

Angewandte Stochastik I

Vorlesungsskript
Sommersemester 2014

von Markus Kunze

Inhaltsverzeichnis

1	Wahrscheinlichkeiten	1
1.1	Beschreibung von Zufallsexperimenten	1
1.2	Wahrscheinlichkeitsräume	3
1.3	Laplace Experimente	6
1.4	Bedingte Wahrscheinlichkeit und Unabhängigkeit	10
2	Zufallsvariablen und ihre Verteilungen	15
2.1	Definitionen	15
2.2	Diskrete Verteilungen	17
2.3	Absolutstetige Verteilungen	20
2.4	Zufallsvektoren	22
2.5	Transformation von Zufallsvariablen	25
3	Momente von Zufallsvariablen	29
3.1	Der Erwartungswert	29
3.2	Varianz und höhere Momente	33
3.3	Kovarianz und Korrelation	34
4	Grenzwertsätze	37
4.1	Das Gesetz der großen Zahlen	37
4.2	Der zentrale Grenzwertsatz	39

Kapitel 1

Wahrscheinlichkeiten

1.1 Beschreibung von Zufallsexperimenten

Die Stochastik beschäftigt sich mit der mathematischen Analyse zufälliger Vorgänge. Unter einem zufälligen Vorgang verstehen wir dabei einen Vorgang der bei Wiederholung unter identischen (oder zumindest ähnlichen) Voraussetzungen nicht immer zum selben Ergebnis führt. Man geht hierbei davon aus, dass alle *möglichen* Ergebnisse des Vorgangs bekannt sind. Diese werden in der Grundmenge Ω zusammengefasst, d.h. Ω besteht aus allen möglichen Versuchsausgängen.

Beispiel 1.1.1. (a) Ein Würfel wird geworfen. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Hierbei bezeichnet $j \in \Omega$ das Ergebnis “Es wird eine j geworfen”.

(b) Eine Münze wird geworfen. $\Omega = \{K, Z\}$. Hierbei bezeichnet K das Versuchsergebnis “Kopf liegt oben”, Z das Elementarereignis “Zahl liegt oben”.

(c) Ein Würfel wird solange geworfen bis zum ersten mal eine Sechs geworfen wurde. Da es keine natürliche Obergrenze für die Anzahl der benötigten Würfe gibt, bietet sich als Grundmenge die Menge der natürlichen Zahlen an: $\Omega = \mathbb{N}$. Hierbei bezeichnet $n \in \Omega$ den Versuchsausgang “im n -ten Wurf fällt zum ersten Mal eine Sechs”.

(d) Die Lebensdauer einer Glühbirne wird ermittelt. Hier wählen wir $\Omega = [0, \infty)$, die Menge der positiven, reellen Zahlen. Hier bezeichnet $t \in \Omega$ das Ergebnis “Die Glühbirne erlischt nach t Sekunden”.

(e) Ein Pfeil wird auf eine Dartscheibe geworfen. Um den Ausgang dieses Zufallsexperiments zu beschreiben legen wir ein Kartesisches Koordinatensystem in den Mittelpunkt der Dartscheibe und wählen die Einheiten so, dass der Radius der Dartscheibe 1 ist. Dann können wir $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ wählen, wobei (x, y) den Versuchsausgang “Der Pfeil landet im Punkt mit Koordinaten (x, y) ” bezeichnet. Wollen wir berücksichtigen dass es möglich ist, die Scheibe zu verfehlen, so können wir als Grundmenge $\tilde{\Omega} := \Omega \cup \{V\}$ wählen, wobei V das Versuchsergebnis “Scheibe verfehlt” bezeichnet.

Häufig ist man bei einem Zufallsexperiment nicht an dem tatsächlichen Ausgang interessiert sonder nur daran, ob das Ergebnis zu einer vorgegebenen Menge von Ergebnissen interessiert. Im Beispiel 1.1.1(e) interessiert man sich etwa lediglich dafür, wieviele Punkte man für den Wurf erhält, ob man also beispielsweise in die Region für “18 Punkte” getroffen hat; wo genau man diese Region getroffen hat ist zweitrangig.

Eine Teilmenge A von Ω nennt man *Ereignis*. Liefert die Durchführung eines Zufallsexperiments das Ergebnis $\omega \in \Omega$ und liegt ω in A , so sagt man das Ereignis A sei *eingetreten*. Insbesondere ist die *leere Menge* \emptyset ein Ereignis. Es tritt nie ein und heißt daher *unmögliches Ereignis*. Andererseits ist auch Ω selbst ein Ereignis. Es tritt immer ein und heißt *sicheres Ereignis*. Ein einelementiges Ereignis nennt man auch *Elementarereignis*.

Wir geben einige Beispiele von Ereignissen in den Situationen von Beispiel 1.1.1:

- Beispiel 1.1.2.** (a) $A = \{2, 4, 6\}$ ist das Ereignis “Eine gerade Zahl wurde gewürfelt”.
- (b) Hier gibt es neben Elementarereignissen $\{K\}$ und $\{Z\}$ nur noch das unmögliche Ereignis \emptyset und das sichere Ereignis Ω .
- (c) Das Ereignis $A = \{1, 2, 3\}$ (“Innerhalb der ersten drei Würfe fällt eine Sechs”) ist beim Mensch-Ärgere-Dich-Nicht von Interesse.
- (d) $A = (86400, \infty)$ ist das Ereignis “Die Glühbirne brennt länger als einen Tag”.
- (e) Ist der Radius des Bull’s Eye $r \in (0, 1)$, so bezeichnet $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$ das Ereignis “Bull’s Eye!”.

Mittels Mengenoperationen können Ereignisse zu neuen Ereignissen verknüpft werden. Der *Schnitt* $A \cap B$ ist das Ereignis, dass sowohl A als auch B eintritt. Die *Vereinigung* $A \cup B$ ist das Ereignis “ A oder B tritt ein”. Die *Differenz* $A \setminus B$ ist das Ereignis “ A , nicht aber B tritt ein”; insbesondere bezeichnet das *Komplement* $A^c := \Omega \setminus A$ das Ereignis “ A tritt nicht ein”. Ist $A \cap B = \emptyset$ so sagen wir A und B sind *unvereinbar*. Die Menge aller Ereignisse ist $\mathcal{P}(\Omega)$, die Potenzmenge von Ω .

Beispiel 1.1.3. Eine Münze wird drei Mal geworfen. Wir wählen

$$\Omega = \{KKK, KKZ, KZK, ZKK, KZZ, ZKZ, ZZK, ZZZ\}.$$

Wobei ein K (Z) an Position j anzeigt dass beim j -ten Wurf Kopf (Zahl) fällt.

Es sei A_j das Ereignis “Im j -ten Wurf fällt Kopf”. Als Teilmenge von Ω ist dann beispielsweise $A_1 = \{KKK, KKZ, KZK, KZZ\}$.

Die Menge $A_1 \cup A_2 \cup A_3$ ist das Ereignis “Es fällt mindestens einmal Kopf”, während $A_1 \cap A_2 \cap A_3$ das einelementige Ereignis $\{KKK\}$ bezeichnet.

Das Ereignis “Es fällt mindestens zweimal Kopf” lässt sich schreiben als

$$(A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3)$$

Als nächstes wollen wir Ereignissen in einem Zufallsexperiment eine *Wahrscheinlichkeit* zuordnen. Diese Wahrscheinlichkeiten sollen in gewissem Sinne die Rolle von relativen Häufigkeiten widerspiegeln: Dass ein Ereignis A Wahrscheinlichkeit p hat soll bedeuten, dass man bei “genügend häufiger Wiederholung des Experiments in p Prozent der Fälle das Ereignis A eintritt”.

Leider ist dies keine einwandfreie Definition, da die relative Häufigkeit selbst zufallsbehaftet ist, also vom Zufallsexperiment abhängt. Daher führen wir im nächsten Abschnitt Wahrscheinlichkeiten *axiomatisch* ein, d.h. wir definieren eine Wahrscheinlichkeit als eine Abbildung mit bestimmten Eigenschaften (die durch Eigenschaften der relativen Häufigkeit motiviert sind).

Wir werden später (im Gesetz der großen Zahlen) zeigen, dass die relative Häufigkeit eines Ereignisses in der Tat in einem gewissen Sinn gegen die Wahrscheinlichkeit dieses Ereignisses konvergiert.

1.2 Wahrscheinlichkeitsräume

Aus mathematischen Gründen ist es nicht immer möglich, Wahrscheinlichkeiten für *alle* Teilmengen einer Grundmenge Ω zu definieren. Daher schränkt man sich hier auf sogenannte σ -Algebren ein.

Definition 1.2.1. Es sei $\Omega \neq \emptyset$. Eine Teilmenge $\Sigma \subset \mathcal{P}(\Omega)$ heißt σ -Algebra, falls

- (i) $\Omega \in \Sigma$.
- (ii) Falls $A \in \Sigma$, so ist auch $A^c \in \Sigma$.
- (iii) Sind $(A_n)_{n \in \mathbb{N}} \subset \Sigma$, so ist auch $\bigcup_{n \in \mathbb{N}} A_n \in \Sigma$.

Das Paar (Ω, Σ) heißt *Messraum*, eine Menge $A \in \Sigma$ *messbar*.

Beispiel 1.2.2. (a) Ist $\Omega \neq \emptyset$ so ist die Potenzmenge $\mathcal{P}(\Omega)$ stets eine σ -Algebra auf Ω . Sie ist die größte σ -Algebra auf Ω . Es gibt auch eine *kleinste* σ -Algebra auf Ω , nämlich $\{\emptyset, \Omega\}$.

(b) Da der Durchschnitt von σ -Algebren wieder eine σ -Algebra ist, gibt es zu jeder Teilmenge $\mathcal{A} \subset \mathcal{P}(\Omega)$ stets eine kleinste σ -Algebra, die \mathcal{A} enthält. Diese wird mit $\sigma(\mathcal{A})$ bezeichnet und heißt *die von \mathcal{A} erzeugte σ -Algebra*. Ist $\Sigma = \sigma(\mathcal{A})$, so heißt \mathcal{A} ein *Erzeuger* von Σ .

(c) Ist $\mathcal{A} = \{A\}$ für eine einzelne Menge A mit $A \neq \emptyset$, $A \neq \Omega$, so ist $\sigma(\mathcal{A}) = \{\emptyset, A, A^c, \Omega\}$.

(d) Ist $\Omega = \mathbb{R}^d$, und $\mathcal{A} = \{B(x, r) : x \in \mathbb{R}^d, r > 0\}$ die Menge aller offenen Kugeln $B(x, r) := \{y \in \mathbb{R}^d : \|x - y\| < r\}$, so heißt $\sigma(\mathcal{A})$ die *Borel σ -Algebra* und wird mit $\mathcal{B}(\mathbb{R}^d)$ bezeichnet.

(e) Ist $\Omega \subset \mathbb{R}^d$ so sei $\mathcal{B}(\Omega) := \{A \cap \Omega : A \in \mathcal{B}(\mathbb{R}^d)\}$. Es ist leicht zu sehen, dass dies eine σ -Algebra ist, die sogenannte *Borel σ -Algebra* auf Ω .

Bemerkung 1.2.3. Es folgt aus den Axiomen einer σ -Algebra und den deMorgan'schen Gesetzen $(\bigcup A_k)^c = \bigcap A_k^c$, dass eine σ -Algebra mit einer Folge von Mengen auch deren Durchschnitt enthält.

Definition 1.2.4. (Kolmogorov 1933)

Ein *Wahrscheinlichkeitsraum* ist ein Tripel $(\Omega, \Sigma, \mathbb{P})$, bestehend aus einer Menge Ω , einer σ -Algebra Σ und einer Abbildung $\mathbb{P} : \Sigma \rightarrow [0, 1]$, derart, dass

- (i) $\mathbb{P}(\Omega) = 1$ (Normiertheit)
- (ii) Sind A_1, A_2, A_3, \dots Elemente von Σ und paarweise unvereinbar, so ist

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \quad (\sigma\text{-Additivität}).$$

Eine Abbildung \mathbb{P} mit diesen Eigenschaften heißt *Wahrscheinlichkeitsmaß* auf Σ .

In abstrakten Wahrscheinlichkeitsräumen im Sinne von Definition 1.2.4 ist es üblich, eine Teilmenge $A \subset \Omega$ nur dann Ereignis zu nennen, wenn $A \in \Sigma$, d.h. falls A messbar ist. Bevor wir Beispiele besprechen, zeigen wir zunächst einige Eigenschaften von Wahrscheinlichkeitsmaßen.

Proposition 1.2.5. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $A, B \in \Sigma$.*

(1) $\mathbb{P}(\emptyset) = 0$.

(2) *Ist $A \subset B$, so ist $\mathbb{P}(A) \leq \mathbb{P}(B)$. Weiter gilt $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. Insbesondere ist $\mathbb{P}(B^c) = 1 - \mathbb{P}(B)$.*

(3) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Seien nun $A_1, \dots, A_n \in \Sigma$.

(4) *Es ist*

$$\mathbb{P}\left(\bigcup_{j=1}^n A_j\right) \leq \sum_{j=1}^n \mathbb{P}(A_j).$$

(5) *Sind A_1, \dots, A_n paarweise unvereinbar, so ist*

$$\mathbb{P}\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n \mathbb{P}(A_j).$$

Beweis. (1) Es sei $A_n := \emptyset$ für alle $n \in \mathbb{N}$. Dann sind die A_n paarweise unvereinbar und ihre Vereinigung ist auch die leere Menge. Daher folgt aus der σ -Additivität, dass $\mathbb{P}(\emptyset) = \sum_{n=1}^{\infty} \mathbb{P}(\emptyset)$. Dies impliziert, dass $\mathbb{P}(\emptyset) = 0$.

(5) folgt nun aus der σ -Additivität, indem man $A_{n+1} = A_{n+2} = \dots = \emptyset$ wählt.

(2) Ist $A \subset B$, so ist $B = A \cup (B \setminus A)$ und die letzten beiden Ereignisse sind unvereinbar. Also ist $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ und daher $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. Weil $\mathbb{P}(B \setminus A) \geq 0$ ist folgt $\mathbb{P}(A) \leq \mathbb{P}(B)$. Die Formel für das Komplement folgt weil $\mathbb{P}(\Omega) = 1$ ist.

(3) Sei $C := A \cap B$. Dann ist $C \subset A$ und $C \subset B$ und $A \cup B = (A \setminus C) \cup (B \setminus C) \cup C$. Diese Ereignisse sind paarweise unvereinbar. Also gilt nach den bereits gezeigten Teilen (5) und (2)

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus C) + \mathbb{P}(B \setminus C) + \mathbb{P}(C) = \mathbb{P}(A) - \mathbb{P}(C) + \mathbb{P}(B) - \mathbb{P}(C) + \mathbb{P}(C). \quad \square$$

(4) zeigen wir induktiv. Für $n = 2$ folgt die Ungleichung sofort aus (3). Wenn die behauptete Ungleichung jedoch für ein $n \in \mathbb{N}$ gilt, so gilt für Mengen A_1, \dots, A_n, A_{n+1}

$$\mathbb{P}(A_1 \cup \dots \cup A_n \cup A_{n+1}) \leq \mathbb{P}(A_1 \cup \dots \cup A_n) + \mathbb{P}(A_{n+1}) \leq \sum_{k=1}^{n+1} \mathbb{P}(A_k),$$

was aus der Induktionshypothese und der Ungleichung für $n = 2$ folgt.

Beispiel 1.2.6. (Endliche Wahrscheinlichkeitsräume)

Wir betrachten die Situation wo $\Omega = \{\omega_1, \dots, \omega_n\}$ endlich ist und $\Sigma = \mathcal{P}(\Omega)$, d.h. jede Menge ist messbar. In diesem Fall definieren wir $p_k = \mathbb{P}(\{\omega_k\}) \geq 0$. Wegen der endlichen Additivität ist dann

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{k:\omega_k \in A} \{\omega_k\}\right) = \sum_{k:\omega_k \in A} \mathbb{P}(\{\omega_k\}) = \sum_{k:\omega_k \in A} p_k. \quad (1.1)$$

Insbesondere folgt $\sum_{k=1}^n p_k = \mathbb{P}(\Omega) = 1$. Sind umgekehrt p_1, \dots, p_n gegeben mit $p_k \geq 0$ für alle $k = 1, \dots, n$ und $\sum_{k=1}^n p_k = 1$, so ist es leicht zu sehen, dass durch (1.1) ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{P}(\Omega))$ definiert ist.

Beispiel 1.2.7. Die Ecken eines Würfels werden gleichmäßig abgeschliffen, sodass der Würfel auch auf jeder der acht Ecken liegen bleiben kann. Dabei ist die Wahrscheinlichkeit einer jeden Ecke nur $1/4$ so groß wie die Wahrscheinlichkeit einer jeden Seite. Wir wählen als Grundmenge $\Omega = \{s_1, \dots, s_6, e_1, \dots, e_8\}$ wobei s_j das Elementarereignis bezeichnet auf der Seite mit der Zahl j liegen zu bleiben. Die Ecken wurden von 1 bis 8 durchnummeriert und e_j bezeichnet das Ergebnis auf der Ecke mit der Zahl j liegen zu bleiben.

Nach obigen Angaben ist $\mathbb{P}(\{s_1\}) = \dots = \mathbb{P}(\{s_6\}) = p$ für eine unbekannte Wahrscheinlichkeit p , während $\mathbb{P}(\{e_1\}) = \dots = \mathbb{P}(\{e_8\}) = p/4$ ist. Weil die Summe über die Wahrscheinlichkeiten der Elementarereignisse 1 ergeben muss, muss also

$$1 = 6p + 8 \cdot \frac{p}{4} = 8 \cdot p \quad \text{also} \quad p = \frac{1}{8}.$$

Die Wahrscheinlichkeit, auf einer der Seiten liegen zu bleiben, also das Ereignis S gegeben durch $S := \{s_1, s_2, s_3, s_4, s_5, s_6\}$, hat Wahrscheinlichkeit $6 \cdot \frac{1}{8} = \frac{3}{4}$. Das Ereignis $E := \{e_1, e_2, e_3, e_4, e_5, e_6\}$ auf einer Ecke liegen zu bleiben ist das komplementäre Ereignis zu S : $E = S^c$. Demnach ist $\mathbb{P}(E) = 1 - \mathbb{P}(S) = 1/4$. Natürlich kann man dies auch durch Aufsummieren der Wahrscheinlichkeiten der Elementarereignisse sehen: $\mathbb{P}(E) = 8 \cdot p/4 = 1/4$.

Bemerkung 1.2.8. Im Allgemeinen, insbesondere wenn die Grundmenge Ω überabzählbar ist, ist es schwierig, ein Wahrscheinlichkeitsmaß explizit anzugeben, häufig ist selbst die σ -Algebra Σ nicht bekannt. Stattdessen kennt man einen Erzeuger \mathcal{A} der σ -Algebra und kennt die Wahrscheinlichkeiten $\mathbb{P}(A)$ für $A \in \mathcal{A}$. In der Maßtheorie konstruiert man (unter geeigneten Voraussetzungen) zu vorgegebenen Mengensystemen \mathcal{A} und Abbildungen $p : \mathcal{A} \rightarrow [0, 1]$ Wahrscheinlichkeitsmaße \mathbb{P} auf $\sigma(\mathcal{A})$ mit $\mathbb{P}(A) = p(A)$. Wir gehen hier nicht ins Detail, sondern stellen lediglich das wichtigste Beispiel vor.

Beispiel 1.2.9. (Lebesguemaß)

Es gibt genau ein Wahrscheinlichkeitsmaß \mathbb{P} auf $\mathcal{B}([0, 1])$ mit $\mathbb{P}([a, b]) = b - a$ für alle $0 \leq a < b \leq 1$. Das Maß \mathbb{P} heißt Lebesguemaß und wird (vor allem in der Analysis) auch mit λ bezeichnet.

Wir geben noch eine wichtige Eigenschaft von Wahrscheinlichkeitsmaßen an, die häufig *Stetigkeit* des Wahrscheinlichkeitsmaßes genannt wird.

Lemma 1.2.10. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A_n eine Folge in Σ mit $A_1 \subset A_2 \subset A_3 \subset \dots$ und $A := \bigcup_{n \in \mathbb{N}} A_n$. (Wir sagen: A_n wächst gegen A und schreiben $A_n \uparrow A$. Dann ist*

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Ist C_n eine Folge in Σ mit $C_1 \supset C_2 \supset \dots$ und $C := \bigcap_{n \in \mathbb{N}} C_n$ (Wir sagen C_n fällt gegen C und schreiben $C_n \downarrow C$), so gilt ebenfalls

$$\mathbb{P}(C) = \lim_{n \rightarrow \infty} \mathbb{P}(C_n).$$

Beweis. Es sei $B_1 := A_1$ und $B_k := A_k \setminus A_{k-1}$ für $k \geq 2$. Dann sind die Mengen B_k paarweise disjunkt und $B_1 \cup \dots \cup B_n = A_n$. Weiterhin ist $\bigcup_{n \in \mathbb{N}} B_n = A$. Somit ist wegen der σ -Additivität von \mathbb{P}

$$\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{P}(B_k) = \lim_{n \rightarrow \infty} \mathbb{P}(B_1 \cup \dots \cup B_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Ist $C_n \downarrow C$, so ist $C_n^c \uparrow C^c$ und es folgt mit dem ersten Teil

$$1 - \mathbb{P}(C) = \mathbb{P}(C^c) = \lim_{n \rightarrow \infty} \mathbb{P}(C_n^c) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(C_n).$$

□

Beispiel 1.2.11. Wir betrachten den Wahrscheinlichkeitsraum $([0, 1], \mathcal{B}([0, 1]), \lambda)$, wobei λ das Lebesguemaß ist. In diesem Fall ist $\lambda(\{x\}) = 0$ für alle $x \in [0, 1]$. Es ist nämlich $[x - \frac{1}{n}, x + \frac{1}{n}] \downarrow \{x\}$ und daher, wegen der Stetigkeit des Maßes,

$$\lambda(\{x\}) = \lim_{n \rightarrow \infty} \lambda\left(\left[x - \frac{1}{n}, x + \frac{1}{n}\right]\right) = \lim_{n \rightarrow \infty} \frac{2}{n} = 0.$$

Zusammen mit der σ -Additivität folgt, dass $\lambda(A) = 0$ für jede abzählbare Menge $A = \{x_n : n \in \mathbb{N}\}$. Es ist nämlich

$$\lambda(A) = \lambda\left(\bigcup_{n \in \mathbb{N}} \{x_n\}\right) = \sum_{n \in \mathbb{N}} \lambda\{x_n\} = \sum_{n \in \mathbb{N}} 0 = 0.$$

Insbesondere ist $\lambda(\mathbb{Q} \cap [0, 1]) = 0$.

1.3 Laplace Experimente

Wir haben gesehen, dass auf einem endlichen Grundraum ein Wahrscheinlichkeitsmaß durch die Wahrscheinlichkeiten der Elementarereignisse eindeutig festgelegt ist. Ein besonders einfacher Fall liegt dann vor, wenn alle Elementarereignisse gleich wahrscheinlich sind. In diesem Fall spricht man von einem *Laplaceschen Wahrscheinlichkeitsraum*. Ein Zufallsexperiment welches durch einen Laplaceschen Wahrscheinlichkeitsraum beschrieben wird nennt man häufig *Laplacesches (Zufalls-)Experiment*.

Weil sich die Wahrscheinlichkeiten der Elementarereignisse zu 1 summieren müssen, hat jedes Elementarereignis Wahrscheinlichkeit $(\#\Omega)^{-1}$, wobei $\#M$ die Anzahl der Elemente der Menge M bezeichnet. In einem Laplaceschen Wahrscheinlichkeitsraum gilt also

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

Man sagt, die Wahrscheinlichkeit einer Menge A ist die Anzahl der *günstigen Ergebnisse* (also derer, bei denen A eintritt) geteilt durch die Anzahl der *möglichen Ergebnisse* (also aller Elemente von Ω).

Ob es angemessen ist, von einem Laplace Experiment auszugehen ist keine mathematische Frage sondern hängt von Erfahrungswerten und/oder Beobachtungen ab.

Beispiel 1.3.1. (a) Der klassische Münzwurf und auch das Werfen eines Würfels werden gewöhnlich als Laplace Experiment angesehen. Das liegt an der Symmetrie der geworfenen Objekte: Es gibt keinen Grund, warum eine Seite der Münze (eine Seite des Würfels) bevorzugt werden sollte.

(b) Werfen wir eine Reißzwecke auf einen Betonboden, so kann sie entweder auf der flachen Seite liegen bleiben (Elementarereignis F) oder aber mit der Spitze schräg nach unten (Elementarereignis S). Man kann also als Grundraum $\Omega = \{F, S\}$ wählen. Es ist nicht klar, ob wir dieses Experiment als Laplace Experiment modellieren können. Hier ist man auf statistische Methoden angewiesen um ein geeignetes Modell zu finden.

Um Wahrscheinlichkeiten in Laplaceschen Wahrscheinlichkeitsräumen zu berechnen muss man die Anzahl gewisser Mengen bestimmen. In Anwendungen sind allerdings die Mengen gewöhnlich so groß, dass man sie nicht mehr explizit aufschreiben kann (oder besser, will). In diesem Fall bedient man sich der Kombinatorik um die Elemente einer Menge abzuzählen.

Grundlage vieler kombinatorischer Überlegungen ist folgender Abzählsatz:

Theorem 1.3.2. *Es sei k eine natürliche Zahl. Hat man k Fächer F_1, \dots, F_k zu belegen und hat man n_1 Möglichkeiten Fach F_1 zu belegen, n_2 Möglichkeiten Fach F_2 zu belegen, ..., n_k Möglichkeiten Fach F_k zu belegen, so hat man insgesamt $n_1 \cdot \dots \cdot n_k$ Möglichkeiten die k Fächer zu belegen.*

Beispiel 1.3.3. Thorsten hat ein Bewerbungsgespräch und stellt dazu ein passendes Outfit zusammen. Neben typischen Studentenklamotten (die er seinem zukünftigen Arbeitgeber lieber ersparen will) findet er 3 Hemden, 4 Krawatten und 2 Anzüge in seinem Kleiderschrank. Er kann damit (von modischen Überlegungen abgesehen) $3 \cdot 4 \cdot 2 = 24$ verschiedene Outfits zusammenstellen.

Beispiel 1.3.4. Sabine bewahrt ihre Socken einzeln und bunt gemischt in einer Schublade auf. Morgens zieht sie zufällig zwei Socken heraus und zieht diese unbesehen an (Nur Spiesser ziehen passende Socken an!). Was ist die Wahrscheinlichkeit, dass Sabine zwei passende Socken zieht, wenn sich insgesamt 10 (verschiedene) Paar Socken, also 20 einzelne Socken in der Schublade befinden.

Lösung: Als Grundraum Ω wählen wir die Menge von Paaren von Socken. Wir haben 20 Möglichkeiten eine erste Socke zu ziehen und anschliessend noch 19 Möglichkeiten eine zweite Socke zu ziehen. Also hat Ω genau $20 \cdot 19 = 380$ Elemente. Wir interessieren uns für das Ereignis A dass beide gezogenen Socken ein zusammengehörendes Paar bilden. Um in A zu liegen haben wir wiederum 20 Möglichkeiten für die erste Socke. Bei der zweiten haben wir jedoch keine Wahl mehr, wir müssen die eine passende Socke wählen. Also hat die Menge A genau $20 \cdot 1 = 20$ Elemente. Demnach erhalten wir

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{20}{380} = \frac{1}{19} \approx 0,0526$$

Das Zufallsexperiment in Beispiel 1.3.4 gehört zu einer Klasse von Laplace Experimenten bei denen sich die Anzahl der Elemente von Mengen mittels sogenannter *Urnenmodellen* bestimmen lassen. Bei solchen Modellen haben wir eine Urne mit n Kugeln die von 1 bis n durchnummeriert sind. Aus diesen Kugeln ziehen wir nacheinander k Kugeln zufällig. Dabei müssen wir unterscheiden ob wir gezogene Kugeln wieder zurücklegen (ziehen mit/ohne Zurücklegen) und ob wir die Reihenfolge, in der wir die Kugeln ziehen, beachten (ziehen mit/ohne Beachten der Reihenfolge). In Beispiel 1.3.4 haben wir ohne Zurücklegen und mit Beachten der Reihenfolge gezogen.

Ziehen mit Zurücklegen mit Beachtung der Reihenfolge:

In diesem Fall haben wir in jedem Zug n Möglichkeiten, nach dem Abzählsatz also insgesamt n^k Möglichkeiten.

Als *Beispiel* betrachten wir das 100-malige Werfen eines Würfels. Wir ziehen aus den Zahlen von 1 bis 6 ($n = 6$) mit Zurücklegen (geworfene Zahlen dürfen wieder geworfen werden) 100 Zahlen heraus ($k = 100$). Wir beachten die Reihenfolge, denn wir merken uns welche Zahl wir als erstes, zweites, usw. gewürfelt haben. Insgesamt gibt es also 6^{100} Möglichkeiten.

Ziehen ohne Zurücklegen mit Beachtung der Reihenfolge:

Beachte dass in diesem Fall $k \leq n$ sein muss. Hier haben wir für die erste Position n Möglichkeiten, für die zweite noch $n - 1$, für die dritte noch $n - 2$, ..., für die k -te noch $n - (k - 1)$. Wir haben also $n \cdot (n - 1) \cdot \dots \cdot (n - (k - 1))$ Möglichkeiten.

Ist $n = k$, so ist das Ergebnis des Ziehens ohne Zurücklegen gerade eine gewisse Reihenfolge der Zahlen 1 bis n . Dafür gibt es $n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1 =: n!$ (Lies n Fakultät) Möglichkeiten. Es gibt also $n!$ Möglichkeiten die Zahlen von 1 bis n anzuordnen.

Beachte, dass $n \cdot \dots \cdot (n - (k - 1)) = \frac{n!}{(n-k)!}$. Manchmal wird schreibt man $(n)_k := \frac{n!}{(n-k)!}$.

Ein *Beispiel* dieser Art des Ziehens tritt beim Elfmeterschiessen im Fussball auf. Es müssen aus den elf Spielern einer Fussballmannschaft fünf Elfmeterschützen (in einer bestimmten Schuss-Reihenfolge) ausgewählt werden hierfür gibt es $(11)_5 = 55.440$ Möglichkeiten.

Ziehen ohne Zurücklegen ohne Beachtung der Reihenfolge:

Um diese Möglichkeiten abzuzählen, ziehen wir zunächst mit Beachtung der Reihenfolge. Dafür gibt es genau $(n)_k$ Möglichkeiten. Allerdings haben wir alle Möglichkeiten mehrfach gezählt und zwar immer dann, wenn die gleichen Zahlen in lediglich anderer Reihenfolge gezogen wurden. Da es $k!$ Möglichkeiten gibt k Zahlen anzuordnen, haben wir also jedes mögliche Ergebnis $k!$ mal gezählt. Also gibt es

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

Möglichkeiten aus n Zahlen k ohne Zurücklegen und ohne Beachtung der Reihenfolge zu ziehen.

Beispielsweise gibt es $\binom{49}{6}$ Mögliche Ergebnisse beim Lotto "6 aus 49", denn es werden 6 Zahlen ohne Zurücklegen aus 49 gezogen und es kommt beim Ergebnis nicht auf die Reihenfolge an, in der die Zahlen gezogen wurden.

Ziehen mit Zurücklegen ohne Beachtung der Reihenfolge:

Dieses Urnenmodell ist äquivalent dazu, unsere k Züge auf die n Möglichkeiten zu verteilen oder, anders ausgedrückt, k Teilchen auf n Zellen zu verteilen. Stellen wir die k Teilchen durch x 'e dar, und trennen wir die Teilchen durch das Symbol $|$ so müssen wir also $n - 1$ Trennsymbole in die x 'e einfügen. Ist etwa $k = 3$ und $n = 4$, so bedeutet die Kette $xx||x|$, dass sich zwei Teilchen in der ersten und ein Teilchen in der dritten Zelle befindet, während die Zellen 2 und 4 leer sind. In der ursprünglichen Interpretation wurde Kugel 1 zweimal und Kugel 3 einmal gezogen.

Wir haben also von insgesamt $k + n - 1$ Stellen $n - 1$ als Trennsymbole auszuwählen. Hierfür gibt es

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

Möglichkeiten.

Dieses Art des Urnenmodells liegt etwa im Falle des 100-maligen Würfels vor, wenn anschliessend nur mitgeteilt wird, wie oft jede Zahl gewürfelt wurde. In diesem Fall ist $n = 6$ und $k = 100$, es gibt also $\binom{105}{5} = 96560646$ mögliche Ergebnisse bei diesem Experiment.

Beispiel 1.3.5. (Urnenmodelle in der statistischen (Quanten-)Mechanik)

Die oben betrachteten Urnenmodelle treten auch in natürlicher Weise in der statistischen Mechanik auf. Wir betrachten ein "Gas" aus k Teilchen, die jeweils einen von n Zuständen einnehmen können. In der klassischen Mechanik können wir die Teilchen unterscheiden und jedes Teilchen kann jeden Zustand einnehmen. Wir ziehen also für jedes unserer k Teilchen einen der n Zustände. Weil wir die Teilchen unterscheiden können, spielt hierbei

die Reihenfolge eine Rolle. Insgesamt gibt es also n^k verschiedene Zustände. Dies ist die *Maxwell–Boltzmann Statistik*.

Bei quantenmechanischen Modellen ist dies anders. Hierbei sind die Teilchen nicht unterscheidbar, beim Auswählen der Zustände kommt es also nicht auf die Reihenfolge des Ziehens an. Weiter muss man hier unterscheiden, ob es sich um Fermionen (für die das Pauli-Prinzip gilt) handelt, oder um Bosonen, (für die das Pauli-Prinzip nicht gilt). Bei Fermionen kann jeder Zustand höchstens von einem Teilchen belegt werden, ein bereits gezogener Zustand darf also nicht nochmals gezogen werden. Hierfür gibt es $\binom{n}{k}$ Möglichkeiten (*Fermi–Dirac Statistik*). Bei Bosonen können Zustände mehrfach besetzt werden, es wird also mit Zurücklegen gezogen. Hierfür gibt es $\binom{n+k-1}{k}$ Möglichkeiten (*Bose–Einstein Statistik*).

Beachte, dass es keine physikalischen Teilchen gibt, die unterscheidbar sind und gleichzeitig dem Pauli-Prinzip unterliegen. Daher kommt das Urnenmodell “Ziehen mit Beachtung der Reihenfolge ohne Zurücklegen” in der statistischen Mechanik nicht vor.

Wir verwenden nun unsere kombinatorischen Modelle um in einigen Laplace Experimenten konkrete Wahrscheinlichkeiten auszurechnen.

Beispiel 1.3.6. Auf einer Party sind n Personen anwesend. Mit welcher Wahrscheinlichkeit haben mindestens zwei von ihnen am gleichen Tag Geburtstag?

Wir nehmen vereinfachend an, das Jahr hat immer 365 Tage (lassen also den 29. Februar unter den Tisch fallen). Wir nehmen weiterhin vereinfachend an, dass alle Tage als Geburtstage gleichwahrscheinlich sind (Dies ist nicht realistisch, allerdings kann man zeigen, dass bei unterschiedlichen Wahrscheinlichkeiten für die verschiedenen Tage die Wahrscheinlichkeit, dass mindestens zwei Personen am gleichen Tag Geburtstag haben höchstens größer wird). Zu guter letzt nehmen wir noch an, dass weniger als 365 Personen auf der Party sind (sonst haben *sicher* mindestens zwei am selben Tag Geburtstag).

Als Grundmenge Ω wählen wir alle möglichen n -Tupel von Tagen. Dabei bezeichne der j -te Eintrag gerade den Geburtstag der j -ten Person. Die Geburtstage werden mit Zurücklegen gezogen, Ω hat also 365^n Elemente. Es ist einfacher, statt des Ereignisses A , dass mindestens zwei Personen am gleichen Tag Geburtstag haben, das komplementäre Ereignis A^c zu betrachten. A^c besagt gerade, dass alle n Personen an verschiedenen Tagen Geburtstag haben. Dies entspricht gerade Ziehen *ohne* Zurücklegen (ein bereits verwendeter Geburtstag darf nicht wieder gezogen werden), A^c hat also $(365)_n$ Elemente.

Folglich ist

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{(365)_n}{365^n} = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - k + 1)}{365 \cdot 365 \cdot \dots \cdot 365}.$$

Die folgende Tabelle enthält die Werte von $\mathbb{P}(A)$ (gerundet auf 4 Nachkommastellen) für einige Werte von n :

n	2	7	10	23	50
$\mathbb{P}(A)$	0,0027	0,0562	0,1169	0,5073	0,9704

Beispiel 1.3.7. Wie groß ist die Wahrscheinlichkeit für “4 Richtige” beim Lotto 6 aus 49?

Wir hatten bereits gesehen, dass Grundmenge Ω aller 6-elementiger Teilmengen der Zahlen von 1 bis 49 gerade $\binom{49}{6}$ Elemente hat. Wir interessieren uns für das Ereignis A “4 Richtige”. Die Menge A besteht aus denjenigen 6-elementigen Teilmengen der Zahlen von 1 bis 49, die genau 4 der gezogenen 6 Zahlen (und 2 der nichtgezogenen 43 übrigen Zahlen) erhalten. Es gibt $\binom{6}{4}$ Möglichkeiten 4 der gezogenen Zahlen auszuwählen und zu jeder dieser Möglichkeiten

gibt es $\binom{43}{2}$ Möglichkeiten diese mit nichtgezogenen Zahlen zu einem vollständigen “Tipp” aufzufüllen. Es ist also

$$\mathbb{P}(A) = \frac{\binom{6}{4} \binom{43}{2}}{\binom{49}{6}} = \frac{13.545}{13.983.816} \approx 0,00097$$

Das letzte Beispiel gehört zu den sogenannten *Hypergeometrischen Modellen*: Es wird aus einer Menge von Objekten gezogen, die wiederum in mehrere Klassen aufgeteilt sind (im Beispiel: Von den 49 Zahlen wurden 6 als “gezogen” markiert, die brigen 43 als “nicht gezogen”).

Allgemeiner betrachtet man eine Urne mit n Kugeln, die eine von r verschiedenen Merkmalen (“Farben”) aufweisen. Es gibt n_1 Kugeln der Farbe 1, n_2 Kugeln der Farbe 2, ..., n_r Kugeln der Farbe r . Natürlich soll gelten dass $n_1 + n_2 + \dots + n_r = n$ ist. Nun interessiert man sich für das Ereignis beim Ziehen von k Kugeln genau k_j Kugeln der Farbe j zu ziehen, wobei $k_1 + \dots + k_r = k$ ist. Es gibt

$$\binom{n_1}{k_1} \binom{n_2}{k_2} \cdot \dots \cdot \binom{n_r}{k_r}$$

Möglichkeiten genau k_j Kugeln der Farbe j zu ziehen. Die Wahrscheinlichkeit dieses Ereignisses ist also

$$\frac{\binom{n_1}{k_1} \binom{n_2}{k_2} \cdot \dots \cdot \binom{n_r}{k_r}}{\binom{n}{k}}.$$

Beispiel 1.3.8. Wie hoch ist die Wahrscheinlichkeit beim Poker ein “Full House” (d.h. zieht man zufällig 5 Karten aus Spiel mit 52 Karten, wie hoch ist die Wahrscheinlichkeit, 3 Karten einer Sorte und 2 Karten einer anderen Sorte zu ziehen).

Insgesamt gibt es $\binom{52}{5} = 2.598.960$ verschiedene Pokerhände. Wir berechnen zunächst die Wahrscheinlichkeit ein Full House mit 3 Assen und 2 Königen zu bekommen. Dazu unterteilen wir die 52 Karten in 3 Klassen:ASSE, Könige und sonstige Karten. Es gibt $\binom{4}{3} \binom{4}{2} = 24$ Möglichkeiten genau 3 Assen und 2 Könige zu ziehen. Die Wahrscheinlichkeit ein Full House mit 3 Assen und 2 Königen zu erhalten ist demnach $24/2.598.960$.

Offensichtlich ist dies auch die Wahrscheinlichkeit für ein Full House mit 3 Königen und zwei Assen, sowie die Wahrscheinlichkeit für ein Full House mit 3 Damen und 2 Achtern etc. Bezeichnet also A das Ereignis “Full House” und A_{ij} für

$$i, j \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, \text{Bube, Dame, König, Ass}\} =: K$$

mit $i \neq j$ das Ereignis “Full House mit 3 i und 2 j ” so ist wegen der Additivität (beachte, $A_{ij} \cap A_{kl} = \emptyset$ für $ij \neq kl$)

$$\mathbb{P}(A) = \sum_{ij \in K, i \neq j} \mathbb{P}(A_{ij}) = (13)_2 \frac{\binom{4}{3} \binom{4}{2}}{\binom{52}{5}} \approx 0,0014$$

Denn es gibt $(13)_2$ Möglichkeiten zwei Elemente aus K ohne Zurückzulegen zu ziehen.

1.4 Bedingte Wahrscheinlichkeit und Unabhängigkeit

In vielen Situationen hat man schon bevor das Ergebnis eines Zufallsexperiments bekannt ist eine gewisse Teilinformation. Ein Kartenspieler, beispielsweise, kennt seine eigenen Karten,

nicht aber die der anderen Mitspieler. Natürlich wird er diese Information bei der Abwägung von Wahrscheinlichkeiten berücksichtigen. Beispielsweise wird er, wenn er selbst 2 Asses besitzt, die Wahrscheinlichkeit, dass sein linker Nachbar mindestens ein Ass hat, geringer einschätzen als wenn er selbst kein Ass besitzt.

Um diese Idee zu formalisieren möchte man eine “Wahrscheinlichkeit *gegeben eine Zusatzinformation*” definieren. Die Zusatzinformation ist das Wissen, dass ein gewisses Ereignis B (z.B. “ich selbst habe 2 Asses”) eingetreten ist. Im Falle eines Laplace Wahrscheinlichkeitsraums kann man diese Idee verwirklichen, indem man den Grundraum einschränkt: Man berücksichtigt nur noch Elementarereignisse, die in B liegen. Schreibt man $\mathbb{P}(A|B)$ (lies: Die Wahrscheinlichkeit von A gegeben B), so erhält man

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

Berücksichtigt man noch dass $\mathbb{P}(A \cap B) = \#(A \cap B)/\#\Omega$ und $\mathbb{P}(B) = \#B/\#\Omega$ ist, so erhält man

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

Diese Gleichung ist auch noch sinnvoll, wenn wir auf einem beliebigen endlichen Wahrscheinlichkeitsraum sind. Wir definieren also

Definition 1.4.1. Es sei $(\Omega, \Sigma, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum und B ein Ereignis mit positiver Wahrscheinlichkeit. Für $A \in \Sigma$ ist *bedingte Wahrscheinlichkeit* $\mathbb{P}(A|B)$ von A gegeben B ist definiert als

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

Der Name *bedingte Wahrscheinlichkeit* ist gerechtfertigt:

Lemma 1.4.2. Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $B \subset \Omega$ mit $\mathbb{P}(B) > 0$. Dann ist $\mathbb{P}(\cdot|B)$ ein Wahrscheinlichkeitsmaß auf (Ω, Σ) .

Beweis. Weil $\emptyset \subset A \cap B \subset B$ ist folgt aus Proposition 1.2.5 (a), dass $0 = \mathbb{P}(\emptyset) \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ und somit $0 \leq \mathbb{P}(A|B) \leq 1$ für alle $A \subset \Omega$ ist. Weil $\Omega \cap B = B$ gilt trivialerweise $\mathbb{P}(\Omega|B) = 1$. Sei nun eine Folge A_k von paarweise unvereinbaren Ereignissen gegeben. Dann sind auch die Ereignisse $B_k := A_k \cap B$ paarweise unvereinbar und es gilt $\bigcup B_k = B \cap \bigcup A_k$. Aus der σ -Additivität folgt somit

$$\mathbb{P}\left(B \cap \bigcup A_k\right) = \sum \mathbb{P}(B \cap A_k).$$

Teilt man durch $\mathbb{P}(B)$ so folgt die σ -Additivität von $\mathbb{P}(\cdot|B)$. □

Die Definition der bedingten Wahrscheinlichkeit lässt sich wie folgt umformulieren: $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Diese Formel lässt sich auch auf endlich viele Ereignisse ausdehnen:

Proposition 1.4.3. (*Multiplikationsregel*) Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \Sigma$ Ereignisse mit $\mathbb{P}(A_1 \cap \dots \cap A_n) > 0$. Dann ist

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) \cdot \mathbb{P}(A_{n-1}|A_1 \cap \dots \cap A_{n-2}) \cdot \dots \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_1).$$

Beispiel 1.4.4. Eine Urne enthalte 10 rote und 10 blaue Kugeln. Nun wird eine Kugel gezogen und diese wird dann, zusammen mit 5 Kugeln der gleichen Farbe, zurückgelegt. Dies wird insgesamt 3 mal wiederholt. Es sei A_j das Ereignis, dass im j -ten Versuch eine rote Kugel gezogen wurde.

Mit der Multiplikationsregel ist

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_3|A_2 \cap A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_1) = \frac{2}{3} \cdot \frac{3}{5} \cdot \frac{1}{2} = \frac{1}{5}.$$

In der Tat, wurden im ersten und zweiten Versuch rote Kugeln gezogen, so befinden sich im dritten Versuch 20 rote und 10 blaue Kugeln in der Urne, sodass die Wahrscheinlichkeit eine rote zu ziehen $2/3$ ist. Ähnlich sieht man, dass $\mathbb{P}(A_2|A_1) = 3/5$ ist.

Nun kommen wir zu zwei wichtigen Aussagen über bedingte Wahrscheinlichkeiten. In der Formulierung verwenden wir den Begriff der *Partition*. Eine Partition einer Menge Ω ist eine Folge $\Omega_1, \Omega_2, \Omega_3, \dots$, derart, dass (i) die Ω_j paarweise disjunkt sind, also $\Omega_i \cap \Omega_j = \emptyset$ für $i \neq j$ und (ii) die Vereinigung aller Ω_j gerade Ω ist, also $\bigcup_{j=1}^{\infty} \Omega_j = \Omega$. Wählt man $\emptyset = \Omega_{n+1} = \Omega_{n+2} = \dots$, so kann man auch endliche Partitionen betrachten.

Theorem 1.4.5. Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ und eine Partition (Ω_k) von Ω mit $\Omega_j \in \Sigma$ für alle $j \in \mathbb{N}$.

(1) (Satz von der totalen Wahrscheinlichkeit) Für $A \in \Sigma$ gilt

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} \mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j)$$

wobei wir für $\mathbb{P}(\Omega_j) = 0$ das Produkt $\mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j)$ als 0 festlegen.

(2) (Satz von Bayes) Für $A \in \Sigma$ mit $\mathbb{P}(A) > 0$ gilt

$$\mathbb{P}(\Omega_j|A) = \frac{\mathbb{P}(\Omega_j)\mathbb{P}(A|\Omega_j)}{\sum_{k=1}^{\infty} \mathbb{P}(A|\Omega_k)\mathbb{P}(\Omega_k)} = \frac{\mathbb{P}(\Omega_j)\mathbb{P}(A|\Omega_j)}{\mathbb{P}(A)}.$$

Beweis. (1) Weil Ω die disjunkte Vereinigung der Ω_j ist, ist A die disjunkte Vereinigung der $A \cap \Omega_j$. Wegen der σ -Additivität von \mathbb{P} folgt

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{j \in \mathbb{N}} (A \cap \Omega_j)\right) = \sum_{j=1}^{\infty} \mathbb{P}(A \cap \Omega_j) = \sum_{j=1}^{\infty} \mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j).$$

(2) folgt aus (1) und der Gleichheit $\mathbb{P}(A \cap \Omega_j) = \mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j)$. \square

Beispiel 1.4.6. Ein Unternehmen bezieht ein elektronisches Bauteil von drei verschiedenen Zulieferern I, II und III. Dabei stammen von I 50% der Bauteile, von II und III jeweils 25% der Bauteile. Aus Erfahrung ist bekannt, dass die Wahrscheinlichkeit dass ein Bauteil defekt ist bei Lieferant I 1%, bei Lieferant II 2% und bei Lieferant III sogar 4% beträgt.

Mit welcher Wahrscheinlichkeit ist ein zufällig ausgewähltes Bauteil defekt? Es sei A das Ereignis "Das Bauteil ist defekt" und B_j das Ereignis "Das Bauteil stammt von Lieferant j ". Nach dem Satz über die totale Wahrscheinlichkeit gilt

$$\mathbb{P}(A) = \mathbb{P}(A|B_I)\mathbb{P}(B_I) + \mathbb{P}(A|B_{II})\mathbb{P}(B_{II}) + \mathbb{P}(A|B_{III})\mathbb{P}(B_{III})$$

$$= 0,01 \cdot 0,5 + 0,02 \cdot 0,25 + 0,04 \cdot 0,25 = 0,02.$$

Mit welcher Wahrscheinlichkeit stammt ein defektes Bauteil von Lieferant I? Nach dem Satz von Bayes ist

$$\mathbb{P}(B_I|A) = \frac{\mathbb{P}(A|B_I)\mathbb{P}(B_I)}{\mathbb{P}(A)} = \frac{0,01 \cdot 0,5}{0,02} = \frac{1}{4}.$$

Beispiel 1.4.7. Ein Test auf eine bestimmte Krankheit erkennt mit 99% Wahrscheinlichkeit eine erkrankte Person als krank (= "Sensitivität"). Die Wahrscheinlichkeit, dass eine nicht erkrankte Person als gesund erkannt wird betrage 95% (= "Spezifität"). Es ist bekannt, dass 1% der Bevölkerung an dieser Krankheit leiden. Wie groß ist die Wahrscheinlichkeit an der Krankheit zu leiden, wenn der Test zu dem Ergebnis kommt, dass man krank ist?

Es sei A das Ereignis "Person ist krank" und B das Ereignis "Test ist positiv". Nach obigen Angaben ist also $\mathbb{P}(A) = 0,01$, $\mathbb{P}(B|A) = 0,99$ und $\mathbb{P}(B^c|A^c) = 0,95$, also $\mathbb{P}(B|A^c) = 0,05$. Nach dem Satz von Bayes ist

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{0,99 \cdot 0,01}{0,99 \cdot 0,01 + 0,05 \cdot 0,99} = \frac{1}{6}.$$

Schliesslich diskutieren wir noch ein klassisches Beispiel, das sogenannte *Ziegenproblem*.

Beispiel 1.4.8. Bei einer Spielshow kommt folgende Situation vor. Der Kandidat sieht drei (verschlossene) Tore. Hinter einem Tor befindet sich der Hauptpreis, ein Auto. Hinter den beiden anderen Toren befindet sich eine Ziege. Der Kandidat wählt ein Tor aus. Danach öffnet der Moderator ein Tor, hinter dem sich eine Ziege verbirgt. Er bietet sodann dem Kandidaten an, doch noch das Tor zu wechseln. Sollte der Kandidat das Tor wechseln oder bei seiner ursprünglichen Wahl bleiben?

Um diese Frage zu beantworten, vergleichen wir die Gewinnchancen bei den beiden Strategien "wechseln" und "nicht wechseln". Zunächst die Strategie "nicht wechseln". Da die drei Tore gleichberechtigt sind, ist die Wahrscheinlichkeit zu Beginn das richtige Tor zu wählen $1/3$. Bei der Strategie "nicht wechseln" ist dies auch die Wahrscheinlichkeit am Ende zu gewinnen, denn die zusätzliche Information, wo sich eine Ziege befindet, wird ja gar nicht berücksichtigt.

Kommen wir nun zur Strategie "wechseln". Bezeichnet A das Ereignis "der Kandidat hat zu Beginn das richtige Tor gewählt" und B das Ereignis "der Kandidat gewinnt am Ende", so ist, bei Verwendung der Strategie "wechseln", $\mathbb{P}(B|A) = 0$. Hat der Kandidat nämlich das richtige Tor gewählt, so stehen hinter beiden anderen Toren Ziegen, weshalb der Kandidat beim Wechseln automatisch zu einer Ziege wechselt. Hat der Kandidat sich andererseits zunächst für ein falsches Tor entschieden, so befindet sich hinter dem einen verbleibenden Tor der Hauptpreis, hinter dem anderen die zweite Ziege. Da aber der Moderator das Tor mit der Ziege öffnet, muss das verbleibende Tor das Auto verbergen. Es ist also $\mathbb{P}(B|A^c) = 1$. Nach dem Satz von der totalen Wahrscheinlichkeit ist

$$\mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Somit ist die Wahrscheinlichkeit zu gewinnen bei der Strategie "wechseln" doppelt so hoch, wie bei der Strategie "nicht wechseln".

Als nächstes definieren wir den wichtigen Begriff der *Unabhängigkeit*.

Definition 1.4.9. Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum $A, B \in \Sigma$ Ereignisse. Dann heißen A und B *unabhängig*, wenn $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Allgemeiner heißt eine Familie $(A_i)_{i \in I}$ von Ereignissen *unabhängig*, falls für paarweise verschiedene Indices $i_1, \dots, i_n \in I$ stets

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_n}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_n})$$

gilt.

Sind A und B unabhängig und ist $\mathbb{P}(A) > 0$, so ist

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

die bedingte Wahrscheinlichkeit von B gegeben A ist also gleich der Wahrscheinlichkeit von B . Mit anderen Worten, die Kenntnis dass A eingetreten ist lässt keine Rückschlüsse darüber zu, ob B eingetreten ist.

Beispiel 1.4.10. Wir werfen eine faire Münze zweimal. Als Grundmenge wählen wir $\Omega = \{KK, KZ, ZK, ZZ\}$; wir unterstellen, dass alle Elementarereignisse gleich wahrscheinlich sind. Wie betrachten folgende Ereignisse:

$$A_1 = \{ZZ, ZK\} = \text{Im ersten Wurf fällt Zahl}$$

$$A_2 = \{KK, ZK\} = \text{Im zweiten Wurf fällt Kopf}$$

$$A_3 = \{KZ, ZK\} = \text{Verschiedene Ausgänge im ersten und zweiten Wurf.}$$

Dann ist $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = 1/2$. Weiter ist $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(\{ZK\}) = \mathbb{P}(A_2 \cap A_3) = 1/4 = 1/2 \cdot 1/2$. Also sind A_1 und A_2 unabhängig und auch A_2 und A_3 unabhängig. Weil $\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(\{ZK\}) = 1/4 = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3)$ ist, sind auch A_1 und A_3 unabhängig. Allerdings ist $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(\{ZK\}) = 1/4 \neq 1/8 = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$. Daher sind A_1, A_2, A_3 nicht unabhängig.

Kapitel 2

Zufallsvariablen und ihre Verteilungen

Bei vielen Zufallsexperimenten interessiert nicht der eigentliche Versuchsausgang $\omega \in \Omega$, sondern lediglich eine “Kenngröße” die jedoch vom Versuchsausgang ω abhängt. Man denke etwa an die Körpergröße eines zufällig ausgewählten Ulmers. Häufig sind sogar die genauen Vorgänge, “die zum Versuchsausgang ω führen” (also das Wahrscheinlichkeitsmaß \mathbb{P}), nicht bekannt. Über Wahrscheinlichkeiten die die erwähnte Kenngröße betreffen kann man jedoch sehr wohl Aussagen machen.

Aus diesem Grund betrachtet man sogenannte *Zufallsvariablen*, d.h. Abbildungen von Ω nach \mathbb{R} (die gewisse technische Bedingungen erfüllen). Diese Zufallsvariablen ermöglichen es uns, uns auf wesentliche Aspekte eines Zufallsexperiments zu beschränken und diese genauer zu beschreiben.

2.1 Definitionen

Definition 2.1.1. Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine *Zufallsvariable* ist eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ derart, dass

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \Sigma$$

für alle $B \in \mathcal{B}(\mathbb{R})$. Eine Abbildung mit dieser Eigenschaft nennt man *messbar*.

Bemerkung 2.1.2. Gelegentlich betrachtet man auch Abbildungen $X : \Omega \rightarrow \mathbb{C}$ oder auch $X : \Omega \rightarrow \mathbb{R}^d$. Diese nennt man messbar, falls alle Komponenten $\operatorname{Re} X$ und $\operatorname{Im} X$ bzw. X_1, \dots, X_d messbar sind. Man spricht von *komplexen Zufallsvariablen* bzw. von *Zufallsvektoren*.

Bemerkung 2.1.3. Ist $\Sigma = \mathcal{P}(\Omega)$, was bei endlicher Grundmenge Ω häufig der Fall ist, so ist *jede* Abbildung $X : \Omega \rightarrow \mathbb{R}$ messbar und somit eine Zufallsvariable.

Beispiel 2.1.4. (Indikatorfunktionen)

Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Weiter sei für $A \subset \Omega$ die *Indikatorfunktion* $\mathbb{1}_A$ von A gegeben als

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

Dann ist $\mathbb{1}_A$ eine Zufallsvariable genau dann, wenn $A \in \Sigma$ liegt. Es ist nämlich

$$\mathbb{1}_A^{-1}(B) = \begin{cases} \emptyset, & B \cap \{0, 1\} = \emptyset \\ A^c, & 0 \in B, 1 \notin B \\ A, & 1 \in B, 0 \notin B \\ \Omega, & \{0, 1\} \subset B. \end{cases}$$

Die Mengen auf der rechten Seite sind messbar, genau dann, wenn $A \in \Sigma$.

Ohne Beweis geben wir folgende Eigenschaften von Zufallsvariablen an.

Proposition 2.1.5. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Weiter seien $X, Y : \Omega \rightarrow \mathbb{R}$ Abbildungen, $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ stetig und $\alpha, \beta \in \mathbb{R}$.*

- (a) *X ist genau dann eine Zufallsvariable, wenn $\{X \leq x\} := \{\omega : X(\omega) \leq x\} \in \Sigma$ für alle $x \in \mathbb{R}$.*
- (b) *Sind X, Y Zufallsvariablen, so ist auch $\alpha X + \beta Y$ eine Zufallsvariable.*
- (c) *Ist X eine Zufallsvariable, so ist auch $\varphi \circ X$ eine Zufallsvariable.*
- (d) *Ist X eine Zufallsvariable, so ist die Abbildung $\mu_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$, definiert durch $\mu_X(B) = \mathbb{P}(X^{-1}(B))$ ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Das Wahrscheinlichkeitsmaß μ_X aus Proposition 2.1.5(d) heißt *Verteilung von X* . Es stellt sich die Frage nach der Beschreibung von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Wie schon im Falle des Lebesguemaßes sind die Wahrscheinlichkeitsmaße durch ihre Werte auf gewissen Systemen von Mengen eindeutig bestimmt.

Definition 2.1.6. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable. Dann heißt die Funktion $F_X : \mathbb{R} \rightarrow \mathbb{R}$, definiert durch $F_X(x) := \mathbb{P}(X \leq x)$ *Verteilungsfunktion von X* .*

Beispiel 2.1.7. *Ist $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X = \mathbb{1}_A$ für ein $A \in \Sigma$, so ist $F_X(x) = \mathbb{P}(A^c)\mathbb{1}_{[0,1)}(x) + \mathbb{1}_{[1,\infty)}(x)$.*

Proposition 2.1.8. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann hat die Verteilungsfunktion F_X folgende Eigenschaften:*

- (i) *F_X ist monoton wachsend.*
- (ii) *F_X ist rechtsseitig stetig, d.h. ist $x_n \downarrow x$ so ist $F_X(x_n) \rightarrow F_X(x)$.*
- (iii) *Es gilt $\lim_{x \rightarrow -\infty} F_X(x) = 0$ und $\lim_{x \rightarrow \infty} F_X(x) = 1$.*

Beweis. Ist $x \leq y$, so ist $\{X \leq x\} \subset \{X \leq y\}$ und somit $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$. Dies zeigt (i).

Sei nun $x_n \downarrow x$. Dann ist $\{X \leq x_n\} \downarrow \{X \leq x\}$. Offensichtlich ist nämlich $\{X \leq x_{n+1}\} \subset \{X \leq x_n\}$ und $\{X \leq x\} \subset \bigcap_n \{X \leq x_n\}$. Ist umgekehrt $\omega \in \bigcap_n \{X \leq x_n\}$, so ist $X(\omega) \leq x_n$ für alle $n \in \mathbb{N}$ und somit auch $X(\omega) \leq \inf_n x_n = x$. Nun folgt (ii) aus der Stetigkeit des Maßes (Lemma 1.2.10).

Auch (iii) folgt aus der Stetigkeit des Maßes und der Beobachtung, dass einerseits für $x_n \downarrow -\infty$ stets $\{X \leq x_n\} \downarrow \emptyset$ und für $x_n \uparrow \infty$ andererseits $\{X \leq x_n\} \uparrow \Omega$ gilt. \square

Umgekehrt ist jede Funktion die Eigenschaften (i) bis (iii) in Proposition 2.1.8 erfüllt bereits eine Verteilungsfunktion. Genauer gilt folgender Satz, den wir hier nicht beweisen.

Satz 2.1.9. *Es sei $F : \mathbb{R} \rightarrow \mathbb{R}$ eine monoton wachsende, rechtsseitig stetige Funktion mit $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$. Dann gibt es ein Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $\mu((-\infty, x]) = F(x)$. Weiter gibt es einen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ und eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ mit Verteilungsfunktion F .*

Mit Hilfe der Verteilungsfunktion lassen sich weitere Wahrscheinlichkeiten für die Zufallsvariable X berechnen.

Lemma 2.1.10. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion F . Weiter seien $a, b, x \in \mathbb{R}$ mit $a < b$ gegeben.*

$$(a) \mathbb{P}(a < X \leq b) = F(b) - F(a).$$

$$(b) \mathbb{P}(X = x) = F(x) - F(x-), \text{ wobei } F(x-) = \lim_{y \uparrow x} F(y) \text{ der linksseitige Grenzwert von } F \text{ an der Stelle } x \text{ ist.}$$

$$(c) \mathbb{P}(X \in [a, b]) = F(b) - F(a-).$$

$$(d) \mathbb{P}(X < x) = F(x-).$$

Beweis. (a) Es ist $(a, b] = (-\infty, b] \setminus (-\infty, a]$ und somit $\mathbb{P}(X \in (a, b]) = \mathbb{P}(X \in (-\infty, b]) - \mathbb{P}(X \in (-\infty, a]) = F(b) - F(a)$.

(b) Ist $y_n \uparrow x$, so ist $(y_n, x] \downarrow \{x\}$. Aufgrund der Stetigkeit des Maßes folgt $\mathbb{P}(X = x) = \lim_{n \rightarrow \infty} \mathbb{P}(X \in (y_n, x]) = \lim_{n \rightarrow \infty} F(x) - F(y_n) = F(x) - F(x-)$.

(c) und (d) (ÜA). □

2.2 Diskrete Verteilungen

Wir nennen eine Zufallsvariable X *diskret* wenn sie höchstens abzählbar viele Werte annimmt, d.h. entweder nimmt X nur endlich viele Werte an oder X nimmt abzählbar unendlich viele Werte an. Genauer gibt es Zahlen $(x_k)_{k \in N}$ wobei $N = \{1, \dots, n\}$ endlich ist oder $N = \mathbb{N}$ abzählbar unendlich ist sodass $\mathbb{P}(X \notin \{x_k : k \in N\}) = 0$. Beachte, dass in diesem Fall $A_k := \{X = x_k\}$ messbar sein muss (denn $A_k = X^{-1}(\{x_k\})$). Weiter gilt in diesem Fall $p_k := \mathbb{P}(X = x_k) = \mathbb{P}(A_k) \in [0, 1]$ und $\sum_{k \in N} p_k = 1$.

Die Verteilung μ_X von X ist also auf der diskreten Menge $\{x_k : k \in N\}$ konzentriert. Die Situation ist hier ähnlich wie in Beispiel 1.2.6. Häufig nennt man die Zahlen p_k auch die *Zähldichte* der Zufallsvariablen X bzw. der Verteilung μ_X . Die Verteilungsfunktion einer diskreten Verteilung ist gegeben als $F_X(x) = \sum_{k: x_k \leq x} p_k$. Beachte, dass dies eine reine Sprungfunktion ist. An der Stelle x_k tritt ein Sprung in Höhe p_k auf.

Wir stellen nun einige Wichtige diskrete Verteilungen vor.

Die Binomialverteilung

Eine Zufallsvariable X heißt *binomialverteilt* mit Parametern n und p , falls $X(\Omega) = \{0, 1, \dots, n\}$ und

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

für alle $k = 0, 1, \dots, n$ gilt. Wir schreiben $X \sim \mathbf{b}_{n,p}$. Ist $n = 1$ so sagt man auch, die Zufallsvariable sei *Bernoulli verteilt*.

Die *Interpretation* ist wie folgt. Bei einem *Bernoulli Experiment* gibt es nur zwei Mögliche Ergebnisse: Erfolg oder Misserfolg. Die Wahrscheinlichkeit für einen Erfolg (die *Erfolgswahrscheinlichkeit*) ist p , die Wahrscheinlichkeit für einen Misserfolg ist dementsprechend $1 - p$. Eine Bernoulli-verteilte Zufallsvariable ist nun 1 bei Erfolge (Wahrscheinlichkeit hierfür ist p) und 0 bei Misserfolg (Wahrscheinlichkeit $1 - p$).

Nun wiederholt man dieses Bernoulli Experiment *unabhängig voneinander* n -mal. Wegen der Unabhängigkeit multiplizieren sich die Wahrscheinlichkeiten für die Ausgänge miteinander. Zum Beispiel ist, beim 4maligen Wiederholen, die Wahrscheinlichkeit für Erfolg im 1. und 4. Versuch und Misserfolg im 2. und 3. Versuch gerade $p \cdot (1 - p) \cdot (1 - p) \cdot p = p^2(1 - p)^2$. Die Wahrscheinlichkeit in n Versuchen k Erfolge zu haben ist gerade $\binom{n}{k} p^k (1 - p)^{n-k}$. Dies sieht man wie folgt: Man wählt von den n Versuchen k aus. Dafür gibt es $\binom{n}{k}$ Möglichkeiten. Die Wahrscheinlichkeit in genau diesen Versuchen Erfolg und in den anderen Misserfolg zu haben ist $p^k (1 - p)^{n-k}$. Insgesamt ist also die Wahrscheinlichkeit für k Erfolge $\binom{n}{k} p^k (1 - p)^{n-k}$.

Die Anzahl der Erfolge beim n -maligen Wiederholen eines Bernoulli Experiments mit Erfolgswahrscheinlichkeit p ist Binomialverteilt mit Parametern n und p .

Die hypergeometrische Verteilung

Es seien $m, n, k \in \mathbb{N}$ mit $k \leq m + n$. Eine Zufallsvariable X heißt *hypergeometrisch verteilt* mit Parametern m, n und k , falls $X(\Omega) = \{0, 1, \dots, k\}$ und

$$\mathbb{P}(X = j) = \frac{\binom{m}{j} \binom{n}{k-j}}{\binom{m+n}{k}}$$

für $j = 0, 1, \dots, k$ gilt. Hierbei setzen wir $\binom{a}{b} := 0$ falls $b < 0$ oder $b > a$ ist.

Wir schreiben $X \sim \text{hg}_{m,n,k}$. Die hypergeometrische Verteilung beschreibt folgendes Experiment:

In einer Urne befinden sich m schwarze und n weiße Kugeln. Es werden k Kugeln ohne Zurücklegen und ohne Beachten der Reihenfolge gezogen. Dann hat die Anzahl der schwarzen Kugeln unter den gezogenen gerade Verteilung $\text{hg}_{m,n,k}$. Damit eine Wahrscheinlichkeitsverteilung vorliegt muss natürlich

$$1 = \sum_{j=0}^k \mathbb{P}(X = j) = \sum_{j=0}^k \frac{\binom{m}{j} \binom{n}{k-j}}{\binom{m+n}{k}}$$

sein. Äquivalent hierzu ist

$$\binom{m+n}{k} = \sum_{j=0}^k \binom{m}{j} \binom{n}{k-j}. \quad (2.1)$$

Ruft man sich die Interpretation der hypergeometrischen Verteilung ins Gedächtnis, so ist die Gleichheit (2.1) klar:

Es gibt genau $\binom{m}{j} \binom{n}{k-j}$ Möglichkeiten, aus m schwarzen Kugeln j und aus n weißen Kugeln $k - j$ auszuwählen. Summiert man über j , so hat man alle Möglichkeiten abgezählt, aus der Gesamtheit der $m + n$ Kugeln k auszuwählen, also gerade $\binom{m+n}{k}$.

Die geometrische Verteilung

Eine Zufallsvariable X heißt *geometrisch verteilt* mit Parameter $p \in (0, 1)$ falls $X(\Omega) = \mathbb{N}$ ist und $\mathbb{P}(X = k) = p(1 - p)^{k-1}$ ist. Wir schreiben $X \sim \text{geom}_p$.

Eine geometrische Verteilung beschreibt beim unabhängigen Wiederholen eines Zufallsexperimentes mit Erfolgswahrscheinlichkeit p die Anzahl der Versuche bis zum ersten Erfolg.

In der Tat ist die Wahrscheinlichkeit in den ersten $k-1$ Versuchen Misserfolg und im k -ten Versuch Erfolg zu haben gerade $(1-p)^{k-1}p$. Beachte, dass mit der Formel für die geometrische Reihe $\sum_{j=0}^{\infty} q^j = (1-q)^{-1}$ folgt, dass

$$\sum_{k=1}^{\infty} (1-p)^{k-1}p = p \sum_{j=0}^{\infty} (1-p)^j = p \frac{1}{1-(1-p)} = \frac{p}{p} = 1.$$

Die Poisson Verteilung

Eine Zufallsvariable X heißt *Poisson verteilt* mit Parameter $\lambda > 0$, falls $X(\Omega) = \mathbb{N}_0$ ist und $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$. Beachte, dass

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Wir schreiben $X \sim \text{Pois}_{\lambda}$.

Die Poisson Verteilung tritt bei der Approximation der Binomialverteilung auf. Wir betrachten die \mathbf{b}_{n,p_n} Verteilung, wobei wir annehmen, dass $\lambda_n := np_n \rightarrow \lambda$ für $n \rightarrow \infty$ konvergiert. Dann ist für festes k

$$\begin{aligned} \mathbf{b}_{n,p_n}(\{k\}) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} n(n-1) \cdots (n-k+1) \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{1}{k!} \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \lambda_n^k \left(1 - \frac{\lambda_n}{n}\right)^n \cdot \left(1 - \frac{\lambda_n}{n}\right)^{-k} \\ &\rightarrow \frac{1}{k!} \cdot 1 \cdot 1 \cdots 1 \cdot \lambda^k e^{-\lambda} (1-0)^{-k}, \end{aligned}$$

wobei wir verwendet haben, dass $(1 + x_n/n)^n \rightarrow e^x$ konvergiert, falls $x_n \rightarrow x$. Somit haben wir gezeigt:

Proposition 2.2.1. *Ist p_n eine Folge in $(0,1)$ mit $np_n \rightarrow \lambda > 0$, so gilt*

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

für jedes $k = 0, 1, 2, \dots$

Bemerkung 2.2.2. In den Anwendungen hat man in der Regel keine Folge p_n von Erfolgswahrscheinlichkeiten in Binomialexperimenten gegeben. Man interpretiert Proposition 2.2.1 daher wie folgt:

Für *kleine Werte* von p und *große Werte* von n kann die Binomialverteilung mit Parametern n und p mit der Poisson verteilung mit Parameter $\lambda = np$ approximiert werden. Daher nennt man die Poisson Verteilung auch die Verteilung der *seltenen Ereignissen*.

Wir verdeutlichen dies an einem Beispiel.

Beispiel 2.2.3. (Radioaktiver Zerfall)

Radioaktiven Zerfall kann man wie folgt modellieren:

Jeder einzelne Atomkern hat (innerhalb einer gewissen Zeit, etwa 10 Sekunden) eine gewisse Wahrscheinlichkeit p zu zerfallen oder auch nicht. Man kann jedoch nicht einzelne Atomkerne betrachten. Stattdessen betrachtet man eine bestimmte Menge eines radioaktiven

Materials und zählt die Zerfälle in einer Sekunde. Diese Anzahl ist dann binomialverteilt mit Parametern n (=Anzahl der Atomkerne) und p . Typischerweise ist n sehr groß, (die Einheit mol der Stoffmenge ist die Anzahl der Atome in 12 Gramm eines bestimmten Kohlenstoffisotops. 1 mol sind ungefähr $6 \cdot 10^{23} = \text{viel}$) die Anzahl der Zerfälle die man in Versuchen beobachtet ist aber vergleichsweise klein (Größenordnung 5 Zerfälle je 10 Sekunden).

Demnach ist es plausibel, eine Poissonverteilung zu unterstellen.

2.3 Absolutstetige Verteilungen

Definition 2.3.1. Es sei X eine Zufallsvariable mit Verteilungsfunktion F_X . Wir sagen, X hat *absolutstetige Verteilung* (gelegentlich auch X sei *absolutstetig*), falls es eine nichtnegative Funktion f auf \mathbb{R} mit

$$\int_{-\infty}^{\infty} f(t) dt = 1$$

gibt (hierbei soll das Integral wohldefiniert sein), sodass

$$F_X(x) = \int_{-\infty}^x f(t) dt$$

In diesem Fall heißt f *Dichte* der Verteilung von X .

Bemerkung 2.3.2. Ist f eine nichtnegative Funktion auf \mathbb{R} mit $\int_{-\infty}^{\infty} f(t) dt = 1$, so kann man zeigen, dass $F(x) = \int_{-\infty}^x f(t) dt$ eine monoton wachsende, rechtsseitig stetige Funktion mit $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$ ist. Somit ist F eine Verteilungsfunktion und f die Dichte dieser Verteilungsfunktion. Man nennt daher jede nichtnegative Funktion f mit $\int_{\mathbb{R}} f(t) dt = 1$ bereits *Dichte*.

Beispiel 2.3.3. Setzen wir

$$f(t) = \begin{cases} \frac{1}{t^2} & t \geq 1 \\ 0 & t < 1 \end{cases} = \frac{1}{t^2} \mathbb{1}_{[1, \infty)}$$

so ist f Dichte einer Verteilungsfunktion. Es ist nämlich

$$\int_{-\infty}^{\infty} f(t) dt = \int_1^{\infty} \frac{1}{t^2} dt = \left[\frac{-1}{t} \right]_1^{\infty} = 0 - (-1) = 1.$$

Ist f die Dichte der Verteilungsfunktion von X , so ist

$$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 \frac{1}{t^2} dt = \left[\frac{-1}{t} \right]_1^2 = -\frac{1}{2} + 1 = \frac{1}{2},$$

also nimmt X mit Wahrscheinlichkeit $1/2$ einen Wert zwischen 1 und 2 an.

Bemerkung 2.3.4. In praktisch allen in Anwendungen relevanten absolutstetigen Verteilungen ist die Dichte f stückweise stetig. In diesem Fall ist die Verteilungsfunktion $F(x) = \int_{-\infty}^x f(t) dt$ als Riemann-Integral definiert. Umgekehrt ist in diesem Fall die Dichte (bis auf endlich viele Ausnahmepunkte) gerade die Ableitung der Verteilungsfunktion. Im Allgemeinen muss man jedoch auf die (hier nicht näher diskutierte) Lebesgue'sche Integrations- und Ableitungstheorie zurückgreifen.

Wir geben nun einige wichtige absolutstetige Verteilungen an.

Stetige Gleichverteilung

Wir sagen X ist *gleichverteilt auf dem Interval* (a, b) und schreiben $X \sim U(a, b)$, falls X die Dichte

$$f(t) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(t) = \begin{cases} \frac{1}{b-a}, & a < t < b \\ 0, & \text{sonst} \end{cases}$$

besitzt.

Ein Beispiel, bei dem diese Verteilung auftritt ist beim Drehen eines ‘‘Glücksrades’’. Der Winkel in dem das Rad im Vergleich zur Ausgangslage zum stehen kommt ist gleichverteilt in $(0, 2\pi)$.

Die Verteilungsfunktion der Gleichverteilung auf (a, b) ist gegeben durch

$$F(x) = \int_{-\infty}^x \mathbb{1}_{(a,b)}(t) dt = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x > b. \end{cases}$$

Exponentialverteilung

Eine Zufallsvariable X heißt exponentialverteilt mit Parameter $\lambda > 0$, wenn X die Dichte

$$f_\lambda(t) = \lambda e^{-\lambda t} \mathbb{1}_{(0,\infty)}(t)$$

besitzt. Wir schreiben in diesem Fall $X \sim \exp_\lambda$. Beachte, dass

$$\int_{-\infty}^{\infty} f_\lambda(t) dt = \int_0^{\infty} \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t} \right]_0^{\infty} = 0 - (-1) = 1$$

sodass f_λ in der Tat eine Dichte ist.

Die Exponentialverteilung ist in gewisser Weise die ‘‘zeitstetige’’ Variante der geometrischen Verteilung und tritt bei sogenannten Wartezeitproblemen auf. Wichtige Beispiele sind: Die Lebensdauer von Glühbirnen oder die Wartezeit auf den nächsten Anruf in einem Callcenter.

Die Exponentialverteilung hat eine wichtige Eigenschaft, die man *Gedächtnislosigkeit* nennt. Ist nämlich $X \sim \exp_\lambda$, so ist

$$\mathbb{P}(X \geq a + b | X \geq a) = \mathbb{P}(X \geq b). \quad (2.2)$$

Bei Wartezeitproblemen interpretiert man diese Gleichheit wie folgt:

Die Wahrscheinlichkeit noch mindestens die Zeitspanne b warten zu müssen, wenn man bereits a gewartet hat ist genau so groß, wie von Anfang an mindestens b warten zu müssen. Um Gleichung (2.2) zu zeigen, rufen wir uns die Definition der bedingten Wahrscheinlichkeit ins Gedächtnis: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Hier haben wir $A = \{X \geq a + b\}$ und $B = \{X \geq a\}$. Beachte, dass $A \subset B$ und daher $A \cap B = A$. Nun beachten wir, dass für $x \in (0, \infty)$

$$\mathbb{P}(X \geq x) = \int_x^{\infty} \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t} \right]_x^{\infty} = e^{-\lambda x}.$$

Somit ergibt sich

$$\frac{\mathbb{P}(X \geq a + b, X \geq a)}{\mathbb{P}(X \geq a)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = \mathbb{P}(X \geq b)$$

wie behauptet.

Normalverteilung

Es seien $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Eine Zufallsvariable X heißt *normalverteilt* mit Parametern μ und σ^2 , falls X die Dichte

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

besitzt. Wir schreiben $X \sim N_{\mu,\sigma^2}$. Ist $\mu = 0$ und $\sigma^2 = 1$, so sagen wir, X ist *standardnormalverteilt*.

Bei der Normalverteilung ist es nicht einfach nachzurechnen, dass die ‘‘Dichte f ’’ wirklich eine Dichte ist. Das liegt daran, dass die Funktion $t \mapsto e^{-t^2}$ keine durch elementare Funktionen ausdrückbare Stammfunktion besitzt. Es gilt jedoch

Lemma 2.3.5. (*Gauß’sches Fehlerintegral*)

Es ist

$$\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}.$$

Substituiert man nun im Gauß’schen Fehlerintegral $t = \frac{s-\mu}{\sigma}$, so ist $\frac{dt}{ds} = \frac{1}{\sigma}$, also $dt = \frac{ds}{\sigma}$ und daher

$$\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{\infty} e^{-\frac{(s-\mu)^2}{2\sigma^2}} \frac{ds}{\sigma}.$$

Es folgt, dass

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds = 1,$$

also ist f tatsächlich eine Dichte.

Die Motivation für die Normalverteilung liegt im *Zentralen Grenzwertsatz* (den wir später besprechen), demzufolge viele Zufallsvariablen zumindest ‘‘annähernd’’ normalverteilt sind.

Die Verteilungsfunktion von Standardnormalverteilten Zufallsvariablen ist tabelliert.

2.4 Zufallsvektoren

Wir hatten bereits bemerkt, dass wir einen Vektor $X = (X_1, \dots, X_d)$ der aus Zufallsvariablen besteht als *Zufallsvektor* bezeichnen. Die *Verteilung* von X ist das Wahrscheinlichkeitsmaß μ_X auf $\mathcal{B}(\mathbb{R}^d)$, gegeben durch $\mu_X(B) = \mathbb{P}(X \in B)$.

Die *gemeinsame Verteilungsfunktion* des Zufallsvektor X ist die Funktion $F_X : \mathbb{R}^d \rightarrow \mathbb{R}$, gegeben durch

$$F_X(x_1, \dots, x_d) := \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Die gemeinsame Verteilungsfunktion hat ähnliche Eigenschaften wie ihr eindimensionales Pendant. Diese werden jedoch für uns nicht von Bedeutung sein, daher geben wir hier keine Details.

Für uns von besonderem Interesse sind wiederum die Fälle, dass X diskret ist oder dass X absolutstetig ist. Hierbei nennen wir X diskret, falls es eine abzählbare Menge M gibt mit $\mathbb{P}(X \in M) = 1$. Wir nennen X absolutstetig, falls es eine Funktion $f : \mathbb{R}^d \rightarrow [0, \infty)$ gibt mit

$$\int_{\mathbb{R}^d} f(t_1, \dots, t_d) dt_1 \dots dt_d = 1$$

derart, dass

$$\mu_X(B) = \int_B f(t_1, \dots, t_d) dt_1 \dots dt_d = 1$$

In diesem Fall nennen wir die Funktion f die *gemeinsame Dichte* des Zufallsvektor X .

Wir diskutieren einige Beispiele. Zunächst betrachten wir den diskreten Fall. Beachte, dass in diesem Fall auch die einzelnen Komponenten X_j diskret sind, d.h. $X_j(\Omega)$ ist eine diskrete Menge. Weiterhin ist die Verteilung bereits durch die Wahrscheinlichkeiten $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$, wobei $x_j \in X_j(\Omega)$ ist, vollständig beschrieben. Diese Wahrscheinlichkeiten nennt man auch *gemeinsame Zähldichte*.

Beispiel 2.4.1. Eine Urne enthält 10 rote, 10 blaue und 5 weiße Kugeln. Aus dieser Urne werden mit einem Griff zwei Kugeln gezogen. Die Zufallsvariable R gebe die Anzahl der gezogenen roten Kugeln an, die Zufallsvariable W die Anzahl der gezogenen weißen Kugeln.

Beachte, dass R und W Werte in $\{0, 1, 2\}$ annehmen. Um die Verteilung des Vektors (R, W) zu bestimmen müssen wir also für $i, j = 0, 1, 2$ die Wahrscheinlichkeiten $\mathbb{P}(R = i, W = j)$ bestimmen. Hierzu verwenden wir das hypergeometrische Modell. Beipielsweise gilt

$$\mathbb{P}(R = 2, W = 0) = \frac{\binom{10}{2} \binom{10}{0} \binom{5}{0}}{\binom{25}{2}} = \frac{3}{20}.$$

Auf ähnliche Weise erhält man die anderen Werte in folgender Tabelle.

$W \backslash R$	0	1	2	$W \downarrow$
0	$\frac{3}{20}$	$\frac{1}{3}$	$\frac{3}{20}$	$\frac{19}{30}$
1	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
2	$\frac{1}{30}$	0	0	$\frac{1}{30}$
\xrightarrow{R}	$\frac{21}{60}$	$\frac{1}{2}$	$\frac{3}{20}$	1

In dieser Tabelle geben die Spalten 0, 1, 2 den Wert für R an, die Zeilen 0, 1, 2 den Wert für W . Die letzte Spalte enthält die Zähldichte von W an. Beipielsweise ist $\mathbb{P}(W = 0) = 19/30$. Beachte, dass die Werte in dieser Spalte genau die Summe der Wahrscheinlichkeiten in der Zeile davor sind. Dies folgt aus der Disjunktheit der Zerlegung

$$\{R = j\} = \{R = j, W = 0\} \cup \{R = j, W = 1\} \cup \{R = j, W = 2\}$$

und der Additivität des Maßes. In der letzten Zeile findet sich die Wahrscheinlichkeitsverteilung von R . Man nennt die Verteilungen von W und von R auch die *Randverteilungen* des Vektors (W, R) .

Wir geben nun ein absolutstetiges Beispiel. Beachte, dass in diesem Fall die einzelnen Komponenten ebenfalls absolutstetig sind. Man erhält die Dichte einer Komponente durch "ausintegrieren der anderen Variablen". In der Tat gilt ja

$$\begin{aligned} \mathbb{P}(X_1 \leq x) &= \mathbb{P}((X_1, \dots, X_n) \in (-\infty, x] \times \mathbb{R} \times \dots \times \mathbb{R}) \\ &= \int_{-\infty}^x \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(t_1, \dots, t_d) dt_d \dots dt_1 = \int_{-\infty}^x f(t) dt \end{aligned}$$

für

$$f(t) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(t, t_2, \dots, t_d) dt_2 \dots dt_d.$$

Beispiel 2.4.2. Es sei (X, Y) ein Vektor mit Dichte $f(x, y) = (x + 2xy)\mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,1)}(y)$. Beachte, dass dies in der Tat eine Dichte ist, es ist nämlich

$$\begin{aligned}\int_{\mathbb{R}^2} f(x, y) dx dy &= \int_0^1 \int_0^1 x + 2xy dx dy = \int_0^1 \left[\frac{1}{2}x^2 + x^2y \right]_0^1 dy \\ &= \int_0^1 \frac{1}{2} + y dy = \left[\frac{1}{2}y + \frac{1}{2}y^2 \right]_0^1 = 1.\end{aligned}$$

Die Dichten der Zufallsvariablen X und Y erhält man wie folgt:

$$f_X(t) = \int_{\mathbb{R}} f(t, y) dy = \int_0^1 t + 2ty dy \mathbb{1}_{(0,1)}(t) = 2t \mathbb{1}_{(0,1)}(t)$$

und

$$f_Y(t) = \int_{\mathbb{R}} f(x, t) dx = \int_0^t x + 2xt dx \mathbb{1}_{(0,1)}(t) = \left(\frac{1}{2} + t \right) \mathbb{1}_{(0,1)}(t)$$

Mittels der gemeinsamen Dichte kann man die Wahrscheinlichkeit berechnen, dass $X \leq Y$ ist. Ist nämlich $A = \{(x, y) \in \mathbb{R}^2 : x \leq y\}$, so ist

$$\begin{aligned}\mathbb{P}(X \leq Y) &= \mathbb{P}((X, Y) \in A) = \int_A f(x, y) dx dy = \int_0^1 \int_x^1 x + 2xy dy dx \\ &= \int_0^1 \left[xy - xy^2 \right]_{y=x}^{y=1} dx = \int_0^1 2x - x^2 - x^3 dx \\ &= 1 - \frac{1}{3} - \frac{1}{4} = \frac{5}{12}.\end{aligned}$$

Beispiel 2.4.3. (Multivariate Normalverteilung)

Es sei $\mu = (\mu_1, \dots, \mu_d) \in \mathbb{R}^d$ und $C \in \mathbb{R}^{d \times d}$ symmetrisch und positiv definit, d.h. $x^T C x > 0$ für alle $x \neq 0$. Der Vektor X heißt *multivariat Normalverteilt* mit Parametern μ und C falls X die Dichte

$$f(t) := \frac{1}{\sqrt{(2\pi)^d \det C}} e^{-\frac{1}{2}(t-\mu)^T C^{-1}(t-\mu)}$$

besitzt.

Ist $d = 2$, so kann man jede positiv definite Matrix C schreiben als

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

wobei $\sigma_1, \sigma_2 > 0$ und $\rho \in (-1, 1)$ gilt. In diesem Fall ist $\det C = (1 - \rho^2)\sigma_1^2\sigma_2^2$ und

$$C^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

Daher hat die Dichte hat die Gestalt

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left(-\frac{1}{2(1-\rho^2)} \left(\frac{(t_1 - \mu_1)^2}{\sigma_1^2} - \frac{\rho(t_1 - \mu_1)(t_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(t_2 - \mu_2)^2}{\sigma_2^2} \right) \right).$$

Definition 2.4.4. Eine Familie von Zufallsvariablen $(X_i)_{i \in I}$ heißt *unabhängig*, falls für alle $n \in \mathbb{N}$, paarweise verschiedene Indices i_1, \dots, i_n und Borelmengen $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ die Ereignisse $\{X_{i_1} \in B_1\}, \dots, \{X_{i_n} \in B_n\}$ unabhängig sind, d.h.

$$\mathbb{P}((X_{i_1}, \dots, X_{i_n}) \in B_1 \times \dots \times B_n) = \prod_{k=1}^n \mathbb{P}(X_{i_k} \in B_k)$$

Um die Unabhängigkeit von endlich vielen Zufallsvariablen zu überprüfen haben wir folgendes Kriterium.

Proposition 2.4.5. *Es sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor mit Verteilungsfunktion F_X . Weiter sei F_{X_j} die Verteilungsfunktion von X_j . Folgende Aussagen sind äquivalent:*

(a) X_1, \dots, X_n sind unabhängig.

(b) Für $x_1, \dots, x_n \in \mathbb{R}$ ist $F_X(x_1, \dots, x_n) = \prod_{k=1}^n F_{X_k}(x_k)$.

Sind X_1, \dots, X_n absolutstetig mit dichten f_1, \dots, f_n so sind (a) und (b) äquivalent zu

(c) Die gemeinsame Dichte f erfüllt $f(t_1, \dots, t_n) = f_1(t_1) \cdots f_n(t_n)$ (fast überall).

Sind X_1, \dots, X_n diskret, so sind (a) und (b) äquivalent zu

(c') $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{k=1}^n \mathbb{P}(X_k = x_k)$ für alle $x_1, \dots, x_n \in \mathbb{R}$.

Bemerkung 2.4.6. Die Aussagen (c) und (c') sagen gerade aus, dass die gemeinsame (Zähl-)Dichte das Produkt der Rand(zähl)dichten ist.

Beispiel 2.4.7. Die Zufallsvariablen R und W in Beispiel 2.4.1 sind nicht unabhängig, denn die gemeinsame Zähldichte ist nicht das Produkt der Randzähldichten. Zum Beispiel ist

$$\mathbb{P}(R = 2, W = 2) = 0 \neq \frac{3}{20} \cdot \frac{1}{30} = \mathbb{P}(R = 2)\mathbb{P}(W = 2).$$

Die Zufallsvariablen X und Y in Beispiel 2.4.2 sind unabhängig, denn es gilt

$$f_X(x)f_Y(y) = 2x\mathbb{1}_{(0,1)}(x) \cdot \left(\frac{1}{2} + y\right)\mathbb{1}_{(0,1)}(y) = (x + 2xy)\mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,1)}(y) = f_{(X,Y)}(x, y).$$

2.5 Transformation von Zufallsvariablen

Es sei nun ein Zufallsvektor $X : \Omega \rightarrow \mathbb{R}^d$ und eine Borel-messbare Funktion $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ gegeben. Dabei heißt φ Borel-messbar, falls $\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R}^d)$ für alle $A \in \mathcal{B}(\mathbb{R}^m)$. Zum Beispiel ist jede stetige Funktion Borel-messbar. Wir wollen nun die Frage untersuchen, ob (und wie) man von Eigenschaften von X auf Eigenschaften von $\varphi(X)$ schliessen kann.

Wir beginnen mit der Unabhängigkeit.

Lemma 2.5.1. *Ist die Familie $(X_i)_{i \in I}$ unabhängig und ist $(\varphi_i)_{i \in I}$ eine Familie von Borel-messbaren Funktionen, so ist die Familie $(\varphi_i(X_i))_{i \in I}$ unabhängig.*

Beweis. Es sei $n \in \mathbb{N}$ und $i_1, \dots, i_n \in I$ paarweise verschieden. Zu vorgegebenen $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ sei $B_k := \varphi_{i_k}^{-1}(A_k)$. Wegen der Borel-Messbarkeit von φ_{i_k} ist $B_k \in \mathcal{B}(\mathbb{R})$. Weiter gilt

$$\mathbb{P}(\varphi_{i_1}(X_{i_1}) \in A_1, \dots, \varphi_{i_n}(X_{i_n}) \in A_n) = \mathbb{P}(X_{i_1} \in B_1, \dots, X_{i_n} \in B_n)$$

$$\begin{aligned}
&= \prod_{k=1}^n \mathbb{P}(X_{i_k} \in B_k) \\
&= \prod_{k=1}^n \mathbb{P}(\varphi_{i_k}(X_{i_k}) \in A_k). \quad \square
\end{aligned}$$

Unsere nächste Frage betrifft die Verteilung von $\varphi(X)$. Im eindimensionalen Fall haben wir folgende Aussagen.

Proposition 2.5.2. *Es sei X eine Zufallsvariable mit Verteilungsfunktion F_X .*

- (a) *Ist $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ streng monoton wachsend, so ist φ Borel-messbar. Weiter ist die Verteilungsfunktion von $\varphi(X)$ gegeben als $F_{\varphi(X)}(x) = F_X(\varphi^{-1}(x))$.*
- (b) *Ist $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ streng monoton fallend, so ist φ Borel-messbar. Weiter ist die Verteilungsfunktion von $\varphi(X)$ gegeben als $F_{\varphi(X)}(x) = 1 - F_X(\varphi^{-1}(x)) + \mathbb{P}(X = \varphi^{-1}(x))$.*
- (c) *Die Verteilungsfunktion von X^2 ist gegeben durch*

$$F_{X^2}(x) = \begin{cases} F_X(\sqrt{x}) - F_X(-\sqrt{x}) + \mathbb{P}(X = -\sqrt{x}), & x \geq 0 \\ 0, & \text{sonst.} \end{cases}$$

Beweis. (a) folgt sofort aus der Beobachtung, dass $\{\varphi(X) \leq x\} = \{X \leq \varphi^{-1}(x)\}$ ist.

(b) folgt aus

$$\{\varphi(X) \leq x\} = \{X \geq \varphi^{-1}(x)\} = \{X \leq \varphi^{-1}(x)\}^c \cup \{X = \varphi^{-1}(x)\}.$$

(c) Ist eine Konsequenz der Gleichheit

$$\begin{aligned}
\{X^2 \leq x\} &= \{|X| \leq \sqrt{x}\} = \{X \leq \sqrt{x}\} \setminus \{X < -\sqrt{x}\} \\
&= \{X \leq \sqrt{x}\} \setminus \{X \leq -\sqrt{x}\} \cup \{X = -\sqrt{x}\}. \quad \square
\end{aligned}$$

Korollar 2.5.3. *Ist X absolutstetig mit Dichte f und sind $a, b \in \mathbb{R}$ mit $a \neq 0$, so ist auch $aX + b$ absolutstetig mit Dichte*

$$f_{aX+b}(x) = \frac{1}{|a|} f\left(\frac{x-b}{a}\right)$$

Beweis. Aus Proposition 2.5.2 und der Tatsache, dass $\mathbb{P}(X = y) = 0$ für alle $y \in \mathbb{R}$ folgt

$$F_{aX+b}(x) = \begin{cases} F_X\left(\frac{x-b}{a}\right), & \text{falls } a > 0 \\ 1 - F_X\left(\frac{x-b}{a}\right) & \text{falls } a < 0. \end{cases}$$

Die Aussage über die Dichte folgt nun mit Substitution. □

Insbesondere gilt folgendes:

Ist $X \sim \mathcal{N}_{\mu, \sigma^2}$ so ist $aX + b \sim \mathcal{N}_{a\mu+b, b^2\sigma^2}$, denn $aX + b$ hat Dichte

$$\frac{1}{|a|} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\frac{t-b}{a} - \mu\right)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi a^2 \sigma^2}} \exp\left(-\frac{t - (a\mu + b)}{2b^2\sigma^2}\right).$$

Insbesondere ist $\frac{X-\mu}{\sigma} \sim \mathcal{N}_{0,1}$.

Wir geben einige weitere Beispiele.

Beispiel 2.5.4. Ist $X \sim N_{\mu, \sigma^2}$, so heißt die Verteilung von $Y := e^X$ *log-Normalverteilung* mit Parametern μ und σ^2 . Diese Verteilung ist absolutstetig mit Dichte

$$f(t) := \frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right) \mathbb{1}_{(0, \infty)}(t)$$

und findet in den Wirtschaftswissenschaften Anwendung.

Beispiel 2.5.5. Ist $X \sim N_{0,1}$ so heißt die Verteilung von $Y := X^2$ *χ^2 -Verteilung mit einem Freiheitsgrad*. Wir schreiben $Y \sim \chi_1^2$. Die Verteilung χ_1^2 ist absolutstetig mit Dichte

$$f(t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{t}{2}} \mathbb{1}_{(0, \infty)}(t).$$

Wegen Proposition 2.5.2 ist nämlich für $x > 0$

$$\begin{aligned} \mathbb{P}(Y \leq x) &= \int_{-\infty}^{\sqrt{x}} e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} - \int_{-\infty}^{-\sqrt{x}} e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} \\ &= 2 \int_0^{\sqrt{x}} e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} \\ &= 2 \int_0^x e^{-\frac{s}{2}} \frac{1}{2\sqrt{s}} \frac{ds}{\sqrt{2\pi}} = \int_0^x e^{-\frac{s}{2}} \frac{ds}{\sqrt{2\pi s}}. \end{aligned}$$

Als nächstes betrachten wir Summen von unabhängigen Zufallsvariablen.

Proposition 2.5.6. *Es seien X, Y unabhängige Zufallsvariablen.*

(a) *Sind X und Y diskret mit Werten in \mathbb{Z} , so ist*

$$\mathbb{P}(X + Y = n) = \sum_{k \in \mathbb{Z}} \mathbb{P}(X = k) \mathbb{P}(Y = n - k).$$

(b) *Sind X und Y absolutstetig mit Dichten f_X bzw. f_Y , so ist auch $X + Y$ absolutstetig mit Dichte*

$$f_{X+Y}(t) = \int_{\mathbb{R}} f_X(r) f_Y(t - r) dr$$

Beweis. (a) Es ist $\{X + Y = n\} = \bigcup_{k \in \mathbb{Z}} \{X = k, Y = n - k\}$. Wegen der σ -Additivität ist also

$$\mathbb{P}(X + Y = n) = \sum_{k \in \mathbb{Z}} \mathbb{P}(X = k, Y = n - k) = \sum_{k \in \mathbb{Z}} \mathbb{P}(X = k) \mathbb{P}(Y = n - k),$$

wobei wir im zweiten Schritt die Unabhängigkeit verwendet haben.

(b) Weil X und Y unabhängig sind, gilt $f_{(X,Y)}(t, s) = f_X(t) f_Y(s)$, siehe Proposition 2.4.5. Wir substituieren $t + s = u$ und $t = v$

$$\begin{aligned} \mathbb{P}(X + Y \leq x) &= \iint_{\{t+s \leq x\}} f_{(X,Y)}(t, s) ds dt \\ &= \iint_{\{t+s \leq x\}} f_X(t) f_Y(s) ds dt \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f_X(v) f_Y(u - v) dv du \end{aligned} \quad \square$$

Wir diskutieren einige Beispiele:

Beispiel 2.5.7. Es seien X, Y unabhängige Zufallsvariablen mit $X \sim \text{Pois}_{\lambda_1}$ und $Y \sim \text{Pois}_{\lambda_2}$. Dann ist $X + Y \sim \text{Pois}_{\lambda_1 + \lambda_2}$. In der Tat folgt aus Proposition 2.5.6, dass

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k) = \sum_{k=0}^n e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda_1^k \lambda_2^{n-k} = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}. \end{aligned}$$

Nun kommen wir zur Summe unabhängiger, normalverteilter Zufallsvariablen.

Proposition 2.5.8. *Es seien X, Y unabhängig mit $X \sim \mathbf{N}_{\mu_1, \sigma_1^2}$ und $Y \sim \mathbf{N}_{\mu_2, \sigma_2^2}$. Dann ist $X + Y \sim \mathbf{N}_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2}$. Insbesondere ist die Summe unabhängiger, normalverteilter Zufallsvariablen normalverteilt.*

Beweis. Wir nehmen zunächst an, dass $\mu_1 = \mu_2 = 0$ ist. Dann hat X Dichte $f(t) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp(-\frac{t^2}{2\sigma_1^2})$ und Y hat Dichte $g(t) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp(-\frac{t^2}{2\sigma_2^2})$. Die Faltung ist gegeben durch

$$(g * f)(t) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{(t-s)^2}{\sigma_1^2} + \frac{s^2}{\sigma_2^2}\right)} ds$$

Wir setzen $\sigma^2 := \sigma_1^2 + \sigma_2^2$ und substituieren s durch

$$\frac{\sigma_1\sigma_2}{\sigma} z + \frac{\sigma_2^2}{\sigma^2} t.$$

Es ist also $ds = \frac{\sigma_1\sigma_2}{\sigma} dz$. Weiter ist

$$\begin{aligned} \left(\frac{(t-s)^2}{\sigma_1^2} + \frac{s^2}{\sigma_2^2}\right) \Big|_{s=\frac{\sigma_1\sigma_2}{\sigma}z + \frac{\sigma_2^2}{\sigma^2}t} &= \frac{1}{\sigma_1^2} \left(\frac{\sigma_1^2}{\sigma^2}t - \frac{\sigma_1\sigma_2}{\sigma}z\right)^2 + \frac{1}{\sigma_2^2} \left(\frac{\sigma_2^2}{\sigma^2}t + \frac{\sigma_1\sigma_2}{\sigma}z\right)^2 \\ &= \frac{1}{\sigma_1^2} \left(\frac{\sigma_1^4}{\sigma^4}t^2 - 2\frac{\sigma_1^2\sigma_1\sigma_2}{\sigma^3}tz + \frac{\sigma_1^2\sigma_2^2}{\sigma^2}z^2\right) \\ &\quad + \frac{1}{\sigma_2^2} \left(\frac{\sigma_2^4}{\sigma^4}t^2 + 2\frac{\sigma_2^2\sigma_1\sigma_2}{\sigma^3}tz + \frac{\sigma_1^2\sigma_2^2}{\sigma^2}z^2\right) \\ &= \frac{\sigma_1^2 + \sigma_2^2}{\sigma^4}t^2 + \frac{\sigma_1^2 + \sigma_2^2}{\sigma^2}z^2 \\ &= \frac{1}{\sigma^2}t^2 + z^2. \end{aligned}$$

Somit ergibt sich

$$(g * f)(t) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{1}{\sigma^2}t^2 + z^2\right)} \frac{\sigma_1\sigma_2}{\sigma} dz = \frac{1}{2\pi\sigma} e^{-\frac{t^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}.$$

Dies zeigt die Behauptung in diesem Fall. Im allgemeinen Fall setzen wir $\tilde{X} = X - \mu_1$ und $\tilde{Y} = Y - \mu_2$. Dann nach obigem ist $\tilde{X} + \tilde{Y} \sim \mathbf{N}_{0, \sigma_1^2 + \sigma_2^2}$. Daher ist $X + Y = \tilde{X} + \tilde{Y} + \mu_1 + \mu_2 \sim \mathbf{N}_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2}$. \square

Kapitel 3

Momente von Zufallsvariablen

Wichtige Charakteristiken von Zufallsvariablen sind ihre *Momente* insbesondere Erwartungswert und Varianz. Die formale Definition dieser Größen beruht auf der allgemeinen Lebesgue'schen Integrationstheorie und würde den Rahmen dieser Vorlesung sprengen. Wir diskutieren daher einige einführende Beispiele und motivieren damit Regeln zur Berechnung des Erwartungswerts für diskrete und für absolutstetige Zufallsvariablen. Diese dienen dann als "Definition" für uns.

3.1 Der Erwartungswert

Wir betrachten zunächst sogenannte *einfache Zufallsvariablen*, d.h. Zufallsvariablen die nur endlich viele Werte annehmen. Jede einfache Zufallsvariable X läßt sich in der Form

$$X = \sum_{k=1}^n x_k \mathbb{1}_{A_k}$$

darstellen, wobei $\{x_1, \dots, x_n\}$ der Wertebereich von X ist und $A_k := X^{-1}(\{x_k\})$ die Menge ist, auf der X den Wert x_k annimmt. Beachte, dass A_k messbar ist, weil X eine Zufallsvariable ist. Für solche Zufallsvariablen definieren wir den *Erwartungswert* als gewichtetes Mittel der Werte:

$$\mathbb{E}(X) := \sum_{k=1}^n x_k \mathbb{P}(A_k) = \sum_{k=1}^n x_k \mathbb{P}(X = x_k).$$

Beispiel 3.1.1. Es wird folgendes Spiel angeboten:

Der Spieler zahlt einen Einsatz von $E \in \mathbb{E}$ (welcher zu bestimmen ist). Anschliessend zieht er eine Karte aus einem Kartenspiel (französisches Blatt, d.h. 32 Karten). Zieht er ein Ass, so werden ihm $10 \in \mathbb{E}$ ausgezahlt. Zieht er eine Bildkarte (Bube, Dame, König) so erhält er $2 \in \mathbb{E}$ ausgezahlt. Ansonsten erhält er nichts. Für welchen Wert von E ist diese Spiel *fair*, d.h. der erwartete Gewinn ist $0 \in \mathbb{E}$.

Es bezeichne X den Gewinn des Spielers. Dann nimmt X die Werte $10 - E$ ("Ass"), $2 - E$ ("Bildkarte") und $-E$ ("sonstige Karte") mit Wahrscheinlichkeiten $4/32$, $12/32$ resp. $16/32$ an. Es ist also

$$\mathbb{E}X = (10 - E) \frac{4}{32} + (2 - E) \frac{12}{32} - E \frac{16}{32} = 2 - E.$$

Demnach ist das Spiel genau dann fair, wenn der Einsatz $2 \in \mathbb{E}$ beträgt.

Die grundlegende Idee bei der allgemeinen Definition des Erwartungswerts ist es nun, eine allgemeine Zufallsvariable X durch eine Folge X_n von einfachen Zufallsvariablen zu approximieren. Sodann möchte man $\mathbb{E}X$ als Grenzwert $\lim \mathbb{E}X_n$ definieren. Hierbei muss man sicherstellen, dass der Grenzwert nicht von der Folge X_n abhängt.

Wir verdeutlichen dies an einem Beispiel:

Beispiel 3.1.2. Die *alternierende harmonische Reihe*

$$\sum_{k=1}^{\infty} (-1)^k \frac{1}{k}$$

konvergiert (und zwar gegen $\log 2$). Allerdings konvergiert sie nicht *absolut*, genauer gilt folgendes:

Ist $c \in \mathbb{R}$ vorgegeben, so gibt es eine Umordnung $(a_n)_{n \in \mathbb{N}}$ der Folge $(-1)^n \frac{1}{n}$, d.h. es gibt eine bijektive Abbildung $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ derart, dass $a_n = (-1)^{\sigma(n)} \frac{1}{\sigma(n)}$, derart, dass

$$\sum_{k=1}^n a_k \rightarrow c \quad \text{für } n \rightarrow \infty.$$

Nun betrachten wir folgende Situation. Wir betrachten einen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ in dem es eine Folge $(A_k) \subset \Sigma$ von paarweise disjunkten Mengen gibt mit $\mathbb{P}(A_k) = 2^{-k}$. Es folgt, dass $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \sum_{k=1}^{\infty} 2^{-k} = \frac{1}{1-\frac{1}{2}} - 1 = 1$. Nun betrachte die Zufallsvariable (!) X , definiert durch

$$X(\omega) := (-1)^k \frac{2^k}{k} \quad \text{für } \omega \in A_k.$$

Wir können X durch die Folge $X_n := \sum_{k=1}^n (-1)^k \frac{2^k}{k} \mathbb{1}_{A_k}$ approximieren. Somit würde folgen, dass

$$\mathbb{E}X_n = \sum_{k=1}^n (-1)^k \frac{2^k}{k} 2^{-k} = \sum_{k=1}^n (-1)^k \rightarrow \log 2,$$

wir würden also $\mathbb{E}X = \log 2$ definieren. Allerdings approximiert für eine Umordnung $a_n = (-1)^{\sigma(n)} \frac{1}{\sigma(n)}$ auch die Folge $\tilde{X}_n := \sum_{k=1}^n (-1)^{\sigma(k)} \frac{2^{\sigma(k)}}{\sigma(k)}$ die Zufallsvariable X . Es ist aber

$$\mathbb{E}\tilde{X}_n = \sum_{k=1}^n (-1)^k \frac{2^{\sigma(k)}}{\sigma(k)} 2^{\sigma(k)} = \sum_{k=1}^n a_k$$

und durch Wahl einer geeigneten Umordnung kann hier *jede reelle Zahl* als Grenzwert auftreten.

Dies zeigt, dass man in diesem Beispiel *keinen sinnvollen Mittelwert definieren kann*.

Beachte, dass bei einer *absolut konvergenten Reihe*, also einer konvergenten Reihe $\sum_{k=1}^{\infty} a_k$ für die auch $\sum_{k=1}^{\infty} |a_k| < \infty$ ist, auch jede Umordnung der Reihe gegen den gleichen Grenzwert konvergiert. Wir treffen nun folgende Definition:

Definition 3.1.3. Es sei X eine diskrete Zufallsvariable mit Wertebereich $\{x_k : k \in \mathbb{N}\}$. Wir sagen X hat *endlichen Erwartungswert* (oder *der Erwartungswert von X existiert*, oder X ist *integrierbar*), falls $\sum_{k=1}^{\infty} |x_k| \mathbb{P}(X = x_k) < \infty$ ist. In diesem Fall definieren wir den Erwartungswert $\mathbb{E}X$ als

$$\mathbb{E}X := \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k).$$

Beispiel 3.1.4. Ist $X \sim \text{Pois}_\lambda$ so ist $X(\Omega) = \mathbb{N}_0$. Weiter gilt

$$\sum_{k=0}^{\infty} |k| \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda e^\lambda = \lambda.$$

Somit existiert der Erwartungswert von X und es gilt $\mathbb{E}X = \lambda$.

Sei nun X eine absolutstetige Zufallsvariable mit Dichte f . Wir nehmen zunächst vereinfachend an, dass X nur Werte zwischen $-T$ und T annimmt, wobei $T > 0$ eine Konstante ist. Wir approximieren X wie folgt. Wir unterteilen $[-T, T]$ in Intervalle $I_k := (-T + k \frac{2T}{n}, -T + (k+1) \frac{2T}{n}]$, $k = 0, \dots, n-1$. Sodann setzen wir $A_k := X^{-1}(I_k)$. Auf A_k nimmt also X einen Wert zwischen $-T + k \frac{2T}{n}$ und $-T + (k+1) \frac{2T}{n}$ an. Jeder dieser Werte ist also ein guter Näherungswert für X . Wählen wir etwa den kleinsten Wert $\xi_k^n := -T + k \frac{2T}{n}$ als Näherungswert, so können wir X durch die Folge

$$X_n := \sum_{k=0}^{n-1} \xi_k^n \mathbb{1}_{A_k}$$

von einfachen Zufallsvariablen approximieren. Beachten wir weiter, dass

$$\mathbb{P}(A_k) = \mathbb{P}(\xi_k^n < X \leq \xi_{k+1}^n) = \int_{\xi_k^n}^{\xi_{k+1}^n} f(t) dt \approx f(\xi_k^n)(\xi_{k+1}^n - \xi_k^n),$$

so gilt

$$\mathbb{E}X_n = \sum_{k=0}^{n-1} \xi_k^n \mathbb{P}(A_k) \approx \sum_{k=0}^{n-1} \xi_k^n f(\xi_k^n)(\xi_{k+1}^n - \xi_k^n) \rightarrow \int_{-T}^T t f(t) dt.$$

Dies motiviert folgende Definition im absolutstetigen Fall:

Definition 3.1.5. Es sei X eine absolutstetige Zufallsvariable mit Dichte f . Wir sagen, X hat *endlichen Erwartungswert* (oder *der Erwartungswert von X existiert* oder X ist *integrierbar*), falls $\int_{\mathbb{R}} |t| f(t) dt < \infty$. In diesem Fall definieren wir den Erwartungswert von X durch

$$\mathbb{E}X := \int_{\mathbb{R}} t f(t) dt.$$

Beispiel 3.1.6. Ist $X \sim \text{exp}_\lambda$, so ist X absolutstetig mit Dichte $f_\lambda(t) := \lambda e^{-\lambda t} \mathbb{1}_{(0, \infty)}(t)$. In diesem Fall gilt

$$\int_{\mathbb{R}} |t| f_\lambda(t) dt = \int_{\mathbb{R}} t f_\lambda(t) dt = \int_0^\infty \lambda t e^{-\lambda t} dt = [-t e^{-\lambda t}]_0^\infty + \int_0^\infty e^{-\lambda t} dt = 0 - \left[\frac{1}{\lambda} e^{-\lambda t} \right]_0^\infty = \frac{1}{\lambda}.$$

Es folgt, dass der Erwartungswert von X existiert und $\mathbb{E}X = \frac{1}{\lambda}$.

Beispiel 3.1.7. Ist X absolutstetig mit Dichte $f(t) = \frac{1}{t^2} \mathbb{1}_{(1, \infty)}(t)$ (siehe Beispiel 2.3.3), so existiert der Erwartungswert von X nicht. Es gilt nämlich

$$\int_{\mathbb{R}} |t| f(t) dt = \int_1^\infty t \frac{1}{t^2} dt = [\log t]_1^\infty = \infty.$$

Wir stellen nun einige Eigenschaften des Erwartungswerts zusammen.

Proposition 3.1.8. Sind X, Y Zufallsvariablen mit endlichem Erwartungswert und $\alpha, \beta \in \mathbb{R}$, so gilt

- (a) (Linearität) Auch $\alpha X + \beta Y$ hat endlichen Erwartungswert und es gilt $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$.
- (b) (Monotonie) Ist $X \leq Y$, so ist $\mathbb{E}X \leq \mathbb{E}Y$. Insbesondere ist $|\mathbb{E}X| \leq \mathbb{E}|X|$.
- (c) Ist Z eine Zufallsvariable mit $|Z| \leq X$, so existiert der Erwartungswert von Z .
- (d) Sind X und Y unabhängig, so hat XY endlichen Erwartungswert und $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Beweis. Da wir den Erwartungswert nicht allgemein definiert haben, können wir auch den Beweis nicht formal führen. Alle Beweise folgen dem Schema “Das ist wahr für einfache Zufallsvariablen und überträgt sich auf Grenzwerte”. Wir zeigen hier lediglich, dass Aussage (d) für einfache Zufallsvariablen gilt.

Ist nämlich $X = \sum_{k=1}^n x_k \mathbb{1}_{A_k}$ und $Y = \sum_{l=1}^m y_l \mathbb{1}_{B_l}$, so ist

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E} \sum_{k=1}^n \sum_{l=1}^m x_k y_l \mathbb{1}_{A_k \cap B_l} = \sum_{k=1}^n \sum_{l=1}^m \mathbb{E} x_k y_l \mathbb{1}_{A_k \cap B_l} \\ &= \sum_{k=1}^n \sum_{l=1}^m x_k y_l \mathbb{P}(A_k \cap B_l) = \sum_{k=1}^n \sum_{l=1}^m x_k y_l \mathbb{P}(A_k) \mathbb{P}(B_l) \\ &= \left(\sum_{k=1}^n x_k \mathbb{P}(A_k) \right) \left(\sum_{l=1}^m y_l \mathbb{P}(B_l) \right) = \mathbb{E}(X)\mathbb{E}(Y). \quad \square \end{aligned}$$

Beispiel 3.1.9. Ist $X \sim N_{0,1}$, so hat X Dichte $\varphi(t) := (\sqrt{2\pi})^{-1} e^{-\frac{t^2}{2}}$. Es ist

$$\int_0^\infty t\varphi(t) dt = \int_0^\infty t e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} = \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \right]_0^\infty = \frac{1}{\sqrt{2\pi}} < \infty.$$

Da $\varphi(-x) = \varphi(x)$ folgt mit der Substitution $t = -s$, dass

$$\int_{-\infty}^0 t\varphi(t) dt = - \int_0^\infty s\varphi(s) ds.$$

Es folgt, dass

$$\int_{\mathbb{R}} |t|\varphi(t) dt = \int_{-\infty}^0 (-t)\varphi(t) dt + \int_0^\infty t\varphi(t) dt = 2 \int_0^\infty t\varphi(t) dt = \frac{2}{\sqrt{2\pi}} < \infty.$$

Somit existiert der Erwartungswert von X . Weiter ist $\mathbb{E}X = 0$, was sofort aus obiger Beobachtung folgt.

Sei nun $X \sim N_{\mu, \sigma^2}$. Dann ist $Y := \frac{X-\mu}{\sigma} \sim N_{0,1}$ und umgekehrt, $X = \sigma Y + \mu$. Wegen der Linearität des Erwartungswerts existiert der Erwartungswert von X und

$$\mathbb{E}X = \sigma \mathbb{E}(Y) + \mathbb{E}(\mu) = \mu,$$

da die Konstante $\mu = \mu \mathbb{1}_\Omega$ Erwartungswert μ hat.

3.2 Varianz und höhere Momente

Wir hatten bereits bemerkt, dass für eine Zufallsvariable X und eine Borel-messbare Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ auch $\varphi(X)$ eine Zufallsvariable ist. Für den Erwartungswert einer derart transformierten Zufallsvariable gilt folgender *Transformationssatz*.

Satz 3.2.1. *Es sei X eine Zufallsvariable und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ Borel-messbar.*

(a) *Ist X diskret mit Werten (x_k) , so existiert $\mathbb{E}\varphi(X)$ genau dann, wenn*

$$\sum_k |\varphi(x_k)| \mathbb{P}(X = x_k) < \infty.$$

In diesem Fall ist

$$\mathbb{E}\varphi(X) = \sum_k \varphi(x_k) \mathbb{P}(X = x_k).$$

(b) *Ist X absolutstetig mit Dichte f , so existiert $\mathbb{E}\varphi(X)$ genau dann, wenn*

$$\int_{\mathbb{R}} |\varphi(t)| f(t) dt < \infty.$$

In diesem Fall ist

$$\mathbb{E}\varphi(x) = \int_{\mathbb{R}} \varphi(t) f(t) dt.$$

Definition 3.2.2. Es sei X eine Zufallsvariable und $k \in \mathbb{N}$. Wir sagen X besitzt *endliche Momente k -ter Ordnung*, falls X^k endlichen Erwartungswert hat. In diesem Fall heißt $\mathbb{E}X^k$ das *k -te Moment* von X und $\mathbb{E}(X - \mathbb{E}(X))^k$ das *k -te zentrierte Moment* von X .

Das zweite zentrierte Moment von X heißt *Varianz von X* und wird mit $\text{Var}(X)$ bezeichnet. Die Wurzel aus $\text{Var}(X)$ heißt *Standardabweichung von X* und wird mit $\sigma(X)$ bezeichnet. In

Bemerkung 3.2.3. Besitzt X endliche Momente k -ter Ordnung, so besitzt X auch endliche Momente l -ter Ordnung für $l \leq k$. Dies folgt aus der Monotonie des Erwartungswertes und der Tatsache dass $|t|^l \leq c(1 + |t|^k)$ für eine geeignete Konstante c gilt. Insbesondere existiert der Erwartungswert von X sofern X endliche Momente k -ter Ordnung für ein $k \geq 1$ besitzt.

Hat X^k endlichen Erwartungswert, so auch $(X - \mathbb{E}(X))^k$, denn wegen des Binomischen Satzes ist

$$(X - \mathbb{E}(X))^k = \sum_{l=0}^k \binom{k}{l} X^l (\mathbb{E}X)^{k-l},$$

also $(X - \mathbb{E}(X))^k$ eine Linearkombination von Momenten niedrigerer Ordnung.

Die Varianz einer Zufallsvariablen ist die mittlere quadratische Abweichung von ihrem Erwartungswert. Sie ist somit ein Maß für die *Streuung* einer Zufallsvariable um ihren Erwartungswert. Beachte, dass falls bei der Interpretation der Zufallsvariable X eine Einheit hinzugefügt werden muss (etwa *Meter* m oder *Stunden* h), so trägt die Varianz von X diese Einheit *zum Quadrat*. Die Standardabweichung jedoch die *gleiche* Einheit.

Lemma 3.2.4. *Es sei X eine Zufallsvariable mit endlichen Momenten zweiter Ordnung. Dann gilt*

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

Beweis. Wir schreiben $\mu := \mathbb{E}X$. Dann ist $(X - \mu)^2 = X^2 - 2\mu X + \mu^2$. Wegen der Linearität des Erwartungswerts gilt

$$\text{Var}(X) = \mathbb{E}(X - \mu)^2 = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \quad \square$$

Wir geben einige Beispiele.

Beispiel 3.2.5. Sei $X \sim \text{Pois}_\lambda$. Wir berechnen zunächst $\mathbb{E}(X(X - 1))$. Beachte, dass $X(X - 1) \geq 0$ ist. Somit

$$\mathbb{E}(X(X - 1)) = e^\lambda \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} = e^\lambda \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.$$

Es folgt, dass $\mathbb{E}(X^2) = \mathbb{E}(X(X - 1)) + \mathbb{E}(X) = \lambda^2 + \lambda$ ist und somit $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \lambda$.

Beispiel 3.2.6. Ist $X \sim N_{0,1}$ so folgt aus partieller Integration, dass

$$\sqrt{2\pi}\mathbb{E}(X^2) = \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = \left[-te^{-\frac{t^2}{2}} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}.$$

Wegen $\mathbb{E}X = 0$ folgt $\text{Var}(X) = \mathbb{E}(X^2) = 1$.

Ist $X \sim N_{\mu,\sigma^2}$, so ist $X = \sigma Y + \mu$ für eine $N_{0,1}$ -verteilte Zufallsvariable Y . Somit ist

$$\text{Var}(X) = \mathbb{E}(X - \mu)^2 = \mathbb{E}(\sigma Y)^2 = \sigma^2 \mathbb{E}(Y) = \sigma^2.$$

3.3 Kovarianz und Korrelation

Wir beginnen mit folgender Beobachtung:

Haben X und Y endliche Momente zweiter Ordnung, so existiert der Erwartungswert von XY . Es ist nämlich

$$|XY| \leq \frac{1}{2}(X^2 + Y^2),$$

was aus den Binomischen Formeln $(a \pm b)^2 = a^2 + \pm 2ab + b^2$ folgt. Somit können wir definieren:

Definition 3.3.1. Sind X und Y Zufallsvariablen mit endlichen Momenten zweiter Ordnung so heißt

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

die *Kovarianz* von X und Y . Die Zahl

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

heißt *Korrelationskoeffizient* von X und Y . Die Zufallsvariablen X und Y heißen *unkorreliert*, falls $\text{Cov}(X, Y) = 0$ gilt.

Zur Berechnung von Kovarianz und Korrelation kann man folgenden mehrdimensionalen Transformationsatz verwenden.

Satz 3.3.2. Es sei $X = (X_1, \dots, X_d)$ ein Zufallsvektor und $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ Borel-messbar.

(a) Ist X diskret, so existiert $\mathbb{E}\varphi(X)$ genau dann, wenn

$$\sum |\varphi(x_1, \dots, x_d)| \mathbb{P}(X_1 = x_1, \dots, X_d = x_d) < \infty.$$

In diesem Fall ist

$$\mathbb{E}\varphi(X) = \sum \varphi(x_1, \dots, x_d) \mathbb{P}(X_1 = x_1, \dots, X_d = x_d).$$

In beiden Fällen wird über alle möglichen Werte summiert.

(b) Ist X absolutstetig mit Dichte f , so existiert $\mathbb{E}\varphi(X)$ genau dann, wenn

$$\int_{\mathbb{R}^d} |\varphi(t_1, \dots, t_d)| f(t_1, \dots, t_d) dt_1 \dots, dt_d < \infty.$$

In diesem Fall ist

$$\mathbb{E}\varphi(x) = \int_{\mathbb{R}^d} \varphi(t_1, \dots, t_d) f(t_1, \dots, t_d) dt_1 \dots, dt_d$$

Beispiel 3.3.3. Es sei $c \in [0, \frac{1}{2}]$. Die Verteilung des Vektors (X, Y) sei gegeben durch

	Y		
	0	1	
X			X ↓
0	c	$\frac{1}{2} - c$	$\frac{1}{2}$
1	$\frac{1}{2} - c$	c	$\frac{1}{2}$
Y ↘	$\frac{1}{2}$	$\frac{1}{2}$	1

In diesem Beispiel hängen die Randverteilungen nicht von c ab: X und Y nehmen die Werte 0 und 1 jeweils mit Wahrscheinlichkeit $1/2$ an. Es ist $\mathbb{E}X = \mathbb{E}(Y) = \frac{1}{2}$ und $\text{Var}X = \text{Var}Y = \frac{1}{4}$. Weiter ist

$$\begin{aligned} \text{Cov}(X, Y) &= \left(-\frac{1}{2}\right)\left(-\frac{1}{2}\right) \cdot c + \left(-\frac{1}{2}\right)\frac{1}{2}\left(\frac{1}{2} - c\right) + \frac{1}{2}\left(-\frac{1}{2}\right)\left(\frac{1}{2} - c\right) + \frac{1}{2} \cdot \frac{1}{2}c \\ &= c - \frac{1}{4} \end{aligned}$$

Somit sind X und Y unkorreliert genau dann, wenn $c = 1/4$ ist. Anhand dieses Beispiels kann man eine Interpretation der Korrelation geben. Ist die Korrelation positiv ($c > 1/4$) so ist es wahrscheinlicher, dass X und Y in die gleiche Richtung von ihrem jeweiligen Erwartungswert abweichen. In diesem Falle sind die Ergebnisse $(X, Y) = (1, 1)$ (positive Abweichung vom Erwartungswert $(\frac{1}{2}, \frac{1}{2})$) und $(X, Y) = (0, 0)$ (negative Abweichung vom Erwartungswert) wahrscheinlicher als die Ausgänge $(1, 0)$ und $(0, 1)$ (unterschiedliche Abweichungen vom Erwartungswert). Ist die Korrelation negativ (also $c < 1/4$) so ist es genau andersherum.

Diese Interpretation gilt auch allgemein.

Satz 3.3.4. Es seien X und Y Zufallsvariablen mit endlichen zweiten Momenten.

(a) $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

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

(c) Sind X und Y unabhängig, so sind X und Y unkorreliert, also $\text{Cov}(X, Y) = 0$.

(d) Sind $\sigma(X), \sigma(Y) > 0$, so ist $\rho(X, Y) \in [-1, 1]$.

Beweis. Sei $\mathbb{E}(X) = \mu$ und $\mathbb{E}(Y) = \nu$. (a) Es gilt

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}(X - \mu)(Y - \nu) = \mathbb{E}[XY - \mu Y - \nu X + \mu\nu] \\ &= \mathbb{E}(XY) - \mu\mathbb{E}(Y) - \nu\mathbb{E}(X) + \mu\nu = \mathbb{E}(XY) - \mu\nu.\end{aligned}$$

(b) Sei $\tilde{X} = X - \mu$ und $\tilde{Y} = Y - \nu$. Dann ist

$$\begin{aligned}\text{Var}(X \pm Y) &= \mathbb{E}(\tilde{X} \pm \tilde{Y})^2 = \mathbb{E}(\tilde{X})^2 \pm 2\mathbb{E}(\tilde{X}\tilde{Y}) + \mathbb{E}(\tilde{Y})^2 \\ &= \text{Var}(X) \pm 2\text{Cov}(X, Y) + \text{Var}(Y).\end{aligned}$$

(c) Sind X und Y unabhängig, so ist $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) = \mu\nu$ und somit $\text{Cov}(X, Y) = 0$ wegen (a).

(d) Sei nun $\tilde{X} = \frac{X-\mu}{\sigma(X)}$ und $\tilde{Y} = \frac{Y-\nu}{\sigma(Y)}$. Dann ist $\text{Var}(\tilde{X}) = \text{Var}(\tilde{Y}) = 1$. Wegen (b) gilt

$$0 \leq \text{Var}(\tilde{X} \pm \tilde{Y}) = 1 \pm 2\text{Cov}(\tilde{X}, \tilde{Y}) + 1 = 2 \pm 2 \frac{\mathbb{E}(X - \mu)(Y - \nu)}{\sigma(X)\sigma(Y)} = 2 \pm 2\rho(X, Y).$$

Dies ist äquivalent zur Behauptung. □

Kapitel 4

Grenzwertsätze

4.1 Das Gesetz der großen Zahlen

Bei der Definition der Wahrscheinlichkeiten haben wir uns von Eigenschaften der *relativen Häufigkeiten* inspirieren lassen. Es stellt sich die Frage, ob (und wie) die relativen Häufigkeiten unsere abstrakt definierten Wahrscheinlichkeiten approximieren. Wir wenden uns nun dieser Frage zu.

Dazu betrachten wir ein Zufallsexperiment welches durch den Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ beschrieben werde. Weiter sei A ein Ereignis mit Wahrscheinlichkeit $\mathbb{P}(A) =: p$. Wir wiederholen nun dieses Zufallsexperiment n -mal unabhängig und zählen, wie oft das Ereignis A eintritt. Die Anzahl S_n der Versuche, in denen A eintritt ist Binomialverteilt mit Parametern n und p . Wir können S_n auch schreiben als $S_n = \sum_{k=1}^n X_k$ wobei X_k Bernoulliverteilt ist mit Erfolgswahrscheinlichkeit p und $X_k = 1$, falls A im k -ten Versuch eintritt. Die *relative Häufigkeit* von A ist also $\frac{1}{n}S_n = \frac{1}{n} \sum_{k=1}^n X_k$. Und die Frage die wir uns stellen ist, ob und in welchem Sinn $\frac{1}{n} \sum_{k=1}^n X_k \rightarrow p$.

Definition 4.1.1. Es seien X, X_1, X_2, X_3, \dots Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$.

- (i) Wir sagen X_n konvergiert *in Wahrscheinlichkeit* gegen X , falls $\mathbb{P}(|X_n - X| \geq \varepsilon) \rightarrow 0$ für alle $\varepsilon > 0$.
- (ii) Wir sagen X_n konvergiert *fast sicher* gegen X , falls $\mathbb{P}(X_n \rightarrow X) = 1$ ist, d.h. die Menge $\{\omega : \lim X_n(\omega) = X(\omega)\}$ hat Wahrscheinlichkeit 1.

Beispiel 4.1.2. Wir betrachten den Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$. Für $n = 2^k + l$, wobei $k \in \mathbb{N}$ und $l \in \{0, 1, 2, \dots, 2^k - 1\}$, sei

$$X_n := \mathbb{1}_{\left[\frac{l}{2^k}, \frac{l+1}{2^k}\right)}.$$

Dann ist $\mathbb{P}(|X_{2^k+l} - 0| \geq \varepsilon) = \mathbb{P}\left(\left[\frac{l}{2^k}, \frac{l+1}{2^k}\right)\right) = 2^k$ und es folgt, dass $X_n \rightarrow 0$ in Wahrscheinlichkeit. Andererseits konvergiert $X_n(\omega)$ für *kein* $\omega \in [0, 1)$, sodass X_n nicht fast sicher konvergiert.

Wir können nun das schwache Gesetz der großen Zahlen in seiner einfachsten Form formulieren.

Satz 4.1.3. (*Schwaches Gesetz der großen Zahlen*)

Es sei X_n eine Folge unabhängiger Zufallsvariablen mit gleichem Erwartungswert μ und gleicher Varianz σ^2 . Dann konvergiert $\frac{1}{n} \sum_{k=1}^n X_k$ in Wahrscheinlichkeit gegen μ .

Bevor wir uns dem Beweis zuwenden, notieren wir folgendes Korollar.

Korollar 4.1.4. *Ein Zufallsexperiment werde n mal unabhängig wiederholt. Dann konvergieren für $n \rightarrow \infty$ die relativen Häufigkeiten eines Ereignisses A gegen seine Wahrscheinlichkeit.*

Beweis. Es genügt das schwache Gesetz der großen Zahlen auf eine Folge unabhängiger Bernoulliverteilter Zufallsvariablen mit Erfolgswahrscheinlichkeit $p = \mathbb{P}(A)$ anzuwenden. Dann sind die Voraussetzungen von Satz 4.1.3 erfüllt mit $\mu = p$ und $\sigma^2 = p(1-p)$. \square

Im Beweise von Satz 4.1.3 verwenden wir folgende Ungleichung, die auf Tschebyscheff zurückgeht.

Lemma 4.1.5. *Es sei X eine Zufallsvariable mit Erwartungswert μ und Varianz σ^2 . Dann gilt für $\varepsilon > 0$ stets*

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

Beweis. Es ist $Y := \varepsilon^2 \mathbb{1}_{\{|X - \mu| \geq \varepsilon\}} \leq |X - \mu|^2$ und daher, wegen der Monotonie des Erwartungswerts,

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

Das ist äquivalent zur Behauptung. \square

Beweis von Satz 4.1.3. Es sei $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$. Wegen der Linearität des Erwartungswertes ist $\mathbb{E}\bar{X} = \frac{1}{n} \sum_{k=1}^n \mathbb{E}X_k = \mu$. Wegen der Unabhängigkeit ist

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{\sigma^2}{n}.$$

Aus der Tschebyscheff Ungleichung folgt für $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| \geq \varepsilon\right) = \mathbb{P}(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\text{Var}\bar{X}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$$

für $n \rightarrow \infty$. \square

Beispiel 4.1.6. (Monte-Carlo Simulation zur numerischen Integration)

Es sei $g : [0, 1]^d \rightarrow \mathbb{R}$ eine stetige Funktion. Wir möchten

$$\int_{[0,1]^d} g(x) dx$$

berechnen. Hierzu gehen wir wie folgt vor. Wir erzeugen (mit einem geeigneten Zufallszahlengenerator – eine Geschichte für sich–) eine Folge X_1, X_2, X_3, \dots von unabhängigen, $U_{[0,1]^d}$ -verteilten Zufallsvektoren. Wegen Lemma 2.5.1 sind auch die Zufallsvariablen $Y_k := g(X_k)$ unabhängig. Weiter ist aufgrund der Transformationsformel 3.3.2

$$\mathbb{E}Y_k = \int_{[0,1]^d} g(x) dx$$

Weiter ist $\text{Var}(Y_k) \equiv \sigma^2$ für ein geeignetes σ^2 . Wir können also das schwache Gesetz der großen Zahlen auf die Folge Y_k anwenden und erhalten, dass $\frac{1}{n} \sum_{k=1}^n Y_k \rightarrow \int_{[0,1]^d} g(x) dx$ in Wahrscheinlichkeit. In der Praxis verwendet man daher (für genügend großes n) $\frac{1}{n} \sum_{k=1}^n g(X_k)$ als "Schätzwert" für das gesuchte Integral.

Die oben formulierte Version des schwachen Gesetz der großen Zahlen ist eine relative einfache (wie man auch an dem relativ einfachen Beweis sieht). Man kann zeigen, dass auf die Existenz der zweiten Momente verzichtet werden kann. Bei einem *starken Gesetz der großen Zahlen* wird statt Konvergenz in Wahrscheinlichkeit fast sichere Konvergenz gezeigt. Wir geben ohne Beweis folgendes starke Gesetz der großen Zahlen an.

Satz 4.1.7. (*Starkes Gesetz der großen Zahlen*)

Es sei X_n eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit endlichem Erwartungswert $\mu := \mathbb{E}X_1 \equiv \mathbb{E}X_n$. Dann konvergiert $\frac{1}{n} \sum_{k=1}^n X_k$ fast sicher gegen μ .

4.2 Der zentrale Grenzwertsatz

Nun kommen wir zu einem zentralen Resultat, welches die Bedeutung der Normalverteilung erklärt.

Theorem 4.2.1. (*Zentraler Grenzwertsatz*)

Es sei X_1, X_2, \dots eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit endlichen Momenten zweiter Ordnung. Wir setzen $0 < \sigma^2 := \text{Var}X_1$ und $\mu := \mathbb{E}X_1$. Weiter sei $S_n := \sum_{k=1}^n X_k$. Dann ist

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt =: \Phi(x).$$

Interpretation: Ist $M_n := \frac{1}{n}S_n$ das Mittel der ersten n Zufallsvariablen, so ist $\mathbb{E}M_n = \mu$ und $\text{Var}M_n = \frac{\sigma^2}{n}$. Folglich hat $\frac{\sqrt{n}}{\sigma}(M_n - \mu) = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ Varianz 1 und Erwartungswert 0.

Der Zentrale Grenzwertsatz besagt gerade, dass die Verteilungsfunktion dieser standardisierten Mittel gegen die Verteilungsfunktion der Standardnormalverteilung konvergiert.

Bemerkung 4.2.2. Beachte, dass $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$ ist. Es folgt aus dem Zentralen Grenzwertsatz, dass

$$\mathbb{P}\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \rightarrow \int_a^b e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}}.$$

Man darf hier in der Wahrscheinlichkeit links statt $<$ auch \leq schreiben.

Dank des zentralen Grenzwertsatzes können wir (unter Kenntnis der Werte von Φ , die tabelliert sind) Wahrscheinlichkeiten approximieren. Wir geben hierzu ein Beispiel:

Beispiel 4.2.3. Ein Würfel werde 600 mal geworfen. Was ist die Wahrscheinlichkeit zwischen 90 und 100 Sechsen zu werfen.

Hierbei handelt es sich um das 600-fache (unabhängige) Wiederholen eines Bernoulli Experiments mit Erfolgswahrscheinlichkeit $p = \frac{1}{6}$. Damit ist die Anzahl der Erfolge $\mathbf{b}_{600, \frac{1}{6}}$ -verteilt und die gesuchte Wahrscheinlichkeit ist

$$\sum_{k=90}^{100} \binom{600}{k} \frac{1}{6}^k \frac{5}{6}^{600-k}.$$

Allerdings ist diese Summe relativ schwierig zu berechnen. Wir verwenden den zentralen Grenzwertsatz, um die Wahrscheinlichkeit zu approximieren.

Seien hierzu X_1, \dots, X_{600} unabhängige, Bernoulli verteilte Zufallsvariablen mit Erfolgswahrscheinlichkeit $\frac{1}{6}$. Es ist also $\mu := \mathbb{E}X_1 = p = \frac{1}{6}$ und $\sigma^2 = p(1-p) = \frac{5}{36}$. Es ist also $n\mu = 100$ und $\sqrt{n}\sigma = \sqrt{500/6} \approx 9,13$.

Die Anzahl der Erfolge bei 600 Versuchen ist $S_{600} = \sum_{k=1}^{600} X_k$. Nach dem zentralen Grenzwertsatz gilt

$$\begin{aligned} \mathbb{P}(90 \leq S_{600} \leq 100) &= \mathbb{P}\left(\frac{90 - 100}{9,13} \leq \frac{S_{600} - 600\mu}{\sqrt{600}\sigma} \leq \frac{100 - 100}{9,13}\right) \\ &\approx \Phi(0) - \Phi(-1,095) = 0,5 - (1 - \Phi(1,095)) \approx 0,36, \end{aligned}$$

also ungefähr 36 Prozent.

Bemerkung 4.2.4. Der zentrale Grenzwertsatz macht keine Angaben darüber, wann die Normalverteilung eine gute Annäherung an die Verteilung eines standardisierten Mittels ist bzw. darüber, wie gut diese Annäherung ist. Bei der Approximation der Binomialverteilung $b_{n,p}$ hat sich als Faustregel etabliert, dass die Approximation gut ist, falls $np(1-p) \geq 9$.

Bemerkung 4.2.5. Bei der Approximation der Binomialverteilung durch die Normalverteilung ist folgendes zu beachten:

Die Binomialverteilung ist eine diskrete Verteilung, genauer nimmt sie nur natürliche Zahlen als Werte an. Demgegenüber ist die Normalverteilung absolutstetig, die Wahrscheinlichkeit eine bestimmte Zahl anzunehmen (insbesondere also eine gegebene natürliche Zahl) ist 0. Will man nun mittels des zentralen Grenzwertsatzes die Wahrscheinlichkeit $\mathbb{P}(a \leq S_n \leq b)$, dass eine Zufallsvariable $S_n \sim b_{n,p}$ zwischen den natürlichen Zahlen a und b liegt approximieren, so nimmt man häufig folgende *Stetigkeitskorrektur* vor, um die Wahrscheinlichkeit, dass $S_n = a$ oder $S_n = b$ ist besser zu approximieren:

$$\begin{aligned} \mathbb{P}(a \leq S_n \leq b) &= \mathbb{P}\left(a - \frac{1}{2} \leq S_n \leq b + \frac{1}{2}\right) = \mathbb{P}\left(\frac{a - \frac{1}{2} - np}{\sigma\sqrt{n}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{b + \frac{1}{2} - np}{\sigma\sqrt{n}}\right) \\ &\approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sigma\sqrt{n}}\right) \end{aligned}$$

Verwenden wir die Stetigkeitskorrektur in Beispiel 4.2.3, so erhalten wir

$$\begin{aligned} \mathbb{P}(90 \leq S_n \leq 100) &\approx \Phi\left(\frac{100 + \frac{1}{2} - 100}{9,13}\right) - \Phi\left(\frac{90 - \frac{1}{2} - 100}{9,13}\right) \\ &\approx \Phi(0,055) - \Phi(-1,15) \approx 0,397. \end{aligned}$$

Vergleichen wir diesen Wert mit dem exakten Wert 0,4025 (der in Tafeln enthalten ist), so sehen wir, dass die Stetigkeitskorrektur einen Wert liefert, der näher am exakten Wert liegt. Zudem zeigt dieses Beispiel, dass die Stetigkeitskorrektur selbst bei $n = 600$ noch einen merkbaren Unterschied macht.

Wir geben nun noch ein etwas anderes Anwendungsbeispiel.

Beispiel 4.2.6. Der Airbus A380 hat gewöhnlich 526 Sitzplätze. Aus Erfahrungen ist bekannt, dass ein verkauftes Ticket mit einer Wahrscheinlichkeit von 0,1 storniert wird. Wie viele Tickets kann man für einen Flug verkaufen, wenn dieser mit einer Wahrscheinlichkeit von höchstens 2% überbucht sein soll?

Es sei X_1, X_2, \dots eine Folge unabhängiger $b_{1,0.9}$ -verteilter Zufallsvariablen und $S_n = \sum_{k=1}^n X_k$. Wir suchen eine Zahl n , sodass $\mathbb{P}(S_n \geq 526) = 0,02$. Es sei hierzu $x_n := \frac{526 - n \cdot 0.9}{\sqrt{n \cdot 0.1 \cdot 0.9}}$. Dann ist

$$0.02 \stackrel{!}{=} \mathbb{P}(S_n \geq 526) = \mathbb{P}\left(\frac{S_n - n \cdot 0.9}{\sqrt{n \cdot 0.1 \cdot 0.9}} \geq x_n\right) \approx 1 - \Phi(x_n)$$

Wir suchen also ein x_n mit $\Phi(x_n) = 0.98$. Dies kann über eine Tabelle der Werte von Φ^{-1} , der Quantilfunktion der Normalverteilung, geschehen. Man schlägt nach, dass $\Phi^{-1}(0,98) \approx 2,05$. Man will also $x_n = 2,05$ wählen. Wir schreiben $m = \sqrt{n}$ und erhalten die Gleichung

$$\frac{526 - m^2 \cdot 0,9}{m \sqrt{0,09}} = 2,05 \Leftrightarrow 526 - m^2 \cdot 0,9 = m \cdot 0,3 \cdot 2,05 \Leftrightarrow m^2 + 0.683 \cdot m = 584,4$$

Löst man diese quadratische Gleichung, so erhält man $n = 552$.