Kapitel 3

Wahrscheinlichkeiten und Zufallsgrößen

Zufällige Ereignisse unterscheiden sich im Grad der Gewissheit ihres Eintretens, d. h., in der *Wahrscheinlichkeit* ihres Eintretens.

Es ist eine Erfahrungssache, dass sich die relative Häufigkeit, mit der ein Ereignis in einer langen Reihe von Versuchen, die immer wieder neu unter im Wesentlichen gleichartigen Bedingungen ausgeführt werden, um einen festen Wert stabilisiert. Diesen Wert könnte man als Grad der Gewissheit des Eintretens des Ereignisses in einem einzelnen Versuch ansehen. Ausgehend von dieser Vorstellung formulieren wir einige plausible Eigenschaften von Wahrscheinlichkeiten, die sich dann auch wieder finden im *Axiomensystem der Wahrscheinlichkeitstheorie*.

Zu den mathematisch übersichtlichsten zufälligen Versuchen gehören die *Laplace-Versuche*. Sie besitzen nur endlich viele und dabei gleichwahrscheinliche Ausgänge. Die Bestimmung von Wahrscheinlichkeiten der mit ihnen zusammenhängenden Ereignisse läuft auf das Abzählen gewisser Fälle, häufig unter Verwendung *kombinatorischer Formeln*, hinaus.

Der Begriff der *Zufallsgröße* gehört ebenfalls zum Grundbestand der Wahrscheinlichkeitstheorie. Zufallsgrößen vermitteln stets Teilaspekte eines zufälligen Versuchs und fungieren als beobachtbare (bzw. interessierende) Größen, wenn der Ausgang des Versuches selbst nicht beobachtbar ist (bzw. nicht von Interesse ist). Von Wichtigkeit sind die von ihnen induzierten *Wahrscheinlich-*

keitsverteilungen.

3.1 Axiomensystem und erste Folgerungen

Wir betrachten einen zufälligen Versuch und nehmen an, Ω sei die Menge seiner möglichen Versuchsausgänge ω und $\mathfrak{A}$ die σ-Algebra der mit diesem Versuch verbundenen Ereignisse, d. h. gemäß Abschnitt 2.4 eine σ-Algebra von Teilmengen von Ω.

Es sei A irgend ein Ereignis aus $\mathfrak{A}$. Wir wissen, dass A bei der Versuchsausführung eintreten kann, aber nicht eintreten muss. Sein Eintreten ist nicht gewiss. Unterschiedliche Ereignisse können sich allerdings im *Grad der Gewissheit* ihres Eintretens unterscheiden.

Beispiel 3.1 (Werfen einer Zündholzschachtel)

Mögliche Versuchsausgänge sind die der drei Seiten, auf denen die Schachtel zu liegen kommen kann: Große Seite (Ober- bzw. Unterseite) / Mittlere Seite (Seiten mit Reibflächen) / Kleine Seite (Stirn- bzw. Hinterseite): Als Ω wählen wir die Menge $\Omega = \{G, M, K\}$.
Wir bemessen den Grad der Gewissheit des Eintreffens jeder der möglichen Fälle aus der Erfahrung oft wiederholter Versuche. Dieser Grad wird umso höher eingeschätzt, je häufiger bei längerer Versuchsreihe die Schachtel auf der entsprechenden Seite zu liegen kommt.

Betrachten wir die Situation von Beispiel 3.1 etwas allgemeiner. Es sei A ein Ereignis, das mit einem zufälligen Versuch verbunden ist. Der Versuch werde n mal durchgeführt, jedes Mal unter im Wesentlichen gleichartigen Bedingungen und unabhängig voneinander.

In $n(A)$ Fällen trete A ein. Dann zeigt die *relative Häufigkeit* des Eintretens von A in n Versuchen, nämlich $\frac{n(A)}{n}$, mit wachsendem n eine bemerkenswerte Stabilität: $\frac{n(A)}{n}$ verändert sich immer weniger, sie scheint gegen einen Grenzwert zu konvergieren, wenn n unbegrenzt wächst.
Wir nennen diese Erscheinung das *empirische Gesetz der großen Zahlen.*

Das folgende Beispiel verdeutlicht die Stabilisierung der relativen Häufigkeiten.

Beispiel 3.2 (Werfen eines Kronenverschlusses)

Relative Häufigkeit dafür, dass die offene Seite nach oben zeigt:

Zahl der Versuche

100	200	300	400	500	600	700
0,7300	0,7750	0,7767	0,7750	0,7800	0,7900	0,7943
800	900	1000				
0,8012	0,7967	0,7910				

(Nach Nawrotzki, K., Lehrmaterial zur Ausbildung von Diplomlehrern Mathematik, Jena 1984)

Mit dem nächsten Beispiel wird deutlich, dass das empirische Gesetz der großen Zahlen häufig auch unbewusst angewandt wird.

Beispiel 3.3 (Skatspiel)

Die relativen Häufigkeiten bestimmter Konstellationen prägen sich beim Spieler ein. Zwei Buben im Skat sind z. B. relativ selten. Daraus wird geschlossen, dass auch im nächsten Spiel nur mit geringer Chance zwei Buben im Skat zu finden sein werden. Es ergibt sich eine Verhaltensgrundlage: "Auf den Skat reizt man nicht".

Aus dem genannten Gesetz der großen Zahlen leitet man die Überzeugung ab, dass es zu jedem zufälligen Ereignis A eine Zahl $P(A)$ gibt, die *Wahrscheinlichkeit* von A, die den Grad der Gewissheit des Eintretens von A (in einem einzelnen Versuch) ausdrückt.

Für lange Versuchsreihen sollte das eine Zahl sein, um die sich $\frac{n(A)}{n}$ stabilisiert:

$$\frac{n(A)}{n} \approx P(A).$$

Daraus ergeben sich plausible Eigenschaften für $P(A)$:

$$0 \leq P(A) \leq 1,$$

$$P(\Omega) = 1, \quad P(\emptyset) = 0,$$

$$P(A \cup B) = P(A) + P(B), \text{falls } A \cap B = \emptyset.$$

Aus diesen Vorstellungen hat sich ein Axiomensystem entwickelt, das 1933 A.N. Kolmogorov in einer berühmten Arbeit eingeführt hat. (Kolmogorov, Grundlagen der Wahrscheinlichkeitstheorie, Springer, Berlin 1933).

Dieses Axiomensystem ist heute die Grundlage der Wahrscheinlichkeitstheorie und Mathematischen Statistik und lautet wie folgt.

Axiomensystem der Wahrscheinlichkeitstheorie

Wir stellen uns wieder $(\Omega, \mathfrak{A})$ als einen zufälligen Versuch vor, die Elemente ω von Ω bilden die möglichen Versuchsausgänge, $\mathfrak{A}$ sei die σ-Algebra der mit dem Versuch verbundenen Ereignisse, also eine σ-Algebra von Teilmengen von Ω (siehe Abschnitt 2.4).

Als *Wahrscheinlichkeitsverteilung* $P(\cdot)$ auf der σ-Algebra $\mathfrak{A}$ von Teilmengen einer nichtleeren Menge Ω bezeichnet man jede Abbildung P von $\mathfrak{A}$ in $[0, 1]$ mit

A1. $P(\Omega) = 1$ und $P(\emptyset) = 0$,

A2. Für jedes $n \geq 2$ und jede Folge $(A_k, k = 1, \cdots, n)$ aus $\mathfrak{A}$ mit

$A_k \cap A_l = \emptyset, k \neq l$ (paarweise Unvereinbarkeit) gilt

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$$

(Endliche Additivität der Wahrscheinlichkeitsverteilung P)

A2.' Für jede abzählbar unendliche Folge $(A_k, k \geq 1)$ aus $\mathfrak{A}$ mit

$A_k \cap A_l = \emptyset \quad , \quad k \neq l$ (paarweise Unvereinbarkeit) gilt

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

(σ-Additivität der Wahrscheinlichkeitsverteilung P)

$(\Omega, \mathfrak{A}, P)$ ist mit dieser Definition ein (normierter) Maßraum, P heißt ein *Wahrscheinlichkeitsmaß* auf $\mathfrak{A}$. Statt *Wahrscheinlichkeitsverteilung* $P(\cdot)$ *auf* $\mathfrak{A}$, sprechen wir einfach auch von einer *Verteilung* $P(\cdot)$ *auf* $\mathfrak{A}$.

Definition 3.4 *Sind* Ω *eine nichtleere Menge,* $\mathfrak{A}$ *eine* σ*-Algebra von Teilmengen von* Ω *und* P *eine Abbildung von* $\mathfrak{A}$ *in* $[0, 1]$ *mit den Eigenschaften A1., A2. und A2'., so heißt das Tripel* $(\Omega, \mathfrak{A}, P))$ *ein* Wahrscheinlichkeitsraum*.*

Bemerkung 3.5

Jeder Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ ist das mathematische Modell eines zufälligen Versuches. Ω enthält dabei die Menge der möglichen Versuchsergebnisse, $\mathfrak{A}$ entspricht der Menge der mit dem Versuch verbundenen Ereignisse, P ist die Wahrscheinlichkeitsverteilung des Versuches. P legt fest, mit welcher Wahrscheinlichkeit $P(A)$ jedes mit dem Versuch verbundene Ereignis $A \in \mathfrak{A}$ bei der Versuchsdurchführung eintritt.

Folgerungen 3.6

1. Für jedes $A \in \mathfrak{A}$ ergibt sich aus A1. und A2. wegen $A \cup \bar{A} = \Omega$ und $A \cap \bar{A} = \emptyset$ die Gleichung

$$P(\bar{A}) = 1 - P(A). \tag{3.1}$$

   In den zwei folgenden Punkten seien A und B irgend zwei Ereignisse aus $\mathfrak{A}$.

2. Stets gilt

$$P(B) = P(B \cap A) + P(B \backslash A) \text{ und} \tag{3.2}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{3.3}$$

3. Ist $A \subseteq B$, so folgt aus (3.2)

$$P(B \backslash A) = P(B) - P(A) \tag{3.4}$$

   und somit

$$P(A) \leq P(B) \text{ (Monotonie der Verteilung } P)$$

4. Für alle $A_1, A_2, ldots, \in \mathfrak{A}$ gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{3.5}$$

5. Für alle $A_1, A_2, \dots, \in \mathfrak{A}$ gilt

$$P\left(\bigcup_{1}^{n} A_k\right) \le \sum_{1}^{n} P(A_k) \text{ (endliche Subadditivität)} \tag{3.6}$$

Das ergibt sich aus (3.3)mittels vollständiger Induktion.

Das Axiom A2.' ermöglicht es, die Wahrscheinlichkeiten von Ereignissen zu bestimmen, die im Zusammenhang mit unendlichen Folgen von Ereignissen stehen.
Das nächste Lemma und seine Folgerungen stellen zu $A2$. äquivalente Eigenschaften bereit.

Lemma 3.7 *(σ-Stetigkeit von Wahrscheinlichkeitsverteilungen)*

Wenn für die Abbildung P von $\mathfrak{A}$ in $[0,1]$ die Axiome A1. und A2. gelten, so ist A2.' äquivalent mit jeder der folgenden Eigenschaften:

a) Für jede monoton fallende Folge $(A_n, n \ge 1)$ aus $\mathfrak{A}$ mit $\bigcap_{n=1}^{\infty} A_n = \emptyset$ gilt $\lim_{n\to\infty} P(A_n) = 0$

b) Für jede monoton wachsende Folge $(A_n, n \ge 1)$ aus $\mathfrak{A}$ mit $\bigcup_{n=1}^{\infty} A_n = \Omega$ gilt $\lim_{n\to\infty} P(A_n) = 1$

Beweis:

A2.' $\Longrightarrow$ a): Mit $B_n = A_n \setminus A_{n+1},\ n \ge 1$ ist $(B_n, n \ge 1)$ eine Folge paarweise disjunkter Ereignisse mit $\bigcup_{n=m}^{\infty} B_n = A_m, m \ge 1$. Folglich gilt mit Axiom A2' die Gleichung

$$P(A_m) = \sum_{n=m}^{\infty} P(B_n), \quad m \ge 1.$$

Somit haben wir wegen $\sum_{n=1}^{\infty} P(B_n) = P(A_1) \leq 1$ die Beziehung $\lim_{m\to\infty} P(A_m) = 0$, also gilt a).

a) $\Longrightarrow$ b): $(\overline{A}_n, n \geq 1)$ ist monoton fallend mit $\bigcap_{n=1}^{\infty} \overline{A}_n = \emptyset$, somit gilt $\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty}(1 - P(\overline{A}_n)) = 1 - 0 = 1$. Somit haben wir b) gezeigt.

b) $\Longrightarrow$ A2.': Ist $(C_n, n \geq 1)$ eine Folge paarweise disjunkter Ereignisse, so definieren wir $C'_n = \bigcup_{m=1}^{n} C_m, C'_\infty = \bigcup_{m=1}^{\infty} C_m, A_n = C'_n \cup \overline{C'}_\infty, n \geq 1$. Damit folgt $A_n \subseteq A_{n+1}, n \geq 1$ und $\bigcup_{n=1}^{\infty} A_n = \Omega$, und deshalb nach Voraussetzung $\lim_{n\to\infty} P(A_n) = 1$. Wegen Axiom A2. ergibt sich
$1 = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P(C'_n) + P(\overline{C'}_\infty)$, also $P(C'_\infty) = 1 - P(\bar{C}'_\infty) = \lim_{n\to\infty} P(C'_n) = \lim_{n\to\infty} \sum_{m=1}^{n} P(C_m) = \sum_{m=1}^{\infty} P(C_m)$. Damit ist $A2'$ nachgewiesen. $\square$

Folgerungen 3.8

1. Ist (A_n) eine monoton fallende (monoton wachsende) Folge aus $\mathfrak{A}$, so gilt

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) \text{ bzw. } \lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n)\right).$$

   Beweis: Man wende das Lemma 3.7 auf $\left(A_n \setminus \left(\bigcap_{k=1}^{\infty} A_k\right)\right)$

   bzw. auf $\left(A_n \cup \overline{\left(\bigcup_{k=1}^{\infty} A_k\right)}\right)$ an. $\square$

2. Für jede Folge (A_n) aus $\mathfrak{A}$ gilt:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leqq \sum_{k=1}^{\infty} P(A_n) \text{ (abzählbare Subadditivität)} \tag{3.7}$$

Beweis: $B_n = \bigcup_{1}^{n} A_k \uparrow B = \bigcup_{1}^{\infty} A_k$, beachte die Ungleichung (3.5). $\square$

Mittels der eben bewiesenen Folgerungen 1. und 2. ergibt sich das

Lemma 3.9 *(Erstes Borel-Cantelli-Lemma)*
Falls $A_n \in \mathfrak{A}, n \geq 1,$ *und* $\sum_{n=1}^{\infty} P(A_n) < \infty$, *so gilt*

$$P\left(\limsup_{n\to\infty} A_n\right) = 0$$

Beweis: $P(\lim_{n\to\infty} \sup A_n) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m\geq n} A_m\right) = \lim_{n\to\infty} P\left(\bigcup_{m\geq n} A_m\right)$

$$\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) = 0 \text{ wegen Ungleichung (3.6).}$$

$\square$

In Worten kann man dieses Lemma wie folgt fassen: Gilt $\sum_{n=1}^{\infty} P(A_n) < \infty$, so ist die Wahrscheinlichkeit dafür, dass unendlich viele der Ereignisse A_n eintreten, gleich Null. Anders ausgedrückt, mit Wahrscheinlichkeit Eins treten höchstens endlich viele der Ereignisse A_n ein.

Wir geben noch eine nützliche Formel zur Berechnung von $P\left(\bigcup_{k=1}^{\infty} A_k\right)$ an, bei der die $(A_k, k = 1, \cdots, n)$ nicht paarweise disjunkt sein müssen. Sie ist eine Verallgemeinerung der Formel 3.3.

Aussage 3.10 (Ein- und Ausschlussformel)

Für alle $n \geq 2$ und alle $A_1, A_2, \ldots, A_n \in \mathfrak{A}$ gilt

$$
\begin{aligned}
P\Big(\bigcup_{k=1}^{n} A_k\Big) &= \sum_{k=1}^{n} P(A_k) - \sum_{1 \leq i < k \leq n} P(A_i \cap A_k) \\
&+ \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \ldots (-1)^{n+1} P(A_1 \cap \cdots \cap A_n) \\
&= \sum_{r=1}^{n} (-1)^{r+1} \sum_{\substack{J_r \subseteq \{1,\ldots,n\} \\ card J_r = r}} P(A_{k_1} \cap \ldots \cap A_{k_r}) \qquad (3.8)
\end{aligned}
$$

wobei $\mathfrak{J}_r = \{k_1, k_2, \ldots, k_r\}$ alle r-elementigen Teilmengen von $\{1, \cdots, n\}$ durchläuft.

Beweis mittels vollst. Induktion, siehe z. B. Henze (2006), Kap. 11.

Die Ein- und Ausschlussformel vereinfacht sich wesentlich, falls die Wahrscheinlichkeiten $P(A_{k_1} \cap \cdots \cap A_{k_r})$ nur von r und nicht von der Wahl des Tupels $(k_1, \cdots, k_r)$ abhängen. Wir definieren:

Definition 3.11 *Es seien $(\Omega, \mathfrak{A}, P)$ ein W-Raum und $A_1, \cdots, A_n$ Ereignisse aus $\mathfrak{A}$. Diese Ereignisse heißen (untereinander)* austauschbar, *falls $P(A_{k_1} \cap \cdots \cap A_{k_r}) = P(A_1 \cap \cdots \cap A_r)$ gilt für alle r mit $1 \leq r \leq n$ und alle r-elementigen Teilmengen $\{k_1, \cdots, k_r\}$ von $\{1, \cdots, n\}$ gilt.*

Aussage 3.12 *Sind $A_1, \cdots, A_n$ austauschbar, so gilt*

$$
P\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} P(A_1 \cap \cdots \cap A_r). \qquad (3.9)
$$

Beweis: Es gibt $\binom{n}{r}$ Teilmengen $\mathscr{J}_r$ von $\{1, \cdots, n\}$ mit r Elementen. □

Aussage 3.13 (Bonferroni-Ungleichungen)

Falls $A_i \in \mathfrak{A}, i = 1, \cdots, n,$ *dann gilt*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) \tag{3.10}$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j)$$

$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \tag{3.11}$$

Beweis: mittels vollständiger Induktion.

Als Ergänzung erwähnen wir schließlich folgende Formel.

Aussage 3.14 *Für alle* $n \geq 2$ *und alle* $A_1, A_2, \cdots, A_n$ *aus* $\mathfrak{A}$ *und*

$A_{n,m}$: = *"Es treten genau* m *der* n *Ereignisse* $A_1, A_2, \cdots, A_n$ *ein"* $1 \leq m \leq n$, *gilt:*

$$P(A_{n,m}) = \sum_{r=m}^{n} (-1)^{r-m} \binom{r}{m} \sum_{\substack{J_r \subseteq \{1,\cdots,n\} \\ card J_r = r}} P(A_{k_1} \cap \cdots \cap A_{k_r}) \tag{3.12}$$

Der Beweis erfolgt ebenfalls mittels vollständiger Induktion.

3.2 Laplace-Experimente

In diesem Abschnitt werden wir erste konkrete mathematische Modelle zufälliger Versuche kennen lernen, die sogenannten *Laplace-Experimente*. Sie zeichnen sich durch besondere Einfachheit aus und sind dennoch in Anwendungen häufig anzutreffen.

Definition 3.15 *Als* Laplace-Experiment (kurz: L-Experiment) *bezeichnet man einen zufälligen Versuch mit:*

1. *der Versuch hat nur endlich viele* $(= N)$ *mögliche Ausgänge.*
2. *Alle Ausgänge haben die gleiche Wahrscheinlichkeit.*

Als mathematisches Modell eines L-Experimentes wählt man einen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ *mit:*

1. Ω *ist endlich:* $\Omega = \{1, 2, \ldots, N\}$ *(allgmeiner, aber äquivalent dazu:*
$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\} \text{ mit } \omega_i \neq \omega_j, i \neq j),$$
2. $\mathfrak{A} = \mathfrak{P}(\Omega)$ *und alle Versuchsausgänge sind gleichwahrscheinlich:*
$$P(\{\omega\}) \equiv: p\ ,\ \omega \in \Omega. \tag{3.13}$$

N *heißt der* Parameter *des L-Experimentes.*

Auf Grund (3.11) gilt wegen

$$1 = P(\Omega) = P(\bigcup_{k=1}^{N}\{\omega_k\}) = \sum_{k=1}^{N} P(\{\omega_k\}) = N \cdot p$$

die Beziehung $P(\{\omega\}) \equiv p = \frac{1}{N}$ und
für jede Teilmenge A von Ω ergibt sich mit $N(A) =$ Anzahl der Elemente von A:

$$P(A) = P(\bigcup_{\omega \in A}\{\omega\}) = \sum_{\omega \in A} P(\{\omega\}) = p \cdot N(A) = \frac{N(A)}{N} = \tag{3.14}$$

$$= \frac{\text{Anzahl der (für A) günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}. \tag{3.15}$$

Die Gesamtwahrscheinlichkeit 1 (d.h., die Wahrscheinlichkeit des sicheren Ereignisses "Irgendein Versuchsausgang tritt ein") ist bei einem Laplace-Experiment gleichmäßig auf die Versuchsausgänge ω verteilt: Man nennt P auch die *gleichmäßige Wahrscheinlichkeitsverteilung* auf $\{1, 2, \ldots, N\}$ oder einfach die *Gleichverteilung* auf $\{1, 2, \ldots, N\}$ bzw. auf $\Omega = \{\omega_1, \ldots, \omega_N\}$.

Bei Laplace-Experimenten spricht man auch davon, dass das Versuchsergebnis *"auf gut Glück"* oder *"rein zufällig"* ausgewählt wird, um die Gleichwahrscheinlichkeit aller möglichen Versuchsausgänge hervorzuheben.

Beispiel 3.16 Der zufällige Versuch bestehe im Werfen zweier regulärer Würfel und im Registrieren, welche Augenzahl der erste und welche der zweite Würfel zeigt. Wir setzen alos $\Omega = \{(i, j) : i, j \in \{1, 2, \ldots, 6\}\}$ mit $i =$ Augenzahl des ersten Würfels und $j =$ Augenzahl des zweiten Würfels.

Alle Ergebnisse sind aus Symmetriegründen (reguläre Würfel) gleichwahrscheinlich, also gilt:

$P((i,j) \text{ tritt auf }) = \frac{1}{36} \quad \text{für alle} \quad i, j \in \{1, \cdots, 6\}.$

Mit $A :=$ "Die Augensumme ist gleich 6" haben wir

$$P(A) = \frac{N(A)}{N} = \frac{5}{36},$$

und für $B :=$ "Die Augenzahlen sind verschieden" erhalten wir

$$P(B) = \frac{N(B)}{N} = \frac{30}{36} = \frac{5}{6}.$$

In Anwendungsbeispielen mit endlichem Ω muss man genau prüfen, ob es sich tatsächlich um ein Laplace-Experiment handelt.

Beobachtet man zum Beispiel im obigen Beispiel nur die Augensumme, so ist dies ein neuer zufälliger Versuch. Man wählt $\Omega = \{2, \ldots, 12\}$. Jetzt sind aber nicht alle Ausgänge gleichberechtigt, d.h. gleichwahrscheinlich:
"Augensumme $= 2$" hat eine kleinere Wahrscheinlichkeit $\left(= \frac{1}{36}\right)$ als "Augensumme $= 4$" $\left(= \frac{1}{12}\right)$, denn das erste Ereignis tritt nur beim Versuchsausgang (1,1), das zweite dagegen bei jedem der Ausgänge (1,3), (2,2), (3,1) ein.

Wir kehren zurück zum Modell des n-maligen Münzenwurfes aus Abschnitt 2.5.

3.3 Münzenwurf, zum Zweiten

Wir setzen hier das Studium des zufälligen Versuches "n-maliges Werfen einer Münze" aus Abschnitt 2.5 fort. Die Münze, die wir für das Spiel verwenden, sei regulär, d. h. symmetrisch. Das bedeute, beide Seiten erscheinen bei einem Wurf mit gleicher Chance. Dann ist das n-malige Werfen ein Laplace-Experiment mit 2^n gleichwahrscheinlichen Ausgängen. Der entsprechende Wahrscheinlichkeitsraum ist $(\Omega, \mathfrak{P}(\Omega), P)$ mit $P(\{\omega\}) = 2^{-n}, \omega \in \Omega$ und

$$P(A) = \sum_{\omega \in A} 2^{-n} = \frac{N(A)}{2^n}, A \subseteq \Omega. \tag{3.16}$$

Wir berechnen die Wahrscheinlichkeit dafür, dass nach n Würfen der zugehörige Pfad $(k, S_k(\omega)), k \geq 0$, mit $S_0(\omega) := 0$ bei r endet: $P(\{\omega \in \Omega | S_n(\omega) = r\})$.

Aussage 3.17 *Für $P(S_n = r)$ gelten die Formeln*

a) Ist n gerade, also $n = 2m$ für ein $m \geq 1$, so gilt

$P(S_{2m} = 2l) = \binom{2m}{m+l} 2^{-2m}$ *falls* $|l| \leq m$,
$P(S_{2m} = r) = 0$ *für alle anderen ganzzahligen r.*

b) Ist n ungerade, also $n = 2m + 1$ für ein $m \geq 1$, so ist

$P(S_{2m+1} = 2l + 1) = \binom{2m+1}{m+l+1} 2^{-2m-1}$ *falls* $-m - 1 \leq l \leq m$,
$P(S_{2m+1} = r) = 0$ *für alle anderen ganzzahligen r.*

Bemerkung: In beiden Fällen handelt es sich um eine um den Nullpunkt symmetrische Verteilung mit Null bzw. ± 1 als Punkte maximaler Wahrscheinlichkeit, vgl. Abschnitt 4.

Die Folge $(S_0, S_1, \ldots, S_n)$ heißt auch eine symmetrische Irrfahrt.

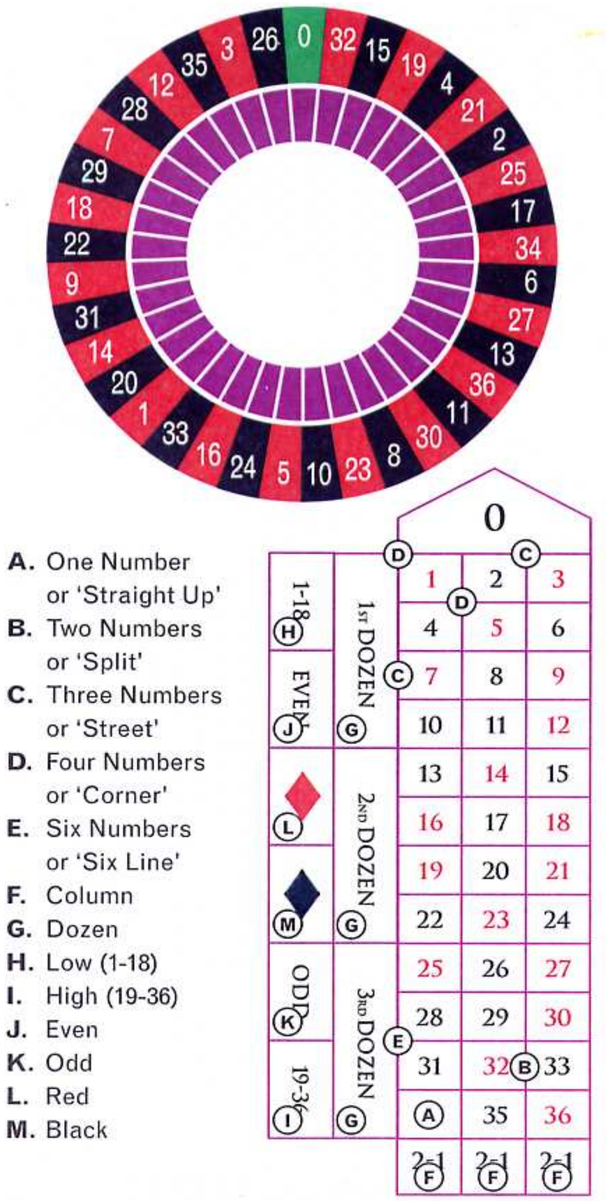

Abbildung 3.1: Quantile der Standardnormalverteilung

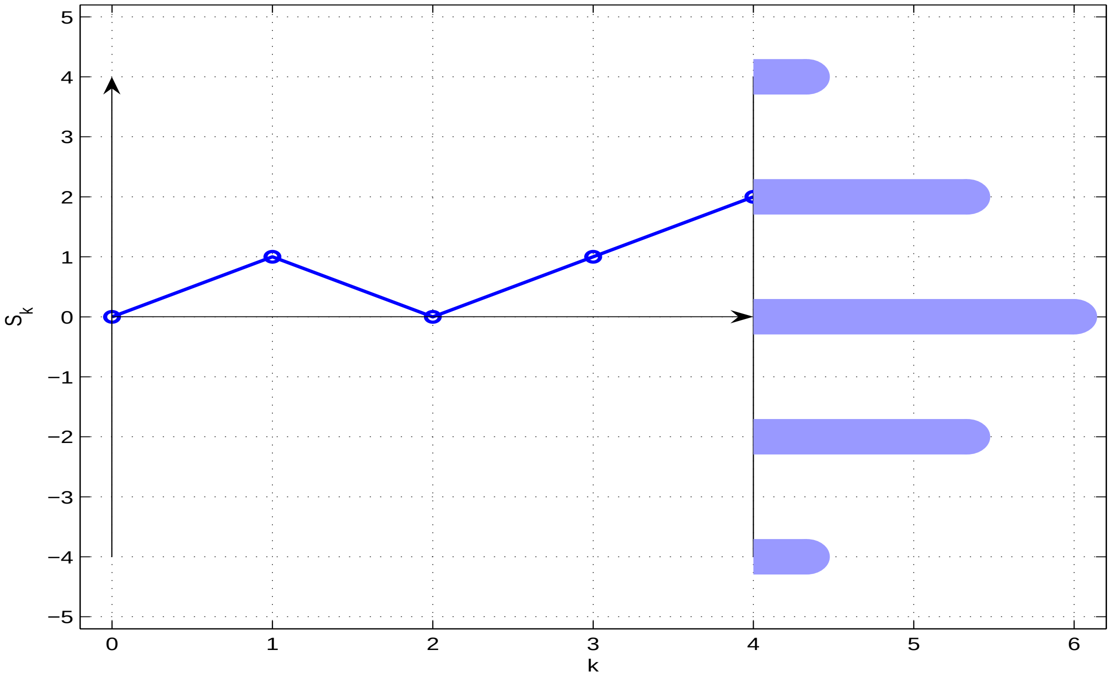

Abbildung 3.2: Beispiel des Pfades der Länge vier und der Orte der Pfadenden

Beweis:

a) $\{S_{2m} = 2l\}$ tritt genau dann ein, wenn in $\omega = (x_1, \cdots, x_{2m})$ genau $l + m$ mal die Eins enthalten ist.

b) $\{S_{2m+1} = 2l + 1\}$ tritt genau dann ein, wenn $\omega = (x_1, \cdots, x_{2m+1})$ genau $m + l + 1$ Einsen enthält.

Für den Spieler A (siehe Abschnitt 2.5) ist es von Interesse, mit welcher Wahrscheinlichkeit er wann zum ersten Mal ein negatives Guthaben hat.

Aussage 3.18 *Die Wahrscheinlichkeit dafür, dass Spieler A zur Zeit n zum ersten Mal ein negatives Gutachten hat, ist für ungerades $n = 2m + 1, m \geq 1$, gleich dem Wert*

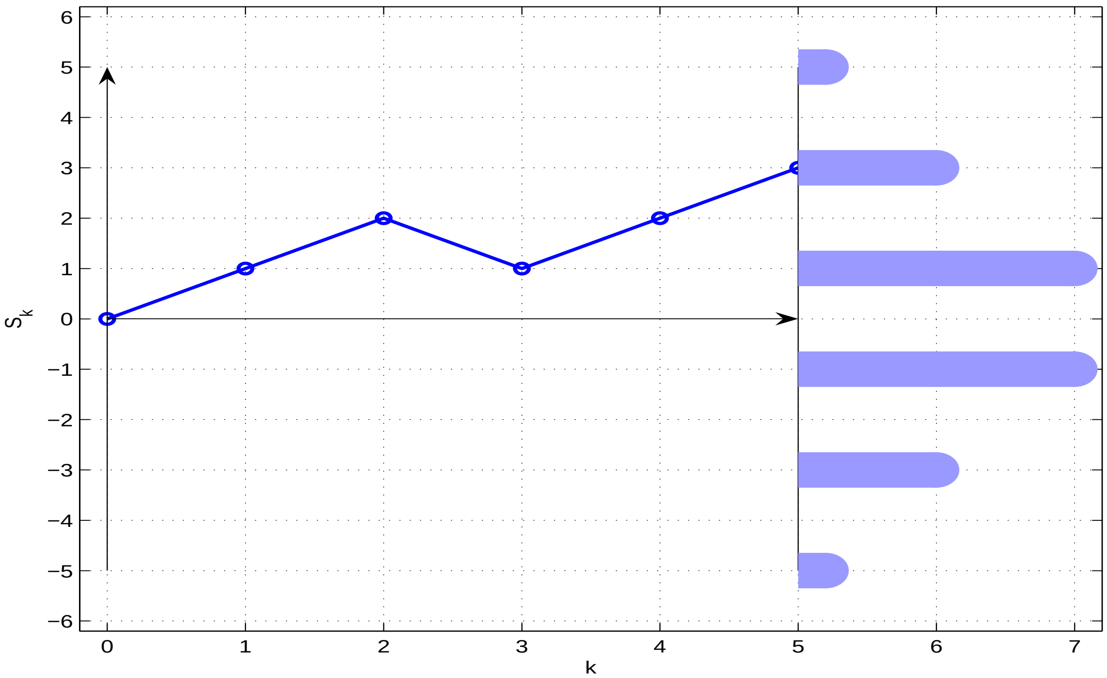

Abbildung 3.3: Beispiel des Pfades der Länge fünf und der Orte der Pfadenden

$$\begin{aligned} &P(\{\omega : s_1 \geq 0, s_2 \geq 0, \ldots, s_{n-1} \geq 0, s_n = -1\}) \\ &= P(S_1 \geq 0, S_2 \geq 0, \ldots, S_{n-1} \geq 0, S_n = -1) = \frac{(2m-1)!!}{(m+1)!2^{m+1}} \end{aligned} \tag{3.17}$$

mit $(2m-1)!! = (2m-1)(2m-3)\cdots 3\cdot 1$ *und für gerades* n *gleich Null.*

Beweis: Für jeden der für das betrachtete Ereignis günstigen Pfade $s = (s_0, s_1, \ldots, s_n)$ gilt $s_{n-1} = 0$. Wir bestimmen deshalb die Zahl aller Pfade von $(0,0)$ nach $(2m, 0)$ die -1 niemals berühren und beachten dabei die Eigenschaft (2.1) jedes Pfades.

Es gibt insgesamt $\binom{2m}{m}$ Pfade von $(0,0)$ nach $(2m, 0)$. Zur Berechnung der gesuchten Zahl der Pfade bedienen wir uns des sogenannten *Spiegelungsprinzips*. Jedem durch ein ω erzeugten Pfad $s = (s_0, s_1, \cdots, s_n)$, der die Zahl -1 jemals vor n erreicht, wird der Pfad s' zugeordnet, der bei -2 startet und bis zur Zeit $T_{-1}(\omega) = \min\{k \geq 1 | s_k = -1\}$ spiegelbildlich bezüglich der Horizontalen der Höhe -1 zu $(s_1, s_2, \cdots s_n)$ verläuft, sowie danach mit $(s_1, s_2, \cdots, s_n)$ übereinstimmt.
Die Zuordnung ist eineindeutig. Folglich ist die Zahl der Pfade, die -1 vor der Zeit n berühren und zur Zeit $n-1 = 2m$ bei Null sind gleich der Zahl der Pfade, die bei -2 starten und zur Zeit $n-1 = 2m$ in Null enden.

Davon gibt es $\binom{2m}{m+1}$ Exemplare. Somit ist die gesuchte Anzahl gleich

$$\binom{2m}{m} - \binom{2m}{m+1} = \binom{2m}{m}\frac{1}{m+1}.$$

Die gesuchte Wahrscheinlichkeit ist also $2^{-2m-1}\binom{2m}{m}\frac{1}{m+1}$. Eine einfache Umformung liefert 3.16. Zu geraden Zeiten n kann S_n nicht zum ersten Mal negativ sein. $\square$

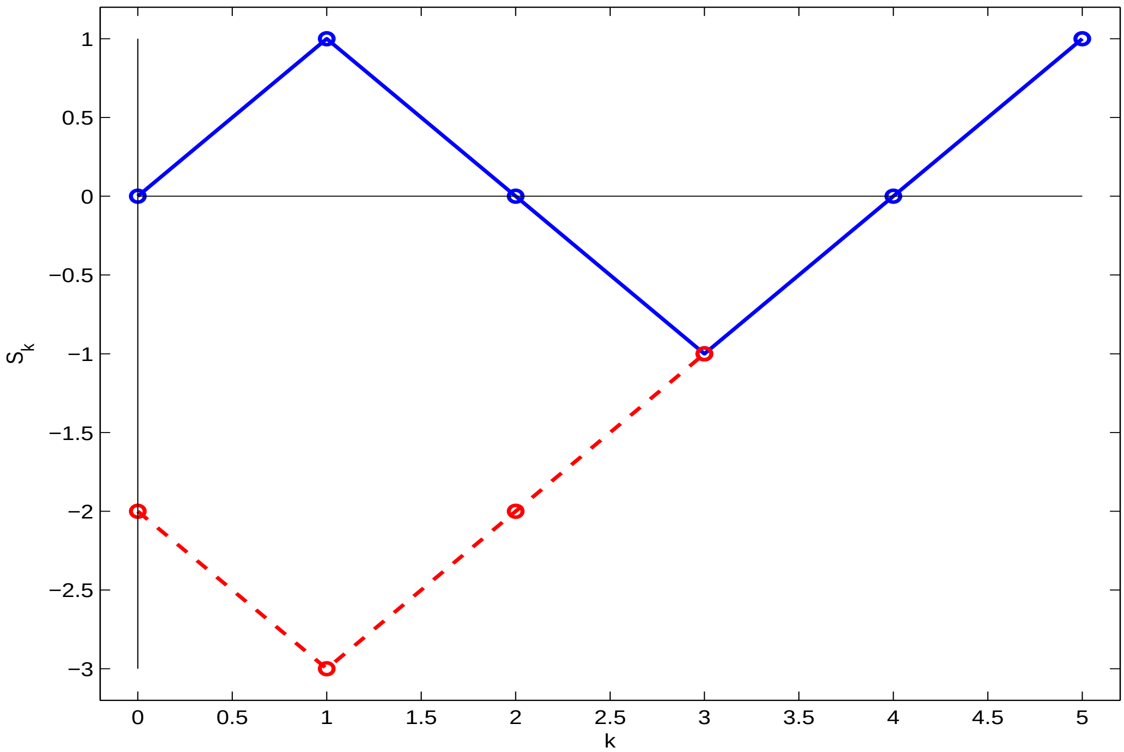

Abbildung 3.4: Gespiegelter Pfad

3.4 Was sagen uns Wahrscheinlichkeiten?

Welche anschauliche Bedeutung hat die Wahrscheinlichkeit $P(A)$ eines zufälligen Ereignisses A? Es gibt zwei grundlegende Erfahrungen im Umgang mit dem Zufall.

a) Empirisches Gesetz der großen Zahlen: In einer langen Reihe gleichartiger, voneinander unabhängiger Versuche ist die relative Häufigkeit $\frac{n(A)}{n}$ des Eintretens von A etwa gleich $P(A)$. Wenn $P(A) > \frac{1}{2}$ gilt, so kann man auf das Eintreten von A Wetten abschließen und wird bei fortlaufenden Wetten dieser Art schließlich im Vorteil sein.

b) Es ist neben dem empirischen Gesetz der großen Zahlen eine zweite Erfahrungstatsache, dass zufällige Ereignisse mit sehr kleinen Wahrscheinlichkeiten bei einmaliger Versuchsdurchführung praktisch nicht eintreten.

   Genauer gesagt: Man muss bei einem zufälligen Versuch, den man einmal durchführt, mit dem Eintreten von A nicht rechnen, falls $P(A)$ sehr klein ist. Diese Erfahrung hat jeder Mensch verinnerlicht.

Beispiel für a):

1) Werfen Sie einen regulären Spielwürfel mehrere Mal und beobachten Sie das Verhalten der relativen Häufigkeit des Auftretens einer "Sechs" im Verlaufe der Würfe. Sie tendiert zu $\frac{1}{6}$.

Beispiele für b):

1) Man erhält keinen Kredit von der Bank, wenn man als Sicherheit anbietet, dass man auf seinen wöchentlichen Tippschein im Lotto "6 aus 49" innerhalb eines Jahres einen "Sechser" erzielt. Die Wahrscheinlichkeit dieses Ereignisses ist so gering, dass man mit seinem Eintreten nicht wirklich rechnet.

2) Man rechnet nicht damit, dass jemand durch maximal dreimaliges zufälliges Raten der PIN bei einer EC-Karte die richtige PIN errät.

$$P(A_1 \cup A_2 \cup A_3) \approx \frac{3}{10^4} = 0,0003$$

3) Wenn ein vorgegebenes Ereignis, das eine sehr kleine Wahrscheinlichkeit hat, tatsächlich eintritt, so zweifelt man mitunter daran, dass der "reine Zufall" zu diesem Ereignis geführt hat. Man stellt eher die Richtigkeit der zugrunde liegenden Annahmen in Frage und prüft sie sorgfältig: Den Ausspruch "Das kann kein Zufall sein!" hat jeder schon mal gehört.

   Beispiel seltener Ziehungen beim Lotto "6 aus 49" wie (1, 2, 3, 4, 5, 6) führen regelmäßig zur Aufmerksamkeit der Medien und der Frage, ob hier nicht der Zufall außer Kraft gesetzt sei. Diese Zahlenkombination hat aber die gleiche (geringe) Wahrscheinlichkeit wie jede andere.

4) Aus Dorothy L. Sayers "Keines natürlichen Todes", rororo, 1991:

   a) S. 58_8: Ein merkwürdiger Zufall, sagte er (der Chef von Scotland Yard) geduldig, und ich kann verstehen, dass Sie sich darüber aufregen.
   b) S. 62_{14}: Schon wieder ein Reinfall sagte Winsey: "Aber ein sonderbarer Zufall ist das schon."

5) Wenn jemand beim Skatspiel dreimal hintereinander alle vier Buben erhält, glaubt man nicht mehr an reinen Zufall, obwohl dieses Ereignis eine positive, wenn auch sehr kleine Wahrscheinlichkeit hat.

3.5 Elemente der Kombinatorik*

Bei der Abzählung der "günstigen" Fälle bei L-Experimenten erweisen sich Formeln der Kombinatorik häufig als günstig.
Wir geben hier vier Grundaufgaben der Kombinatorik an, sie werden häufig

auch in Form sogenannter Urnenprobleme formuliert.

Wir beginnen mit einer elementaren aber wichtigen Feststellung.

Aussage 3.19 *Es seien $M_1, \ldots, M_m$ m Mengen mit $m_1, \ldots, m_m$ Elementen. Dann hat die Menge M aller m-Tupel $(i_1, \ldots, i_m)$ mit $i_k \in M_k$ für $k = 1, \ldots, m$ genau $m_1, m_2, \ldots, m_m$ Elemente.*

Beweis: Mittels vollständiger Induktion

Als Nächstes kommen wir zu den vier angekündigten Aufgaben der Kombinatorik.
Aus einer Menge $M = \{a_1, a_2, \ldots, a_m\}$ von m Elementen $(m \geq 1)$ werden r Elemente ausgewählt, $r \geq 1$. Man spricht von einer Ŝtichprobe vom Umfang r aus der Menge M. Die Entnahme von Stichproben kann auf unterschiedliche Weise erfolgen:
Mit Wiederholung oder ohne Wiederholung
Mit Berücksichtigung der Reihenfolge oder ohne Berücksichtigung der Reihenfolge.
(d. h. geordnet oder ungeordnet)
Dementsprechend unterscheiden wir vier Fälle.

Anzahl von möglichen Stichproben des Umfanges r aus einer Menge M vom Umfang m

(In den Fällen ohne Wiederholung ist $r \leq m$ vorauszusetzen.)

	mit Wiederholung	ohne Wiederholung
geordnet	r-Tupel $(a_1, \ldots, a_r)$ mit $a_i \in M, i = 1, \ldots, r$ r-Permutation mit W, m^r **A1**	r-Tupel $(a_1, \ldots, a_r)$ mit $a_i \in M, i \in 1, \ldots, r)$, paarw. verschieden r-Permutation ohne W. $m(m-1)\ldots(m-r+1) =: (m)_r$ **A2**
ungeordnet	$[a_1, a_2, \ldots, a_r]$, Anordnung von r Elementen $\{a_1, \ldots, a_r\} \subset M$ $a_i \in M, i = 1, \ldots, r$ r-Kombination mit W. $\binom{m+r-1}{r}$ **A3**	 Teilmenge vom Umfang r r-Kombination ohne W. $\binom{m}{r} = \frac{(m)_r}{r!}$ **A4**

Die Fälle A1, A2 und A4 sind leicht zu beweisen.

Der Fall A3:
Jede ungeordnete Stichprobe vom Umfang r mit Wiederholung aus der Menge M ist eindeutig charakterisiert durch eine Folge $(i_1, i_2, \ldots, i_m)$ natürlicher Zahlen $i_k \geq 0$ mit $\sum_{k=1}^{m} i_k = r$, wobei i_k angibt, wie oft das Element a_k aus M in der Stichprobe vorkommt.

Diese Vektoren $(i_1, \ldots, i_m)$ lassen sich eineindeutig auf die Menge aller Anordnungen der Form $\bullet\bullet\bullet|\bullet\bullet|||\bullet\bullet\bullet$ von r Punkten und $(m-1)$ Strichen abbilden, wobei vor dem ersten Strich i_1 Punkte stehen, zwischen dem k-ten und $(k+1)$-ten Strich i_{k+1} Punkte stehen, und nach dem $(m-1)$-ten Strich i_m Punkte platziert sind. Insgesamt gibt es $\binom{m+r-1}{r}$ solcher Anordnungen.

Zu jedem der vier Fälle der Entnahme von Stichproben vom Umfang r aus einer Menge vom Umfang m gibt es ein sogenanntes "duales Problem" der Verteilung von r Kugeln auf m Urnen.

A_1: Duales Problem: r unterscheidbare Kugeln werden auf m Urnen verteilt, wobei in jeder Urne beliebig viele Kugeln liegen dürfen. Auf wie viel verschiedene Weisen ist dies möglich? Jede Verteilung der Kugeln ist charakterisiert durch die r Nummern $(a_1, a_2, \ldots, a_r)$ der Urnen, in die die erste zweite, ..., r-te Kugel zu liegen kommt. Für jede dieser Nummern gibt es m Möglichkeiten. Im Ergebnis entsteht wieder eine Stichprobe $(a_1, a_2, \ldots, a_r)$ vom Umfang r mit Wiederholung aus einer Menge vom Umfang m.

A_2: Duales Problem: r unterscheidbare Kugeln werden auf m Urnen verteilt, wobei in jeder Urne höchstens eine Kugel liegen darf. Auf wie viel verschiedene Weisen ist dies möglich? Jede Verteilung der Kugel ist charakterisiert durch die r Nummern $(a_1, a_2, \ldots, a_r)$ der Urnen, in die die erste, zweite k-te, r-te Kugel zu liegen kommt. Für die erste Kugel gibt es m, für die zweite $m-1$, für die r-te Kugel $m-r+1$ Möglichkeiten. Im Ergebnis entsteht wieder eine Stichprobe $(a_1, a_2, \ldots, a_r)$ vom Umfang r ohne Wiederholung aus einer Menge vom Umfang m.

A_3: Duales Problem: r ununterscheidbare Kugeln werden auf m Urnen aufgeteilt, wobei jede Urne auch mehrfach besetzt werden kann.
Jede Aufteilung ist charakterisiert durch die Anzahl i_k der Kugeln, die in die k-te Urne fallen, $k = 1, \ldots, m, \sum_{k=1}^{m} i_k = r$.

A_4: Duales Problem: r ununterscheidbare Kugeln sind auf m Urnen so aufzuteilen, dass in jeder Urne höchstens eine Kugel zu liegen kommt. Die Aufteilung ist charakterisiert durch die Menge $\{a_1, a_2, \ldots, a_r\}$ der Urnen, die durch eine Kugel besetzt werden, also durch eine r-elementige Teilmenge von M.

Anzahl der Aufteilungen von r Kugeln auf m Urnen

	mit Mehrfach- -besetzung	ohne Mehrfach- -besetzung
unterscheidbare Kugeln	m^r **A1**	$m(m-1)\dots(m-r+1)$ **A2**
ununterscheidbare Kugeln	$\binom{m+r-1}{r}$ **A3**	$\binom{m}{r}$ **A4**

Beispiele:

1. Wie groß ist die Wahrscheinlichkeit, beim Werfen von vier Würfeln mindestens eine "Sechs" zu erzielen?
   Der zufällige Versuch "Werfen von vier Würfeln" ist ein Laplace-Experiment mit 6^4 möglichen Ausgängen. Es bezeichne A das Ereignis "Es erscheint mindestens eine Sechs". Dann gibt es 5^4 günstige Ausgänge für das komplementäre Ereignis $\overline{A}$ = "Es erscheint keine Sechs". Also gilt

   $$P(A) = 1 - \left(\frac{5}{6}\right)^4 = 0,52.$$

2. Es werden k Kugeln auf n Urnen aufgeteilt, $k \leq n$. Jede Kugel habe die gleiche Wahrscheinlichkeit in jede Urne zu gelangen. Wie groß ist die Wahrscheinlichkeit für das Ereignis A, dass es nach der Aufteilung Urnen gibt, in der mehr als eine Kugel liegt?

   Lösung: Es gibt n^k Möglichkeiten der geschilderten Aufteilung und $(n)_k$ Möglichkeiten, die günstig sind für das komplementäre Ereignis $\overline{A}$ = "In keiner Urne liegt mehr als eine Kugel". Daraus folgt

   $$P(A) = 1 - \frac{(n)_k}{n^k}.$$

3. In einem Raum mögen sich k Personen befinden. Wie groß ist die Wahrscheinlichkeit für das Ereignis A, dass mindestens zwei dieser Personen am gleichen Tag Geburtstag haben? (Jeder Tag des Jahres komme bei

jeder Person mit gleicher Wahrscheinlichkeit als Geburtstag in Frage, Schaltjahre bleiben unberücksichtigt.)

Lösung: $\overline{A}$ = "Alle k Personen haben an verschiedenen Tagen Geburtstag". Es gibt $N = (365)^k$ möglich Fälle für Geburtstage und $N(\overline{A}) = (365)_k$ für $\overline{A}$ günstige Fälle, somit ist $P(A) = 1 - \frac{(365)_k}{365^k}$.

Diese Wahrscheinlichkeit wächst mit k und ist gleich $0,507$ für $k = 23$.

4. Koinzidenzproblem:
   n Briefe werden auf rein zufällige Weise in n adressierte Umschläge gesteckt. Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Brief in den richtigen Umschlag kommt?

   Lösung: Mögliche Versuchsausgänge $\omega = (a_1, \ldots, a_n)$ sind die Permutation von $(1, \ldots, n)$ mit a_k gleich der Nummer des Umschlages, in den der k-te Brief kommt. Wir setzen

$$A_k := \{\omega | a_k = k\}, k = 1, \ldots, n.$$

   Das interessierende Ereignis A ist gleich $\bigcup_{k=1}^{n} A_k$. Zur Anwendung der Ein- und Ausschlussformel berechnen wir

$$P(A_{k_1} \cap \ldots \cap A_{k_r}) = P\big(\{\omega = (a_1, \ldots, a_n) | a_{k_1} = k_1, \ldots, a_{k_r} = k_r\}\big) =$$

$$\frac{card\{\omega : a_{k_1} = k_1, \ldots, a_{k_r} = k_r\}}{n!} = \frac{(n-r)!}{n!}.$$

   Die Ein- und Ausschlussformel liefert

$$P(A) = \sum_{r=1}^{n} (-1)^{r+1} \binom{n}{r} \frac{(n-r)!}{n!} = \sum_{r=1}^{n} (-1)^{r+1} \frac{1}{r!} \underset{n\to\infty}{\longrightarrow} 1 - \frac{1}{\mathrm{e}} = 0,632$$

Für $n \geq 7$ ist die Näherung auf drei Stellen genau. (Weitere Bemerkungen zum Koinzidenzproblem und anderen kombinatorischen Aufgaben findet man in Henze, Kap. 9-11.)

5. Am Eröffnungstag eines großen Kongresses mit 500 Teilnehmern soll jeder teilnehmenden Person, die an diesem Tag Geburtstag hat, ein Blumenstrauß überreicht werden. Wie viele Sträuße braucht man mindestens, wenn man mit Sicherheit ausschließen will, dass man zu wenige Sträuße hat?
   Wie groß muss die Zahl der Sträuße mindestens sein, wenn man mit der Wahrscheinlichkeit von 0,95 diesen blamablen Fall vermeiden will?
   Wir nehmen näherungsweise an, dass für jede Person die Wahrscheinlichkeit, an einem bestimmten Tag Geburtstag zu haben, gleich ist für alle Tage des Jahres. Schaltjahre werden nicht berücksichtigt. Dann ist die Feststellung der Geburtstage aller Teilnehmer ein Laplace-Experiment mit den möglichen Ausgängen $\omega = (i_1, \ldots, i_{500})$, wobei i_k die Nummer des Tages angibt, an denen der k-te Teilnehmer Geburtstag hat. Die Menge aller möglichen Versuchsausgänge hat den Umfang $N = 365^{500}$. Es gibt nämlich $N = 365^{500}$ Möglichkeiten der Verteilung der Geburtstage der 500 Personen auf das Jahr.

   Für das Ereignis $A_k :=$ "Genau k Personen haben am Eröffnungstag Geburtstag" gibt es $N(A_k) = \binom{500}{k} \cdot 364^{500-k}$ "günstige" Versuchsausgänge.

   Es gilt $P(A_k) = \frac{N(A_k)}{N}$, und folglich ergibt sich

k	0	1	2	3	4	5
$P(A_k)$	0,2532	0,3478	0,2384	0,1087	0,0371	0,0101

   Deshalb ist

$$P\left(\bigcup_{k=1}^{5} A_k\right) = \sum_{k=1}^{5} P(A_k) = 0,9953.$$

Das heißt, mit an Eins grenzender Wahrscheinlichkeit haben höchstens fünf der Personen am Eröffnungstag Geburtstag.
(Zur Berechnung wurde die Näherung

$$P(A_k) \approx \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \text{ mit } \lambda = \frac{500}{365}$$

benutzt.)

3.6 Rein zufällige Wahl eines Punktes aus $[0, 1)$

Das Laplace-Experiment lässt sich nicht unmittelbar auf das in der Überschrift genannte Problem anwenden, da $[0, 1)$ unendlich viele Punkte enthält. Wir müssen hier den Begriff der "rein zufälligen Wahl" etwas modifizieren.

Rein zufällige Wahl soll bedeuten, dass für jedes Intervall $[a, b) \subseteq [0, 1)$ die Wahrscheinlichkeit, dass der gewählte Punkt aus $[a, b)$ stammt, unabhängig von der Lage des Intervalls sein soll. Das heißt

$$P([a, b)) = P([a + x, b + x)) \tag{3.18}$$

für alle x mit $a + x \geq 0, b + x \leq 1$.

Daraus folgt

$$P([a, b)) = b - a. \tag{3.19}$$

(Beweisen Sie (3.19).)

Es existiert allerdings kein Wahrscheinlichkeitsmaß P auf der Potenzmenge von $[0,1)$ mit den Eigenschaften (3.18) und (3.19), siehe Elstrodt, III. 2. Man kann aber zeigen, dass ein Wahrscheinlichkeitsmaß P mit (3.18) und (3.19) existiert auf der kleinsten σ-Algebra $\mathscr{B}_{[0,1)}$ von Teilmengen von $[0,1)$, die alle Intervalle der Form $[a,b)$ mit $0 \leq a < b \leq 1$ enthält (σ-Algebra der Borelmengen aus $[0,1)$). Dieses Maß ist eindeutig bestimmt und heißt Lebesgue-Borel-Maß auf $([0,1), \mathscr{B}_{[0,1)})$ oder einfach Lebesguemaß. Wir werden es mit $\lambda_{[0,1)}$ bezeichnen.

Die Tatsache, dass man $\lambda_{[0,1)}$ unter Beibehaltung von (3.18) nicht auf $\mathfrak{P}([0,1))$ erweitern kann, führt zu der auf den ersten Blick eigenartigen Situation, dass man nicht jede Teilmenge C von $[0,1)$ als zufälliges Ereignis bei der rein zufälligen Wahl eines Punktes aus $[0,1)$ ansehen kann.

Der Wahrscheinlichkeitsraum $([0,1), \mathscr{B}_{[0,1)}, \lambda_{[0,1)})$ ist das mathematische Modell des zufälligen Versuches, einen Punkt aus dem Intervall "rein zufällig" oder "auf gut Glück" auszuwählen.

Die Wahrscheinlichkeitsverteilung $\lambda_{[0,1)}$ "verteilt" die Gesamtwahrscheinlichkeit Eins "gleichmäßig" auf das Intervall $[0,1)$. Sie heißt *gleichmäßige Verteilung* auf $[0,1)$. Wir werden sie mit $U([0,1))$ bezeichnen. Insbesondere hat dann auch jeder Punkt $x \in [0,1)$ als Ereignis $\{x\}$ die gleiche Wahrscheinlichkeit, die folglich gleich Null sein muss. Das folgt auch aus $\{x\} = \bigcap_{k=1}^{\infty} \left[x, x + \frac{1}{k}\right)$ und der σ-Stetigkeit von $\lambda_{[0,1)}$.

In dem eben eingeführten Wahrscheinlichkeitsraum gibt es Ereignisse, die nicht unmöglich (bzw. nicht sicher) sind, aber dennoch die Wahrscheinlichkeit Null (bzw. Eins) haben. Das führt uns auf folgende Definition.

Definition 3.20 *Es sei* $(\Omega, \mathfrak{A}, P)$ *ein Wahrscheinlichkeitsraum. Jedes Ereignis* $A \in \mathfrak{A}$ *mit* $P(A) = 1 (P(A) = 0)$ *heißt* fast sicheres Ereignis *(bzw.* fast unmögliches Ereignis*).*

Bei der rein zufälligen Wahl eines Punktes aus $[0,1)$ ist das Ereignis $A :=$ "Es wird ein irrationaler Punkt gewählt" ein fast sicheres Ereignis und $\overline{A} =$ "Es wir ein rationaler Punkt gewählt" ein fast unmögliches Ereignis.

3.7 Zufallsgrößen

Unter einer Zufallsgröße versteht man umgangssprachlich eine Größe, die im Rahmen gewisser zufälliger Erscheinungen einen Wert annimmt, der nicht von vornherein feststeht, sondern vom Zufall abhängt. Beispiele findet man überall. In der Natur (Wetter), der Wirtschaft (Aktienkurse), der Technik (Ausfallzeitpunkte von Konsumgütern). Ihre mathematische Erfassung und Untersuchung ist ein zentraler Punkt der Wahrscheinlichkeitstheorie.

Im Allgemeinen sind zufällige Erscheinungen von sehr komplexer Natur. Man denke nur an das Wetter oder das Geschehen an einer Aktienbörse. Durch die Konzentration auf Zufallsgrößen, wie Tageshöchsttemperatur, monatliche Niederschlagsmenge bzw. Aktientagesschlusskurse oder wöchentliche Rendite bestimmter Unternehmen werden Teilaspekte der zugrunde liegenden zufälligen Prozesse herausgestellt, für die man sich besonders interessiert.

Es seien $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und X eine Abbildung von Ω in eine Menge E, z.B. $E = R_1$, $E = R_n$ oder $E = N_0 = \{0, 1, 2, \cdots, n, ...\}$. Indem man nicht den Versuchsausgang $\omega \in \Omega$ zur Kenntnis nimmt, sondern nur den Wert $X(\omega)$ beobachtet, den die Funktion X in Abhängigkeit von ω annimmt, ist ein neues zufälliges Experiment definiert mit möglichen Ausgängen $x = X(\omega)$, die aus E stammen. Die mit diesem neuen Experiment verbundenen Ereignisse sind nunmehr Teilmengen von E. Sie bilden eine σ-Algebra $\mathfrak{E}$ von Teilmengen von E. Das Ereignis B aus $\mathfrak{E}$ tritt für den neuen Versuch offenbar genau dann ein, wenn der ursprüngliche Versuch zu einem ω führt, für das $X(\omega) \in B$ gilt.

Beispiel 3.21 Wir betrachten das Laplace-Experiment des gleichzeitigen Werfens zweier regulärer Würfel und wählen

$$\Omega = \{\omega = (i, j) : i, j \in \{1, 2, \ldots, 6\}\}, \quad \mathfrak{A} = \mathfrak{P}(\Omega),$$

$$P(\{\omega\}) = 36^{-1}, \quad \omega \in \Omega,$$

$X(\omega) = i + j \quad , \quad \omega = (i, j) \in \Omega.$

Hier wählt man den Bildraum E als die Menge $\{2, 3, \cdots, 12\}$ und für $\mathfrak{E}$ die Potenzmenge $\mathfrak{P}(E)$.

Die Funktion X gibt also die Augensumme der zwei geworfenen Würfel an.

Das Ereignis "Augensumme ist gleich 4" entspricht der Menge $\{4\}$ aus E und tritt genau dann ein, wenn ein $\omega = (i, j)$ mit $i + j = 4$ Ergebnis des Würfelns ist, also wenn $(1, 3), (2, 2)$ oder $(3, 1)$ gewürfelt wurde.

Wir kehren zurück zum allgemeinen Fall und wollen auf $(E, \mathfrak{E})$ eine Wahrscheinlichkeitsverteilung P^X einführen, die den Ereignissen des neuen Versuches ihre Wahrscheinlichkeiten zuordnet. Das geschieht durch

$$P^X(B) := P(\{\omega \in \Omega | X(\omega) \in B\}) = P(X^{-1}(B)),\ B \in \mathfrak{E} \tag{3.20}$$

Üblicherweise ist $\mathfrak{E}$ zusammen mit E von vornherein festgelegt. Damit die Definition von P^X dann sinnvoll ist, müssen wir eine Forderung an X stellen, die wir in der nächsten Definition formulieren.

Definition 3.22 *Die Abbildung X von $(\Omega, \mathfrak{A})$ in $(E, \mathfrak{E})$ heißt eine* Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ mit Werten in $(E, \mathfrak{E})$*, falls gilt*

$$X^{-1}(B) := \{\omega \in \Omega | X(\omega) \in B\} \in \mathfrak{A}, \textit{für alle } B \textit{ aus } \mathfrak{E}, \tag{3.21}$$

m.a.W., falls die Abbildung X (in der Sprache der Maßtheorie)eine $\mathfrak{A} - \mathfrak{E}$-messbare Abbildung ist.

Die Eigenschaft (3.21) kann man kurz schreiben als

$$X^{-1}(\mathfrak{E}) \subseteq \mathfrak{A}. \tag{3.22}$$

Notation 3.23 Für $\{\omega \in \Omega | X(\omega) \in B\}$ schreiben wir häufig kürzer $\{X \in B\}$, und statt $P^X(B)$ bzw. $P(X^{-1}(B))$ verwenden wir die Schreibweise $P(X \in B)$.

Aussage 3.24 *Durch*

$$P^X(B) := P(X \in B), \qquad B \in \mathfrak{E} \tag{3.23}$$

ist auf $\mathfrak{E}$ *eine Wahrscheinlichkeitsverteilung* P^X *gegeben. Die Verteilung* P^X *nennt man die* Wahrscheinlichkeitsverteilung der Zufallsgröße X *oder die* durch X induzierte Wahrscheinlichkeitsverteilung.

Zum Beweis prüft man die Axiome A1.-A2.' nach.

Das zufällige Experiment "Beobachtung der Zufallsgröße X" wird also mathematisch modelliert durch den Wahrscheinlichkeitsraum $(E, \mathfrak{E}, P^X)$.

Beispiel 3.25 (Fortsetzung von Beispiel 3.21)
Für die Wahrscheinlichkeitsverteilung P^X der Zufallsgröße X gelten mit

$$E = \{2, \ldots, 12\}, \quad \mathfrak{E} = \mathfrak{P}(E)$$

die Gleichungen

$$P^X(\{k\}) = P(\{\omega \in \Omega | X(\omega) = k\}) =$$

$$\frac{\#\{\omega = (i,j) | i + j = k\}}{36} = \frac{6 - |7 - k|}{36}, \quad k \in E$$

und

$$P^X(B) = \sum_{k \in B} P^X(\{k\}), \quad B \in \mathfrak{E}. \tag{3.24}$$

Die Forderung $X^{-1}(\mathfrak{E}) \subseteq \mathfrak{A}$ ist in diesem Fall natürlich erfüllt, da $\mathfrak{A} = \mathfrak{P}(\{1, 2, \ldots, 6\}^2)$.

Aus vorgegebenen Zufallsgrößen kann man durch eine Vielzahl von Operationen neue Zufallsgrößen bilden. Exemplarisch erwähnen wir hier einige Fälle. Sie sind in den beiden folgenden Aussagen enthalten.

Aussage 3.26 *Ist* $(E, \mathfrak{E}) = (R_n, \mathfrak{B}_n)$*, so sind mit Zufallsgrößen* $X, X_1, X_2, \cdots$ *über* $(\Omega, \mathfrak{A}, P)$ *auch die Vielfachen* $aX, a \in R_1$*, die Summen* $\sum_{k=1}^{n} X_k$*, die Funktionen* $\max(X_1, X_2, \cdots, X_n) \min(X_1, X_2, \ldots, X_n)$*, der Grenzwert* $\lim_{n\to\infty} X_n$ *(sofern er existiert) wieder Zufallsgrößen. (Die Operationen verstehen sich dabei punktweise, also* ω*-weise.)*
Diese Tatsache ergibt sich sofort aus den entsprechenden Eigenschaften messbarer Funktionen, die in der Maßtheorie bewiesen werden.

Aussage 3.27 *Es seien* $(\Omega, \mathfrak{A}, P)$ *ein Wahrscheinlichkeitsraum,* $(E, \mathfrak{E})$ *und* $(F, \mathfrak{F})$ *messbare Räume.*
Ist X *eine Zufallsgröße über* $(\Omega, \mathfrak{A}, P)$ *mit Werten in* $(E, \mathfrak{E})$ *und ist* Y *eine Zufallsgröße über* $(E, \mathfrak{E}, P^X)$ *mit Werten in* $(F, \mathfrak{F})$*, so ist* $Z = Y \circ X$ *eine Zufallsgröße über* $(\Omega, \mathfrak{A}, P)$ *mit Werten in* $(F, \mathfrak{F})$ *und ihre Wahrscheinlichkeitsverteilung* P^Z *auf* $\mathfrak{F}$ *ist gegeben durch*

$$\begin{aligned} P^Z(C) = P(Z \in C) = P(X \in Y^{-1}(C)) = \\ P^X(Y^{-1}(C)) = P(X^{-1}(Y^{-1}(C))), C \in \mathfrak{F}. \end{aligned} \tag{3.25}$$

Beweis: Nach Definition gilt $Y^{-1}(\mathfrak{F}) \subseteq \mathfrak{E}$ und $X^{-1}(\mathfrak{E}) \subseteq \mathfrak{A}$, folglich ist $Z^{-1}(\mathfrak{F}) = X^{-1}(Y^{-1}(\mathfrak{F})) \subseteq \mathfrak{A}$. Also (siehe (3.22)) ist Z eine Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ mit Werten in $(F, \mathfrak{F})$. Die Formel (3.25) ergibt sich unmittelbar aus der Definition der Verteilung von Z. Man beachte die Notation 3.23. □

Definition 3.28 *Jede Zufallsgröße über einem Wahrscheinlichkeitsraum* $(\Omega, \mathfrak{A}, P)$ *mit Werten in* $(R_1, \mathscr{B}_1)((R_n, \mathscr{B}_n))$ *heißt eine* reellwertige Zufallsgröße *(bzw. ein* n-dimensionaler zufälliger Vektor).
$\mathfrak{B}_1$ *bzw.* $\mathfrak{B}_n$ *bezeichnen dabei die* σ*-Algebren der Borelmengen aus* R_1 *bzw.* R_n*.*

Beispiel 3.29 *In Abschnitt 3.3 haben wir die symmetrische Irrfahrt* $(S_0, S_1, \ldots, S_n)$ *kennen gelernt. Definiert man die reellwertige Zufallsgröße* $T{-}1$ *durch*

$$T_{-1}(\omega) = \min\{K \leq n : S_k(\omega) = -1\},\ \omega \in \Omega = \{-1, -1\}^n$$

,

mit $\min\emptyset := \infty$*, so hat* T_{1-} *die möglichen Werte* $1, 3, 5, \ldots, 2\left[\frac{n+1}{2}\right] - 1, \infty$ *und es gilt (siehe (3.17))*

$$\left.\begin{array}{l} P(T_{1-} = 2m+1) = q_m, \qquad m = 1, 2, \ldots \left[\frac{n+1}{2}\right] - 1 \\ \textit{mit } q_m = \frac{(2m-1)!!}{(m+1)!2^{m+1}},\ m \geq 1, \textit{ und} \\ P(T_{1-} = 1) = \frac{1}{2},\ P(T_{-1} = \infty) = \sum\limits_{m \geq \left[\frac{n+1}{2}\right]} q_m. \end{array}\right\} \tag{3.26}$$

Die Zufallsgröße T_{-1} gibt den Zeitpunkt an, zu dem der Spieler A zum ersten Mal einen negativen Gewinn verbucht, also den Zeitpunkt seines Ruines, falls er kein zusätzliches Kapital besitzt. Das Ereignis $\{T_{-1} = \infty\}$ tritt ein, falls er nach n-maligem Werfen der Münze noch nicht "ruiniert" ist.

Wenn er unbegrenzt lange spielt, ergibt sich

$$P(T_{-1} < \infty) = \sum_{m=0}^{\infty} P(T_{-1} = 2m+1) = \sum_{m=0}^{\infty} q_m = 1.$$

(Den Beweis der letzten Gleichung führen wir später.)

Das bedeutet, bei unbegrenzter Fortführung des Münzwurfes wird der Spieler A mit Wahrscheinlichkeit Eins irgendwann "ruiniert", d. h. sein Guthaben wird irgendwann negativ.

Völlig analog kann man aber auch schlussfolgern, dass er mit Wahrscheinlichkeit Eins irgendwann mindestens einen Betrag der Größe Eins auf seinem Konto hat. (Wir setzen dabei voraus, dass er in der Zwischenzeit, wenn sein Guthaben im Negativen ist, immer genügend Finanzmittel besitzt, das Spiel fortzusetzen.) Wenn er die Strategie verfolgt, in dem Moment aufzuhören zu spielen, wenn er das erste Mal einen Gesamtgewinn der Höhe Eins hat, so gewinnt er bei dem vereinbarten durchaus fairen Spiel des Münzenwurfes ohne Zeitlimit mit Wahrscheinlichkeit Eins eine Geldeinheit. Ein Paradoxon.

3.8 Verteilungsfunktionen

Verteilungsfunktionen auf R_1

Definition 3.30 *Ist Q eine Wahrscheinlichkeitsverteilung auf $R_1, \mathscr{L}_1)$, so bezeichnet man die durch*

$$F(x) := Q((-\infty, x]), \quad x \in R_1 \tag{3.27}$$

auf R_1 gegebene Funktion F als Verteilungsfunktion der Verteilung Q.
Ist X eine reellwertige Zufallsgrß̈e über einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$, so nennt man die zu P^X (siehe (3.19) und (3.22)) gehörende Verteilungsfunktion F die Verteilungsfunktion der Zufallsgröße X. Gegebenenfalls schreibt man F_Q bzw. F_X an Stelle F.

Es gilt

$$F(b) - F(a) = Q((a, b]),\ a < b, \tag{3.28}$$

$$F_X(x) = P^X((-\infty, x]) = P(X \leq x) \tag{3.29}$$

Es sei Q eine Wahrscheinlichkeitsverteilung auf $(R_1, \mathscr{L}_1)$.

Aussage 3.31 *Die Verteilungsfunktion F der Verteilung Q hat folgende Eigenschaften:*

1. *F ist monoton nichtfallend: $x \leq y \Longrightarrow F(x) \leq F(y)$,*
2. *$\lim\limits_{x \to -\infty} F(x) = 0, \lim\limits_{x \to \infty} F(x) = 1$,*
3. *F ist an jeder Stelle $x \in R_1$ von rechts stetig:*

   $F(x+0) := \lim\limits_{y \downarrow x} F(y) = F(x)$,

4. *Für jedes $x \in R_1$ gilt mit $F(x-0) := \lim\limits_{y \uparrow x} F(y)$*

   $$F(x) - F(x-0) = Q\{x\}).$$

Beweis: Unter Verwendung von (3.27), Lemma 3.6 und Folgerung 3.7 haben wir:

1. $x \leq y \Longrightarrow (-\infty, x] \subseteq (-\infty, y] \Longrightarrow F(x) \leq F(y)$,
2. $x_n \downarrow -\infty \Longrightarrow (-\infty, x_n] \downarrow \emptyset \Longrightarrow \lim\limits_{n\to\infty} F(x_n) = 0$,

   $x_n \uparrow \infty \Longrightarrow (-\infty, x_n] \uparrow R_1 \Longrightarrow \lim\limits_{n\to\infty} F(x_n) = 1$,
3. Wegen $(-\infty, x] = \bigcap\limits_n (-\infty, x_n]$ für jede Folge (x_n) mit $x_n \downarrow x$ folgt $\Longrightarrow$ $F(x_n) \downarrow F(x)$,
4. $(x_n, x] \downarrow \{x\}$ für jede Folge (x_n) mit $x_n \uparrow x$.

   $$\Longrightarrow F(x) - F(x-0) := \lim_{n\to\infty} (F(x) - F(x_n)) = \lim_{n\to\infty} Q((x_n, x]) = Q(\{x\}).$$ □

Definition 3.32 *Jede Funktion F und R_1 mit den Eigenschaften 1. - 3. aus Aussage 3.31 heißt eine* Verteilungsfunktion auf R_1*.*

Es sei F eine Verteilungsfunktion auf R_1, d. h. eine Funktion mit den Eigenschaften 1. - 3. aus Aussage 3.31.

Aussage 3.33 *Es gibt eine eindeutig bestimmte Wahrscheinlichkeitsverteilung Q auf $(R_1, \mathscr{B}_1)$, die F als Verteilungsfunktion besitzt.*

Beweis: Wir setzen

$$Q((a,b]) := F(b) - F(a), \quad a < b.$$

Das Mengensystem $\gamma = \{(a,b] | -\infty < a < b < \infty\}$ ist ein Semiring, und die da durch definierte nichtnegative Mengenfunktion Q ist auf γ σ-stetig (d. h., ist

($(a_n, b_n]$ eine monoton fallende Folge halboffener Intervalle mit $\bigcap_{n\geq 1}(a_n, b_n] = \emptyset$, so haben wir $\lim_{n\to\infty} Q((a_n, b_n]) = 0$).
(Der Beweis ist sehr technisch, siehe z. B. Siraev, II § 3, Punkt 1). Außerdem gilt $Q(R_1) = 1$.

In der Maßtheorie wird gezeigt, dass man jedes Q mit diesen Eigenschaften auf eine und nur eine Weise zu einem Wahrscheinlichkeitsmaß, das wir ebenfalls mit Q bezeichnen, auf die σ-Algebra $\mathfrak{B}_1 = \sigma(\gamma)$ der Borelmengen aus R_1 erweitern kann. Wegen

$$Q((-\infty, x]) = \lim_{n\to\infty} Q((-n, x]) = \lim_{n\to\infty} (F(x) - F(-n)) = F(x)$$

folgt die Behauptung. □

Beispiel 3.34

a) Ist $(\Omega, \mathfrak{A}, P)$ ein Laplace-Experiment mit $\Omega = \{w_1, w_2, \cdots w_N\}, \mathfrak{A} = \mathfrak{P}(\Omega)$,
$P(\{\omega\}) = \frac{1}{N}, \omega \in \Omega$, und ist X eine reellwertige Zufallsgröße über $(\Omega, \mathfrak{A}, P)$ mit $X(\omega_k) = x_k, k = 1, \cdots, N, \quad x_k \neq x_j$ für $k \neq j$, so lautet die Verteilungsfunktion $F = F_X$ wie folgt:

$$F(x) = \tfrac{1}{N} \sum_{k:x_k\leq x} \mathbb{1} = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{(-\infty, x_k]}(x), x \in R_1$$

F ist in diesem Fall eine stückweise konstante rechtsseitig stetige, nichtfallende Funktion auf R_1 mit den Sprungstellen x_k, den Sprunghöhen $\frac{1}{N}$ und $F(x) = 0$ für $x < \min_{k=1,\cdots,N} x_k$ und $F(x) = 1$ für $x \geq \max_{k=1,\ldots,N} x_k$.

b) Ist $(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}_{[0,1)}, \lambda_{[0,1)})$ der Wahrscheinlichkeitsraum, der die rein zufällige Wahl eines Punktes ω aus $[0,1)$ modelliert, und ist $X(\omega) = \omega, \omega \in \Omega$, so gilt für $F = F_X$

$$F(x) = (x \wedge 1) \vee 0 = \begin{cases} 0 & x \leq 0 \\ x & x \in [0,1] \\ 1 & x \geq 1 \end{cases}$$

Eine Möglichkeit, sich von der Lage einer Wahrscheinlichkeitsverteilung auf R_1 und ihre Ausbreitung eine Vorstellung zu verschaffen, besteht in der Berechnung ihrer Quantile.

Definition 3.35 *Es seien Q eine Wahrscheinlichkeitsverteilung auf $(R_1, \mathfrak{B}_1)$ und F ihre Verteilungsfunktion:*

$$F(x) = Q((-\infty, x]), \quad x \in R_1.$$

Weiterhin sei p irgendeine Zahl mit $0 < p < 1$.
*Als p-*Quantil *der Verteilung Q bezeichnet man jede Zahl $q_p \in R_1$ mit*

$$F(q_p - 0) \leq p \leq F(q_p). \tag{3.30}$$

Aussage 3.36 *Für jedes $p \in (0,1)$ ist die Menge aller p-Quantile von Q ist nichtleer und bildet ein beschränktes abgeschlossenes Intervall. Sie ist einelementig genau dann, wenn es keine zwei Zahlen $x < y$ gibt mit $F(x) = F(y) = p$.*

Definition 3.37 *Jedes Quartil $q_{\frac{1}{2}}$ heißt* Median. *Jedes Quartil $q_{\frac{1}{4}}$ $q_{\frac{3}{4}}$ heißt* unteres *(oberes) Quartil.*

Die Differenz $q_{\frac{3}{4}} - q_{\frac{1}{4}}$ ist ein Maß für die "Ausbreitung" der Verteilung Q. Es gilt $Q([q_{\frac{1}{4}}, q_{\frac{3}{4}}]) = Q((-\infty, q_{\frac{3}{4}}]) - Q((-\infty, q_{\frac{1}{4}})) \geq \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$.

Das heißt, zwischen $q_{\frac{1}{4}}$ und $q_{\frac{3}{4}}$ befindet sich mindestens die Q-"Wahrscheinlichkeitsmaße" $\frac{1}{2}$.

Ein Median ist ein Wert, den man als zentrum der Wahrscheinlichkeitsverteilung Q bezeichnen kann.

Aussage 3.38 *Ist F eine streng wachsende (nicht notwendige stetige) Verteilungsfunktion auf R_1, so existiert zu jedem $p \in (0,1)$ das eindeutig bestimmte Quantil q_p und es gilt*

$$q_p = F^{-1}(p) = (p \in (0,1)) \tag{3.31}$$

wobei $F^{-1}(p) := \inf\{x|F(x) > p\}$ (rechtsstetige Inverse) gesetzt wird.

Verteilungsfunktionen auf R_n

Es seien $n \geq 2$ und Q eine Wahrscheinlichkeitsverteilung auf $(R_n, \mathscr{B}_n$.

Definition 3.39 *Mit der Bezeichnung $x = (x_1, x_2, \ldots, x_n)^T \in R_n$ und $(-\infty, x] = \prod_{k=1}^{n}{}^{\otimes} (-\infty, x_k]$ ist durch*

$$F(x) = Q((-\infty, x]), \quad x \in R_n$$

eine Funktion auf R_n definiert, die Verteilungsfunktion der Wahrscheinlichkeitsverteilung Q.
Ist $X = (X_1, X_2, \ldots, X_n)$ ein zufälliger Vektor über einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Werten in R_n, so nennt man die Verteilungsfunktion F der Verteilung P^X auch Verteilungsfunktion von X. In diesem Fall gilt für alle $x = (x_1, \ldots, x_n)^T \in R_n$ die Beziehung

$$F(x) = P^X((-\infty, x]) = P(X \in (-\infty, x]) =$$

$$P(X_1 \in (-\infty, x_1], \ldots, X_n \in (-\infty, x_n]) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \tag{3.32}$$

Aussage 3.40 *Die Verteilungsfunktion F der Wahrscheinlichkeitsverteilung Q hat folgende Eigenschaften:*

1. $0 \leq F(x) \leq 1, x = (x_1, x_2, \ldots, x_n)^r \in R_n$,
2. $\lim\limits_{x_k \downarrow -\infty} F(x_1, \cdots, x_n) = 0$ *für jedes $k = 1, \cdots, n$,*

3. $\lim\limits_{x_1,\cdots,x_n\uparrow\infty} F_X(x_1,\cdots,x_n) = 1$

4. *F ist an jeder Stelle $x = (x_1, x_2, \ldots, x_n)^T \in R_n$ von rechts stetig:*
   $\lim\limits_{\substack{h_i\downarrow 0\\ i=1,\cdots,n}} F(x_1+h_1,\cdots,x_n+h_n) = F(x_1,\cdots,x_n)$,

5. *Mit der Definition*
   $\triangle_{h_i}F(x) = F(x_1,\cdots,x_{i-1},x_i+h_i,x_{i+1},\cdots,x_n) - F(x_1,\cdots,x_n)$ *gilt*
   $0 \le \triangle_{h_1}\cdot\ldots\cdot\triangle_{h_n}F(x) x \in R_n, h_i \ge 0, i_1,\ldots,n$
   (Verallgemeinerung der Monotonie im Fall $n = 1$).

Bemerkung 3.41 Für $n = 2$ lautet die Eigenschaft 5. wie folgt:

$$\begin{aligned} F_X(x_1+h_1,x_2+h_2) \; &- \; F_X(x_1,x_2+h_2) - F_X(x_1+h_1,x_2) \\ &+ \; F_X(x_1,x_2) \ge 0. \end{aligned} \tag{3.33}$$

Definition 3.42 *Jede Funktion F auf R_n mit den Eigenschaften 1.-5. aus Aussage 3.40 nennen wir eine* Verteilungsfunktion *auf R_n.*

Zu jeder Verteilungsfunktion F auf R_n definieren wir für alle $a = (a_1, a_2, \ldots, a_n)^T \in R_n, h_1, h_2, \ldots, h_n > 0$ durch

$$Q\Big(\prod_{k=1}^{n}{}^{\otimes}(a_k, a_k+h_k]\Big) = \triangle_{h_1},\triangle_{h_2}\ldots\triangle_{h_n}F(a) \tag{3.34}$$

eine Mengenfunktion Q auf dem Semiring γ_n aller n-dimensionalen "Quader" $\prod\limits_{k=1}^{n}{}^{\otimes}(a_k, b_k] \subseteq R_n$.

Aussage 3.43 *Die durch (3.34) definierte Mengenfunktion Q ist auf γ_n σ-additiv und lässt sich auf eine und nur eine Weise zu einer Wahrscheinlichkeitsmaverteilung auf $\mathscr{B}_n$ fortsetzen, das wir wiederum mit Q bezeichnen. Die Verteilung Q besitzt F als Verteilungsfunktion.*

Zum Beweis, der rein maßtheoretischer Natur ist, sei auch hier auf Siraev (...), II, §3, verwiesen.

Beispiel 3.44

a) Die Funktion F, definiert durch

$$F(x_1, x_2) = [(x_1 \wedge x_2) \wedge 1] \vee 0, \quad (x_1, x_2)^T \in R_2, \tag{3.35}$$

ist eine Verteilungsfunktion auf R_2.

Beweis: Die Eigenschaften 1.-4. aus Aussage 3.40 sind offensichtlich. Zum Nachweis von 5. bemerken wir zunächst, dass für jedes Rechteck $(x_1, x_1 + h_1] \times (x_2, x_2 + h_2]$, das disjunkt zu $\{(x, x) : 0 < x \leq 1\}$ ist, gilt $x_2 \geq x_1 + h_1$ oder $x_1 \geq x_2 + h_2$. Daraus folgt für diese Rechtecke

$$\triangle_{h_1} \triangle_{h_2} F(x_1, x_2) = 0.$$

Andererseits ist für jedes Rechteck $(x, x + h] \times (x, x + h]$

$$\triangle_h \triangle_h F(x, x) = (x + h - x - x + x) = h > 0$$

.

Mit Hilfe der Additivität von Q auf γ_2 ergibt sich 5. □

b) Sind $F_k, k = 1, 2, \ldots, n$ Verteilungsfunktionen auf R_1, so ist F definiert auf R_n durch

$$F(x) = \prod_{k=1}^{n} F_k(x_k), \quad x = (x_1, x_2, \ldots, x_n)^T \in R_n$$

eine Verteilungsfunktion auf R_n.

Es seien Q eine Wahrscheinlichkeitsverteilung auf $(R_n, \mathscr{B}_n)$ und F ihre Verteilungsfunktion.

Definition 3.45 *Für jede r-elementige Teilmenge $\mathscr{J}_r = \{k_1, k_2, \ldots, k_r\}$ von $\{1, 2, \ldots, n\}$ bezeichne $\prod_{\mathscr{J}_r}$ den Projektionsoperator, definiert durch*

$$\prod_{\mathscr{J}_r} x = (x_{k_1}, x_{k_2}, \ldots, x_{k_r})^T \in R_r, \; x = (x_1, \ldots, x_n)^T \in R_n.$$

Offenbar ist $\Pi_{\mathscr{J}_r}$ eine $\mathscr{B}_n - \mathscr{B}_r$-meßbare Abbildung.

Die durch

$$Q_{\mathscr{J}_r}(B) := Q(\prod_{\mathscr{J}_r}^{-1}(B)), \quad B \in \mathscr{L}_r \tag{3.36}$$

definierte Mengenfunktion $Q_{\mathscr{J}_r}$ auf $\mathscr{B}_r$ ist eine Wahrscheinlichkeitsverteilung und heißt die zu $\mathscr{J}_r$ gehörende r-dimensionale Randverteilung *von Q.*

Aussage 3.46 *Die Verteilungsfunktion $F_{\mathscr{J}_r}$ der Verteilung $Q_{\mathscr{J}_r}$ hängt mit der Verteilungsfunktion F wie folgt zusammen:*

$$F_{\mathscr{J}_r}(x_{k_1}, x_{k_2}, \ldots, x_{k_r}) =$$

$$F(\infty, \ldots, \infty, x_{k_1}, \infty, \ldots, x_{k_2}, \ldots, \infty, x_{k_r}, \infty, \ldots, \infty) \tag{3.37}$$

Beweis:

$$F_{\mathscr{J}_r}(x_{k_1}, x_{k_2}, \ldots, x_{k_r}) = Q_{\mathscr{J}_r}\Big(\prod_{l=1}^{r}{}^{\otimes} (-\infty, x_{k_l}]\Big) =$$

$$Q(\prod_{\mathscr{J}_r}^{-1}\Big(\prod_{l=1}^{r}{}^{\otimes} (-\infty, x_{k_l}]\Big) = Q\Big(\prod_{m=1}^{n}{}^{\otimes} B_m\Big) \text{ mit}$$

$$B_m = (-\infty, x_{k_l}], \text{ falls } m = k_l \text{ für ein } l = 1, 2, \ldots, r,$$

$$B_m = (-\infty, \infty)\text{falls } m \neq k_l \text{ für alle } l = 1, 2, \ldots, r.$$

$\square$

Aus der Kenntnis der Randverteilungsfunktionen $F_{\mathscr{J}_r}$ mit $r < n$ kann die Verteilungsfunktion F selbst i.a. nicht rekonstruiert werden.

Zum Beispiel haben die beiden Verteilungsfunktionen

$$G(x_1, x_2) = ((x_1 \wedge x_2) \wedge 1) \vee 0 \text{ und}$$

$$H(x_1, x_2) = [(x_1 \wedge 1) \vee 0][(x_2 \wedge 1) \vee 0]$$

(siehe Beispiel 3.42)
die gleichen Randverteilungsfunktionen:

$$G(x_1, \infty) = (x_1 \wedge 1) \vee 0 = H(x_1, \infty)$$

$$G(\infty, x_2) = (x_2 \wedge 1) \vee 0 = H(\infty, x_2).$$

Ist $X = (X_1, \ldots, X_n)^T$ ein n-dimensionaler zufälliger Vektor mit der Verteilungsfunktion F, so ist $F_{\mathscr{J}_r}$ die Verteilungsfunktion des Vektors $(X_{k_1}, \ldots, X_{k_r})^T$ wobei $\mathscr{J}_r = \{k_1, \ldots, k_r\}$ gilt. Das ergibt sich einfach aus (3.37) und

$$F(\infty, \ldots, \infty, x_{k_1}, \infty, \ldots, x_{k_2}, \ldots, x_{k_r}, \infty, \ldots, \infty) =$$

$$P(X_1 < \infty, \ \ldots, X_{k_1} \leq x_{k_1}, \ldots, X_{k_r} \leq x_{k_r+1} < \infty, \ldots, X_n < \infty) =$$

$$P(X_{k_1} \leq x_{k_1}, \ldots, X_{k_n} \leq x_{k_n}).$$

3.9 Verteilungsdichten

Wir haben gesehen, dass sich Wahrscheinlichkeitsverteilungen auf $(R_1, \mathfrak{B}_1)$ bzw. $(R_n, \mathfrak{B}_n)$ durch ihre Verteilungsfunktionen charakterisieren lassen. Diese wiederum besitzen in Spezialfällen eine besonders einfache Struktur. Zum einen handelt es sich dabei um sogenannte *diskrete Verteilungen* bzw. *diskret verteilte Zufallsgrößen*. Diesen Verteilungen ist das Kapitel 4 gewidmet. Zum anderen heben sich Verteilungen mit einer *Verteilungsdichte*, man spricht auch einfach von *Dichten*, heraus. Diese Verteilungen werden wir in voller Allgemeinheit unter Verwendung des Begriffs des Lebesgueintegrals erst in Kapitel 7 behandeln. Vorab wollen wir jedoch einige wichtige Fälle, in denen man mit dem bereits bekannten Riemannintegral auskommt, vorstellen.

Verteilungsdichten auf R_1

Es sei F eine Verteilungsfunktion auf R_1 (vgl. Definition 3.32).

Definition 3.47 *Gibt es eine stückweise stetige Funktion f auf R_1 mit*

1. $$f(x) \geq 0,\ x \in R_1, \tag{3.38}$$

2. $$\int\limits_a^b f(x)dx = F(b) - F(a), \quad a, b \in R_1, \tag{3.39}$$

so heißt f eine Dichte der Verteilungsfunktion F.
Ist X eine reellwertige Zufallsgröße und f_X eine Dichte ihrer Verteilungsfunktion F_X, so heißt f_X auch Dichte der Zufallsgröße X.

Aussage 3.48

a) *Für jede Dichte f gilt*

$$\int\limits_{-\infty}^x f(s)ds = F(x) \tag{3.40}$$

$$\int\limits_{-\infty}^{\infty} f(x)dx = 1, \tag{3.41}$$

b) *besitzt F eine Dichte f, so ist F stetig,*

c) *besitzt F eine Dichte f, die stetig in einer Umgebung von x ist, so ist F differenzierbar in diesem x_1, und es gilt*

$$\frac{d}{dx}F(x) = f(x) \tag{3.42}$$

Beweis:

a) Folgt aus (3.39) für $a \to -\infty, b = x$ bzw. $a \to -infty$ und $b \to \infty$ sowie der Eigenschaft 2. aus Aussage 3.31.

b) $F(x+h) - F(x) = \int\limits_x^{x+h} f(s)ds \underset{h \to 0}{\longrightarrow} 0,$

c) nach Voraussetzung gilt es ein $\delta > 0$, so dass f stetig ist in $(x-\delta, x+\delta)$. Für jedes h mit $|h| < \delta$ gilt

$$F(x+h) - F(x) = \int_x^{x+h} f(s)ds = h \cdot f(\xi)$$

für ein ξ zwischen x und $x+h$. Daraus und aus der Stetigkeit von f in $(x-\delta, x+\delta)$ folgt c). □

Beispiel 3.49

a) Die Verteilungsfunktion $F(x) = (x \wedge 1) \vee 0, x \in R_1$, (siehe Beispiel 3.33b)) besitzt die Dichte

$$f(x) = \mathbb{1}_{[0,1]}(x), \quad x \in R_1$$

Die Verteilungsfunktion $F(x)$ aus Beispiel 3.33a) besitzt keine Dichte.

b) Es sei $\lambda > 0$ und f_λ die durch

$$f_\lambda(x) = \left\{ \begin{array}{cc} 0 & , x < 0 \\ \lambda e^{-\lambda x} & , x \geq 0 \end{array} \right\} = \mathbb{1}_{[0,\infty)}(x) \lambda e^{-\lambda x}, \ x \in R_1$$

definierte Funktion. Dann ist f_λ die Dichte einer Wahrscheinlichkeitsverteilung auf $(R_1, \mathfrak{B}_1)$, die man als *Exponentialverteilung* $EXP(\lambda)$ *mit dem Parameter* λ bezeichnet. Ihre Verteilungsfunktion F_λ lautet

$$F_\lambda(x) = \left\{ \begin{array}{cc} 0 & , x \leq 0, \\ 1 - e^{-\lambda x} & , x > 0. \end{array} \right.$$

c) Die *Normalverteilung* $N(\mu, \sigma^2)$ ist für jedes $\mu \in R_1$ und jedes $\sigma^2 > 0$ definiert als die Verteilung mit der Dichte

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (x-\mu)^2 \right], x \in R_1. \tag{3.43}$$

Die zugehörige Verteilungsfunktion

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(u)du, \quad x \in R_1,$$

ist nicht explizit durch elementare Funktionen ausdrückbar.

Die Verteilung $N(0,1)$ heißt *Standardnormalverteilung*. Die Werte ihrer Verteilungsfunktion $\Phi_{0,1}$ sind vertafelt. Es bestehen folgende Beziehungen ($\sigma = \sqrt{\sigma^2} > 0$)

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma}\,\varphi_{0,1}\left(\frac{x-\mu}{\sigma}\right), \quad \Phi_{\mu,\sigma^2}(x) = \Phi_{0,1}\left(\frac{x-\mu}{\sigma}\right) \tag{3.44}$$

Statt $\varphi_{0,1}$ und $\Phi_{0,1}$ schreiben wir auch einfach φ bzw. Φ, falls keine Verwechslungen möglich sind.

Aussage 3.50 *Die Dichte φ und die Verteilungsfunktion Φ der $N(0,1)$-Verteilung besitzen folgende Eigenschaften*

1. *φ ist bezüglich Null symmetrisch: $\varphi(-x) = \varphi(x), x \in R_1$,*
2. *φ ist unimodal und hat ihr Maximum bei Null,*
3. *φ hat zwei Wendepunkte, und zwar bei $+1$ und -1,*
4. *$1 - \Phi(x) = \Phi(-x), \quad x \in R_1, \quad \Phi(0) = 0,5$,*
5. *$1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \quad (x > 0)$.*

Beweis: Die Eigenschaften 1. - 4. sind offensichtlich, 5. folgt aus

$$1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} \int_{x}^{\infty} s\mathrm{e}^{-\frac{s^2}{2}} ds.$$

□

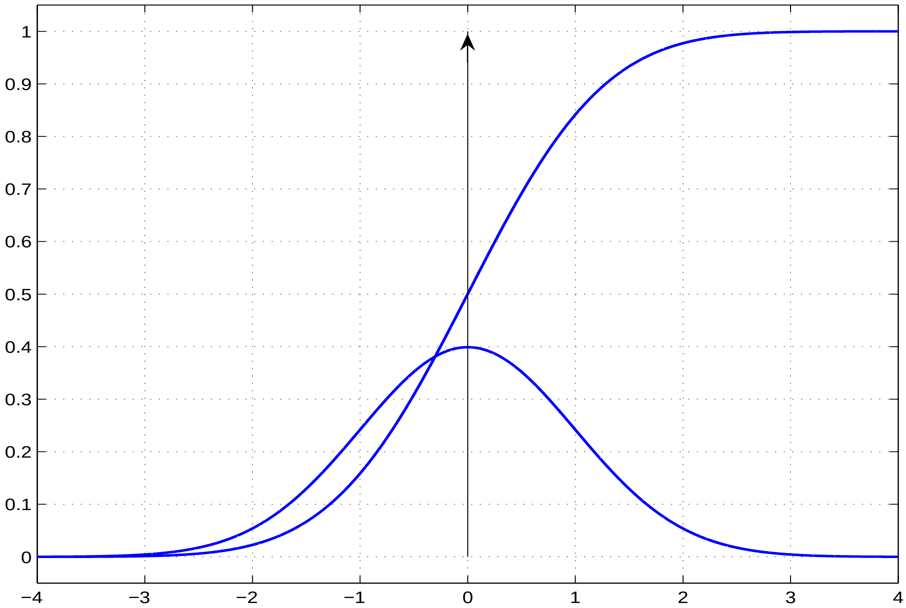

Abbildung 3.5: Verteilungsfunktion und Dichte der Standardnormalverteilung

Bei Anwendungen der Normalverteilung werden häufig ihre Quantile benötigt. Die wichtigsten sind in folgender Tafel zusammengefasst. Sie sind definiert als Lösung der Gleichung $\Phi(q_p^{0,1}) = p$:

p	$0,5$	$0,9$	$0,95$	$0,975$	$0,99$	$0,995$
$q_p^{(0,1)}$	0	$1,282$	$1,645$	$1,960$	$2,326$	$2,576$

(3.45)

Tafel 1 Quantile der $N(0,1)$-Verteilung

Die Quantile q_p^{μ,σ^2} der $N(\mu,\sigma^2)$-Verteilung sind definiert durch $\Phi_{\mu,\sigma^2}(q_p^{\mu,\sigma^2}) = p$ und berechnen sich aus den $q_p^{0,1}$ wie folgt:

$$q_p^{\mu,\sigma^2} = \mu + \sigma q_p^{0,1}, q_{1-p}^{\mu,\sigma^2} = \mu - \sigma q_p^{0,1}. \tag{3.46}$$

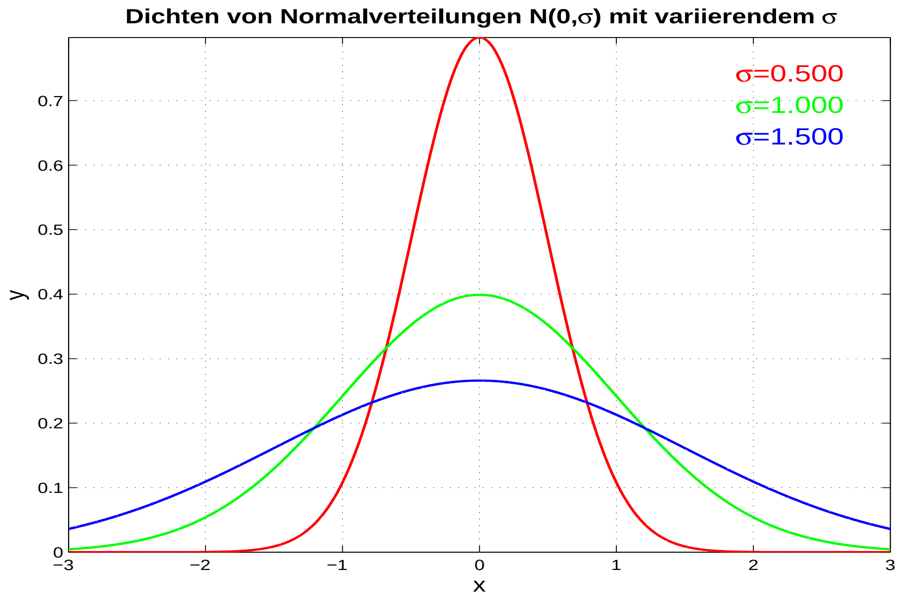

Abbildung 3.6: Normalverteilungsdichten mit verschiedenen Streuungen

Aussage 3.51 *Es sei X eine Zufallsgröße mit einer Dichte f_X, für die $\{x : f_X(x) > 0\}$ ein Intervall (a, b) mit $-\infty \le a < b \le \infty$ bilde. Weiterhin sei ψ eine streng monotone stetig differenzierbare Funktion auf (a, b). Dann besitzt $Y = \psi(X)$ eine Dichte f_Y, die gegeben ist durch*

$$\begin{aligned} f_Y(y) &= f_X(\psi^{-1}(y)) \cdot \left|\frac{d\psi^{-1}}{dy}\right|(y), \quad && y \in Wb(\psi) \\ &= 0 && y \notin Wb(\psi) \end{aligned}$$

Beweis: Zunächst sei ψ streng wachsend. Nach Voraussetzung bildet ψ das Intervall (a, b) eineindeutig auf ein Intervall (c, d) ab. Für $y \in (c, d)$ gilt

$$F_Y(y) = P(Y \le y) = P(X \le \psi^{-1}(y)) = F_X(\psi^{-1}(y)) = \int\limits_a^{\psi^{-1}(y)} f_X(s)ds$$

und somit

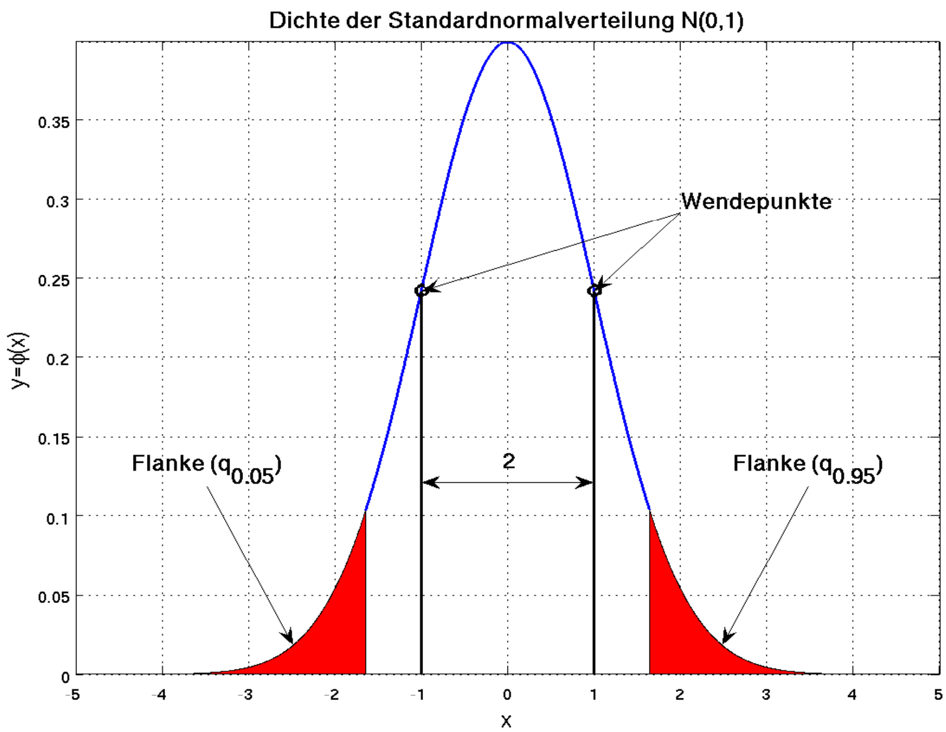

Abbildung 3.7: Quantile der Standardnormalverteilung

$$F_Y(y) = \int\limits_c^y f_X(\psi^{-1}(t)) \, \frac{d\psi^{-1}}{dt} \, (t)dt.$$

Wir setzen

$$f_Y(y) = f_X(\psi^{-1}(y)) \, \frac{d\psi^{-1}}{dy} \, (y) \text{ für } y \in (c, d).$$

Ist $y \notin (c, d)$, so gilt

$$F_Y(y) = P(Y \leq y) \equiv 0 \text{ für } y < c \text{ bzw. für } y \geq d \text{ ist } F_Y(y) = P(Y \geq y) \equiv 1.$$

Somit haben wir $f_Y(y) = 0$ für solche y. Insgesamt ergibt sich damit $F_Y(y) = \int\limits_{-\infty}^{y} f_Y(t)dt, \; y \in R_1$. Ist dagegen ψ streng fallend, so haben wir für $y \in (c, d)$

$$F_Y(y) = P(Y \geq y) = P(X \geq \psi^{-1}(y)) = 1 - P(X < \psi^{-1(y)}) =$$

$$1 - \int\limits_a^{\psi^{-1}(y)} f_X(s)ds = \int\limits_{\psi^{-1}(y)}^{b} f_X(s)ds = - \int\limits_c^y f_X(\psi^{-1}(t) \, \frac{d\psi^{-1}}{dt} \, dt.$$

Für $y \notin (c, d)$ schließen wir wie im Fall wachsender Funktionen ψ. Analog wie oben setzen wir

$$f_Y(y) = f_X(\psi^{-1}(y))|\frac{d\psi^{-1}}{dy}(y)|, \; y \in (c, d), = 0, y \notin (c, d).$$

Damit ist die Aussage bewiesen. □

Beispiel 3.52 X sei gleichmäßig auf $[0, 1)$ verteilt, es sei $\lambda > 0$, und es gelte $\psi(x) = -\frac{1}{\lambda} \, lnx, x \in [0, 1)$, dann hat $Y = -\frac{1}{\lambda} lnX$ die Dichte

$$f(y) = \mathbb{1}_{(0,\infty)}(y)\lambda \mathrm{e}^{-\lambda y}, \; y \in R_1$$

.

Verteilungsdichten auf R_n

Es seie F eine Verteilungsfunktion auf $R_n (n \geq 2)$, siehe Definition 3.42 und Aussage 3.40.

Definition 3.53 *Gibt es eine reellwertige Riemann-integrierbare Funktion f auf R_n mit*

1. $$f(x) \geq 0, x \in R_n, \tag{3.47}$$
2. $$F(x) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(s_1, s_2, \dots, s_n) ds, \dots ds_n, x = (x_1, \dots, x_n)^T, \tag{3.48}$$

so heißt f eine Dichte der Verteilungsfunktion F.

Aussage 3.54 *Ist Q die durch F erzeugte Wahrscheinlichkeitsverteilung, so gilt*

$$Q\left(\prod_{k=1}^{n}{}^{\otimes} (x_k, x_k + h_k]\right) = \triangle_{h_1} \dots, \triangle_{h_n} F(x_1, x_2, \dots, x_n) = \int_{x_n}^{x_n+h_n} \dots \int_{x_1}^{x_1+h_1} f(s_1, \dots, s_n) ds_1 \dots ds_n \tag{3.49}$$

Beweis: Der Beweis folgt aus der Addivität des Integrals. □

Beispiel 3.55 (Fortsetzung von Beispiel 3.44)

a) F hat keine Dichte, die Verteilung Q_F ist auf $\{(x, x) : 0 \leq x \leq 1\}$ konzentriert.

b) Haben die Verteilungsfunktionen F_k die Dichten $f_k, k = 1, \dots, n$ so besitzt F eine Dichte f mit

$$f(x_1, x_2, \dots, x_n) = \prod_{k=1}^{n} f_k(x_k), \quad x = (x_1, \dots, x_n)^T \in R_n. \tag{3.50}$$

Aussage 3.56

a) Für jede Dichte f von F gilt

$$\begin{aligned} F(x) &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(s_1, \dots, s_n) ds_1 \dots ds_n, \\ x &= (x_1, \dots, x_n)^T \in R_n, \end{aligned} \tag{3.51}$$

$$1 = F(\infty, \infty, \dots, \infty) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(s_1, \dots, s_n) ds_1 \dots ds_n \tag{3.52}$$

b) besitzt F eine Dichte, so ist F eine stetige Funktion,

c) Besitzt F eine Dichte f, die stetig in Umgebung von $x \in R_n$ ist, so ist F n-mal differenzierbar in diesem $x = (x_1, \dots, x_n)^T$, und es gilt

$$\frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, \dots, x_n) = f(x_1, \dots, x_n). \tag{3.53}$$

Der Beweis erfolgt analog zum Beweis von Aussage 3.48.

Aussage 3.57 *Besitzt F eine Dichte f, so hat auch jede Randverteilungsfunktion $F_{\mathscr{J}_r}$ mit $\mathscr{J}_r = \{k_1, \dots, k_r\} \subseteq \{1, 2, \dots, n\}$ eine Dichte $f_{\mathscr{J}_r}$, die sich folgendermaßen berechnen lässt:*

$$f_{\mathscr{J}_r}(x_{k_1}, \dots, x_{k_r}) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{(n-r)-mal} f(s_1, \dots, s_{k_1-1}, x_{k_1}, \dots, x_{k_r}, S_{k_1+1}, \dots, S_n) ds_1 \dots ds_n$$

Beweis: Der Beweis ergibt sich aus (3.37) und der Definition 3.53 der Dichte f durch Umordnung der Reihenfolge der entsprechenden n-fachen Integrale. □

Beispiel 3.58

a) Es sei

$$\sum = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\rho_2 & \sigma_2^2 \end{pmatrix} \text{ mit } \sigma_1, \sigma_2 > 0, |\rho| < 1, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in R_2$$

.

Dann ist die Funktion $\psi_{\mu,\sum}$ definiert durch

$$\psi_{\mu,\sum}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

.

$$\exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)\right], (x_1, x_2)^T \in R_2$$

die Dichte einer Wahrscheinlichkeitsverteilung auf $(R_1, \mathscr{B}_1)$ die als Normalverteilung $N_2(\mu, \sum)$ bezeichnet wird.

Die Randverteilungsdichten der Verteilung $N_2(\mu, \sum)$ sind eine $N(\mu_1, \sigma_1^2)$- bzw. eine $N(\mu_2, \sigma_2^2)$-Verteilung. Man beachte, dass in den Randverteilungen der Parameter ρ nicht mehr auftritt.

b) Es seien $\mu \in R_n$ und $\sum$ eine positiv definite symmetrische $n \times n$-Matrix. Dann ist die Funktion $\varphi_{\mu,\sum}$, definiert durch

$$\varphi_{\mu,\sum}(x) = \frac{1}{(2\pi)^{n\backslash 2}\sqrt{\det \sum}} \exp\left[-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)\right], x \in R_n,$$

die Dichte der sogenannten n-dimensionalen Normalverteilung $N_n(\mu, \sum)$.

Zu jeder Teilmenge $\mathscr{J}_r$ von $\{1, 2, \ldots, n\}$ mit $\mathscr{J}_r = (k_1, \ldots, k_r)$ ist die zu $\mathscr{J}_r$ gehörende Randverteilung ebenfalls eine Normalverteilung und zwar gleich $N_r(\Pi_{\mathscr{J}_r}\mu, \Pi_{\mathscr{J}_r} \sum \Pi_{\mathscr{J}_r}^T)$ wobei $\Pi_{\mathscr{J}_r}$ die Projektionsmatrix ist, die

$x = (x_1, x_2, \ldots, x_n)^T$ auf $\Pi_{\mathscr{J}_r} x_n = (x_{k_1}, \ldots, x_{k_r})$ abbildet.

Der Vektor $\Pi_{\mathscr{J}_r}\mu$ ergibt sich aus μ durch Entfernen aller Komponenten x_l mit $l \notin \mathscr{J}_r$ und die Matrix $\Pi_{\mathscr{J}_r} \sum \Pi^T_{\mathscr{J}_r}$ ergibt sich aus $\sum = (s_{ij})$ durch Entfernen aller Elemente s_{ij} mit $i \notin \mathscr{J}_r$ oder $j \notin \mathscr{J}_r$.

Aussage 3.59 (Transformationsformel für n-dimensionale Dichten) *Es sei $X = (X_1, \ldots, X_n)^T$ ein zufälliger n-dimensionaler Vektor mit der Dichte f.*
Weiterhin sei U eine offene Menge aus R_n mit $P^X(U) = 1$ und h eine eineindeutige stetig differenzierbare Funktion von U auf $V \subseteq R_n$, deren Jacobimatrix

$$\mathscr{J}_h(x) := \left(\frac{\delta h_i(x)}{\delta x_j}\right)_{i,j=1,\ldots,n}$$

nirgends auf U singulär ist. Mit g werde die inverse Funktion h^{-1} bezeichnet.

Dann hat der zufällige Vektor $Y := h(X)$ eine Dichte f_Y mit

$$f_Y(y) = \begin{cases} f_X(g(y))|\det \mathscr{J}_g(y)| \text{ , falls } y \in V \\ 0, \text{ falls } y \in R_n \backslash V. \end{cases}$$

Bemerkung: Die soeben formulierte Aussage findet man in der Literatur in unterschiedlicher Form, je nachdem, welche Voraussetzungen man an g stellt. Siehe zum Beispiel Pfanzagl, 1991, Kap. 3.4.

Beispiel 3.60 Es seien A eine reguläre $n \times n$-Matrix und $b \in R_n$. Wir definieren

$$g(x) = Ax + b, \quad x \in R_n.$$

$$Y := g(X).$$

Dann gilt $g^{-1}(y) = A^{-1}(y - b)$, $\mathfrak{J}_{g^{-1}}(y) = A^{-1}$ und Y hat die Dichte

$$f_Y(y) = f_X(A^{-1}(y - b))|\det A^{-1}| \text{ , } y \in R_n.$$