP.\end{aligned}$$

3. Allgemeiner Fall mittels Zerlegung $g = g^+ - g^-$.

$\square$

**Satz 8.24 (Fubini)** $(\Omega, \mathcal{A}, P)$ *sei das Produkt der beiden Wahrscheinlichkeitsräume* $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, 2$, *d.h.*

- $\Omega = \Omega_1 \times \Omega_2$,
- $\mathcal{A} = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$,
- $P = P_1 \times P_2$.

$f : \Omega \to \mathbb{R}$ *sei integrierbar bzgl.* $P$. *Dann gilt:*

*1.*
$$f(\omega_1, \cdot) : \Omega_2 \ni \omega_2 \mapsto f(\omega_1, \omega_2) \in \mathbb{R}$$
*ist für* $P_1$*–fast alle* $\omega_1 \in \Omega_1$ $P_2$*–integrierbar, d.h. es existiert* $N_1 \in \mathcal{A}_1$, $P_1(N_1) = 0$ *und* $\forall \omega_1 \in N_1^c$ *ist* $f(\omega_1, \cdot)$ *eine* $P_2$*–integrierbare Funktion.*

*2.*
$$f(\cdot, \omega_2) : \Omega_1 \ni \omega_1 \mapsto f(\omega_1, \omega_2) \in \mathbb{R}$$
*ist für* $P_2$*–fast alle* $\omega_2 \in \Omega_2$ $P_1$*–integrierbar, d.h. es existiert* $N_2 \in \mathcal{A}_2$, $P_2(N_2) = 0$ *und* $\forall \omega_2 \in N_2^c$ *ist* $f(\cdot, \omega_2)$ *eine* $P_1$*–integrierbare Funktion.*

*3. Die gemäß 1. bzw. 2. bis auf Nullmengen definierten Funktionen*
$$\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \cdot)\,dP_2$$
*und*
$$\omega_2 \mapsto \int_{\Omega_1} f(\cdot, \omega_2)\,dP_1$$

*sind $P_1$– bzw. $P_2$–integrierbar und es gilt:*

$$\begin{aligned}\int_\Omega f\,dP &= \int_{\Omega_1\times\Omega_2} f\,d(P_1\times P_2)\\ &= \int_{\Omega_1}\left(\int_{\Omega_2} f(\omega_1,\cdot)\,dP_2\right)P_1(d\omega_1)\\ &= \int_{\Omega_2}\left(\int_{\Omega_1} f(\cdot,\omega_2)\,dP_1\right)P_2(d\omega_2).\end{aligned}$$

Beweis: Siehe Maßtheorie. □

**Bemerkung** Die bisher entwickelte Integrationstheorie ist auch für beliebige $\sigma$–finite Maße $\mu$ anstelle eines Wahrscheinlichkeitsmaßes $P$ gültig.
$\mu : \mathcal{A} \to [0,\infty]$ ist Maß $:\Leftrightarrow$

1. $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$ für disjunkte $A_n \in \mathcal{A}$,
2. $\mu(\emptyset) = 0$.

$\mu : \mathcal{A} \to [0,\infty]$ ist $\sigma$–finites Maß $:\Leftrightarrow$ $\mu$ ist Maß und es existieren $A_n \in \mathcal{A}$, $n\in\mathbb{N}$, mit $\bigcup_{n\in\mathbb{N}} A_n = \Omega$ und $\mu(A_n) < \infty$, $n \in \mathbb{N}$.

**Definition 8.25** Definiere $\lambda_n^0 : \mathcal{I}_n^0 := \{(\mathbf{a},\mathbf{b}] = \times_{i=1}^n (a_i,b_i] : \mathbf{a},\mathbf{b}\in\mathbb{R}^n\} \to [0,\infty]$ durch

$$\lambda_n^0((\mathbf{a},\mathbf{b}]) := \prod_{1\le i\le n}(b_i - a_i).$$

$\lambda_n^0$ ist $\sigma$–additiv auf $\mathcal{I}_n^0$ und kann eindeutig zu einem ($\sigma$–finiten) Maß auf $\mathbb{B}^n$ fortgesetzt werden. Dieses Maß ist das Lebesgue–Maß, i.Z. $\lambda_n$, vgl. 7.8.

**Satz 8.26** *Es gilt:*

$$\lambda_n = \lambda_1^n,$$

*wobei $\lambda_1^n$ das n–fache Produkt von $\lambda_1$ bezeichnet.*

Beweis: Klar, da

$$\begin{aligned}\lambda_n((\mathbf{a},\mathbf{b}]) &= \lambda_n^0((\mathbf{a},\mathbf{b}])\\ &= \prod_{1\le i\le n}(b_i-a_i)\\ &= \prod_{1\le i\le n}\lambda_1((a_i,b_i])\\ &= \lambda_1^n((\mathbf{a},\mathbf{b}]).\end{aligned}$$

□

**Definition 8.27** Es sei $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{B}^n, \lambda_n)$. Wir setzen:

$$\int f(\mathbf{x})\, d\mathbf{x} := \int f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n := \int f\, d\lambda_n.$$

**Bemerkung 8.28** Setze $(\Omega, \mathcal{A}, P) := ((0,1], \mathbb{B}\cap(0,1], \lambda_1/(0,1])$, $f_n := n1_{(0,1/n]}$, $n \in \mathbb{N}$. Dann gilt:

$$f_n(\omega) \to_{n\to\mathbb{N}} 0, \quad \omega \in \Omega,$$

aber

$$\int f_n\, d\lambda_1 = n\lambda_1((0,1/n]) = 1 \neq \int 0\, d\lambda_1 = 0.$$

Die Monotonie– bzw. Beschränktheitsvoraussetzungen in den Integrationssätzen sind also wesentlich.

# 9 Verteilungen und ihre Charakterisierungen

**Definition 9.1** $(\Omega, \mathcal{A}, P)$ sei Wahrscheinlichkeitsraum, $(\Omega', \mathcal{A}')$ messbarer Raum, $f : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$.
Das Wahrscheinlichkeitsmaß $P * f : \mathcal{A}' \to [0,1]$, definiert durch

$$(P * f)(A) := P(f^{-1}(A)), \qquad A \in \mathcal{A}',$$

heißt Verteilung von $f$, (s. 7.1).

**Bemerkung** Jedes Wahrscheinlichkeitsmaß $P'$ auf einem beliebigen messbaren Raum $(\Omega', \mathcal{A}')$ kann als Verteilung einer geeigneten Zufallsgröße aufgefasst werden: Setze

$$(\Omega, \mathcal{A}, P) := (\Omega', \mathcal{A}', P'),\ f(\omega) := \omega,\ \omega \in \Omega.$$

**Definition 9.2** $f$ sei Zufallsvariable über $(\Omega, \mathcal{A}, P)$, d.h. $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$.
Definiere $F : \mathbb{R} \to [0,1]$ durch

$$\begin{aligned} F(x) &:= P(\{f \le x\}) \\ &= P(\{\omega \in \Omega : f(\omega) \le x\}) \\ &= (P * f)((-\infty, x]), \qquad x \in \mathbb{R}. \end{aligned}$$

$F$ heißt Verteilungsfunktion von $f$ bzw. von $P * f$.

**Beispiel 9.3**

1. $f$ sei das Ergebnis beim Würfeln, d.h. $P(\{i\}) = 1/6$, $i = 1, \ldots, 6$. Dann gilt:

$$F(x) = \begin{cases} 0 & \text{für } x < 1, \\ i/6 & \text{für } i \le x < i+1,\ i = 1, \ldots, 5, \\ 1 & \text{für } x \ge 6. \end{cases}$$

2. $f$ sei gleichverteilt auf $(0,1)$, d.h. $P(\{f \in B\} = \lambda_1(B)$ für $B \in \mathbb{B} \cap (0,1)$. Dann gilt:

$$\begin{aligned} F(x) &= P(\{f \leq x\}) \\ &= P(\{f \in (-\infty, x]\}) \\ &= P(\{f \in (0, x]\}) \\ &= \lambda_1((0,x]) \;=\; x \qquad \text{für } 0 \leq x \leq 1. \end{aligned}$$

**Bemerkung 9.4** Die Verteilungsfunktion $F(x) = (P * f)((-\infty, x])$ einer Zufallsvariablen $f$ hängt offenbar nur von der Verteilung $P * f$ von $f$ ab, nicht von den konkreten Werten von $f$.

**Satz 9.5** *$F$ sei Verteilungsfunktion der Verteilung $Q := P * f$. Dann gilt:*

1. *$F$ ist monoton wachsend.*
2. *$F$ ist rechtsseitig stetig.*
3. $\lim_{x\to\infty} F(x) = 1,\ \lim_{x\to-\infty} F(x) = 0.$

Beweis:

**„1.“**

$$\begin{aligned} & x < y \\ &\Rightarrow\ (-\infty, x] \subset (-\infty, y] \\ &\Rightarrow\ F(x) = Q((-\infty, x]) \leq Q((-\infty, y]) = F(y). \end{aligned}$$

**„2.“** $(-\infty, x] = \bigcap_{n\in\mathbb{N}} (-\infty, x_n]$, falls $x_n \downarrow x$. Damit:

$$\begin{aligned} F(x) &= Q((-\infty, x]) \\ &= Q\left(\bigcap_{n\in\mathbb{N}} (-\infty, x_n]\right) \\ &= \lim_{n\in\mathbb{N}} Q((-\infty, x_n]) \\ &= \lim_{n\in\mathbb{N}} F(x_n). \end{aligned}$$

**„3.“**

$$\begin{aligned} & x_n \uparrow \infty \\ &\Rightarrow\ \mathbb{R} = \bigcup_{n\in\mathbb{N}} (-\infty, x_n] \\ &\Rightarrow\ 1 = Q(\mathbb{R}) = \lim_{n\in\mathbb{N}} Q((-\infty, x_n]) = \lim_{n\in\mathbb{N}} F(x_n); \end{aligned}$$

$$\begin{aligned} & x_n \downarrow -\infty \\ &\Rightarrow\ \emptyset = \bigcap_{n\in\mathbb{N}} (-\infty, x_n] \\ &\Rightarrow\ 0 = Q(\emptyset) = \lim_{n\in\mathbb{N}} Q((-\infty, x_n]) = \lim_{n\in\mathbb{N}} F(x_n). \end{aligned}$$

□

**Satz 9.6** *Eine Verteilung $Q$ ist durch ihre Verteilungsfunktion $F$ eindeutig bestimmt.*
*Das bedeutet: Sind $Q_1, Q_2$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathbb{B})$ mit $Q_1 \neq Q_2$, so existiert $x \in \mathbb{R}$ mit*

$$F_{Q_1}(x) = Q_1((-\infty, x]) \neq Q_2((-\infty, x]) = F_{Q_2}(x).$$

Beweis:

1. Es gilt:
$$Q((x, y]) = F(y) - F(x) \text{ für } x < y.$$

2. 
$$\begin{aligned} Q((x, y)) &= \lim_{n \in \mathbb{N}} Q((x, y - 1/n]) \\ &= \lim_{n \in \mathbb{N}} (F(y - 1/n) - F(x)), \end{aligned}$$
denn $(x, y - 1/n] \uparrow (x, y)$.

3. Jede offene Menge in $\mathbb{R}$ ist Vereinigung von abzählbar vielen disjunkten offenen Intervallen
$\Rightarrow_{2.}$ $Q(G)$ ist für offene Mengen $G$ durch $F$ bestimmt
$\Rightarrow$ $Q(B)$ ist für beliebiges $B \in \mathbb{B}$ bestimmt (Maßtheorie, Fortsetzungssatz).

□

**Satz 9.7** $F : \mathbb{R} \to [0, 1]$ *sei eine Funktion, die 1.–3. von Satz 9.5 erfülle. Dann ist $F$ die Verteilungsfunktion einer geeigneten Zufallsvariablen $f$, d.h. es existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine Zufallsvariable $f$ auf $(\Omega, \mathcal{A}, P)$ mit $F(x) = (P * f)((-\infty, x])$, $x \in \mathbb{R}$.*

Beweis: Setze $\Omega := (0, 1)$, $\mathcal{A} := \mathbb{B} \cap (0, 1)$, $P := \lambda_1/\Omega$,

$$f(\omega) := \inf\{x \in \mathbb{R} : F(x) \geq \omega\}$$

$= \min\{x \in \mathbb{R} : F(x) \geq \omega\}$ wegen der rechtsseitigen Stetigkeit von $F$. Dann gilt:

1. $f$ ist Zufallsvariable, denn: $f$ ist monoton wachsend, d.h. $f(\omega_1) \leq f(\omega_2)$, $\omega_1 \leq \omega_2$, und damit ist $\{f \leq x\}$ ein Intervall (in $\Omega$), $x \in \mathbb{R}$, also in $\mathbb{B} \cap \Omega$.

2.
$$\omega \le F(y) \Leftrightarrow f(\omega) \le y, \quad y \in \mathbb{R},$$

$$\begin{aligned} \Rightarrow \{\omega \in \Omega : \omega \le F(y)\} &= \{\omega \in \Omega : f(\omega) \le y\} \\ \Rightarrow P(\{f \le y\}) &= \lambda_1(\{\omega \in \Omega : \omega \le F(y)\}) \\ &= \lambda_1((0, F(y)]) \\ &= F(y), \quad y \in \mathbb{R}. \end{aligned}$$

□

**Definition 9.8** $F : \mathbb{R} \to [0,1]$ ist Verteilungsfunktion $:\Leftrightarrow$ $F$ erfüllt 1.–3. von Satz 9.5.

**Definition 9.9** $F$ sei Verteilungsfunktion; dann heißt

$$F^{-1}(y) := \inf\{x \in \mathbb{R} : F(x) \ge y\}, \quad y \in (0,1),$$

verallgemeinerte Inverse von $F$ oder Quantilfunktion.

**Satz 9.10** *$f$ sei eine auf $(0,1)$ gleichverteile Zufallsvariable und $F$ eine beliebige Verteilungsfunktion. Dann besitzt die Zufallsvariable*

$$g := F^{-1} \circ f$$

*die Verteilungsfunktion $F$.*

Beweis: Siehe Übungen. □

**Definition 9.11** $f = (f_1, \dots, f_n)$ sei $n$–dimensionaler Zufallsvektor, dann: $F : \mathbb{R}^n \to [0,1]$ mit

$$F(y_1, \dots, y_n) := P(\{f_1 \le y_1, \dots, f_n \le y_n\})$$

ist die ($n$–dimensionale) Verteilungsfunktion oder gemeinsame Verteilungsfunktion von $f_1, \dots, f_n$.

**Bemerkung**

$n = 1$:
$$P(\{f \in (x,y]\}) = F(y) - F(x), \; x < y.$$

$n = 2$: Es sei $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$, $x_i < y_i$, $i = 1, 2$. Dann:

$$\begin{aligned} & P(\{(f_1, f_2) \in (\mathbf{x}, \mathbf{y}]\}) \\ = \quad & F(y_1, y_2) - F(y_1, x_2) - F(x_1, y_2) + F(x_1, x_2). \end{aligned}$$

Diese Gleichungen lassen sich auf beliebige Dimensionen erweitern.

**Satz 9.12** *Die Zufallsvariablen $f_1, \ldots, f_n$ sind genau dann unabhängig, wenn für ihre gemeinsame Verteilungsfunktion $F$ gilt:*

$$F(y_1, \ldots, y_n) = F_1(y_1) \cdots F_n(y_n)$$

*für $(y_1, \ldots, y_n) \in \mathbb{R}^n$, wobei $F_i$ die Verteilungsfunktion zu $f_i$ ist, $i = 1, \ldots, n$.*

Beweis:

**„⇒“**

$$\begin{aligned} F(y_1, \ldots, y_n) &= P\left( \bigcap_{1 \le i \le n} \{f_i \le y_i\} \right) \\ &= \prod_{1 \le i \le n} P(\{f_i \le y_i\}) \\ &= \prod_{1 \le i \le n} F_i(y_i). \end{aligned}$$

**„⇐“** Es gilt für $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$:

$$F(\mathbf{y}) = P\left( \bigcap_{1 \le i \le n} \{f_i \le y_i\} \right) = \prod_{1 \le i \le n} P(\{f_i \le y_i\}),$$

d.h.

$$\begin{aligned} & (P * (f_1, \ldots, f_n))((-\infty, \mathbf{y}]) \\ & = ((P * f_1) \times \cdots \times (P * f_n))((-\infty, \mathbf{y}]) \\ \Rightarrow \quad & (P * (f_1, \ldots, f_n))(B) \\ & = ((P * f_1) \times \cdots \times (P * f_n))(B) \\ & \qquad \text{für } B \in \mathbb{B}^n \text{ (s. Maßtheorie)} \\ \Rightarrow \quad & (P * (f_1, \ldots, f_n))(B_1 \times \cdots \times B_n) \\ & = ((P * f_1) \times \cdots \times (P * f_n))(B_1 \times \cdots \times B_n) \\ & \qquad \text{für } B_i \in \mathbb{B},\ i = 1 \ldots, n, \\ \Rightarrow \quad & P(\{f_i \in B_i, i = 1, \ldots, n\}) \\ & = \prod_{i=1}^{n} P(\{f_i \in B_i\}) \end{aligned}$$

für $B_i \in \mathbb{B}$, $i = 1, \ldots, n$, d.h. $f_1, \ldots, f_n$ sind unabhängig (s. 7.23).

$\square$

**Definition 9.13** Eine Zufallsvariable $f$ ist <u>diskret verteilt</u> $:\Leftrightarrow$ Es existiert eine abzählbare Menge $B = \{b_1, b_2, \ldots\} \subset \mathbb{R}$ mit $P(\{f \in B\}) = 1$.

**Beispiele 9.14** Im folgenden sei $f$ eine Zufallsvariable.

1. $f$ besitzt <u>Bernoulli–Verteilung</u> mit Parameter $p \in [0,1]$, i.Z. $B(1,p)$,
$$:\Leftrightarrow P(\{f=1\}) = p,\ P(\{f=0\}) = 1-p.$$
2. $f$ besitzt <u>Binomialverteilung</u> mit den Parametern $n \in \mathbb{N}$, $p \in [0,1]$, i.Z. $B(n,p)$,
$$\begin{aligned} :\Leftrightarrow P(\{f=k\}) &= B(n,p)(\{k\}) \\ &= \binom{n}{k} p^k (1-p)^{n-k},\ k = 0,1,\ldots,n. \end{aligned}$$
3. $f$ besitzt <u>Poisson–Verteilung</u> mit dem Parameter $\lambda > 0$, i.Z. $P_\lambda$,
$$\begin{aligned} :\Leftrightarrow P(\{f=k\}) &= P_\lambda(\{k\}) \\ &= e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0,1,\ldots \end{aligned}$$
4. $f$ besitzt <u>geometrische Verteilung</u> zum Parameter $p \in [0,1]$
$$:\Leftrightarrow P(\{f=k\}) = p(1-p)^{k-1},\ k = 1,2,\ldots$$

**Bemerkung 9.15** $f_1, f_2, \ldots$ seien unabhängige, zum Parameter $p \in [0,1]$ Bernoulli–verteilte Zufallsvariablen. Dann gilt:

1. $f_1 + \cdots + f_n$ ist $B(n,p)$–verteilt.
2. $f := \inf\{m \in \mathbb{N} : f_m = 1\}$ ist geometrisch verteilt zum Parameter $p$.

Beweis: Siehe Übungen. $\square$

**Definition 9.16**

1. Eine Verteilung $Q$ auf $\mathbb{B}$ heißt <u>absolutstetig</u> $:\Leftrightarrow$ $Q$ besitzt eine Dichte bzgl. $\lambda_1$, d.h. es existiert eine Borel–messbare Funktion $h : \mathbb{R} \to [0,\infty)$ mit
$$Q(B) = \int_B h\, d\lambda_1, \quad B \in \mathbb{B}.$$

2. Eine Zufallsvariable $f$ heißt absolutstetig $:\Leftrightarrow P * f$ ist absolutstetig.

**Satz 9.17** *$f$ sei eine absolutstetige Zufallsvariable mit Dichte $h$ und Verteilungsfunktion $F$. Dann gilt:*

$$h(x) = F'(x) = \frac{\partial F(x)}{\partial x},$$

*falls $h$ an der Stelle $x$ stetig ist.*

Beweis: Es gilt:

$$\begin{aligned}
&\frac{F(x+\varepsilon) - F(x)}{\varepsilon} \\
&= \frac{P(\{f \in (x, x+\varepsilon]\})}{\varepsilon} \\
&= \frac{\int_{(x,x+\varepsilon]} h(y)\, dy}{\varepsilon} \\
&= \frac{\int_{(x,x+\varepsilon]} h(y) - h(x)\, dy}{\varepsilon} + \frac{\int_{(x,x+\varepsilon]} h(x)\, dy}{\varepsilon} \\
&=: I + II.
\end{aligned}$$

Offenbar gilt:

$$II = h(x)\frac{\lambda_1((x, x+\varepsilon])}{\varepsilon} = h(x),$$

sowie

$$\begin{aligned}
|I| &\leq \frac{\int_{(x,x+\varepsilon]} |h(y) - h(x)|\, dy}{\varepsilon} \\
&\leq \frac{\int_{(x,x+\varepsilon]} \sup_{z\in(x,x+\varepsilon]} |h(z) - h(x)|\, dy}{\varepsilon} \\
&= \sup_{z\in(x,x+\varepsilon]} |h(z) - h(x)| \to_{\varepsilon\to 0} 0,
\end{aligned}$$

wegen der Stetigkeit von $h$ in $x$. Hieraus folgt die Behauptung. □

**Definition 9.18** $Q_{a,b}$ ist die Gleichverteilung auf $(a, b) \subset \mathbb{R}$, $-\infty < a < b < \infty :\Leftrightarrow Q_{a,b}$ besitzt die Dichte

$$h_{a,b}(x) := \frac{1}{b-a} 1_{(a,b)}(x), \quad x \in \mathbb{R}.$$

**Beispiel 9.19** *$f$ sei die Lebensdauer eines nicht alternden Objektes. Dabei bedeutet keine Alterung:*

$$P(\{f > s + t | f > s\}) = P(\{f > t\}), \quad t, s \geq 0.$$

*Es gilt also für $t, s, \geq 0$:*

$$\begin{aligned} & P(\{f > s + t\}) = P(\{f > s\})P(\{f > t\}), \\ \Rightarrow \quad & \textit{für } G(y) := P(\{f > y\}) \textit{ gilt:} \\ & \qquad G(s + t) = G(s)G(t) \\ \Rightarrow \quad & \exists \lambda \geq 0 : G(y) = \exp(-\lambda y) \end{aligned}$$

*(als einzige nicht identisch verschwindende monoton fallende Lösung obiger Funktionalgleichung[2])*

$$\begin{aligned} \Rightarrow F(y) &:= P(\{f \leq y\}) \\ &= 1 - G(y) \\ &= 1 - \exp(-\lambda y), \quad y \geq 0. \end{aligned}$$

**Definition 9.20** Eine Zufallsvariable $f$ ist exponentialverteilt zum Parameter $\lambda > 0 :\Leftrightarrow P * f$ besitzt die Dichte

$$h(x) := \begin{cases} 0, & x < 0, \\ \lambda \exp(-\lambda x), & x \geq 0. \end{cases}$$

Offenbar gilt dann: $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0$.

**Definition 9.21** Der Zufallvektor $f = (f_1, \ldots, f_n)$ ist absolutstetig mit der Dichte $h :\Leftrightarrow$

$$\begin{aligned} P(\{f \in B\}) &= \int_B h \, d\lambda^n \\ &= \int_B h(x_1, \ldots, x_n) \, dx_1 \cdots dx_n, \quad B \in \mathbb{B}^n. \end{aligned}$$

**Satz 9.22** *$h : \mathbb{R}^n \to [0, \infty)$ sei integrierbar bzgl. $\lambda^n$ und es gelte für den Zufallsvektor $f$:*

$$\begin{aligned} & P(\{f \leq \mathbf{y}\}) \\ & = \int_{(-\infty, y_1]} \cdots \int_{(-\infty, y_n]} h(x_1, \ldots, x_n) \, dx_1 \cdots dx_n \end{aligned}$$

*für $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Dann ist $h$ eine Dichte von $f$.*

[2] S. 133 in Krengel, U. (2002). *Einführung in die Wahrscheinlichkeitstheorie u. Statistik*, 6. Auflage. Vieweg, Braunschweig.

Beweisskizze: Aus der Voraussetzung folgt nach dem Satz von Fubini für jedes $\mathbf{y} \in \mathbb{R}^n$:

$$P(\{f \in (-\infty, \mathbf{y}]\}) = \int_{(-\infty,\mathbf{y}]} h \, d\lambda^n$$

und hieraus (Maßtheorie, Fortsetzungssatz) für jedes $B \in \mathbb{B}^n$:

$$P(\{f \in B\}) = \int_B h \, d\lambda^n$$

□

**Satz 9.23** *$f_1, \ldots, f_n$ seien Zufallsvariablen mit Dichten $h_1, \ldots, h_n$. Dann gilt:*
*$f_1, \ldots, f_n$ sind unabhängig $\Leftrightarrow$ $f = (f_1, \ldots, f_n)$ besitzt die Dichte*

$$h(x_1, \ldots, x_n) = h_1(x_1) \cdots h_n(x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Beweis:

**„⇒"** Für $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ gilt:

$$\begin{aligned}
&P(\{f \in (-\infty, \mathbf{y}]\}) \\
&\quad = P(\{f_i \le y_i, i = 1, \ldots, n\}) \\
&\quad = \prod_{i=1}^n P(\{f_i \le y_i\}) \\
&\quad = \prod_{i=1}^n \int_{(-\infty,y_i]} h_i(x_i) \, dx_i \\
&\quad = \int_{(-\infty,y_1]} \cdots \int_{(-\infty,y_n]} h_1(x_1) \cdots h_n(x_n) \, dx_n \cdots dx_1 \\
&\quad = \int_{(-\infty,y_1]} \cdots \int_{(-\infty,y_n]} h(x_1, \ldots, x_n) \, dx_n \cdots dx_1 \\
&\quad \Rightarrow \text{ Behauptung aus 9.22.}
\end{aligned}$$

**„⇐"** Für $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ gilt auf Grund des Satzes von Fubini:

$$\begin{aligned}
&P(\{f \in (-\infty, \mathbf{y}]\}) \\
&\quad = \int_{(-\infty,\mathbf{y}]} h \, d\lambda^n \\
&\quad = \int_{(-\infty,y_1]} \cdots \int_{(-\infty,y_n]} h(x_1, \ldots, x_n) \, dx_1 \cdots dx_n \\
&\quad = \int_{(-\infty,y_1]} \cdots \int_{(-\infty,y_n]} h_1(x_1) \cdots h_n(x_n) \, dx_1 \cdots dx_n \\
&\quad = \prod_{i=1}^n \int_{(-\infty,y_i]} h_i(x_i) \, dx_i \\
&\quad = \prod_{i=1}^n P(\{f_i \le y_i\}) \\
&\quad \Rightarrow \text{ Behauptung aus 9.12.}
\end{aligned}$$

$\square$

**Beispiel 9.24 (Buffonsches Nadelproblem)** Eine Nadel der Länge 1 wird zufällig auf ein Raster aus Parallelen mit dem einheitlichen Abstand 1 geworfen. Wie groß ist die Wahrscheinlichkeit, dass die Nadel eine der Geraden schneidet?

LÖSUNG: Annahmen:

1. Der Winkel $\varphi$ zwischen der Geraden und der Nadel ist auf $(0, \pi)$ gleichverteilt.
2. Der Abstand $d$ des Nadelmittelpunktes zur nächsten Geraden ist auf $(0, 1/2)$ gleichverteilt.
3. $\varphi$ und $d$ sind unabhängig.

Dann gilt nach 9.23

$$h(x_1, x_2) = \begin{cases} \frac{2}{\pi} & \text{für } 0 < x_1 < \pi,\ 0 < x_2 < \frac{1}{2}, \\ 0 & \text{sonst} \end{cases}$$

ist gemeinsame Dichte des Zufallsvektors $(\varphi, d)$.
Ferner sei $A$ das Ereignis, dass die Nadel eine der Geraden schneidet, d.h.

$$A = \left\{ \omega \in \Omega : d(\omega) \leq \frac{1}{2} \sin \varphi(\omega) \right\}.$$

Mit

$$A' = \left\{ (x_1, x_2) \in (0, \pi) \times \left(0, \frac{1}{2}\right) : x_2 \leq \frac{1}{2} \sin(x_1) \right\}$$

folgt aus dem Satz von Fubini:

$$\begin{aligned} P(A) &= P(\{(\varphi, d) \in A'\}) \\ &= \int_{A'} h \, d\lambda^2 \\ &= \int_{\left\{(x_1,x_2)\in(0,\pi)\times\left(0,\frac{1}{2}\right) : x_2 \leq \frac{1}{2}\sin(x_1)\right\}} \frac{2}{\pi} \, dx_1 dx_2 \\ &= \int_{(0,\pi)} \left( \int_{(0, \frac{1}{2}\sin(x_1))} \frac{2}{\pi} \, dx_2 \right) dx_1 \\ &= \frac{1}{\pi} \int_0^{\pi} \sin(x_1) \, dx_1 \\ &= \frac{2}{\pi}. \end{aligned}$$

$\square$

**Satz 9.25** *$f$ sei eine absolutstetige Zufallsvariable mit stetiger Dichte $h$, $I \subset \mathbb{R}$ Intervall mit $P(\{f \in I\}) = 1$.*
*$g : I \to \mathbb{R}$ sei differenzierbar mit $g'(x) > 0$ für alle $x \in I$ oder $g'(x) < 0$ für alle $x \in I$. Dann gilt:*
*$g \circ f$ ist absolutstetig mit der Dichte*

$$\tilde{h}(y) = h(g^{-1}(y))|(g^{-1})'(y)|$$

*für alle $y$ mit*

$$\inf_{x \in I} g(x) < y < \sup_{x \in I} g(x)$$

*und $\tilde{h}(y) = 0$ sonst.*

BEWEIS: $g$ ist auf $I$ streng monoton und differenzierbar $\Rightarrow g^{-1}$ ist definiert (auf $g(I)$) und differenzierbar (mit $(g^{-1})'(y) = 1/g'(g^{-1}(y))$).

1. $\forall x \in I$ gelte $g'(x) > 0$, d.h. $g$ ist monoton wachsend

$$\begin{aligned} \Rightarrow F_{g \circ f}(y) &= P(\{g \circ f \leq y\}) \\ &= P(\{f \leq g^{-1}(y)\}) \\ &= F_f(g^{-1}(y)) \\ \Rightarrow F'_{g \circ f}(y) &= h(g^{-1}(y))(g^{-1})'(y). \end{aligned}$$

2. $g' < 0$ analog.

□

**Bemerkung 9.26** Im obigen Satz kann die Stetigkeit von $h$ ersatzlos gestrichen werden.[3]

**Beispiel 9.27** Ein Teilchen trete mit einem Winkel $\varphi$ zur $x$–Achse aus dem Nullpunkt aus, wobei $\varphi$ auf $(-\pi/2, \pi/2)$ gleichverteilt sei. Im Abstand $\lambda$ vom Nullpunkt sei ein Schirm aufgestellt, auf den das Teilchen trifft. Die Koordinaten dieses Punktes seien $(\lambda, g)$ wobei $g$ zufällig ist. Man bestimmt eine Dichte der Zufallsvariablen $g$.

LÖSUNG: $\varphi$ besitzt die Dichte

$$h(y) = \frac{1}{\pi} 1_{\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)}(y),$$

da $\varphi$ auf $(-\pi/2, \pi/2)$ gleichverteilt ist.
Ferner gilt:

$$g = \lambda \tan(\varphi),$$

[3] S. 148 in Krengel, U. (2002). *Einführung in die Wahrscheinlichkeitstheorie u. Statistik*, 6. Auflage. Vieweg, Braunschweig.

wobei $\lambda \tan(x)$ streng monoton ist im Intervall $(-\pi/2, \pi/2)$.
Nach 9.25, 9.26 besitzt $g$ die Dichte

$$\begin{aligned} h_g(y) &= \frac{1}{\pi} \left| \frac{\partial}{\partial y} \left( \arctan\left(\frac{y}{\lambda}\right)\right)\right| \\ &= \frac{1}{\pi} \frac{1}{1+\frac{y^2}{\lambda^2}} \frac{1}{\lambda} \\ &= \frac{1}{\pi} \frac{\lambda}{\lambda^2 + y^2}, \qquad y \in \mathbb{R}. \end{aligned}$$

□

**Definition 9.28** Die Verteilung auf $\mathbb{R}$ mit der Dichte

$$h_\lambda(y) := \frac{1}{\pi} \frac{\lambda}{\lambda^2 + y^2}, \qquad y \in \mathbb{R},$$

heißt Cauchy–Verteilung mit Parameter $\lambda > 0$.

**Satz 9.29** *Ist eine Verteilungsfunktion $F$ auf $(\mathbb{R}, \mathbb{B})$ überall differenzierbar, so ist $F'$ eine Dichte von $F$.*[4]

# 10 Momente

**Definition 10.1** Es sei $f \in \mathcal{L}(\Omega, \mathcal{A}, P)$. Dann heißt

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

Erwartungswert von $f$ (Mittelwert von $P * f$).

**Bemerkung 10.2** Es gilt

$$E(f) = \int_\Omega f \, dP =_{8.17} \int_{\mathbb{R}} x \, (P * f)(dx),$$

d.h. $E(f)$ hängt nur von der Verteilung $P * f$ von $f$ ab. Man schreibt daher auch häufig

$$\int x \, F(dx) := E(f),$$

wobei $F(x) := P(\{f \leq x\})$, $x \in \mathbb{R}$ (s. 9.6 bzw. 9.2).

**Bemerkung 10.3** $f, g$ seien Zufallsvariablen, $\alpha \in \mathbb{R}$, dann:

[4] Natanson, I.P. (1975). *Theorie der Funktionen einer Veränderlichen*, 4. Auflage. Deutsch, Zürich.

1. $E(\alpha f) = \alpha E(f)$, $E(f+g) = E(f) + E(g)$, falls $f, g \in \mathcal{L}(\Omega, \mathcal{A}, P)$.
2. $E(f)$ existiert $\Leftrightarrow$ $E(|f|)$ existiert, und in diesem Fall gilt: $|E(f)| \leq E(|f|)$.
3. Falls $E(f)$ existiert und $|g| \leq |f| \Rightarrow E(g)$ existiert.

Beweis: 8.13. □

**Satz 10.4** *$f$ sei eine diskret verteilte Zufallsvariable mit $P(\{f \in B\}) = 1$ für eine abzählbare Teilmenge $B$ von $\mathbb{R}$. $\varphi : \mathbb{R} \to \mathbb{R}$ sei beliebig, $g := \varphi \circ f$; dann:*

$$E(g) = \sum_{x \in B} \varphi(x) P(\{f = x\}),$$

*falls diese Summe absolut konvergiert.*

Beweis: S. Übungen. □

**Satz 10.5** *$f$ sei absolutstetige Zufallsvariable mit Dichte $h$, $\varphi : \mathbb{R} \to \mathbb{R}$ sei Bairesche Funktion. Dann gilt für $g := \varphi \circ f$:*

$$E(g) = \int_{\mathbb{R}} \varphi(x) h(x)\, dx$$

*falls $\int |\varphi h|\, d\lambda_1 < \infty$.*

Beweis:

$$\begin{aligned} E(g) &= \int_{\Omega} \varphi \circ f\, dP \\ &=_{8.17} \int_{\mathbb{R}} \varphi\, d(P * f) \\ &=_{8.23} \int_{\mathbb{R}} \varphi(x) h(x)\, dx, \end{aligned}$$

$h = d(P * f)/d\lambda_1$. □

**Definition 10.6** $f$ sei Zufallsvariable, $n \in \mathbb{Z}_+$. Dann heißt, falls existent:

$$\mu_n := E(f^n)$$

$\underline{n\text{–tes Moment von } f}$,

speziell

$$\mu := \mu_1 = E(f);$$

$$m_n := E((f-\mu)^n)$$

<u>$n$–tes zentriertes Moment von $f$</u>,

speziell

$$m_2 := V(f) := \mathrm{var}(f) =: \sigma^2(f)$$

<u>Varianz von f</u>;

$$E(|f|^n) := n\text{–tes absolutes Moment;}$$

$$\begin{aligned}\sigma(f) &:= \sqrt{\sigma^2(f)} \\ &= \sqrt{E((f-\mu)^2)} = E((f-\mu)^2)^{1/2}\end{aligned}$$

<u>Standardabweichung</u> oder <u>Streuung</u> von $f$.

**Satz 10.7** *$E(f^n)$ existiere $\Rightarrow$ $E(f^m)$ existiert, $0 \le m \le n$.*

Beweis: Es gilt:

$$|f(\omega)|^m \le \begin{cases} |f(\omega)|^n, & \text{falls } |f(\omega)| \ge 1 \\ 1, & \text{falls } |f(\omega)| \le 1 \end{cases} =: g(\omega);$$

$E(g) \le 1 + E(|f|^n) < \infty \Rightarrow_{8.13}$ Behauptung. $\square$

**Satz 10.8** *$m_1 = 0$, $\sigma^2 = \mu_2 - \mu_1^2$ $(= E(f^2) - E(f)^2)$, $m_3 = \mu_3 - 3\mu\mu_2 + 2\mu^3$.*

Beweis:

$$m_1 = E(f-\mu) = E(f) - E(\mu) = \mu - \mu = 0;$$

$$\begin{aligned}\sigma^2 &= E((f-\mu)^2) \\ &= E(f^2 - 2f\mu + \mu^2) \\ &= E(f^2) - 2\mu E(f) + \mu^2 \\ &= \mu_2 - 2\mu^2 + \mu^2 \\ &= \mu_2 - \mu^2;\end{aligned}$$

$$\begin{aligned}m_3 &= E((f-\mu)^3) \\ &= E(f^3) - E(3f^2\mu) + E(3f\mu^2) - \mu^3 \\ &= \mu_3 - 3\mu\mu_2 + 3\mu^3 - \mu^3 \\ &= \mu_3 - 3\mu\mu_2 + 2\mu^3.\end{aligned}$$

$\square$

**Satz 10.9 (Markoffsche Ungleichung)** *$f$ sei Zufallsvariable, $\varepsilon > 0$. Dann gilt:*

$$P(\{|f| \geq \varepsilon\}) \leq \frac{E(|f|)}{\varepsilon}.$$

BEWEIS: Es gilt:

$$\begin{aligned}
E(|f|) &= \int_\Omega |f(\omega)|\, dP \\
&\geq \int_{\{\omega\in\Omega:\, |f(\omega)|\geq\varepsilon\}} |f(\omega)|\, P(d\omega) \\
&\geq \int_{\{\omega\in\Omega:\, |f(\omega)|\geq\varepsilon\}} \varepsilon\, P(d\omega) \\
&= \varepsilon P(\{|f| \geq \varepsilon\}).
\end{aligned}$$

$\square$

**Korollar 10.10 (Tschebyscheffsche Ungleichung)** *Für $f \in \mathcal{L}(\Omega, \mathcal{A}, P)$ und $\varepsilon > 0$ gilt:*

$$P(\{|f - \mu| \geq \varepsilon\}) \leq \frac{\sigma^2(f)}{\varepsilon^2}.$$

BEWEIS: Es gilt:

$$P(\{|f - \mu| \geq \varepsilon\}) = P(\{(f - \mu)^2 \geq \varepsilon^2\}) \leq_{10.9} \frac{\sigma^2(f)}{\varepsilon^2}.$$

$\square$

**Satz 10.11 (Schwarzsche Ungleichung)** *$f, g$ seien Zufallsvariablen mit $E(f^2) < \infty$, $E(g^2) < \infty$. Dann ist $fg \in \mathcal{L}(\Omega, \mathcal{A}, P)$ und es gilt:*

$$E(fg)^2 \leq E(f^2)E(g^2).$$

BEWEIS:

1. $$(f \pm g)^2 \geq 0 \Rightarrow |2fg| \leq f^2 + g^2 \Rightarrow_{8.13} fg \in \mathcal{L}(\Omega, \mathcal{A}, P)$$

2. Für alle $x \in \mathbb{R}$ gilt:

$$\begin{aligned} \varphi(x) &:= E((xf-g)^2) \\ &= x^2E(f^2) - 2xE(fg) + E(g^2) \geq 0 \end{aligned}$$

$\Rightarrow$ die quadratische Gleichung $\varphi(x) = 0$ besitzt höchstens eine Lösung

$\Rightarrow$ die Diskriminante dieser quadratischen Gleichung kann nicht positiv sein, d.h. es gilt:

$$\left(\frac{E(fg)}{E(f^2)}\right)^2 - \frac{E(g^2)}{E(f^2)} \leq 0$$

$\Rightarrow$ Behauptung.

$\square$

**Satz 10.12** *$f_1, \ldots, f_n$ seien unabhängige Zufallsvariablen, $f_i \in \mathcal{L}(\Omega, \mathcal{A}, P)$, $i = 1, \ldots, n$. Dann ist $\prod_{1\leq i\leq n} f_i \in \mathcal{L}(\Omega, \mathcal{A}, P)$ und es gilt:*

$$E\left(\prod_{i=1}^n f_i\right) = \prod_{i=1}^n E(f_i).$$

Beweis: O.E. sei $n = 2$. Dann gilt:

$$\begin{aligned} &E(|f_1f_2|) \\ &= \int_\Omega |f_1f_2|\, dP \\ &=_{8.17} \int_{\mathbb{R}^2} |xy|\, (P*(f_1,f_2))(d(x,y)) \\ &= \int_{\mathbb{R}^2} |xy|\, ((P*f_1)\times(P*f_2))(d(x,y)) \\ &=_{\text{Fubini}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |x||y|\, (P*f_1)(dx)\right) (P*f_2)(dy) \\ &= \int_{\mathbb{R}} |x|\, (P*f_1)(dx) \int_{\mathbb{R}} |y|\, (P*f_2)(dy) \\ &=_{8.17} E(|f_1|)E(|f_2|). \end{aligned}$$

$\square$

**Definition 10.13** $f, g$ seien quadratintegrierbare Zufallsvariablen.

$$\begin{aligned} \operatorname{cov}(f,g) &:= E\left((f-E(f))(g-E(g))\right) \\ &= E(fg) - E(f)E(g) \end{aligned}$$

ist die Kovarianz von $f$ und $g$.

Ist zusätzlich $\sigma^2(f) > 0$, $\sigma^2(g) > 0$, so heißt

$$\varrho(f,g) := \frac{\operatorname{cov}(f,g)}{\sigma(f)\sigma(g)} \quad \in [0,1]$$

<u>Korrelationskoeffizient</u> von $f$ und $g$.
$f$ und $g$ heißen <u>positiv</u> bzw. <u>un–</u> bzw. <u>negativ korreliert</u>, falls

$$\varrho(f,g) > 0 \text{ bzw. } = 0 \text{ bzw. } < 0.$$

**Bemerkung** $\varrho(f,g) > 0$ $(< 0)$ bedeutet anschaulich, dass $f - E(f)$ und $g - E(g)$ die Tendenz besitzen, das selbe (unterschiedliche) Vorzeichen zu haben.
Ferner kann $\varrho(f,g)$ als Maß für den Grad der „linearen Abhängigkeit“ zwischen $f$ und $g$ angesehen werden (s. folgenden Satz).

**Satz 10.14** *$f, g$ seien Zufallsvariablen mit $0 < \sigma^2(f), \sigma^2(g) < \infty$. Dann gilt:*

1. *Falls $f, g$ unabhängig sind $\Rightarrow \varrho(f,g) = 0$.*
2. *Falls $\varrho(f,g) \in \{-1,1\} \Rightarrow \exists a, b \in \mathbb{R}$ :*
$$P(\{g = af + b\}) = 1.$$
3. *Die <u>mittlere quadratische Abweichung</u> $E((f-(af+b))^2)$ der linearen Approximation von $g$ durch $af + b$ ist genau dann minimal, wenn*
$$a^* = \frac{cov(f,g)}{\sigma^2(f)} \text{ und } b^* = E(g) - a^*E(f).$$
*In diesem Fall gilt:*
$$E((g - (a^*f + b^*))^2) = \left(1 - \varrho(f,g)^2\right)\sigma^2(g).$$

Beweis:

**„1.“** Folgt aus 10.12.

**„3.“** Es gilt:

$$\begin{aligned} & E((g - af - b)^2) \\ &= a^2E(f^2) + b^2 + 2abE(f) \\ &\qquad -2aE(fg) - 2bE(g) + E(g^2) \\ &=: p(a,b), \qquad a, b \in \mathbb{R}. \end{aligned}$$

$p$ ist ein Polynom zweiten Grades in den Variablen $a, b$ und besitzt genau ein Minimum für

$$a^* = \frac{\operatorname{cov}(f,g)}{\sigma^2(f)}, \; b^* = E(g) - a^*E(f)$$

(elementar mittels partieller Ableitungen).

**„2.“** Im Fall $\varrho(f,g) \in \{-1,1\}$ folgt aus 3. $E((g-(a^*f+b^*))^2)=0$, d.h.

$$P(\{|g-(a^*f+b^*)| \neq 0\}) = 0,$$

d.h.

$$P(\{g = a^*f+b^*\}) = 1$$

(s. Übungen).

□

**Bemerkung** $a^* = \text{cov}(f,g)/\sigma^2(f)$ ist ein geeignetes Mittel zur Vorhersage von $g$ aus $f$, sog. Regression von $f$ auf $g$, da nach 10.14, 3.,

$$\hat{g} := E(g) + a^*(f - E(f))$$

die beste lineare Approximation von $g$ durch $f$ darstellt.
$a^*$ heißt (einfacher) Regressionskoeffizient von $g$ auf $f$ und die Gerade

$$m(t) := a^*(t - E(f)) + E(g)$$

heißt Regressionsgerade von $g$ auf $f$. Der Fehler

$$g - \hat{g} = g - m(f)$$

bei dieser Approximation heißt Residuum.
Falls $E(g) = E(f) = 0$ und $\text{var}(f) = \text{var}(g) = 1$, so folgt $a^* = \text{cov}(f,g) = \varrho(f,g) \in [-1,1]$ und damit

$$\hat{g} = a^*f \Rightarrow |\hat{g}| = \varrho(f,g)||f| \leq |f|,$$

daher die Bezeichnung „Regression“ (Rückschritt).

**Satz 10.15** *$f_1, \ldots, f_n$ seien quadratintegrierbare, unkorrelierte Zufallsvariablen (also cov$(f_i, f_j) = 0$, $i \neq j$. Dann gilt:*

$$\sigma^2(f_1 + \cdots + f_n) = \sigma^2(f_1) + \cdots + \sigma^2(f_n).$$

Beweis:

$$\begin{aligned}
&\sigma^2(f_1 + \cdots + f_n)\\
&= E\left(((f_1+\cdots+f_n) - E(f_1+\cdots+f_n))^2\right)\\
&= E\left(((f_1 - E(f_1)) + \cdots + (f_n - E(f_n)))^2\right)\\
&= E\left(\sum_{1\leq i,j\leq n} (f_i - E(f_i))(f_j - E(f_j))\right)\\
&= \sum_{1\leq i,j\leq n} \text{cov}(f_i, f_j)\\
&= \sum_{i=1}^{n} \text{cov}(f_i, f_i)\\
&= \sum_{i=1}^{n} \sigma^2(f_i).
\end{aligned}$$

□

**Korollar 10.16** *$f_1, \dots, f_n$ seien unabhängige, quadratintegrierbare Zufallsvariablen. Dann gilt:*

$$\sigma^2(f_1 + \cdots + f_n) = \sigma^2(f_1) + \cdots + \sigma^2(f_n).$$

# 11 Gesetze der großen Zahlen

**Bemerkung** $f, f_1, f_2, \dots$ seien Zufallsvariablen über $(\Omega, \mathcal{A}, P)$, dann gilt $\{\lim_{n\in\mathbb{N}} f_n = f\} \in \mathcal{A}$, denn:

$$\{\lim_{n\in\mathbb{N}} f_n = f\} = \bigcap_{k\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \{|f_n - f| < 1/k\} \in \mathcal{A}.$$

**Definition 11.1**

$$\begin{aligned} & f_n \to_{n\in\mathbb{N}} f \ P\text{–f.s.} \\ :\Leftrightarrow \quad & P\left(\left\{\lim_{n\in\mathbb{N}} f_n = f\right\}\right) = 1, \end{aligned}$$

($f_n$ konvergiert $P$–fast sicher gegen $f$).

**Satz 11.2 (f.s. Eindeutigkeit des Grenzwertes)** *Es gelte $f_n \to_{n\in\mathbb{N}} f$ $P$–f.s., $f_n \to_{n\in\mathbb{N}} \tilde{f}$ $P$–f.s. $\Rightarrow P(\{f = \tilde{f}\}) = 1$, d.h. $f = \tilde{f}$ $P$–f.s.*

Beweis:

$$\begin{aligned} \{f = \tilde{f}\} \quad &\supset \quad \left\{\lim_{n\in\mathbb{N}} f_n = f\right\} \cap \left\{\lim_{n\in\mathbb{N}} f_n = \tilde{f}\right\} \\ \Rightarrow P(\{f \neq \tilde{f}\}) \quad &\leq \quad P\left(\left\{\lim_{n\in\mathbb{N}} f_n \neq f\right\}\right) \\ &\qquad + P\left(\left\{\lim_{n\in\mathbb{N}} f_n \neq \tilde{f}\right\}\right) \ = \ 0. \end{aligned}$$

□

**Satz 11.3** *$f_{i,n} \to_{n\in\mathbb{N}} f_i$ $P$–f.s., $i = 1, \dots, k$, $g : \mathbb{R}^k \to \mathbb{R}$ stetig*

$$\Rightarrow g(f_{1,n}, \dots, f_{k,n}) \to_{n\in\mathbb{N}} g(f_1, \dots, f_k) \ P\text{–f.s.}$$

Beweis: Nach Voraussetzung existieren $P$–Nullmengen $N_1, \dots, N_k \in \mathcal{A}$ mit $f_{i,n}(\omega) \to_{n\in\mathbb{N}} f_i(\omega)$ für alle $\omega \in N_i^c$, $i = 1, \dots, k$. Für alle $\omega \in (N_1 \cup \cdots \cup N_k)^c$ gilt also

$$(f_{1,n}(\omega), \dots, f_{k,n}(\omega)) \to_{n\in\mathbb{N}} (f_1(\omega), \dots, f_k(\omega))$$

und damit

$$g\left(f_{1,n}(\omega), \dots, f_{k,n}(\omega)\right) \to_{n\in\mathbb{N}} g\left(f_1(\omega), \dots, f_k(\omega)\right).$$

Da $P((N_1 \cup \cdots \cup N_k)^c) = 1$, folgt die Behauptung. □

**Beispiel 11.4** $f_n \to_{n\in\mathbb{N}} f$ $P$–f.s., $g_n \to_{n\in\mathbb{N}} g$ $P$–f.s. $\Rightarrow f_n + g_n \to_{n\in\mathbb{N}} f+g$ $P$–f.s.

**Definition 11.5**

$$\begin{aligned} & f_n \xrightarrow{P} f \\ & :\Leftrightarrow \quad \forall \varepsilon > 0 : \lim_{n\in\mathbb{N}} P(\{|f_n - f| > \varepsilon\}) = 0, \end{aligned}$$

($\underline{f_n \text{ konvergiert in Wahrscheinlichkeit gegen } f}$).

**Satz 11.6 (f.s. Eindeutigkeit des Grenzwertes)** *$f_n \xrightarrow{P} f$, $f_n \xrightarrow{P} \tilde{f} \Rightarrow f = \tilde{f}$ $P$–f.s.*

Beweis: Es gilt für $n \in \mathbb{N}$ und $\varepsilon > 0$

$$\{|f - \tilde{f}| > \varepsilon\} \subset \{|f_n - f| > \varepsilon/2\} \cup \{|f_n - \tilde{f}| > \varepsilon/2\}$$

und damit

$$\begin{aligned} & P(\{|f - \tilde{f}| > \varepsilon\}) \\ & \quad \leq \; P(\{|f_n - f| > \varepsilon/2\}) + P(\{|f_n - \tilde{f}| > \varepsilon/2\}) \end{aligned}$$

$$\Rightarrow P(\{|f - \tilde{f}| > \varepsilon\}) = 0 \text{ für beliebiges } \varepsilon > 0$$

$$\begin{aligned} \Rightarrow P(\{f \neq \tilde{f}\}) \;&=\; P\left(\bigcup_{k\in\mathbb{N}} \{|f - \tilde{f}| > 1/k\}\right) \\ &\leq\; \sum_{k\in\mathbb{N}} P(\{|f - \tilde{f}| > 1/k\}) \;=\; 0. \end{aligned}$$

□

**Satz 11.7** *$f_n \to_{n\in\mathbb{N}} f$ $P$–f.s. $\Rightarrow f_n \xrightarrow{P} f$.*

Beweis: O.b.d.A. sei $f \equiv 0$ (betrachte sonst $\tilde{f}_n := f_n - f$). Mit

$$K := \left\{ \lim_{n\in\mathbb{N}} f_n = 0 \right\} = \bigcap_{k\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \left\{ \sup_{n\geq m} |f_n| \leq 1/k \right\} \in \mathcal{A}$$

gilt:

$$
\begin{aligned}
& f_n \to 0 \ P\text{–f.s.} \\
\Leftrightarrow\ & P(K^c) = 0 \\
\Leftrightarrow\ & P\left( \bigcup_{k\in\mathbb{N}} \underbrace{\bigcap_{m\in\mathbb{N}} \left\{ \sup_{n\geq m} |f_n| > 1/k \right\}}_{\uparrow \text{ in } k} \right) = 0 \\
\Leftrightarrow\ & P\left( \underbrace{\bigcap_{m\in\mathbb{N}} \left\{ \sup_{n\geq m} |f_n| > 1/k \right\}}_{\downarrow \text{ in } m} \right) = 0 \text{ für alle } k \in \mathbb{N} \\
\Leftrightarrow\ & \lim_{m\in\mathbb{N}} P\left( \left\{ \sup_{n\geq m} |f_n| > 1/k \right\} \right) = 0 \\
\Rightarrow\ & \lim_{m\in\mathbb{N}} P(\{|f_m| > 1/k\} = 0 \text{ für alle } k \in \mathbb{N} \\
\Rightarrow\ & \text{ Behauptung.}
\end{aligned}
$$

□

Die Umkehrung von Satz 11.7 ist i.a. nicht richtig, wie folgendes Beispiel zeigt.

**Beispiel 11.8** Setze $(\Omega, \mathcal{A}, P) := ([0,1), \mathbb{B} \cap [0,1), \lambda/\mathbb{B} \cap [0,1))$, $f_1 := 1_{[0,1)}$, $f_2 := 1_{[0,1/2)}, f_3 := 1_{[1/2,1)}, f_4 := 1_{[0,1/3)}, f_5 := 1_{[1/3,2/3)}, \ldots$ Offenbar gilt $P(\{|f_n| > \varepsilon\}) \to_{n\in\mathbb{N}} 0$, $\varepsilon > 0$, aber $f_n(\omega)$ konvergiert für <u>kein</u> $\omega \in \Omega$ gegen 0.

**Satz 11.9 (Schw. Gesetz der großen Zahlen I)** *$f_1, \ldots, f_n$ seien identisch verteilte, quadratintegrierbare und unkorrelierte Zufallsvariablen, d.h. cov$(f_i, f_j)$ $= 0$ für $i \neq j$. Dann gilt mit $\mu := E(f_1)$, $\sigma^2 := \sigma^2(f_1)$ für $\varepsilon > 0$:*

$$
\begin{aligned}
P\left( \left\{ \left| \frac{1}{n} \sum_{i=1}^{n} f_i - \mu \right| \geq \varepsilon \right\} \right) \ & \leq \ \frac{\sigma^2}{n\varepsilon^2} \\
& \to \ 0, \text{ falls } n \to \infty.
\end{aligned}
$$

BEWEIS: Aus der Tschebyscheff–Ungleichung folgt:

$$\begin{aligned} P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^n f_i - \mu\right| \geq \varepsilon\right\}\right) \\ &= P\left(\left\{\left|\sum_{i=1}^n (f_i - \mu)\right| \geq n\varepsilon\right\}\right) \\ &\leq \frac{1}{n^2\varepsilon^2}\mathrm{var}\left(\sum_{i=1}^n (f_i - \mu)\right) \\ &=_{10.15} \frac{n\sigma^2}{n^2\varepsilon^2}. \end{aligned}$$

□

**Korollar 11.10 (Schw. G. d. großen Zahlen II)** $f_n$, $n \in \mathbb{N}$, *sei eine Folge unabhängiger, identisch verteilter und quadratintegrierbarer Zufallsvariablen. Dann gilt:*

$$\frac{1}{n}\sum_{i=1}^n f_i \xrightarrow{P} E(f_1).$$

Wie wir am Ende dieses Kapitels sehen werden, gilt sogar:

$$\frac{1}{n}\sum_{i=1}^n f_i \to_{n\in\mathbb{N}} E(f_1) \quad P\text{–f.s.}$$

Dies ist das Starke Gesetz der großen Zahlen. Die Gesetze der großen Zahlen decken sich völlig mit unserer Anschauung und unserer Erfahrung. Es wird im mathematischen Modell die Erfahrungstatsache bestätigt, dass bei einer großen Anzahl $n$ von unabhängigen Wiederholungen des gleichen Experimentes die relative Häufigkeit

$$h_n(A) = \frac{1}{n}\sum_{i=1}^n 1_A(f_i)$$

des Eintretens eines Ereignisses $A$ in der Nähe der Wahrscheinlichkeit $p := P(\{f_1 \in A\})$ liegt:
$f_1, f_2, \ldots$ seien unabhängig und identisch verteilt. Dann sind $1_A(f_1), 1_A(f_2), \ldots$ unabhängig und identisch verteilt mit $E(1_A(f_1)) = P(\{f_1 \in A\}) = p$ sowie $\sigma^2(1_A(f_1)) = p(1-p)$ und es gilt

$$\begin{aligned} P(\{|h_n(A) - p| \geq \varepsilon\}) \\ &= P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^n (1_A(f_i) - p)\right| \geq \varepsilon\right\}\right) \\ &\leq \frac{p(1-p)}{n\varepsilon^2} \\ &\leq \frac{1}{4n\varepsilon^2} \to_{n\in\mathbb{N}} 0, \end{aligned}$$

unabhängig von $p$. Beachte, dass $p(1-p) \leq 1/4$ für $p \in [0,1]$.

**Satz 11.11 (Kolmogoroffsche Ungleichung)** *Für unabhängige Zufallsvariablen $f_1, \dots, f_k$ mit $E(f_i) = 0$, $i = 1, \dots, k$, und $\varepsilon > 0$ gilt:*

$$P\left(\left\{\max_{1\le j\le k}\left|\sum_{i=1}^{j} f_i\right| \ge \varepsilon\right\}\right) \le \frac{1}{\varepsilon^2}\sum_{i=1}^{k}\sigma^2(f_i).$$

Beweis: O.E. sei $\sigma^2(f_i) < \infty$, $i = 1, \dots, k$. Setze $S_j := \sum_{i=1}^{j} f_i$ für $j = 1, \dots, k$ und

$$A_j := \{|S_1| < \varepsilon, \dots, |S_{j-1}| < \varepsilon, |S_j| \ge \varepsilon\}.$$

Die Ereignisse $A_1, \dots, A_k$ sind disjunkt, und für jedes $j$ sind die Zufallsvariablen $1_{A_j} S_j$ und $S_k - S_j$ unabhängig, da die erste nur von $f_1, \dots, f_j$ und die zweite nur von $f_{j+1}, \dots, f_k$ abhängt. Es folgt:

$$\begin{aligned}
&\sum_{j=1}^{k}\sigma^2(f_j)\\
&=_{10.16} \quad \sigma^2(S_k) = E(S_k^2)\\
&\ge \quad \sum_{j=1}^{k} E(1_{A_j} S_k^2)\\
&= \quad \sum_{j=1}^{k} E(1_{A_j}(S_j + (S_k - S_j))^2)\\
&\ge_{10.12} \quad \sum_{j=1}^{k}\left(E(1_{A_j} S_j^2) + 2E(1_{A_j} S_j)\underbrace{E(S_k - S_j)}_{=0}\right)\\
&= \quad \sum_{j=1}^{k}\int_{A_j} S_j^2\, dP\\
&\ge \quad \sum_{j=1}^{k}\varepsilon^2 P(A_j)\\
&= \quad \varepsilon^2 P\left(\bigcup_{j=1}^{k} A_j\right)\\
&= \quad \varepsilon^2 P\left(\left\{\max_{1\le j\le k}\left|\sum_{i=1}^{j} f_i\right| \ge \varepsilon\right\}\right).
\end{aligned}$$

□

**Satz 11.12** *$f_n$, $n \in \mathbb{N}$, seien unabhängige, quadratintegrierbare Zufallsvariablen. Falls die Reihen $\sum_{n\in\mathbb{N}} E(f_n)$ und $\sum_{n\in\mathbb{N}}\sigma^2(f_n)$ eigentlich konvergent sind, so existiert eine Zufallsvariable $S$ mit*

$$S_n := \sum_{i=1}^{n} f_i \to_{n\in\mathbb{N}} S \quad P\text{–f.s.}$$

Beweis:

1. O.E. sei $E(f_n) = 0$, $n \in \mathbb{N}$ (sonst Übergang zu $\sum_{i=1}^n f_i = \sum_{i=1}^n (f_i - E(f_i)) + \sum_{i=1}^n E(f_i)$).

2. Für $\varepsilon > 0$ und $m \in \mathbb{N}$ gilt;

$$\begin{aligned}
& P\left(\left\{\sup_{n>m} |S_n - S_m| > \varepsilon\right\}\right) \\
&= P\left(\bigcup_{k\in\mathbb{N}} \underbrace{\left\{\max_{m<n\leq m+k} |S_n - S_m| > \varepsilon\right\}}_{\uparrow \text{ in } k}\right) \\
&= \lim_{k\in\mathbb{N}} P\left(\left\{\max_{m<n\leq m+k} |S_n - S_m| > \varepsilon\right\}\right) \\
&\leq_{11.11} \limsup_{k\in\mathbb{N}} \frac{1}{\varepsilon^2} \sum_{n=m+1}^{m+k} \sigma^2(f_n) \\
&= \frac{1}{\varepsilon^2} \sum_{n>m} \sigma^2(f_n).
\end{aligned}$$

Nach Voraussetzung gilt $\lim_{m\in\mathbb{N}} \sum_{n>m} \sigma^2(f_n) = 0$, also

$$\begin{aligned}
& P\left(\left\{\inf_{m\in\mathbb{N}} \sup_{n>m} |S_n - S_m| > \varepsilon\right\}\right) \\
&\leq P\left(\bigcap_{m\in\mathbb{N}} \left\{\sup_{n>m} |S_n - S_m| > \varepsilon\right\}\right) \\
&\leq \liminf_{m\in\mathbb{N}} P\left(\left\{\sup_{n>m} |S_n - S_m| > \varepsilon\right\}\right) \\
&= 0.
\end{aligned}$$

Das Cauchy–Konvergenzkriterium für reelle Zahlenfolgen ergibt nun:

$$\begin{aligned}
& P\left(\{S_n,\ n \in \mathbb{N},\ \text{ist eigentlich konvergent}\}\right) \\
&= P\left(\left\{\forall r \in \mathbb{N}\, \exists m \in \mathbb{N} :\ \sup_{n>m} |S_n - S_m| < \frac{1}{r}\right\}\right) \\
&= P\left(\bigcap_{r\in\mathbb{N}} \left\{\inf_{m\in\mathbb{N}} \sup_{n>m} |S_n - S_m| < \frac{1}{r}\right\}\right) \\
&= 1 - P\left(\bigcup_{r\in\mathbb{N}} \left\{\inf_{m\in\mathbb{N}} \sup_{n>m} |S_n - S_m| \geq \frac{1}{r}\right\}\right) \\
&= 1.
\end{aligned}$$

□

**Korollar 11.13 (St. G. d. gr. Z. v. Kolmogoroff)** *Für jede Folge $f_n$, $n \in \mathbb{N}$, unabhängiger Zufallsvariablen mit $\sum_{n\in\mathbb{N}} \sigma^2(f_n)/n^2 < \infty$ gilt:*

$$\frac{1}{n}\sum_{i=1}^{n}(f_i - E(f_i)) \to_{n\in\mathbb{N}} 0 \quad P\text{–}f.s.$$

Beweis: O.E. gelte $E(f_n) = 0$, $n \in \mathbb{N}$. Nach 11.12 existiert eine Zufallsvariable $T$ mit

$$T_n := \sum_{i=1}^{n} \frac{f_i}{i} \to_{n\in\mathbb{N}} T \quad P\text{–f.s.}$$

Beachte: $\mathrm{var}(f_i/i) = \mathrm{var}(f_i)/i^2$. Damit folgt ($T_0 := 0$):

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} f_i \\
&= \frac{1}{n}\sum_{i=1}^{n} i(T_i - T_{i-1}) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} iT_i - \sum_{i=1}^{n}(i+1)T_i + (n+1)T_n\right) \\
&= \frac{n+1}{n}T_n - \frac{1}{n}\sum_{i=1}^{n} T_i \\
&\to_{n\in\mathbb{N}} \; T - T \;=\; 0 \quad P\text{–f.s.}
\end{aligned}$$

□

**Korollar 11.14** *$f_n$, $n \in \mathbb{N}$, seien unabhängige, identisch verteilte Zufallsvariablen. Falls dann $E(f_1^2) < \infty$, so gilt:*

$$\frac{1}{n}\sum_{i=1}^{n} f_i \to_{n\in\mathbb{N}} E(f_1) \quad P\text{–}f.s.$$

Beweis: Es gilt

$$\sum_{n\in\mathbb{N}} \frac{\sigma^2(f_n)}{n^2} = \sigma^2(f_1)\sum_{n\in\mathbb{N}} \frac{1}{n^2} < \infty.$$

Damit folgt die Behauptung aus 11.13. Beachte:

$$\frac{1}{n}\sum_{i=1}^{n}(f_i - E(f_i)) = \left(\frac{1}{n}\sum_{i=1}^{n} f_i\right) - E(f_1).$$

□

**Satz 11.15** *$f_n$, $n \in \mathbb{N}$, und $f'_n$, $n \in \mathbb{N}$, seien Folgen von Zufallsvariablen mit $\sum_{n\in\mathbb{N}} P(\{f_n \neq f'_n\}) < \infty$; $f$ sei eine Zufallsvariable. Dann gilt:*

$$\frac{1}{n}\sum_{i=1}^{n} f_i \to_{n\in\mathbb{N}} f \quad P\text{–}f.s.$$
$$\Leftrightarrow \quad \frac{1}{n}\sum_{i=1}^{n} f'_i \to_{n\in\mathbb{N}} f \quad P\text{–}f.s.$$

Beweis: Nach dem Lemma von Borel–Cantelli gilt mit $A_n := \{f_n \neq f'_n\}$:

$$\begin{aligned} P(N_1) &:= P(\{f_n \neq f'_n \text{ für unendlich viele } n \in \mathbb{N}\}) \\ &= P(\limsup_{n\in\mathbb{N}} A_n) \ = \ 0. \end{aligned}$$

Gilt nun $n^{-1}\sum_{i=1}^{n} f_i \to_{n\in\mathbb{N}} f$ $P$–f.s., so existiert $N_2 \in \mathcal{A}$ mit $P(N_2) = 0$ und

$$\frac{1}{n}\sum_{i=1}^{n} f_i(\omega) \to_{n\in\mathbb{N}} f(\omega) \text{ für alle } \omega \in N_2^c.$$

Damit gilt für alle $\omega \in N_1^c \cap N_2^c$:

$$\lim_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} f'_i(\omega) = \lim_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^{n} f_i(\omega) = f(\omega).$$

Da $P(N_1^c \cap N_2^c) = 1$, folgt die Behauptung. □

**Satz 11.16 (St. G. d. gr. Z. von Khinchine)** *$f_n$, $n \in \mathbb{N}$, sei eine Folge unabhängiger, identisch verteilter Zufallsvariablen.*

1. *Falls $f_1$ integrierbar ist, so ist $E(f_n) = E(f_1) =: \mu \in \mathbb{R}$ und es gilt:*

$$\frac{1}{n}\sum_{i=1}^{n} f_i \to_{n\in\mathbb{N}} \mu \quad P\text{–}f.s.$$

2. *Falls es ein $c \in \mathbb{R}$ gibt mit der Eigenschaft:*

$$\frac{1}{n}\sum_{i=1}^{n} f_i \to_{n\in\mathbb{N}} c \quad P\text{–}f.s.,$$

*so ist $f_1$ integrierbar, und es gilt $c = E(f_1)$.*

Beweis:

**„1.“** Die durch

$$\begin{aligned} f_n'(\omega) &:= \begin{cases} f_n(\omega), & \text{falls } |f_n(\omega)| \le n \\ 0 & \text{sonst} \end{cases} \\ &= f_n(\omega) 1_{[-n,n]}(f_n(\omega)), \quad n \in \mathbb{N}, \end{aligned}$$

definierten Zufallsvariablen sind unabhängig, und es gilt

$$\begin{aligned} &\sum_{n\in\mathbb{N}} \frac{\sigma^2(f_n')}{n^2} \\ &= \sum_{n\in\mathbb{N}} \frac{E(f_n'^2) - E(f_n')^2}{n^2} \\ &\le \sum_{n\in\mathbb{N}} \frac{E(f_n'^2)}{n^2} \\ &\le \sum_{n\in\mathbb{N}} \frac{1}{n^2} \sum_{k=1}^{n} k^2 P(\{k-1 < |f_1| \le k\}) \\ &= \sum_{k\in\mathbb{N}} kP(\{k-1 < |f_1| \le k\}) \left( k \sum_{n\ge k} \frac{1}{n^2} \right), \end{aligned}$$

wobei

$$\begin{aligned} \sum_{n\ge k} \frac{1}{n^2} &\le 2 \sum_{n\ge k} \frac{1}{n(n+1)} \\ &= 2 \sum_{n\ge k} \left( \frac{1}{n} - \frac{1}{n+1} \right) = \frac{2}{k}, \end{aligned}$$

d.h. insgesamt

$$\begin{aligned} \sum_{n\in\mathbb{N}} \frac{\sigma^2(f_n')}{n^2} &\le 2 \sum_{k\in\mathbb{N}} kP(\{k-1 < |f_1| \le k\}) \\ &\le 2(E(|f_1|) + 1) < \infty, \end{aligned}$$

s. Übungen.

Anwendung von 11.13 liefert:

$$\frac{1}{n} \sum_{i=1}^{n} (f_i' - E(f_i')) \to_{n\in\mathbb{N}} 0 \quad P\text{–f.s.}$$

Nach dem Satz von der dominierten Konvergenz gilt ferner:

$$\begin{aligned} E(f_n') &= E(f_n 1_{[-n,n]}(f_n)) \\ &= E(f_1 1_{[-n,n]}(f_1)) \\ &\to_{n\in\mathbb{N}} E(f_1) = \mu, \end{aligned}$$

also auch $n^{-1} \sum_{i=1}^{n} E(f_i') \to_{n\in\mathbb{N}} \mu$ und somit $n^{-1} \sum_{i=1}^{n} f_i' \to_{n\in\mathbb{N}} \mu$ $P$–f.s.
Zusammen mit

$$\begin{aligned} \sum_{n\in\mathbb{N}} P(\{f_n \ne f_n'\}) &= \sum_{n\in\mathbb{N}} P(\{|f_n| > n\}) \\ &= \sum_{n\in\mathbb{N}} P(\{|f_1| > n\}) \\ &< \infty, \end{aligned}$$

(da $f_1$ integrierbar ist, s. Übungen) folgt die Behauptung 1. aus 11.15.

**„2.“** Mit $S_n := \sum_{i=1}^n f_i$ gilt:

$$\begin{aligned} &\frac{S_n}{n} \to_{n\in\mathbb{N}} c \quad P\text{–f.s.} \\ &\Rightarrow \quad \frac{f_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \to_{n\in\mathbb{N}} 0 \quad P\text{–f.s.} \\ &\Rightarrow \quad P\text{–f.s.: } \left|\frac{f_n}{n}\right| > 1 \text{ für höchstens endl. viele } n \in \mathbb{N}, \end{aligned}$$

d.h. in diesem Fall existiert $n_0 \in \mathbb{N}$ so, dass $|f_n/n| \leq 1$ für alle $n \geq n_0$. Also gilt mit $A_n := \{|f_n/n| > 1\}$:

$$P\left(\limsup_{n\in\mathbb{N}} A_n\right) = 0.$$

Da $f_n$, $n \in \mathbb{N}$, eine Folge unabhängiger Zufallsvariablen ist, sind die $A_n$, $n \in \mathbb{N}$, unabhängige Ereignisse, und somit folgt aus dem Lemma von Borel–Cantelli:

$$\infty > \sum_{n\in\mathbb{N}} P(A_n) = \sum_{n\in\mathbb{N}} P(\{|f_1| > n\})$$

und damit, dass $f_1$ integrierbar ist (s. Übungen).

□

**Korollar 11.17** *Bei einer Folge unabhängiger Wiederholungen $f_1, f_2, \ldots$ eines Experimentes konvergiert die relative Häufigkeit $h_n(A) = n^{-1}\sum_{i=1}^n 1_A(f_i)$ des Eintritts eines Ereignisses $A$ fast sicher gegen die Wahrscheinlichkeit $P(\{f_1 \in A\})$.*

**Beispiel 11.18** $f_n$, $n \in \mathbb{N}$, seien unabhängige, identisch verteilte Zufallsvariablen mit der Verteilungsfunktion $F$. Dann gilt für die <u>empirische Verteilungsfunktion</u> oder <u>Stichproben-Verteilungsfunktion</u> zur Stichprobe $f_1, \ldots, f_n$:

$$F_n(t) := \frac{1}{n}\sum_{i=1}^n 1_{(-\infty,t]}(f_i) \to_{n\in\mathbb{N}} F(t) \quad P\text{–f.s.}, \ t \in \mathbb{R}.$$

Beweis: Setze $\tilde{f}_n := 1_{(-\infty,t]}(f_n)$, $n \in \mathbb{N}$. Dann sind $\tilde{f}_n$, $n \in \mathbb{N}$, unabhängig und identisch verteilt mit

$$E(\tilde{f}_n) = E\left(1_{(-\infty,t]}(f_1)\right) = P(\{f_1 \leq t\}) = F(t).$$

Damit folgt aus 11.16:

$$F_n(t) = \frac{1}{n}\sum_{i=1}^n \tilde{f}_i \to_{n\in\mathbb{N}} E(\tilde{f}_1) = F(t) \quad P\text{–f.s.}$$

$\square$

Es gilt sogar die folgende Verschärfung der vorausgegangenen Aussage, die einen Hauptsatz der Stochastik darstellt.

**Satz 11.19 (Glivenko–Cantelli)** *$f_n$, $n \in \mathbb{N}$, seien unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F$. Dann gilt:*

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \to_{n\in\mathbb{N}} 0 \quad P\text{–}f.s.,$$

*d.h. die empirische Verteilungsfunktion konvergiert mit Wahrscheinlichkeit 1 <u>gleichmäßig über $\mathbb{R}$</u> gegen die zugrunde liegende Verteilungsfunktion.*

Beweis: Die $\mathcal{A}$–Messbarkeit von $\sup_{t\in\mathbb{R}} |F_n(t) - F(t)|$ folgt aus der rechtsseitigen Stetigkeit von Verteilungsfunktionen:

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| = \sup_{t\in\mathbb{Q}} |F_n(t) - F(t)|.$$

Setze nun $t_{j,k} := F^{-1}(j/k)$, $j = 1, \ldots, k-1$, $k \in \mathbb{N}$. Dann folgt:

$$F_n(t_{j,k}) \to_{n\in\mathbb{N}} F(t_{j,k}) \quad P\text{–f.s.},$$

sowie

$$F_n(t_{j,k} - 0) = \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,t_{j,k})}(f_i) \to_{n\in\mathbb{N}} F(t_{j,k} - 0)\ P\text{–f.s.},$$

wobei $F(t_{j,k} - 0) = \lim_{\varepsilon\downarrow 0} F(t_{j,k} - \varepsilon) = P(\{f_1 < t_{j,k}\})$.
Damit gilt (mit $F(t_{j,k} + 0) := F(t_{j,k})$):

$$\sup_{j=1,\ldots,k-1} |F_n(t_{j,k} \pm 0) - F(t_{j,k} \pm 0)| \to_{n\in\mathbb{N}} 0 \quad P\text{–f.s.}$$

Sei nun $t_{j,k} < t < t_{j+1,k}$. Dann gilt:

$$\begin{array}{rcccl} F(t_{j,k}) & \le & F(t) & \le & F(t_{j+1,k} - 0), \\ F_n(t_{j,k}) & \le & F_n(t) & \le & F_n(t_{j+1,k} - 0) \end{array}$$

sowie

$$0 \le \underbrace{F(t_{j+1,k} - 0)}_{\le \frac{j+1}{k}} - \underbrace{F(t_{j,k})}_{\ge \frac{j}{k}} \le \frac{1}{k}.$$

Es folgt:

$$\begin{aligned} F_n(t) - F(t) &\le F_n(t_{j+1,k} - 0) - F(t_{j,k}) \\ &\le F_n(t_{j+1,k} - 0) - F(t_{j+1,k} - 0) + \frac{1}{k} \end{aligned}$$

und

$$\begin{aligned} F_n(t) - F(t) &\ge F_n(t_{j,k}) - F(t_{j+1,k} - 0) \\ &\ge F_n(t_{j,k}) - F(t_{j,k}) - \frac{1}{k}. \end{aligned}$$

Insgesamt erhalten wir:

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \\
&\le \sup_{1\le j\le k} |F_n(t_{j,k} \pm 0) - F(t_{j,k} \pm 0)| + \frac{1}{k} \\
&\quad + \sup_{t<t_{1,k}} |F_n(t) - F(t)| + \sup_{t>t_{k-1,k}} |F_n(t) - F(t)|.
\end{aligned}$$

Ferner gilt:

$$\begin{aligned}
&\limsup_{n\in\mathbb{N}} \sup_{t<t_{1,k}} |F_n(t) - F(t)| \\
&\le \limsup_{n\in\mathbb{N}} (F_n(t_{1,k} - 0) + F(t_{1,k} - 0)) \\
&\le \frac{2}{k} \quad P\text{–f.s.}
\end{aligned}$$

sowie

$$\begin{aligned}
&\limsup_{n\in\mathbb{N}} \sup_{t>t_{k-1,k}} |F_n(t) - F(t)| \\
&\le \limsup_{n\in\mathbb{N}} \left( \sup_{t>t_{k-1,k}} |F_n(t) - 1| + \sup_{t>t_{k-1,k}} |1 - F(t)| \right) \\
&\le \limsup_{n\in\mathbb{N}} (1 - F_n(t_{k-1,k}) + 1 - F(t_{k-1,k})) \\
&= 2 \underbrace{\left( 1 - \underbrace{F(t_{k-1,k})}_{\ge \frac{k-1}{k}} \right)}_{\le \frac{1}{k}} \quad P\text{–f.s.}
\end{aligned}$$

Da $k \in \mathbb{N}$ beliebig war, folgt insgesamt die Behauptung. □

**Beispiel 11.20** $f_n$, $n \in \mathbb{N}$, sei eine Folge unabhängiger, identisch verteilter und quadratintegrierbarer Zufallsvariablen. Dann gilt für das Stichprobenmittel:

$$\hat{\mu}_n := \frac{1}{n} \sum_{i=1}^{n} f_i \to_{n\in\mathbb{N}} E(f_1) \quad P\text{–f.s.}$$

sowie für die Stichprobenvarianz:

$$\begin{aligned}
\hat{\sigma}_n^2 \quad &:= \quad \frac{1}{n} \sum_{i=1}^{n} (f_i - \hat{\mu}_n)^2 \\
&= \quad \frac{1}{n} \sum_{i=1}^{n} f_i^2 - \left( \frac{1}{n} \sum_{i=1}^{n} f_i \right)^2 \\
&\to_{n\in\mathbb{N}} \quad E(f_1^2) - E(f_1)^2 \; = \; \sigma^2(f_1) \quad P\text{–f.s.}
\end{aligned}$$

Ist ferner $(f_n, g_n)$, $n \in \mathbb{N}$, eine Folge unabhängiger, identisch verteilter Zufallsvektoren und $f_1, g_1$ quadratintegrierbar mit $0 < \sigma^2(f_1), \sigma^2(g_1)$, so gilt für die Stichprobenkovarianz

$$\begin{aligned}
\widehat{\text{cov}}_n &:= \frac{1}{n}\sum_{i=1}^n f_i g_i - \left(\frac{1}{n}\sum_{i=1}^n f_i\right)\left(\frac{1}{n}\sum_{i=1}^n g_i\right) \\
&\to_{n\in\mathbb{N}} E(f_1 g_1) - E(f_1)E(g_1) \quad P\text{–f.s.} \\
&= \text{cov}(f_1, g_1).
\end{aligned}$$

Damit gilt auch für den Stichprobenkorrelationskoeffizienten

$$\begin{aligned}
\hat{\varrho}_n &:= \frac{\widehat{\text{cov}}_n}{\hat{\sigma}_n(f)\hat{\sigma}_n(g)} \\
&= \frac{\left(\frac{1}{n}\sum_{i=1}^n f_i g_i\right) - \left(\frac{1}{n}\sum_{i=1}^n f_i\right)\left(\frac{1}{n}\sum_{i=1}^n g_i\right)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^n (f_i - \hat{\mu}_n)^2\right)\left(\frac{1}{n}\sum_{i=1}^n (g_i - \hat{\nu}_n)^2\right)}} \\
&\to_{n\in\mathbb{N}} \frac{\text{cov}(f_1, g_1)}{\sigma(f_1)\sigma(g_1)} \quad P\text{–f.s.} \\
&= \varrho(f_1, g_1)
\end{aligned}$$

wobei $\hat{\nu}_n := n^{-1}\sum_{i=1}^n g_i$ das Stichprobenmittel von $g_1, \dots, g_n$ ist.
Somit gilt für den (einfachen) Stichprobenregressionskoeffizienten

$$\begin{aligned}
\hat{a}_n &:= \frac{\widehat{\text{cov}}_n}{\hat{\sigma}_n^2(f_1)} \\
&= \frac{\left(\frac{1}{n}\sum_{i=1}^n f_i g_i\right) - \left(\frac{1}{n}\sum_{i=1}^n f_i\right)\left(\frac{1}{n}\sum_{i=1}^n g_i\right)}{\frac{1}{n}\sum_{i=1}^n f_i^2 - \left(\frac{1}{n}\sum_{i=1}^n f_i\right)^2} \\
&\to_{n\in\mathbb{N}} \frac{\text{cov}(f_1, g_1)}{\sigma^2(f_1)} \quad P\text{–f.s.}
\end{aligned}$$

Die oben aufgeführten Schätzer konvergieren also mit Wahrscheinlichkeit 1 gegen den jeweils zu schätzenden Wert, d.h. sie sind sog. (stark) konsistente Schätzerfolgen.
Ferner können wir mit obigen Schätzern nahe liegend eine Schätzung der Regressionsgeraden von $g_1$ auf $f_1$ definieren:

$$\begin{aligned}
\hat{m}_n(t) &:= \hat{a}_n(t - \hat{\mu}_n) + \hat{\nu}_n \\
&=: \hat{a}_n t + \hat{b}_n \\
&\to_{n\in\mathbb{N}} a^*(t - E(f_1)) + E(g_1) \quad P\text{–f.s.}, \ t \in \mathbb{R}.
\end{aligned}$$

Die Gerade $\hat{m}_n(t)$ erhält man auch, wenn eine Gerade $at + b$ so gewählt wird, dass die Summe der vertikalen Abstände der Datenpunkte $(f_i, g_i)$, $i = 1, \dots, n$, von der Geraden minimal wird, d.h.[5]

$$\sum_{i=1}^n (g_i - \hat{a}_n f_i - \hat{b}_n)^2 = \min_{a,b} \sum_{i=1}^n (g_i - a f_i - b)^2.$$

Dies ist die Methode der kleinsten Quadrate, die auf Gauss zurück geht und zunächst vorwiegend in der (Fehler–) Ausgleichsrechnung Verwendung fand.

[5] S. etwa Abschnitt 13.4 in Krengel, U. (2002). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 6. Auflage. Vieweg, Braunschweig.

**Bemerkung 11.21** Es gilt:

$$\begin{aligned} E(\hat{\sigma}_n^2) &= E\left(\frac{1}{n}\sum_{i=1}^n \left(f_i - \frac{1}{n}\sum_{j=1}^n f_j\right)^2\right) \\ &= E\left(\left(f_1 - \frac{1}{n}\sum_{j=1}^n f_j\right)^2\right) \\ &= \frac{n-1}{n}\sigma^2(f_1), \end{aligned}$$

d.h. $\hat{\sigma}_n^2$ ist kein erwartungstreuer Schätzer, wohl aber

$$\tilde{\sigma}_n^2 := \frac{n}{n-1}\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(f_i - \frac{1}{n}\sum_{j=1}^n f_j\right)^2.$$

**Beispiel 11.22 (Die Monte–Carlo Methode)** Gesetze der großen Zahlen können auch zur approximativen Berechnung von Integralen eingesetzt werden:
Problem: $g : [0,1] \to \mathbb{R}$ sei eine ($\lambda_1$–) integrierbare Funktion; bestimme $\int_{[0,1]} g(x)\,\lambda_1(dx)$.
Dies ist häufig praktisch kaum möglich. Verschaffen wir uns nun eine Stichprobe $f_1, \dots, f_n$ unabhängiger und auf dem Intervall $[0,1]$ gleichverteilter Zufallsvariablen (d.h. $P * f_i = P * f_1 = \lambda_1/[0,1]$), so gilt:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n g(f_i) &\to_{n\in\mathbb{N}} E(g(f_1)) \quad P\text{–f.s.} \\ &=_{8.17} \int_{[0,1]} g(x)\,\lambda_1(dx). \end{aligned}$$

Sind also $x_1, \dots, x_n$ beobachtete Werte von $f_1, \dots, f_n$, d.h. $x_i = f_i(\omega)$, $i = 1, \dots, n$, so gilt:

$$\frac{g(x_1) + \cdots + g(x_n)}{n} \approx \int_{[0,1]} g(x)\,dx.$$

$x_1, \dots, x_n$ heißen auch Zufallszahlen. Zufallszahlen werden in der Praxis üblicherweise vom Computer nach gewissen Algorithmen erzeugt. Da diese somit aber nicht „wirklich“ zufällig sind, spricht man in diesem Fall von Pseudozufallszahlen. Sie sollten sich wie „echte“ Zufallszahlen verhalten, tun das aber nicht immer. . . [6]

# 12 Der Zentrale Grenzwertsatz

$f_n$, $n \in \mathbb{N}$, sei eine Folge unabhängiger, identisch verteilter und quadratintegrierbarer Zufallsvariablen. Setze $\mu := E(f_1)$. Wir hatten in Kapitel 11 gesehen, dass

$$\frac{1}{n}\sum_{i=1}^n f_i \to_{n\in\mathbb{N}} \mu \quad P\text{–f.s.}$$

[6] S. Bemerkung (3.43) in Georgii, H.-O. (2002). *Stochastik.* De Gruyter, Berlin.

bzw.
$$P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f_i - \mu\right| \geq \varepsilon\right\}\right) \to_{n\in\mathbb{N}} 0$$
für alle $\varepsilon > 0$. Wir können nun fragen: Falls wir das feste $\varepsilon$ durch eine Folge $\varepsilon_n > 0$, $n \in \mathbb{N}$, mit $\varepsilon_n \to_{n\in\mathbb{N}} 0$ ersetzen, wie schnell darf dann $\varepsilon_n$ gegen 0 konvergieren, so dass
$$P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f_i - \mu\right| \geq \varepsilon_n\right\}\right) \to_{n\in\mathbb{N}} c \in (0,1)?$$
Wir werden sehen, dass dies für $\varepsilon_n \sim 1/\sqrt{n}$ der Fall ist; genauer wird für $t > 0$ gelten:
$$P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f_i - \mu\right| \geq \frac{\sigma t}{\sqrt{n}}\right\}\right) \to_{n\in\mathbb{N}} 2(1-\Phi(t)),$$
wobei
$$\Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx$$
und $\sigma^2 := \sigma^2(f_1)$. Dies wird eine unmittelbare Folgerung aus dem Zentralen Grenzwertsatz sein, den wir in diesem Abschnitt beweisen werden.

**Satz 12.1** *$f, g$ seien unabhängige Zufallsvariablen mit Verteilungsfunktion $F$ bzw. $G$. Dann besitzt $f + g$ die Verteilungsfunktion*
$$\begin{aligned}
(F * G)(t) &= \int G(t-s)\,(P * f)(ds) \\
\Big( &= \int G(t-s)\,F(ds) \Big) \\
&= \int F(t-s)\,(P * g)(ds) \\
\Big( &= \int F(t-s)\,G(ds) \Big).
\end{aligned}$$
*$F * G$ heißt Faltung von $F$ und $G$.*

Beweis: Es gilt:

$$\begin{aligned}
&(F * G)(t) \\
&\quad = \quad P(\{f + g \le t\}) \\
&\quad = \quad \int_\Omega 1_{(-\infty,t]}(f+g)\, dP \\
&\quad =_{8.17} \quad \int_{\mathbb{R}^2} 1_{(-\infty,t]}(r+s)\, (P * (f,g))(d(r,s)) \\
&\quad = \quad \int_{\mathbb{R}^2} 1_{(-\infty,t]}(r+s)\, ((P * f) \times (P * g))(d(r,s)) \\
&\quad =_{\text{Fubini}} \quad \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{(-\infty,t]}(r+s)(P * g)(ds)(P * f)(dr) \\
&\quad = \quad \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{(-\infty,t-r]}(s)(P * g)(ds)(P * f)(dr) \\
&\quad = \quad \int_{\mathbb{R}} G(t-r)\, (P * f)(dr).
\end{aligned}$$

□

**Satz 12.2** *$f, g$ seien unabhängige Zufallsvariablen mit Dichte $h_1$ bzw. $h_2$. Dann besitzt $f + g$ die Dichte*

$$(h_1 * h_2)(t) := \int_{\mathbb{R}} h_1(t-s)h_2(s)\, ds = \int_{\mathbb{R}} h_2(t-s)h_1(s)\, ds.$$

Beweis: $F$ bzw. $G$ sei die Verteilungsfunktion von $f$ bzw. $g$. Nach 12.1 besitzt $f + g$ die Verteilungsfunktion

$$\begin{aligned}
&(F * G)(t) \\
&\quad = \quad \int G(t-s)\, F(ds) \\
&\quad =_{8.23} \quad \int_{\mathbb{R}} G(t-s)h_1(s)\, ds \\
&\quad = \quad \int_{\mathbb{R}} \left( \int_{(-\infty,t-s]} h_2(r)\, dr \right) h_1(s)\, ds \\
&\quad =_{\text{Subst.}} \quad \int_{\mathbb{R}} \left( \int_{(-\infty,t]} h_2(r-s)\, dr \right) h_1(s)\, ds \\
&\quad =_{\text{Fubini}} \quad \int_{(-\infty,t]} \left( \int_{\mathbb{R}} h_2(r-s)h_1(s)\, ds \right) dr \\
&\quad =_{\text{Def.}} \quad \int_{(-\infty,t]} (h_1 * h_2)(r)\, dr.
\end{aligned}$$

Nach dem Maßerweiterungssatz 6.16, 6.19, 9.6 folgt hieraus die Behauptung.

□

**Definition 12.3** Das Wahrscheinlichkeitsmaß auf $\mathbb{B}_1$ mit der Dichte[7]

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R},$$

heißt Standardnormalverteilung, i.Z. $N(0,1)$. Die Verteilungsfunktion von $N(0,1)$ bezeichnen wir mit $\Phi$, d.h. $\Phi(x) = \int_{(-\infty,x]} \varphi(y)\, dy$.

**Bemerkung 12.4** Die Zufallsvariable $f$ sei nach $N(0,1)$ verteilt. Dann gilt:

1. $E(f) = 0$,
2. $\sigma^2(f) = 1$.

Beweis:

**„1.“** S. Übungen (beachte, dass $\varphi(x) = \varphi(-x)$, $x \in \mathbb{R}$.

**„2.“** Es gilt:

$$\begin{aligned}
&\int_{-\infty}^{\infty} x^2 \varphi(x)\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left( x \exp\left(-\frac{x^2}{2}\right)\right) dx \\
&=_{\text{part. Int.}} \frac{1}{\sqrt{2\pi}} \left[ x \left( -\exp\left(-\frac{x^2}{2}\right)\right)\right]_{-\infty}^{\infty} \\
&\qquad - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 1 \left( -\exp\left(-\frac{x^2}{2}\right)\right) dx \\
&= \int_{-\infty}^{\infty} \varphi(x)\, dx \\
&= 1.
\end{aligned}$$

□

**Satz 12.5** *Die Zufallsvariable $f$ sei $N(0,1)$–verteilt; $\sigma > 0$, $\mu \in \mathbb{R}$. Dann besitzt die Zufallsvariable $g := \sigma f + \mu$ die Dichte*

$$\begin{aligned}
\varphi_{(\mu,\sigma^2)}(x) &:= \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.
\end{aligned}$$

*Die zugehörige Verteilung heißt Normalverteilung mit Mittelwert $\mu$ und Varianz $\sigma^2$, i.Z. $N(\mu,\sigma^2)$,*
*(denn $E(g) = E(\sigma f + \mu) = \mu$, $\sigma^2(g) = E((g-\mu)^2) = E((\sigma f)^2) = \sigma^2 E(f^2) = \sigma^2$).*

[7]S. etwa Satz 19.1. in Bandelow, C. (1989). *Einführung in die Wahrscheinlichkeitstheorie.* BI, Mannheim.

BEWEIS: S. Übungen. □

**Satz 12.6 (Faltungsth. der Normalverteilung)** *$f_1, \dots, f_n$ seien unabhängige Zufallsvariablen mit $P * f_i = N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$. Dann gilt:*

$$P * (f_1 + \dots + f_n) = N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2).$$

BEWEIS: Offenbar genügt es, den Fall $n = 2$ zu betrachten. Zunächst gilt für $x, a, b \in \mathbb{R}$ und $\sigma^2, \tau^2 > 0$:

$$\begin{aligned} & \frac{(x-a)^2}{\sigma^2} + \frac{(x-b)^2}{\tau^2} \\ = \; & \frac{(x-c)^2}{\varrho^2} + \frac{(a-b)^2}{\sigma^2 + \tau^2} \end{aligned} \tag{2}$$

mit

$$c := \frac{a\tau^2 + b\sigma^2}{\sigma^2 + \tau^2}, \quad \varrho^2 := \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}.$$

Ferner besitzt $f_1 + f_2$ nach 12.2 die Dichte

$$\begin{aligned} h(y) &= \int_{-\infty}^{\infty} \varphi_{(\mu_1, \sigma_1^2)}(y - x) \varphi_{(\mu_2, \sigma_2^2)}(x) \, dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{(y - x - \mu_1)^2}{2\sigma_1^2}\right) \\ &\qquad \times \exp\left(-\frac{(x - \mu_2)^2}{2\sigma_2^2}\right) dx \\ &=_{(2)} \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-c)^2}{2\varrho^2}\right) \\ &\qquad \times \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) dx \\ &= \text{const} \, \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \\ &= \varphi_{(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)}(y). \end{aligned}$$

□

**Definition 12.7** $f, f_n$, $n \in \mathbb{N}$, seien Zufallsvariablen über $(\Omega, \mathcal{A}, P)$ mit Verteilungsfunktionen $F, F_n$, $n \in \mathbb{N}$.
$f_n$, $n \in \mathbb{N}$, heißt <u>in Verteilung</u> oder <u>schwach</u> konvergent gegen $f$, falls

$$F_n(t) \to_{n \in \mathbb{N}} F(t)$$

für alle <u>Stetigkeitsstellen</u> von $F$, d.h.

$$P(\{f_n \le t\}) \to_{n \in \mathbb{N}} P(\{f \le t\}),$$

falls $F(t) = P(\{f \le t\})$ in $t$ stetig ist, i.Z.

$$f_n \to_D f \quad \text{(in distribution)}.$$

**Beispiel 12.8** Es gelte $P * f_n = \delta_{1/n}$, $n \in \mathbb{N}$, wobei $\delta_{x_0}$ das Dirac–Maß (Ein–Punkt–Maß) im Punkt $x_0$ bezeichne, d.h. $\delta_{x_0}(B) = 1_B(x_0) = 1$, falls $x_0 \in B$, und 0 sonst, $B \in \mathbb{B}$. Offenbar gilt:

$$F_n(t) = \begin{cases} 1, & t \geq 1/n \\ 0, & t < 1/n. \end{cases}$$

Dann gilt:

$$f_n \to_D f$$

mit Verteilungsfunktion

$$F(t) = 1_{[0,\infty)}(t) = F_{\delta_0},$$

denn offenbar gilt

$$F_n(t) \to_{n\in\mathbb{N}} F(t),\ t \neq 0,\ \text{aber } 0 = F_n(0),\ F(0) = 1.$$

Die Verteilungskonvergenz einer Folge $f_n$, $n \in \mathbb{N}$, ist eine Aussage über die Verteilungen von $f_n$. Sie ist daher von anderem Charakter als die fast sichere Konvergenz oder die Konvergenz in Wahrscheinlichkeit, da sich Zufallsvariablen mit identischer Verteilung beliebig unterscheiden können.
Tatsächlich ist sie die schwächste der drei Konvergenzarten, was die Bezeichnung „schwache Konvergenz“ rechtfertigt.

**Satz 12.9** $f_n \xrightarrow{P} f \Rightarrow f_n \to_D f$.

Beweis: $\forall \varepsilon > 0\, \exists n_0 \in \mathbb{N}\, \forall n \geq n_0 :\ P(\{|f - f_n| > \varepsilon\}) < \varepsilon$. Wegen

$$\begin{aligned} \{f \leq x - \varepsilon\} &\subset \{f_n \leq x\} \cup \{|f_n - f| > \varepsilon\}, \\ \{f_n \leq x\} &\subset \{f \leq x + \varepsilon\} \cup \{|f_n - f| > \varepsilon\} \end{aligned}$$

gilt für alle $n \geq n_0$

$$F_f(x - \varepsilon) \leq F_{f_n}(x) + \varepsilon,\ F_{f_n}(x) \leq F_f(x + \varepsilon) + \varepsilon,$$

also

$$F_f(x - \varepsilon) - \varepsilon \leq F_{f_n}(x) \leq F_f(x + \varepsilon) + \varepsilon,$$

d.h. $F_{f_n}(x) \to_{n\in\mathbb{N}} F_f(x)$ für alle Stetigkeitsstellen von $F_f$. □

**Beispiel 12.10** Setze $\Omega := \{-1, 1\}$, $\mathcal{A} := \mathcal{P}(\Omega)$, $P(A) := |A|/2$, $A \subset \Omega$, $f_n(\omega) := \omega(-1)^n$, $n \in \mathbb{N}$, $f(\omega) := \omega$. Dann gilt:

1. $f_n \to_D f$ (denn $P * f_n = P * f$ für alle $n \in \mathbb{N}$),
2. $f_n \stackrel{P}{\not\to} f$ (denn $P(\{|f_n - f| > 1\}) = 1$ für alle ungeraden $n$).

**Definition 12.11** Setze $\mathcal{K}^\infty$ := Menge der beliebig oft differenzierbaren Funktionen $\psi : \mathbb{R} \to \mathbb{R}$ mit $\psi(x) = 0$ für alle $|x|$ hinreichend groß.

**Satz 12.12** *Für Zufallsvariablen $f, f_n$, $n \in \mathbb{N}$, gilt:*

$$\begin{aligned} & f_n \to_D f \\ \Leftrightarrow \quad & E(\psi(f_n)) \to_{n\in\mathbb{N}} E(\psi(f)) \text{ für alle } \psi \in \mathcal{K}^\infty. \end{aligned}$$

BEWEIS:

**„⇒"** Sei $\psi \in \mathcal{K}^\infty$. Wähle $a, b$ so, dass

$$\{x \in \mathbb{R} : \psi(x) \neq 0\} \subset I := (a, b]$$

mit

$$a, b \notin U := \{x \in \mathbb{R} : x \text{ Unstetigkeitsstelle von } F_f\}.$$

$U$ ist abzählbar (s. Übungen), also ist $U^c$ dicht in $\mathbb{R}$. Daher und wegen der gleichmäßigen Stetigkeit von $\psi$ auf $[a, b]$ existiert zu $\varepsilon > 0$ eine Treppenfunktion $e = \sum_{i=1}^m a_i I_{(t_{i-1},t_i]}$ mit $a = t_0 < t_1 < \cdots < t_m = b$, $t_i \notin U$, $i = 0, 1, \ldots, m$, so dass

$$\sup_{x\in I} |\psi(x) - e(x)| < \varepsilon.$$

Somit gilt:

$$|E(\psi(f_n)) - E(e(f_n))| \leq E(|\psi(f_n) - e(f_n)|) \leq \varepsilon$$

und ebenso

$$|E(\psi(f)) - E(e(f))| \leq E(|\psi(f) - e(f)|) \leq \varepsilon.$$

Für $e$ gilt nun:

$$\begin{aligned} E(e(f_n)) &= \sum_{i=1}^m a_i P(\{f_n \in (t_{i-1}, t_i]\}) \\ &= \sum_{i=1}^m a_i \left(F_{f_n}(t_i) - F_{f_n}(t_{i-1})\right) \\ &\to_{n\in\mathbb{N}} \sum_{i=1}^m a_i \left(F_f(t_i) - F_f(t_{i-1})\right) \\ &= E(e(f)). \end{aligned}$$

Wir erhalten also insgesamt:

$$\begin{aligned} & \limsup_{n\in\mathbb{N}} |E(\psi(f_n)) - E(\psi(f))| \\ = \; & \limsup_{n\in\mathbb{N}} |E(\psi(f_n)) - E(e(f_n)) \\ & \qquad + E(e(f_n)) - E(e(f)) + E(e(f)) - E(\psi(f))| \\ \leq \; & 2\varepsilon. \end{aligned}$$

Hieraus folgt die Behauptung.

„$\Leftarrow$“ Ist $I$ ein beschränktes Intervall, dessen Randpunkte $a, b$ Stetigkeitsstellen von $F_f$ sind, so existieren ein abgeschlossenes Intervall $A$ und ein offenes Intervall $O$ mit

$$A \subset (a,b) \subset I \subset [a,b] \subset O$$

und

$$(P * f)(O \backslash A) < \varepsilon.$$

Ferner existieren[8] $\psi, \xi \in \mathcal{K}^\infty$ mit

$$1_A \leq \psi \leq 1_I \leq \xi \leq 1_O.$$

Damit folgt:

$$\begin{aligned} (P * f_n)(A) &= E(1_A(f_n)) \\ &\leq E(\psi(f_n)) \\ &\leq (P * f_n)(I) \\ &\leq E(\xi(f_n)) \\ &\leq (P * f_n)(O) \end{aligned}$$

sowie

$$\begin{aligned} (P * f)(A) &= E(1_A(f)) \\ &\leq E(\psi(f)) \\ &\leq (P * f)(I) \\ &\leq E(\xi(f)) \\ &\leq (P * f)(O). \end{aligned}$$

Es folgt aus $(P * f)(O) - (P * f)(A) < \varepsilon$:

$$\limsup_{n \in \mathbb{N}} |(P * f_n)(I) - (P * f)(I)| < \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, folgt:

$$(P * f_n)(I) \to_{n \in \mathbb{N}} (P * f)(I).$$

Sei nun $x$ eine Stetigkeitsstelle der Verteilungsfunktion $F_f$ von $P * f$. Ferner seien $x = x_1 > x_2 > \dots$ Stetigkeitsstellen von $F_f$ mit

$$(-\infty, x] = \bigcup_{k \in \mathbb{N}} (x_{k+1}, x_k].$$

Dann gilt mit $Q_n := P * f$, $Q := P * f_n$ und $I_k := (x_{k+1}, x_k]$:

$$\begin{aligned} Q_n((-\infty, x]) &= \sum_{k \in \mathbb{N}} Q_n(I_k) \\ &\geq \sum_{k \leq K} Q_n(I_k) \end{aligned}$$

[8] S. etwa 7.23 (f) in Walter, W. (1991). *Analysis 2*, 3. Auflage. Springer, Berlin.

für ein beliebiges $K \in \mathbb{N}$ und damit

$$\liminf_{n\in\mathbb{N}} Q_n((-\infty, x]) \geq \sum_{k\leq K} Q(I_k),$$

d.h.

$$\liminf_{n\in\mathbb{N}} Q_n((-\infty, x]) \geq \sum_{k\in\mathbb{N}} Q(I_k) = Q((-\infty, x]).$$

Andererseits folgt mit diesen Argumenten

$$\begin{aligned}
&\limsup_{n\in\mathbb{N}} Q_n((-\infty, x]) \\
&\quad = \limsup_{n\in\mathbb{N}} Q_n(1 - (x, \infty)) \\
&\quad = 1 - \liminf_{n\in\mathbb{N}} Q_n((x, \infty)) \\
&\quad \leq 1 - Q((x, \infty)) \\
&\quad = Q((-\infty, x])
\end{aligned}$$

und somit insgesamt

$$\lim_{n\in\mathbb{N}} Q_n((-\infty, x]) = Q((-\infty, x]).$$

□

**Lemma 12.13** *$f_1, f_2, f_3$ seien Zufallsvariablen, $f_3$ sei von $f_1$ und von $f_2$ unabhängig. Dann gilt für alle stetigen und beschränkten Funktionen $\psi : \mathbb{R} \to \mathbb{R}$:*

$$\begin{aligned}
&\left|E\Big((\psi(f_1 + f_3) - \psi(f_2 + f_3))\Big)\right| \\
&\quad \leq \sup_{q\in\mathbb{Q}} |E(\psi(f_1 + q) - \psi(f_2 + q))|.
\end{aligned}$$

Beweis: Es gilt:

$$\begin{aligned}
&\left|E\Big((\psi(f_1 + f_3) - \psi(f_2 + f_3))\Big)\right| \\
&=_{\text{Fubini}} \quad \left|\int E\Big((\psi(f_1 + x) - \psi(f_2 + x))\Big)\,(P * f_3)(dx)\right| \\
&\leq \quad \int \left|E\Big((\psi(f_1 + x) - \psi(f_2 + x))\Big)\right|\,(P * f_3)(dx) \\
&\leq \quad \sup_{x\in\mathbb{R}} |E(\psi(f_1 + x) - \psi(f_2 + x))| \\
&= \quad \sup_{q\in\mathbb{Q}} |E(\psi(f_1 + q) - \psi(f_2 + q))|.
\end{aligned}$$

□

**Satz 12.14 (Zentraler Grenzwertsatz)** *$f_n$, $n \in \mathbb{N}$, sei eine Folge unabhängiger, identisch verteilter und quadratintegrierbarer Zufallsvariablen. Dann gilt mit $\mu := E(f_1)$, $\sigma^2 := \sigma^2(f_1)$ für $t \in \mathbb{R}$:*

$$P\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{f_i-\mu}{\sigma} \le t\right\}\right)$$
$$\to_{n\in\mathbb{N}} \quad \Phi(t) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t \exp\left(-\frac{x^2}{2}\right)\,dx$$

$$\left( \qquad \Leftrightarrow \qquad \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{f_i-\mu}{\sigma} \to_D f,\ P*f = N(0,1).\right)$$

Beweis:

1. O.E. sei $\mu = 0, \sigma^2 = 1$, sonst Übergang zu $\tilde{f}_n := (f_n - \mu)/\sigma$, $n \in \mathbb{N}$.
2. $f_n^*$, $n \in \mathbb{N}$, sei eine Folge unabhängiger, $N(0,1)$–verteilter Zufallsvariablen. Für diese Folge gilt bereits nach 12.6
$$P*\left(\frac{f_1^*+\cdots+f_n^*}{\sqrt{n}}\right) = N(0,1),$$
d.h.
$$P\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i^* \le t\right\}\right) = \Phi(t), \quad t \in \mathbb{R}, n \in \mathbb{N},$$
und damit
$$E(\psi(T_n^*)) = E(\psi(f)) \text{ für alle } \psi \in \mathcal{K}^\infty,$$
wobei
$$T_n^* := \frac{f_1^*+\cdots+f_n^*}{\sqrt{n}}$$
und $f$ eine nach $N(0,1)$–verteilte Zufallsvariable ist.
Zu zeigen ist nun nach 12.12 mit $T_n := (f_1+\cdots+f_n)/\sqrt{n}$:
$$E(\psi(T_n)) \to_{n\in\mathbb{N}} E(\psi(f)) \text{ für alle } \psi \in \mathcal{K}^\infty,$$
d.h.
$$E(\psi(T_n) - \psi(T_n^*)) \to_{n\in\mathbb{N}} 0 \text{ für alle } \psi \in \mathcal{K}^\infty.$$
3. Da die Behauptung lediglich die Verteilung der $f_i$, $i \in \mathbb{N}$, betrifft, können wir annehmen, dass die $f_i$ und die $f_i^*$ auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ definiert sind und sämtlich voneinander unabhängig sind.
Taylor–Entwicklung liefert nun für $x, u \in \mathbb{R}$ und $\psi \in \mathcal{K}^\infty$
$$\begin{aligned}&\psi(x+u)\\ &= \psi(u) + \psi'(u)x + \psi''(u+\vartheta_{x,u}x)\frac{x^2}{2}\\ &= \psi(u) + \psi'(u)x + \psi''(u)\frac{x^2}{2} + r(x,u)x^2,\end{aligned}$$

wobei $0 < \vartheta_{x,u} < 1$ und $r(x,u) := (\psi''(u + \vartheta_{x,u}x) - \psi''(u))/2$.

Da die Funktion $\psi''$ stetig ist und $\psi''(x) = 0$ für $|x|$ hinreichend groß (d.h. $\psi''$ besitzt einen kompakten Träger), ist $\psi''$ beschränkt und gleichmäßig stetig, d.h. es gilt

$$\sup_{x,u} |r(x,u)| < \infty \text{ und } \sup_u |r(x,u)| \to_{x\to 0} 0.$$

4. Es gilt:

$$\begin{aligned}
&\psi(T_n) - \psi(T_n^*) \\
&= \psi\left(\frac{f_1 + \cdots + f_n}{\sqrt{n}}\right) - \psi\left(\frac{f_1^* + \cdots + f_n^*}{\sqrt{n}}\right) \\
&= \sum_{i=1}^n \left(\psi\left(\frac{f_1 + \cdots + f_i + f_{i+1}^* + \cdots + f_n^*}{\sqrt{n}}\right)\right. \\
&\qquad\qquad \left. - \psi\left(\frac{f_1 + \cdots + f_{i-1} + f_i^* + \cdots + f_n^*}{\sqrt{n}}\right)\right) \\
&= \sum_{i=1}^n \left(\psi\left(\frac{f_i}{\sqrt{n}} + U_i\right) - \psi\left(\frac{f_i^*}{\sqrt{n}} + U_i\right)\right),
\end{aligned}$$

wobei

$$U_i := \frac{f_1 + \cdots + f_{i-1} + f_{i+1}^* + \cdots + f_n^*}{\sqrt{n}}$$

von $f_i/\sqrt{n}$ und $f_i^*/\sqrt{n}$ unabhängig ist, $i = 1, \ldots, n$. Nach 12.13 gilt daher

$$\begin{aligned}
&|E(\psi(T_n) - \psi(T_n^*))| \\
&\leq \sum_{i=1}^n \left| E\left(\psi\left(\frac{f_i}{\sqrt{n}} + U_i\right) - \psi\left(\frac{f_i^*}{\sqrt{n}} + U_i\right)\right)\right| \\
&\leq \sum_{i=1}^n \sup_{q\in\mathbb{Q}} \left| E\left(\psi\left(\frac{f_i}{\sqrt{n}} + q\right) - \psi\left(\frac{f_i^*}{\sqrt{n}} + q\right)\right)\right| \\
&= n \sup_{q\in\mathbb{Q}} \left| E\left(\psi\left(\frac{f_1}{\sqrt{n}} + q\right) - \psi\left(\frac{f_1^*}{\sqrt{n}} + q\right)\right)\right|.
\end{aligned}$$

Nach 3. gilt

$$\begin{aligned}
&\psi\left(\frac{f_1}{\sqrt{n}} + q\right) \\
&= \psi(q) + \psi'(q)\frac{f_1}{\sqrt{n}} + \psi''(q)\frac{f_1^2}{2n} + r\left(\frac{f_1}{\sqrt{n}}, q\right)\frac{f_1^2}{n}, \\
&\psi\left(\frac{f_1^*}{\sqrt{n}} + q\right) \\
&= \psi(q) + \psi'(q)\frac{f_1^*}{\sqrt{n}} + \psi''(q)\frac{f_1^{*2}}{2n} + r\left(\frac{f_1^*}{\sqrt{n}}, q\right)\frac{f_1^{*2}}{n}.
\end{aligned}$$

Wegen $E(f_1) = E(f_1^*) = 0$ und $E(f_1^2) = E(f_1^{*2}) = 1$ folgt damit:

$$\begin{aligned}
&|E(\psi(T_n) - \psi(T_n^*))| \\
\leq \quad & n \sup_{q\in\mathbb{Q}} \left| E\left( \frac{f_1^2}{n} r\left(\frac{f_1}{\sqrt{n}}, q\right) - \frac{f_1^{*2}}{n} r\left(\frac{f_1^*}{\sqrt{n}}, q\right)\right)\right| \\
\leq \quad & E\left( f_1^2 \sup_{q\in\mathbb{Q}} \left| r\left(\frac{f_1}{\sqrt{n}}, q\right)\right|\right) \\
& \qquad + E\left( f_1^{*2} \sup_{q\in\mathbb{Q}} \left| r\left(\frac{f_1^*}{\sqrt{n}}, q\right)\right|\right) \\
\to_{n\in\mathbb{N}} \quad & 0
\end{aligned}$$

nach 3. und dem Satz von der dominierten Konvergenz.

□

Dieselben Argumente wie beim Beweis des Satzes von Glivenko–Cantelli führen zu der folgenden Verschärfung des obigen Satzes.

**Korollar 12.15** *Unter den Voraussetzungen von 12.14 gilt:*

$$\sup_{t\in\mathbb{R}} \left| P\left(\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{f_i - \mu}{\sigma} \leq t \right\}\right) - \Phi(t) \right| \to_{n\in\mathbb{N}} 0.$$

**Korollar 12.16 (Satz von Moivre–Laplace)** *Die Zufallsvariablen $f_i$, $i \in \mathbb{N}$, seien unabhängig und $B(1,p)$–verteilt mit $p \in (0,1)$. Dann gilt mit $S_n := \sum_{i=1}^n f_i$ für $-\infty \leq t_1 \leq t_2 \leq \infty$ :*

$$\begin{aligned}
&P\left(\left\{ t_1 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq t_2 \right\}\right) \\
&\qquad \to_{n\in\mathbb{N}} \quad \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} \exp\left(-\frac{x^2}{2}\right) dx,
\end{aligned}$$

*bzw.*

$$\begin{aligned}
&\max_{k_1,k_2\in\{0,\ldots,n\}} \Bigg| \; B(n,p)(\{k_1, k_1+1, \ldots, k_2\}) \\
&\qquad\qquad - \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{k_1 - np}{\sqrt{np(1-p)}}}^{\frac{k_2 - np}{\sqrt{np(1-p)}}} \exp\left(-\frac{x^2}{2}\right) dx \Bigg| \\
&\to_{n\in\mathbb{N}} \quad 0.
\end{aligned}$$

Beweis: Es gilt $E(f_i) = p$, $\sigma^2(f_i) = p - p^2 = p(1-p)$. Die erste Behauptung folgt nun unmittelbar aus dem Zentralen Grenzwertsatz.
Ferner gilt mit $t_1 = (k_1 - np)/(\sqrt{np(1-p)}$, $t_2 = (k_2 - np)/(\sqrt{np(1-p)}$:

$$\begin{aligned} & P\left(\left\{t_1 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq t_2\right\}\right) \\ = \quad & P(\{k_1 \leq S_n \leq k_2\}) \\ = \quad & B(n,p)(\{k_1, k_1+1, \ldots, k_2\}), \end{aligned}$$

s. Übungen. Die zweite Behauptung folgt damit aus 12.15. □

Es ist also nach dem Satz von Moivre–Laplace möglich, die Wahrscheinlichkeit $B(n,p)(\{k_1, k_1{+}1, \ldots, k_2\})$, deren exakte Berechnung auf die Summation unhandlicher Ausdrücke $\binom{n}{k}p^k(1-p)^{n-k}$ hinausläuft, näherungsweise mittels der Verteilungsfunktion $\Phi$ zu berechnen, wenn $n$ groß ist.

**Beispiel 12.17 (Macht entschloss. Minderheit)** An einer Stichwahl zwischen den beiden Kandidaten $A$ und $B$ nehmen 1 Million Wähler teil. 2000 Wähler unterwerfen sich der Parteidisziplin und stimmen geschlossen für Kandidat $A$. Die übrigen 998000 Wähler sind mehr oder weniger unentschlossen und treffen ihre Entscheidung unabhängig voneinander durch Werfen einer (fairen) Münze. Wie groß ist die Wahrscheinlichkeit $p_A$ für einen Sieg von $A$?

Lösung: $A$ siegt genau dann, wenn er mehr als 498000 der Stimmen der 998000 unentschlossenen Wähler erhält. Die Anzahl $f$ der $A$–Stimmen dieser Wähler ist $B(998000, 1/2)$–verteilt. Es folgt:

$$\begin{aligned} p_A &= P(\{f > 498000\}) \\ &= P\left(\left\{\frac{f - 998000\frac{1}{2}}{\sqrt{998000\frac{1}{2}\frac{1}{2}}} > \frac{498000 - 998000\frac{1}{2}}{\sqrt{998000\frac{1}{2}\frac{1}{2}}}\right\}\right) \\ &\approx P\left(\left\{\frac{f - 998000\frac{1}{2}}{\sqrt{998000\frac{1}{2}\frac{1}{2}}} > -2{,}002\right\}\right) \\ &\approx 1 - \Phi(-2{,}002) \\ &= \Phi(2{,}002) \\ &\approx 0{,}977. \end{aligned}$$

(Zum Vergleich: Abschätzung mittels der Tschebyscheff–Ungleichung ergibt:

$$\begin{aligned}
1 - p_A &\approx P\left(\left\{\frac{f - 998000\frac{1}{2}}{\sqrt{998000\frac{1}{2}\frac{1}{2}}} \leq -2,002\right\}\right) \\
&\leq P\left(\left\{\left|\frac{f - 998000\frac{1}{2}}{\sqrt{998000\frac{1}{2}\frac{1}{2}}}\right| \geq 2,002\right\}\right) \\
&\leq \frac{1}{2,002^2} \\
&\approx \frac{1}{4} \\
&\Rightarrow p_A \gtrsim \frac{3}{4}.)
\end{aligned}$$

□

**Beispiel 12.18** Es gilt:

$$\lim_{n \in \mathbb{N}} \left( \exp(-n) \sum_{i=0}^{n} \frac{n^i}{i!} \right) = \frac{1}{2}.$$

Beweis: $f_1, f_2$ seien unabhängige, Poisson–verteilte Zufallsvariablen zu den Parametern $\lambda_1$ bzw. $\lambda_2 > 0$, d.h. es gilt für $i = 1, 2$:

$$P(\{f_i = k\}) = \exp(-\lambda_i) \frac{\lambda_i^k}{k!}, \quad k = 0, 1, \ldots$$

Dann ist $f_1 + f_2$ Poisson–verteilt zum Parameter $\lambda_1 + \lambda_2$:

$$\begin{aligned}
&P(\{f_1 + f_2 = k\}) \\
&= \sum_{i=0}^{\infty} P(\{f_1 + f_2 = k, f_1 = i\}) \\
&= \sum_{i=0}^{k} P(\{f_2 = k - i, f_1 = i\}) \\
&= \sum_{i=0}^{k} P(\{f_2 = k - i\}) P(\{f_1 = i\}) \\
&= \sum_{i=0}^{k} \exp(-\lambda_2) \frac{\lambda_2^{k-i}}{(k-i)!} \exp(-\lambda_1) \frac{\lambda_1^i}{i!} \\
&= \exp(-(\lambda_1 + \lambda_2)) \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \\
&= \exp(-(\lambda_1 + \lambda_2)) \frac{(\lambda_1 + \lambda_2)^k}{k!}, \quad k = 0, 1, \ldots
\end{aligned}$$

Ferner gilt $E(f_1) = \lambda_1$, $\sigma^2(f_1) = \lambda_1$.

Es sei nun $f_i$, $i \in \mathbb{N}$, eine Folge unabhängiger, identisch zum Parameter 1 Poisson–verteilter Zufallsvariablen. Dann gilt nach dem Zentralen Grenzwertsatz:

$$P\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f_i - 1) \leq 0\right\}\right) \to_{n\in\mathbb{N}} \Phi(0) = \frac{1}{2}.$$

Andererseits gilt:

$$\begin{aligned} & P\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f_i - 1) \leq 0\right\}\right) \\ = \; & P\left(\left\{\sum_{i=1}^{n} f_i \leq n\right\}\right) \\ = \; & \exp(-n)\sum_{i=0}^{n}\frac{n^i}{i!}, \end{aligned}$$

da $\sum_{i=0}^{n} f_i$ Poisson–verteilt ist zum Parameter $n$. □

Der folgende Satz macht eine Aussage über die Konvergenzgeschwindigkeit beim Zentralen Grenzwertsatz.

**Satz 12.19 (Berry–Esseen)** *Es seien $f_1, f_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen. Ist $0 < \sigma^2 := Var(f_1) < \infty$, $\gamma := E(|f_1 - \mu|^3) < \infty$ mit $\mu := E(f_1)$, so gilt:*

$$\begin{aligned} & \sup_{x\in\mathbb{R}}\left|P\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f_i - \mu}{\sigma} \leq x\right\}\right) - \Phi(x)\right| \\ \leq \; & 0,8\frac{\gamma}{\sigma^3}\frac{1}{\sqrt{n}}. \end{aligned}$$

Beweis: S. Gänssler, P. und Stute, W. (1977). □

Eine weitere direkte Anwendung des Zentralen Grenzwertsatzes führt zu sog. Konfidenzintervallen (Vertrauensintervallen). Der einfachste Fall ist der folgende: Angenommen, $f_n$, $n \in \mathbb{N}$, ist eine Folge unabhängiger, identisch verteilter und quadratintegrierbarer Zufallsvariablen mit bekannter Varianz $\sigma^2$, aber unbekanntem Mittelwert $\mu$, der geschätzt werden soll.
Mit dem arithmetischen Mittel $\hat{\mu}_n := n^{-1}\sum_{i=1}^{n} f_i$ erhalten wir eine Punktschätzung für $\mu$. Allerdings wird $\hat{\mu}_n$ um den wahren Wert $\mu$ (zufällig) schwanken. Es erscheint daher vernünftig, zusätzlich zur Punktschätzung $\hat{\mu}_n$ ein Intervall

$$I_n := [\hat{\mu}_n - c, \hat{\mu}_n + c]$$

(mit dem Mittelpunkt $\hat{\mu}_n$) anzugeben, von dem man weiß, dass es den unbekannten Mittelwert $\mu$ mit hoher Wahrscheinlichkeit enthält. Dies ist eine Bereichsschätzung von $\mu$.
Problem: Wie soll $c > 0$ gewählt werden?

Einerseits natürlich möglichst klein, um eine gute (Bereichs–)Schätzung für $\mu$ zu erhalten.
Andererseits darf $c$ nicht zu klein gewählt werden, da $I_n$ den Wert $\mu$ mit hoher Wahrscheinlichkeit enthalten soll.
Eine Lösung dieses Zielkonfliktes bietet der Zentrale Grenzwertsatz wie folgt:
Wir wissen, dass für $t \geq 0$ gilt:

$$\begin{aligned} P\left(\left\{-t \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f_i - \mu}{\sigma} \leq t\right\}\right) \\ \to_{n\in\mathbb{N}} \quad & \Phi(t) - \Phi(-t) \\ = \quad & 2\Phi(t) - 1, \end{aligned}$$

wobei

$$\begin{aligned} & P\left(\left\{-t \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f_i - \mu}{\sigma} \leq t\right\}\right) \\ = \quad & P\left(\left\{\hat{\mu}_n - \frac{t\sigma}{\sqrt{n}} \leq \mu \leq \hat{\mu}_n + \frac{t\sigma}{\sqrt{n}}\right\}\right) \\ = \quad & P\left(\left\{\mu \in \left[\hat{\mu}_n - \frac{t\sigma}{\sqrt{n}}, \hat{\mu}_n + \frac{t\sigma}{\sqrt{n}}\right]\right\}\right). \end{aligned}$$

Wählen wir also

$$c := c_n := \frac{t\sigma}{\sqrt{n}} \quad (\to_{n\in\mathbb{N}} 0!),$$

so erhalten wir

$$P(\{\mu \in I_n\}) \to_{n\in\mathbb{N}} 2\Phi(t) - 1.$$

Dabei wählen wir nun $t > 0$ so, dass $2\Phi(t) - 1$ gleich der (hohen) vorgegebenen Wahrscheinlichkeit sein soll, sagen wir $1 - \alpha$, mit der $\mu$ in $I_n$ liegen soll. Ein typischer Wert wäre $1 - \alpha = 0,95$.
Es soll also gelten:

$$2\Phi(t) - 1 = 1 - \alpha \Leftrightarrow t = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

d.h. $t =$(1 − $\alpha$/2)–Quantil der Standardnormalverteilung, i.Z. $u_{\alpha/2}$. Damit erhalten wir

$$P\left(\left\{\mu \in \left[\hat{\mu}_n - \frac{u_{\alpha/2}\sigma}{\sqrt{n}}, \hat{\mu}_n + \frac{u_{\alpha/2}\sigma}{\sqrt{n}}\right]\right\}\right) \to_{n\in\mathbb{N}} 1 - \alpha$$

und

$$I_n(\alpha) := \left[\hat{\mu}_n - \frac{u_{\alpha/2}\sigma}{\sqrt{n}}, \hat{\mu}_n + \frac{u_{\alpha/2}\sigma}{\sqrt{n}}\right]$$

heißt Konfidenzintervall (für $\mu$) zum (asymptotischen) Niveau $1 - \alpha$.
Beachte, dass, wenn $P * f_i = N(\mu, \sigma^2)$, d.h. $f_i$ selbst normalverteilt, $i = 1, 2, \ldots$, aus dem Faltungstheorem der Normalverteilung 12.6 sofort folgt:

$$P(\mu \in I_n(\alpha)) = 1 - \alpha.$$

# Literatur


[1] Georgii, H.–G. (2002). *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*. De Gruyter, Berlin.

[2] Gänssler, P. und Stute, W. (1977). *Wahrscheinlichkeitstheorie*. Springer, Heidelberg.

[3] Krengel, U. (2002). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 6. Auflage. Vieweg, Braunschweig.

[4] Walter, W. (1991). *Analysis 2*. 3. Auflage. Springer, Heidelberg.