

Stochastik für Wirtschaftswissenschaftler

Vorlesungsskript
Wintersemester 2015/16

Jürgen Kampf

basierend auf einem Skript von Markus Kunze

Kapitel 1

Diskrete Wahrscheinlichkeitsräume

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 (im Rahmen des Modells) identischen 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. Die Elemente von Ω bezeichnet man auch als *Elementarereignisse*.

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

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

(c) Die Temperatur an der Ulmer Adenauerbrücke am Schwörmontag um 14.00 Uhr wird ermittelt. Hier wählen wir $\Omega = \mathbb{R}$ wobei $t \in \Omega$ das Elementarereignis “Die Temperatur beträgt t °C” bezeichnet. Der Wahrscheinlichkeitsraum ist sicher etwas zu groß gewählt, denn $t = 100$ kann nicht eintreten. Aber weil die maximale denkbare Temperatur nicht vernünftig ermittelt werden kann, müssen wir das in Kauf nehmen.

(d) 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) das Elementarereignis “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 Elementarereignis “Scheibe verfehlt” bezeichnet.

(e) 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 $\Omega = \mathbb{N} \cup \{\infty\}$ an. Hierbei bezeichnet $n \in \mathbb{N}$ das Elementarereignis “im n -ten Wurf fällt zum ersten Mal eine Sechs” und ∞ das Elementarereignis “es fällt überhaupt keine Sechs”.

Fazit:

1. Bei der Wahl der Grundmenge Ω gibt es kein “richtig” oder “falsch”. Ziel ist, einen einfachen, aber dennoch adequaten Raum zu wählen.
2. Man wählt die Grundmenge Ω lieber zu groß als zu klein.

Häufig ist man bei einem Zufallsexperiment nicht an dem tatsächlichen Ausgang interessiert sondern nur daran, ob das Ergebnis zu einer vorgegebenen Menge von Ergebnissen gehört. Im Beispiel 1.1.1(d) 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. Solche Mengen von Ergebnissen nennt man *Ereignisse*. 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*.

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 den einelementigen Ereignissen $\{K\}$ und $\{Z\}$ nur noch das unmögliche Ereignis \emptyset und das sichere Ereignis Ω .

(c) $A = (25, \infty)$ ist das Ereignis “Die Temperatur beträgt über 25 °C”.

(d) 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!”.

(e) Beim Mensch-Ärgere-Dich-Nicht ist das Ereignis $A = \{1, 2, 3\}$ (“Innerhalb der ersten drei Würfe fällt eine Sechs”) von Interesse.

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*. Ereignisse A_1, \dots, A_n heißen *paarweise unvereinbar*, falls je zwei dieser Ereignisse unvereinbar sind.

Die Menge aller Ereignisse ist der sog. *Ereignisraum* Σ . Oftmals ist $\Sigma = \mathcal{P}(\Omega)$ (die Potenzmenge $\mathcal{P}(\Omega)$ von Ω ist die Menge aller Teilmengen von Ω); die Ausnahmen hiervon sind nicht Gegenstand dieser Vorlesung.

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

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

wobei ein K bzw. Z an Position j anzeigt, dass beim j -ten Wurf Kopf bzw. Zahl fällt.

Es sei A_j das Ereignis “Im j -ten Wurf fällt Kopf”. Dann ist 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 $\{KKK\}$ (“Es fällt dreimal Kopf”) 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 relative Häufigkeiten widerspiegeln: Dass ein Ereignis A Wahrscheinlichkeit p hat soll bedeuten, dass man bei "genügend häufiger Wiederholung des Experiments in $p \cdot 100$ Prozent der Fälle das Ereignis A eintritt". Die Wahrscheinlichkeit soll also gegen die relative Häufigkeit konvergieren.

Leider ist dies keine einwandfreie Definition. Wir führen Wahrscheinlichkeiten daher *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.

Leider gibt es (aus mathematischen Gründen) einige Subtilitäten bei der Definition von Wahrscheinlichkeiten auf unendlichen (insbesondere auf überabzählbaren) Grundmengen. Daher beschränken wir uns zunächst auf den Fall *endlicher* Grundmengen.

Definition 1.1.4. Ein *endlicher Wahrscheinlichkeitsraum* ist ein Paar (Ω, \mathbb{P}) , bestehend aus einer endlichen Menge Ω und einer Abbildung $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$, sodass

- (i) $\mathbb{P}(\Omega) = 1$ (Normiertheit)
- (ii) Sind $A, B \in \mathcal{P}(\Omega)$ unvereinbar, so ist $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (Additivität).

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

Wir notieren einige einfache Schlussfolgerungen.

Proposition 1.1.5. *Es sei (Ω, \mathbb{P}) ein endlicher Wahrscheinlichkeitsraum, A, B Ereignisse.*

- (1) *Ist $A \subseteq B$, so ist $\mathbb{P}(A) \leq \mathbb{P}(B)$. Genauer gilt $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. Insbesondere ist $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ und $\mathbb{P}(\emptyset) = 0$.*
- (2) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Seien nun A_1, \dots, A_n Ereignisse.

- (3) *Es ist*

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

- (4) *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) Ist $A \subseteq B$, 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, indem man $B = \Omega$ setzt, weil $\mathbb{P}(\Omega) = 1$ ist.

- (4) folgt induktiv aus Eigenschaft (ii) eines Wahrscheinlichkeitsmaßes.

(2) Sei $C := A \cap B$. Dann ist $C \subseteq A$ und $C \subseteq B$ und $A \cup B = (A \setminus C) \cup (B \setminus C) \cup C$. Diese Ereignisse sind paarweise unvereinbar. Also gilt nach Teilen (4) und (1)

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

- (3) folgt induktiv aus (2). □

Es folgt aus der endlichen Additivität eines Wahrscheinlichkeitsmaßes, Teil (4) von Proposition 1.1.5, dass ein Wahrscheinlichkeitsmaß bereits durch die Werte auf den Elementarereignissen eindeutig bestimmt ist. In der Tat, ist $\Omega = \{\omega_1, \dots, \omega_n\}$ und ist $p_j := \mathbb{P}(\{\omega_j\})$, so ist

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{\omega_j \in A} \{\omega_j\}\right) = \sum_{\omega_j \in A} \mathbb{P}(\{\omega_j\}) = \sum_{\omega_j \in A} p_j.$$

Beachte, dass notwendigerweise $p_1 + \dots + p_n = 1$ sein muss. Umgekehrt liefert aber auch jede Wahl von p_1, \dots, p_n mit $p_1, \dots, p_n \geq 0$ und $\sum_{i=1}^n p_i = 1$ ein Wahrscheinlichkeitsmaß. In der Tat, definiere

$$\mathbb{P}(A) := \sum_{\omega_j \in A} p_j$$

für jede Menge $A \subseteq \Omega$. Für jedes $A \subseteq \Omega$ gilt nun

$$\mathbb{P}(A) = \sum_{\omega_j \in A} p_j \geq 0 \quad \text{und} \quad \mathbb{P}(A) = \sum_{\omega_j \in A} p_j \leq \sum_{j=1}^n p_j = 1.$$

Also ist \mathbb{P} in der Tat eine Abbildung $\mathcal{P}(\Omega) \rightarrow [0, 1]$. Bedingung (i) aus der Definition des Wahrscheinlichkeitsmaßes ist erfüllt, da

$$\mathbb{P}(\Omega) = \sum_{j=1}^n p_j = 1.$$

Um Bedingung (ii) nachzuprüfen, seien $A, B \subseteq \Omega$ zwei unvereinbare Ereignisse. Dann gilt

$$\mathbb{P}(A \cup B) = \sum_{\omega_j \in A \cup B} p_j = \sum_{\omega_j \in A} p_j + \sum_{\omega_j \in B} p_j = \mathbb{P}(A) + \mathbb{P}(B).$$

Also ist \mathbb{P} ein Wahrscheinlichkeitsmaß.

Beispiel 1.1.6. Ein Würfel wird geworfen. Wir nehmen $\Omega = \{1, 2, 3, 4, 5, 6\}$. Es erscheint plausibel, dass alle Augenzahlen die gleiche Wahrscheinlichkeit haben. Wir können also

$$p_1 = \dots = p_6 = \frac{1}{6}$$

wählen. Die Wahrscheinlichkeit, eine gerade Zahl zu würfeln, ist nun $\mathbb{P}(\{2, 4, 6\}) = p_2 + p_4 + p_6 = 1/2$.

Beachte, dass unsere Definition nicht verlangt, dass alle Elementarereignisse gleich wahrscheinlich sind. In der Tat könnte für einen gezinkten Würfel $p_6 = 0,25$ während $p_1, \dots, p_5 = 0,15$ ist. Für einen solchen Würfel wäre $\mathbb{P}(\{2, 4, 6\}) = 0,15 + 0,15 + 0,25 = 0,55$, es wäre also wahrscheinlicher eine gerade Zahl zu würfeln als eine ungerade Zahl.

Beispiel 1.1.7. Die Ecken eines Würfels werden gleichmäßig abgeschliffen, sodass der Würfel auch auf jeder der acht Ecken liegen bleiben kann. Diese werden von 1 bis 8 durchnummeriert. 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 und e_j das Elementarereignis 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, folgt

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

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: $\mathbb{P}(E) = 8 \cdot p/4 = 1/4$.

1.2 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 1.2.1. (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 ein Laplace Experiment ein adäquates Modell für dieses Experiment darstellt. Dies muss man mittels statistischer Methoden entscheiden.

Es hängt von der Wahl des Grundraums ab, ob ein Experiment als Laplace Experiment betrachtet werden kann:

Beispiel 1.2.2. a) Das Werfen eines Würfels kann auch als Experiment mit Grundraum $\Omega = \{1, \dots, 100\}$ und $p_1 = \dots = p_6 = \frac{1}{6}$ und $p_7 = \dots = p_{100} = 0$ beschrieben werden. Dieses Experiment ist nicht Laplacesch.

b) Ein weniger triviales Beispiel ergibt sich wie folgt:

Zwei (nicht unterscheidbare) Münzen werden gleichzeitig geworfen und das Ergebnis festgestellt. Es können zweimal Kopf (KK), zweimal Zahl (ZZ) oder einmal Kopf und einmal Zahl (KZ) auftreten. Wir können demnach einen dreielementigen Grundraum $\Omega = \{KK, KZ, ZZ\}$ wählen. Allerdings beschreibt ein Laplacescher Wahrscheinlichkeitsraum mit diesem Grundraum das Experiment nicht zutreffend.

Folgende Überlegung zeigt, dass das Elementarereignis KZ die doppelte Wahrscheinlichkeit im Vergleich zu KK (und auch im Vergleich zu ZZ) haben sollte.

Werfen wir nämlich zwei unterscheidbare Münzen, so können wir als Grundraum $\tilde{\Omega} = \{kk, kz, zk, zz\}$ wählen, wobei der erste Buchstabe das Ergebnis von Münze 1 und der zweite Buchstabe das Ergebnis von Münze 2 wiedergibt. Nun ist die Annahme, dass alle Elementarereignisse gleich wahrscheinlich sind plausibel. Weil wir die Münzen nicht unterscheiden können, sind die Elementarereignisse kz und zk zu einem einzigen Elementarereignis KZ verschmolzen, was aber dennoch Wahrscheinlichkeit $1/2$ (und *nicht* Wahrscheinlichkeit $1/3$) haben sollte.

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:

Satz 1.2.3. *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, anschließend in jedem Fall 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 1.2.4. Thorsten hat ein Bewerbungsgespräch und stellt dazu ein passendes Outfit zusammen. Er hat 3 Hemden, 4 Krawatten und 2 Anzüge. Er kann damit (von modischen Überlegungen abgesehen) $3 \cdot 4 \cdot 2 = 24$ verschiedene Outfits zusammenstellen.

Beispiel 1.2.5. 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 1.2.5 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 1.2.5 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. Wir erinnern uns, dass die Fakultät einer Zahl $m \in \mathbb{N}$ definiert ist durch $m! := m \cdot (m - 1) \cdot \dots \cdot 2 \cdot 1$. Es gilt $n \cdot \dots \cdot (n - (k - 1)) = \frac{n!}{(n-k)!}$.

Ist $n = k$, so ist das Ergebnis des Ziehens ohne Zurücklegen gerade eine gewisse Reihenfolge der Zahlen 1 bis n . Dafür gibt es $n!$ Möglichkeiten.

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 $\frac{11!}{6!} = 55.440$ Möglichkeiten.

Ziehen ohne Zurücklegen ohne Beachtung der Reihenfolge:

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

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

Möglichkeiten aus n Zahlen k ohne Zurücklegen und ohne Beachtung der Reihenfolge zu ziehen. Die Zahl $\binom{n}{k}$ heißt *Binomialkoeffizient* "n über k".

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:

Wahrscheinlichkeitsräume, die diesem Modell folgen, sind in der Regel nicht Laplacesch. Da die Diskussion dieses Modells darüberhinaus sehr aufwendig wäre, übergehen wir es. In der Praxis empfiehlt sich meist, die Reihenfolge zu beachten, auch wenn sie unwichtig wäre.

Wir verwenden Urnenmodelle nun in einigen konkreten Laplace Experimenten um Wahrscheinlichkeiten auszurechnen.

Beispiel 1.2.6. 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 , das 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 $(365)!/(365 - n)!$ Elemente.

Folglich ist

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

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 1.2.7. 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”, also dass genau 4 Zahlen des Tipps tatsächlich gezogen werden. Die Menge A besteht aus denjenigen 6-elementigen Teilmengen der Zahlen von 1 bis 49, die genau 4 der 6 angekreuzten Zahlen (und 2 der 43 nicht angekreuzten Zahlen) erhalten. Es gibt $\binom{6}{4}$ Möglichkeiten, 4 der angekreuzten Zahlen auszuwählen, und zu jeder dieser Möglichkeiten gibt es $\binom{43}{2}$ Möglichkeiten, diese mit nicht angekreuzten Zahlen 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 unterschiedlich markiert sind.

Allgemeiner betrachtet man eine Urne mit n Kugeln, die eines 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 $n_1 + n_2 + \dots + n_r = n$ gelten. 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. Wir nehmen an, dass Kugeln auch dann unterscheidbar sind, wenn sie die gleiche Farbe haben. Es gibt

$$\binom{n_1}{k_1} \binom{n_2}{k_2} \cdots \binom{n_r}{k_r}$$

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

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

Beispiel 1.2.8. Wie hoch ist die Wahrscheinlichkeit, beim Poker ein “Full House” zu ziehen?

Beim Poker gibt es 52 Karten, jeder von 13 Werten kommt 4-mal vor. In einem ersten Schritt (den wir hier ausschließlich betrachten) zieht man zufällig 5 Karten. Von einem Full House spricht man, wenn 3 Karten einen Wert und 2 Karten den selben anderen Wert haben.

Insgesamt gibt es $\binom{52}{5} = 2.598.960$ verschiedene Pokerhände (= Möglichkeiten, die 5 Karten zu ziehen, wobei die Reihenfolge keine Rolle spielt, aber verschiedene Karten des selben Werts unterschieden werden können). 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 $(i, j) \neq (k, l)$)

$$\mathbb{P}(A) = \sum_{i, j \in K, i \neq j} \mathbb{P}(A_{ij}) = \frac{13!}{11!} \cdot \frac{\binom{4}{3} \binom{4}{2}}{\binom{52}{5}} \approx 0,0014,$$

denn es gibt $\frac{13!}{11!}$ Möglichkeiten zwei Elemente aus K ohne Zurückzulegen zu ziehen.

1.3 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 Einschätzung von Wahrscheinlichkeiten berücksichtigen. Beispielsweise wird er, wenn er selbst 2 Assen 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 Assen”) 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 1.3.1. Es sei (Ω, \mathbb{P}) ein endlicher Wahrscheinlichkeitsraum und B ein Ereignis mit positiver Wahrscheinlichkeit. Die *bedingte Wahrscheinlichkeit* $\mathbb{P}(A|B)$ von A gegeben B ist definiert als

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

Eine bedingte Wahrscheinlichkeit ist in der Tat eine Wahrscheinlichkeit:

Lemma 1.3.2. Sei (Ω, \mathbb{P}) ein endlicher Wahrscheinlichkeitsraum und $B \subseteq \Omega$ mit $\mathbb{P}(B) > 0$. Dann ist die Abbildung $A \mapsto \mathbb{P}(A|B)$ ein Wahrscheinlichkeitsmaß auf Ω und auch ein Wahrscheinlichkeitsmaß auf B .

Beweis. Weil $\emptyset \subseteq A \cap B \subseteq B$ ist, folgt aus Proposition 1.1.5 (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 $A \subseteq \Omega$ ist.

Weil $\Omega \cap B = B$ gilt trivialerweise $\mathbb{P}(\Omega|B) = 1$. Auch $\mathbb{P}(B|B) = 1$ ist offensichtlich.

Sind schließlich A_1, A_2 unvereinbar, so sind auch $A_1 \cap B$ und $A_2 \cap B$ unvereinbar. Weil aber $(A_1 \cup A_2) \cap B = (A_1 \cap B) \cup (A_2 \cap B)$ ist, folgt aus der Additivität von \mathbb{P} , dass $\mathbb{P}((A_1 \cup A_2) \cap B) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B)$. Also

$$\mathbb{P}(A_1 \cup A_2|B) = \frac{\mathbb{P}((A_1 \cup A_2) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(A_2 \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B).$$

□

Insbesondere gelten also die Aussagen von Proposition 1.1.5 für bedingte Wahrscheinlichkeiten.

Beispiel 1.3.3. Beim Skat werden 32 Karten wie folgt aufgeteilt: Jeder der drei Spieler erhält 10 Karten und die restlichen beiden Karten werden verdeckt auf den Tisch gelegt; man sagt sie "kommen in den Skat". Berücksichtigt man die Reihenfolge der Spieler (Geben-Sagen-Hören), so gibt es insgesamt $\binom{32}{10} \binom{22}{10} \binom{12}{10}$ verschiedene Skat Spiele. (Der Geber erhält 10 von den 32 Karten, der Sager erhält 10 von den verbliebenen 22 Karten, der Hörer erhält 10 von den verbliebenen 12 Karten; offensichtlich ist die Reihenfolge, in der die Karten verteilt werden, unerheblich und wir können die Reihenfolge so wählen, dass die Rechnungen möglichst einfach wird). Unter den 32 Karten befinden sich 4 Assen.

Sei A das Ereignis, dass der Sager mindestens ein Ass erhält, und B das Ereignis, dass der Geber zwei Assen hat. Wir wollen die Wahrscheinlichkeit $\mathbb{P}(A)$, dass der Sager mindestens ein Ass erhält und die bedingte Wahrscheinlichkeit $\mathbb{P}(A|B)$, dass der Sager mindestens ein Ass erhält *gegeben, die Information, dass der Geber (genau) zwei Assen hat*, berechnen. Auch hier ist es leichter, das komplementäre Ereignis A^c , dass der Sager kein Ass erhält, zu betrachten.

Das Ereignis A^c tritt in $\binom{28}{10} \binom{22}{10} \binom{12}{10}$ Fällen ein. Dies berechnet sich wie folgt: Zunächst bekommt der Sager 10 von den 28 Karten, die kein Ass sind. Dann bekommt der Geber

10 Karten, die aus den übrigen 18 Karten plus den 4 Assen ausgewählt wurden. Zuletzt bekommt der Hörer 10 von den dann verbliebenen 12 Karten. Daher

$$\mathbb{P}(A^c) = \frac{\binom{28}{10} \binom{22}{10} \binom{12}{10}}{\binom{32}{10} \binom{22}{10} \binom{12}{10}} = \frac{\binom{28}{10}}{\binom{32}{10}} \approx 0,2034,$$

also $\mathbb{P}(A) = 0,7966$. Der Sager erhält also in etwa 80 % der Fälle mindestens ein Ass.

Das Ereignis B , der Geber hat zwei Assen, tritt in $\binom{4}{2} \binom{28}{8} \binom{22}{10} \binom{12}{10}$ Fällen auf (zunächst geben wir dem Geber 2 der 4 Assen und 8 der 28 nicht-Asse, dann dem Sager 10 der übrigen 22 Karten und dem Hörer 10 der verbliebenen 12). Das Ereignis $B \cap A^c$, dass der Geber 2 Assen, der Sager aber kein Ass erhält, tritt in $\binom{4}{2} \binom{28}{8} \binom{20}{10} \binom{12}{10}$ Fällen auf (der Geber bekommt 2 Assen und 8 nicht-Asse, der Sager bekommt 10 von den verbliebenen 20 nicht-Asse, der Hörer 10 von den verbliebenen 10 nicht-Asse plus 2 Assen). Es ist also

$$\mathbb{P}(A^c|B) = \frac{\binom{4}{2} \binom{28}{8} \binom{20}{10} \binom{12}{10}}{\binom{4}{2} \binom{28}{8} \binom{22}{10} \binom{12}{10}} = \frac{\binom{20}{10}}{\binom{22}{10}} \approx 0,2857.$$

Also ist $\mathbb{P}(A|B) \approx 0,7143$. Wenn der Geber also weiß, dass er zwei Assen hat, wird er die Wahrscheinlichkeit, dass der Sager mindestens ein Ass hat, nur noch auf 71% einschätzen.

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

Proposition 1.3.4. (Multiplikationsregel) *Es seien A_1, \dots, A_n Ereignisse in einem endlichen Wahrscheinlichkeitsraum derart, dass $\mathbb{P}(A_1 \cap \dots \cap A_n) > 0$. 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 1.3.5. 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 dreimal ausgeführt. Wie groß ist die Wahrscheinlichkeit, dass dreimal eine rote Kugel gezogen wird?

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 bedingte Wahrscheinlichkeit $\mathbb{P}(A_3 | A_1 \cap A_2)$, eine rote zu ziehen, $2/3$ ist. Wurde im ersten Versuch eine rote Kugel gezogen, so befinden sich beim zweiten Versuch 15 rote und 10 blaue Kugeln in der Urne, sodass $\mathbb{P}(A_2 | A_1) = 3/5$ 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 $\Omega_1, \dots, \Omega_n$ von Teilmengen von Ω derart, dass

- (i) die Mengen 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 $\bigcup_{j=1}^n \Omega_j = \Omega$.

Die Menge Ω wird also in die Mengen $\Omega_1, \dots, \Omega_n$ “zerlegt”; jedes Element von Ω ist genau in einer der Mengen Ω_j enthalten.

Beispiel 1.3.6. Eine häufig verwendete Partition ist die folgende: Es sei $B \subseteq \Omega$ eine Teilmenge. Dann definiert $\Omega_1 := B$ und $\Omega_2 := B^c$ eine Partition von Ω . Es sind nämlich Ω_1 und Ω_2 disjunkt und es ist $\Omega_1 \cup \Omega_2 = \Omega$.

Satz 1.3.7. *Es sei (Ω, \mathbb{P}) ein endlicher Wahrscheinlichkeitsraum und $\Omega_1, \dots, \Omega_n$ eine Partition von Ω .*

(1) **(Satz von der totalen Wahrscheinlichkeit)** Für $A \subseteq \Omega$ gilt

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

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

(2) **(Formel von Bayes)** Für $A \subseteq \Omega$ mit $\mathbb{P}(A) > 0$ gilt

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

Beweis. Aus der Definition der bedingten Wahrscheinlichkeit folgt $\mathbb{P}(B|C) \cdot \mathbb{P}(C) = \mathbb{P}(B \cap C)$ für Ergebnisse $B, C \subseteq \Omega$.

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

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

(2) Es gilt

$$\mathbb{P}(\Omega_j|A)\mathbb{P}(A) = \mathbb{P}(\Omega_j \cap A) = \mathbb{P}(A|\Omega_j) \cdot \mathbb{P}(\Omega_j),$$

woraus die erste Gleichheit folgt. Die zweite Gleichheit folgt aus (1). \square

Beispiel 1.3.8. 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

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A|B_I)\mathbb{P}(B_I) + \mathbb{P}(A|B_{II})\mathbb{P}(B_{II}) + \mathbb{P}(A|B_{III})\mathbb{P}(B_{III}) \\ &= 0,01 \cdot 0,5 + 0,02 \cdot 0,25 + 0,04 \cdot 0,25 = 0,02. \end{aligned}$$

Es wird ein defektes Bauteil entdeckt. Mit welcher Wahrscheinlichkeit stammt es 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 1.3.9. 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

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

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

Definition 1.3.10. Es sei (Ω, \mathbb{P}) ein endlicher Wahrscheinlichkeitsraum.

- (i) Zwei Ereignisse $A, B \subseteq \Omega$ heißen *unabhängig*, wenn $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- (ii) Allgemeiner heißt eine Familie von Ereignissen *unabhängig*, falls für alle $k = 1, \dots, n$ und alle Indizes $1 \leq i_1 < \dots < i_k \leq n$ stets

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

gilt.

- (iii) Eine Familie von Ereignissen $A_1, \dots, A_n \subseteq \Omega$ heißt *paarweise unabhängig*, falls je zwei Ereignisse A_i und A_j mit $j \neq i$ unabhängig sind.

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 1.3.11. 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:

- $A_1 = \{ZZ, ZK\}$ = Im ersten Wurf fällt Zahl
- $A_2 = \{KK, ZK\}$ = Im zweiten Wurf fällt Kopf
- $A_3 = \{KZ, ZK\}$ = 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 $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(\{ZK\}) = 1/4 = 1/2 \cdot 1/2 = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$. Also sind A_1 und A_2 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. Wegen $\mathbb{P}(A_2 \cap A_3) = 1/4 = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$ sind auch A_2 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.

Paarweise unabhängige Ereignisse müssen also nicht unabhängig sein.

1.4 Wiederholung von Zufallsexperimenten

Wir erinnern uns, dass Wahrscheinlichkeiten in gewissem Sinne relative Häufigkeiten repräsentieren sollen. Relative Häufigkeiten wiederum kann man bilden, wenn man einen zufälligen Vorgang beliebig oft und gleichen Bedingungen wiederholen kann. Es fehlt allerdings noch ein mathematisches Modell für ein wiederholtes Experiment.

Nehmen wir also an, wir haben Modelle $(\Omega_1, \mathbb{P}_1), \dots, (\Omega_n, \mathbb{P}_n)$ für gewisse Zufallsexperimente. Führen wir diese Experimente nacheinander und unabhängig voneinander aus, so bietet sich als Grundmenge für das zusammengesetzte Experiment das kartesische Produkt

$$\bar{\Omega} = \Omega_1 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \Omega_i \text{ für } i = 1, \dots, n\}$$

an. Jedes Versuchsergebnis $\omega \in \bar{\Omega}$ ist also ein n -Tupel $(\omega_1, \dots, \omega_n)$, wobei die i -te Komponente $\omega_i \in \Omega_i$ gerade den Ausgang des i -ten Zufallsexperiments angibt. Das Wahrscheinlichkeitsmaß ist eindeutig festgelegt, wenn wir für Elementarereignisse verlangen, dass

$$\bar{\mathbb{P}}(\{(\omega_1, \dots, \omega_n)\}) = \mathbb{P}_1 \otimes \dots \otimes \mathbb{P}_n(\{(\omega_1, \dots, \omega_n)\}) = \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}_n(\{\omega_n\}).$$

Wir nennen $(\bar{\Omega}, \bar{\mathbb{P}})$ das *Produkt der Wahrscheinlichkeitsräume* $(\Omega_1, \mathbb{P}_1), \dots, (\Omega_n, \mathbb{P}_n)$.

Beispiel 1.4.1. Es sei $\Omega_1 = \{K, Z\}$ und $\mathbb{P}(\{K\}) = \mathbb{P}(\{Z\}) = \frac{1}{2}$. Dann beschreibt (Ω_1, \mathbb{P}_1) das Werfen einer fairen Münze. Es sei weiter $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$ und $\mathbb{P}(\{j\}) = \frac{1}{6}$, sodass (Ω_2, \mathbb{P}_2) das Werfen eines Würfels modelliert.

Das Produkt $(\Omega_1 \times \Omega_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ beschreibt das Experiment, zunächst eine Münze zu werfen und dann einen Würfel zu werfen. Wir haben

$$\begin{aligned} \Omega_1 \times \Omega_2 = \{ & (K, 1), (K, 2), (K, 3), (K, 4), (K, 5), (K, 6) \\ & (Z, 1), (Z, 2), (Z, 3), (Z, 4), (Z, 5), (Z, 6)\}. \end{aligned}$$

Zusammen mit dem Produktmaß $\mathbb{P}_1 \otimes \mathbb{P}_2$ ist $\Omega_1 \times \Omega_2$ ein Laplacescher Wahrscheinlichkeitsraum. In der Tat hat $\Omega_1 \times \Omega_2$ genau 12 Elemente und es ist $(\mathbb{P}_1 \otimes \mathbb{P}_2)(\{i, j\}) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$ für $i \in \{K, Z\}$ und $j \in \{1, 2, \dots, 6\}$ nach Definition des Produktmaßes.

Wir haben behauptet, das Produkt von Wahrscheinlichkeitsräumen modelliert die *unabhängige* Ausführung von Zufallsexperimenten. Wir wollen dies nun präzisieren. Ist A_i eine Teilmenge von Ω_i , so wird das Ereignis “Im i -ten Experiment tritt Ereignis A_i ein” im zusammengesetzten Modell beschrieben durch die Menge

$$\{(\omega_1, \dots, \omega_{i-1}, \omega, \omega_{i+1}, \dots, \omega_n) : \omega_i \in A_i\} = \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n =: \bar{A}_i.$$

Sind nun $A_1 \subseteq \Omega_1, \dots, A_n \subseteq \Omega_n$ gegeben, so sind die Ereignisse $\bar{A}_1, \dots, \bar{A}_n$ unabhängig.

Wir überprüfen dies im Fall $n = 2$. Der allgemeine Fall ist ähnlich, aber komplizierter aufzuschreiben. Es ist

$$\bar{A}_1 \cap \bar{A}_2 = \{(\omega_1, \omega_2) : \omega_1 \in A_1, \omega_2 \in A_2\} = A_1 \times A_2.$$

Daher ist

$$\begin{aligned} \bar{\mathbb{P}}(\bar{A}_1 \cap \bar{A}_2) &= \sum_{\omega \in A_1 \times A_2} \bar{\mathbb{P}}(\{\omega\}) = \sum_{(\omega_1, \omega_2) \in A_1 \times A_2} \mathbb{P}_1(\{\omega_1\}) \cdot \mathbb{P}_2(\{\omega_2\}) \\ &= \sum_{\omega_1 \in A_1} \mathbb{P}_1(\{\omega_1\}) \sum_{\omega_2 \in A_2} \mathbb{P}_2(\{\omega_2\}) = \mathbb{P}_1(A_1) \mathbb{P}_2(A_2) \end{aligned}$$

$$= \bar{\mathbb{P}}(\bar{A}_1)\bar{\mathbb{P}}(\bar{A}_2),$$

was die behauptete Unabhängigkeit zeigt. Beachte, dass insbesondere $\bar{\mathbb{P}}(A_1 \times A_2) = \mathbb{P}_1(A_1) \times \mathbb{P}_2(A_2)$ gilt.

Wir betrachten nun die n -malige Wiederholung eines Zufallsexperiments und berechnen einige interessante Wahrscheinlichkeiten.

Sei also (Ω, \mathbb{P}) ein endlicher Wahrscheinlichkeitsraum und betrachte $\bar{\Omega} := \Omega^n := \Omega \times \dots \times \Omega$ (n -mal) und $\bar{\mathbb{P}} := \mathbb{P}^{\otimes n} := \mathbb{P} \otimes \dots \otimes \mathbb{P}$.

Beispiel 1.4.2. Wir betrachten ein Ereignis $A \subseteq \Omega$ und interessieren uns dafür, wie oft dieses Ereignis beim n -maligen Wiederholen des Versuches auftritt. Wir schreiben $p = \mathbb{P}(A)$. Bezeichnet E_k das Ereignis (in $\bar{\Omega}$), dass bei n Versuchen das Ereignis A genau k -mal auftritt, so suchen wir die Wahrscheinlichkeit von E_k .

Es ist $E_0 = A^c \times \dots \times A^c$ und daher $\bar{\mathbb{P}}(E_0) = \mathbb{P}(A^c) \cdot \dots \cdot \mathbb{P}(A^c) = (1-p)^n$. Für $k = 1$ zerlegen wir das Ereignis E_1 wie folgt:

$$E_1 = (A \times A^c \times \dots \times A^c) \cup (A^c \times A \times A^c \times \dots \times A^c) \cup \dots \cup (A^c \times \dots \times A^c \times A).$$

Wir haben also E_1 disjunkt zerlegt in n Mengen, die jeweils Produkte der Mengen A und A^c sind. Dabei steht in der j -ten Menge genau an der Stelle j ein Faktor A und alle anderen Faktoren sind A^c . Die Wahrscheinlichkeit einer solchen Menge ist $p(1-p)^{n-1}$. Daher ist $\bar{\mathbb{P}}(E_1) = np(1-p)^{n-1}$.

Allgemeiner sieht man, dass

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

Man argumentiert hier wie folgt: Man zerlegt E in Produkte wie oben, wobei in den Produkten an k Stellen A und an den restlichen $n-k$ Stellen A^c steht. Ein solches kartesisches Produkt hat (unter $\bar{\mathbb{P}}$) Wahrscheinlichkeit $p^k(1-p)^{n-k}$. Da es aber gerade $\binom{n}{k}$ Möglichkeiten gibt, aus n Positionen k für die A 's auszuwählen, hat man E_k in $\binom{n}{k}$ disjunkte Mengen mit jeweils Wahrscheinlichkeit $p^k(1-p)^{n-k}$ zerlegt.

In der Situation von Beispiel 1.4.2 verwendet folgende Terminologie: Der Parameter p heißt *Erfolgswahrscheinlichkeit*. Ein einmaliges Durchführen des Experiments liefert entweder "Erfolg" (oben: A tritt ein) und das mit Wahrscheinlichkeit p oder "Misserfolg" (oben: A^c tritt ein) und das mit Wahrscheinlichkeit $1-p$. Ein solches Experiment heißt *Bernoulli-Experiment*.

Führt man ein Bernoulli-Experiment n mal durch und zählt die Anzahl der Erfolge, so erhält man eine Zahl in der Menge $\{0, 1, \dots, n\}$. Dabei ist die Wahrscheinlichkeit $\bar{\mathbb{P}}(\{k\})$, dass man genau k -mal Erfolg hatte, gerade $\binom{n}{k} p^k (1-p)^{n-k}$. Nun ist $\bar{\mathbb{P}}$ ein Wahrscheinlichkeitsmaß, da $\bar{\mathbb{P}}(k) \geq 0$ für $k \in \{0, \dots, n\}$ und

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

wegen des Binomialsatzes. Dieses Maß nennt man *Binomialverteilung* mit Parametern n und p , in Zeichen *Bin*(n, p)-Verteilung.

Eine weitere interessante Frage beim Wiederholen von Zufallsexperimenten hatten wir in Beispiel 1.1.1(c) gesehen. Wir können ein Bernoulli-Experiment wiederholen bis wir zum

ersten Mal Erfolg hatten, und fragen wie lange das dauert. Beachte, dass in diesem Fall die Grundmenge $\Omega = \mathbb{N} \cup \{\infty\}$ nicht endlich ist, und wir deshalb bisher keine Wahrscheinlichkeitsmaße auf ihr definieren können. Wir wollen nun den Begriff des Wahrscheinlichkeitsraums erweitern und auch "kleine" unendliche Grundmengen wie \mathbb{N} und \mathbb{Z} zulassen, nicht aber so große wie \mathbb{R} .

Definition 1.4.3. Eine Menge M heißt (höchstens) *abzählbar*, falls es eine surjektive Abbildung $\alpha : \mathbb{N} \rightarrow M$ gibt.

Beispiel 1.4.4. • Jede endliche Menge ist abzählbar.

- \mathbb{N} ist abzählbar; wähle z.B. die Identität als α .
- \mathbb{Z} ist abzählbar; wähle z.B.

$$\alpha(n) := \begin{cases} -\frac{n-1}{2} & \text{falls } n \text{ ungerade} \\ \frac{n}{2} & \text{falls } n \text{ gerade} \end{cases}$$

- \mathbb{Q} ist abzählbar.
- Teilmengen abzählbarer Mengen sind abzählbar.
- Vereinigungen von abzählbar vielen abzählbaren Mengen sind abzählbar.
- \mathbb{R} ist nicht abzählbar.
- Kein Intervall (a, b) oder $[a, b]$ mit $a < b$ ist abzählbar.

Definition 1.4.5. Ein *diskreter Wahrscheinlichkeitsraum* ist ein Paar (Ω, \mathbb{P}) , bestehend aus einer höchstens abzählbaren Menge Ω und einer Abbildung $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ mit folgenden Eigenschaften:

- (i) $\mathbb{P}(\Omega) = 1$ (Normiertheit)
- (ii) Ist $(A_k)_{k \in \mathbb{N}}$ eine Folge paarweise disjunkter Teilmengen von Ω , so ist

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

Diese Eigenschaft heißt *σ -Additivität*.

Wir müssen uns überlegen, dass das Konzept des diskreten Wahrscheinlichkeitsraums in der Tat das Konzept des endlichen Wahrscheinlichkeitsraums verallgemeinert.

Proposition 1.4.6. a) Jeder endliche Wahrscheinlichkeitsraum (Ω, \mathbb{P}) ist auch diskreter Wahrscheinlichkeitsraum.

b) Ein diskreter Wahrscheinlichkeitsraum (Ω, \mathbb{P}) ist ein endlicher Wahrscheinlichkeitsraum, falls Ω endlich viele Elemente hat.

Beweis. a) Sei (Ω, \mathbb{P}) ein endlicher Wahrscheinlichkeitsraum. Da jede endliche Menge abzählbar ist, ist Ω abzählbar, und Bedingung (i) aus der Definition des diskreten Wahrscheinlichkeitsraums stimmt mit Bedingung (i) aus der Definition des endlichen Wahrscheinlichkeitsraum überein.

Weisen wir Bedingung (ii) nach. Sei $(A_k)_{k \in \mathbb{N}}$ eine Folge paarweise disjunkter Teilmengen von Ω . Dann sind alle A_k außer endlich vielen leer: Es gibt ein $n \in \mathbb{N}$ mit $A_k = \emptyset$ für alle $k \geq n$. Nun folgt aus Proposition 1.1.5(1) und (4), dass

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

Also ist (Ω, \mathbb{P}) abzählbarer Wahrscheinlichkeitsraum.

b) Sei (Ω, \mathbb{P}) ein abzählbarer Wahrscheinlichkeitsraum, für den Ω endlich ist. Zeigen wir, dass (Ω, \mathbb{P}) endlicher Wahrscheinlichkeitsraum ist. Bedingung (i) überträgt sich direkt. Zum Nachweis von (ii) seien $A, B \subseteq \Omega$ unvereinbar. Wir setzen $A_1 = A$, $A_2 = B$ und $A_k = \emptyset$ für $k \geq 3$. Nun folgt Bedingung (ii) aus Bedingung (i) für abzählbare Wahrscheinlichkeitsräume. \square

Bemerkung 1.4.7. Ist (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und ist Ω abzählbar unendlich, so kann man die Elemente von Ω aufzählen, es ist also $\Omega = \{\omega_k : k \in \mathbb{N}\}$. Setzt man nun $\mathbb{P}(\{\omega_k\}) =: p_k$ so erhält man aus der σ -Additivität, dass

$$\mathbb{P}(A) = \sum_{\omega_k \in A} p_k. \quad (1.1)$$

Ist umgekehrt eine Folge p_k mit $\sum_{k=1}^{\infty} p_k = 1$ gegeben, so definiert (1.1) ein Wahrscheinlichkeitsmaß auf Ω . Die σ -Additivität folgt hierbei aus dem "großen Umordnungssatz" der Analysis.

Beispiel 1.4.8. Kommen wir nun auf die Frage zurück, wie viele Versuche bis zum ersten Erfolg eines Bernoulli Experiment nötig sind. Die Wahrscheinlichkeit des Elementarereignisses k ("Im k -ten Versuch hat man zum ersten Mal Erfolg") ist $p(1-p)^{k-1}$, denn dies ist bei k -maliger Wiederholung des Versuches die Wahrscheinlichkeit dafür, in Versuchen $1, 2, \dots, k-1$ Misserfolg und in Versuch k Erfolg zu beobachten.

Beachte, dass für $q \in (0, 1)$ gerade $\sum_{k=0}^{\infty} q^k = (1-q)^{-1}$ (Geometrische Reihe). Damit erhält man

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

und somit $\mathbb{P}(\{\infty\}) = 0$. Es wäre also nicht notwendig gewesen, das Ergebnis ∞ , dass der Erfolg überhaupt nie eintritt, in den Grundraum einzuschließen. Wir haben ein Wahrscheinlichkeitsmaß auf \mathbb{N} via

$$\mathbb{P}(A) = \sum_{k \in A} p(1-p)^{k-1}$$

definiert. Dieses Wahrscheinlichkeitsmaß heißt *geometrische Verteilung mit Erfolgswahrscheinlichkeit p* , in Zeichen $\text{Geom}(p)$.

1.5 Zufallsvariablen und ihre Momente

Bei vielen Zufallsexperimenten interessiert nicht so sehr der tatsächliche Ausgang ω des Experiments, sondern vielmehr eine bestimmte Größe $X(\omega)$, die vom Ausgang des Experimentes abhängt. Spielt man beispielsweise Lotto (das Ergebnis ω ist also eine 6-elementige Teilmenge der Zahlen von 1 bis 49), so ist man primär am *Gewinn* interessiert (der aber natürlich von ω) abhängt.

Besteht das Zufallsexperiment darin, zufällig einen Bewohner Ulms auszuwählen, so sind (gerade bei statistischen Untersuchungen) lediglich bestimmte Aspekte der Person interessant, z.B. das Alter oder das Einkommen der Person.

Definition 1.5.1. Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum. Eine *Zufallsvariable* ist eine Abbildung $X : \Omega \rightarrow \mathbb{R}$.

Beachte, dass der Wertebereich $X(\Omega) := \{X(\omega) : \omega \in \Omega\}$ ebenfalls abzählbar ist. Die *Verteilung von X* ist das Wahrscheinlichkeitsmaß \mathbb{P}_X auf $X(\Omega)$ gegeben durch

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}).$$

Manchmal verwendet man auch den Begriff *Zähldichte* für diese Wahrscheinlichkeiten und schreibt $f_X(x) := \mathbb{P}_X(\{x\})$. Beachte, dass $f_X(x) \neq 0$ für höchstens abzählbar viele x . Daher ist

$$\mathbb{P}_X(A) = \sum_{x \in A} f_X(x).$$

Beispiel 1.5.2. Wir werfen einen Würfel zweimal. X beschreibe die Summe der gewürfelten Augenzahlen. Das zweimalige Werfen eines Würfels kann als Laplace Experiment mit Grundmenge $\Omega := \{(i, j) : i, j = 1, \dots, 6\}$ beschrieben werden, wobei (i, j) für das Elementarereignis “ i im ersten Wurf, j im zweiten Wurf” steht.

Es ist $X : \Omega \rightarrow \mathbb{R}$ gegeben durch $X((i, j)) = i + j$. Demnach ist der Wertebereich $X(\Omega) = \{2, 3, \dots, 12\}$. Um die Verteilung zu bestimmen, muss man die Ereignisse $\{\omega \in \Omega : X(\omega) = x\}$ bestimmen. Beispielsweise erhalten wir

$$\begin{aligned} \{\omega \in \Omega : X(\omega) = 2\} &= \{(1, 1)\} \\ \{\omega \in \Omega : X(\omega) = 3\} &= \{(1, 2), (2, 1)\} \\ \{\omega \in \Omega : X(\omega) = 4\} &= \{(1, 3), (2, 2), (3, 1)\} \\ \{\omega \in \Omega : X(\omega) = 5\} &= \{(1, 4), (2, 3), (3, 2), (4, 1)\} \\ &\dots \end{aligned}$$

Die Zähldichte von X , und somit die Verteilung von X , wird in folgender Tabelle zusammengefasst.

x	2	3	4	5	6	7	8	9	10	11	12
$f_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Durch eine Zufallsvariable wird gewissermaßen der Wahrscheinlichkeitsraum verkleinert: Statt (Ω, \mathbb{P}) wird nur noch $(X(\Omega), \mathbb{P}_X)$ betrachtet.

Betrachten wir nochmals Beispiel 1.4.2 (Anzahl der Erfolge bei n -maliger Wiederholung eines Bernoulli-Experiments). Auch hier ist der zugrundeliegende Wahrscheinlichkeitsraum $\bar{\Omega} := \Omega^n$ von untergeordnetem Interesse (und unter Umständen sehr groß, man denke etwa an

n -maliges Lotto spielen). Allerdings interessieren wir uns nur dafür, wie oft ein bestimmtes Ereignis (etwa “6 Richtige”) eintritt. Dies wird durch eine Zufallsvariable $X : \bar{\Omega} \rightarrow \mathbb{R}$ angegeben. In Beispiel 1.4.2 haben wir gerade die Verteilung von X bestimmt. Es ist nämlich $X(\bar{\Omega}) = \{0, 1, \dots, n\}$ und für $k \in X(\bar{\Omega})$

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

Wir führen nun noch einige Notationen ein. In Ausdrücken wie $\{\omega \in \Omega : X(\omega) = x\}$ tragen einige Teile keine Information. Deshalb lässt man “ $\omega \in \Omega$ ” und alle im Folgenden auftretenden “ ω weg und schreibt stattdessen z.B. $\{X = x\}$. Gibt man Wahrscheinlichkeiten an, so lässt man meist auch noch die geschweiften Klammern weg und schreibt $\mathbb{P}(X = x)$ anstelle von $\mathbb{P}(\{X = x\})$ (was wiederum eine Abkürzung für $\mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$ ist).

Wir kommen nun zu einer zentralen Definition:

Definition 1.5.3. Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Wir sagen *der Erwartungswert von X ist endlich*, falls

$$\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < \infty.$$

In diesem Fall heißt $\mathbb{E}X$, definiert durch

$$\mathbb{E}X := \sum_{x \in X(\Omega)} x \mathbb{P}(X = x)$$

der *Erwartungswert* von X .

Bemerkung 1.5.4. (a) Beachte, dass $X(\Omega)$ eine abzählbare Menge ist. Die Summen oben sind also entweder endliche Summen (in diesem Fall ist der Erwartungswert von X stets endlich, denn die Summe endlich vieler reeller Zahlen ist stets endlich) oder aber die Summe hat abzählbar unendlich viele Summanden, d.h. es liegt eine Reihe vor.

(b) Die Bedingung, dass $\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < \infty$ ist, bedeutet gerade dass die Reihe $\sum_{x \in X(\Omega)} x \mathbb{P}(X = x)$ *absolut konvergiert*; insbesondere konvergiert die Reihe (und der Wert der Reihe ist unabhängig von der Reihenfolge der Summation). Es gibt aber auch konvergente Reihen, die nicht absolut konvergieren. Beispielsweise konvergiert die alternierende harmonische Reihe

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

Sie konvergiert aber nicht absolut, denn die harmonische Reihe $\sum_{n=1}^{\infty} \frac{1}{n}$ divergiert.

Beachte, dass der Erwartungswert nicht definiert ist, falls $\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) = \infty$. Denn dann hängt der Wert von $\sum_{x \in X(\Omega)} x \mathbb{P}(X = x)$ von der Reihenfolge der Summation ab.

- (c) Der Erwartungswert ist das (mit den Wahrscheinlichkeiten) gewichtete Mittel der Werte von X und gibt an, welchen Wert die Zufallsvariable “im Schnitt” annimmt.
- (d) Die obige Definition des Erwartungswert hängt nur von der Verteilung von X , genauer vom diskreten Wahrscheinlichkeitsraum $(X(\Omega), \mathbb{P}_X)$ ab.

Beispiel 1.5.5. Wir betrachten die Situation von Beispiel 1.5.2, d.h. X ist die Augensumme beim zweimaligen Werfen eines Würfels. Da die Zufallsvariable nur endlich viele Werte annimmt, ist der Erwartungswert von X endlich. Er berechnet sich wie folgt:

$$\begin{aligned} \mathbb{E}X &= \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) = \sum_{j=2}^{12} j \mathbb{P}(X = j) \\ &= 2 \frac{1}{36} + 3 \frac{2}{36} + 4 \frac{3}{36} + 5 \frac{4}{36} + 6 \frac{5}{36} + 7 \frac{6}{36} + 8 \frac{5}{36} + 9 \frac{4}{36} + 10 \frac{3}{36} + 11 \frac{2}{36} + 12 \frac{1}{36} \\ &= \frac{252}{36} = 7 \end{aligned}$$

Beispiel 1.5.6. 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 (4 der 32 Karten), so werden ihm $10 \in \mathbb{E}$ ausgezahlt. Zieht er eine Bildkarte (Bube, Dame, König; je 4 Karten), so erhält er $2 \in \mathbb{E}$ ausgezahlt. Ansonsten erhält er nichts. Für welchen Wert von E ist dieses 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 $4/32$, $12/32$ bzw. $16/32$ 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.

Wir zeigen nun einige wichtige Eigenschaften des Erwartungswerts. Zunächst zeigen wir eine alternative Darstellung des Erwartungswerts, bei der über die Elemente von Ω (nicht die von $X(\Omega)$) summiert wird.

Lemma 1.5.7. *Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit endlichem Erwartungswert. Dann ist*

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}).$$

Hierbei konvergiert die Summe absolut.

Beweis. Da $X(\Omega)$ abzählbar ist, gibt es Zahlen x_j , $j \in J$, mit $X(\Omega) = \{x_j : j \in J\}$, wobei die Indexmenge J entweder endlich oder abzählbar unendlich ist. Es sei A_j das Ereignis $\{X = x_j\}$.

Nun ist

$$\begin{aligned} \mathbb{E}X &= \sum_{j \in J} x_j \mathbb{P}(A_j) = \sum_{j \in J} x_j \sum_{\omega \in A_j} \mathbb{P}(\{\omega\}) = \sum_{j \in J} \sum_{\omega \in A_j} X(\omega) \mathbb{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}). \end{aligned}$$

Hier haben wir beim zweiten Gleichheitszeichen die σ -Additivität von \mathbb{P} verwendet und beim dritten, dass für $\omega \in A_j$ die Gleichheit $x_j = X(\omega)$ gilt. Bei der letzten Gleichheit haben wir verwendet, dass Ω die disjunkte Vereinigung der A_j , $j \in J$, ist, sowie den "großen Umordnungssatz" der Analysis, der es erlaubt die Summanden einer absolut konvergenten Reihe beliebig zu gruppieren und aufzusummieren. \square

Nun folgt sofort:

Proposition 1.5.8. *Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum, X, Y Zufallsvariablen und $\alpha, \beta \in \mathbb{R}$.*

- (1) *Ist $0 \leq Y \leq X$ und hat X endlichen Erwartungswert, so hat auch Y endlichen Erwartungswert und es gilt $\mathbb{E}Y \leq \mathbb{E}X$.*
- (2) *Haben X und Y endlichen Erwartungswert, so auch $\alpha X + \beta Y$ und es gilt $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}X + \beta \mathbb{E}Y$.*

Beweis. Wir verwenden Lemma 1.5.7.

- (1) Aufgrund der Monotonie von Reihen mit positiven Einträgen ist

$$\sum_{\omega \in \Omega} Y(\omega) \mathbb{P}(\{\omega\}) \leq \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) < \infty$$

nach Voraussetzung.

- (2) Beachte, dass

$$|\alpha X + \beta Y| \leq |\alpha| |X| + |\beta| |Y|$$

ist. Aufgrund der Rechenregeln für Reihen mit positiven Einträgen ist

$$\sum_{\omega \in \Omega} (|\alpha| |X(\omega)| + |\beta| |Y(\omega)|) \mathbb{P}(\{\omega\}) = |\alpha| \mathbb{E}|X| + |\beta| \mathbb{E}|Y| < \infty.$$

Nach Teil (1) hat also $\alpha X + \beta Y$ endliche Erwartung. Nun gilt

$$\begin{aligned} \mathbb{E}[\alpha X + \beta Y] &= \sum_{\omega \in \Omega} (\alpha X(\omega) + \beta Y(\omega)) \mathbb{P}(\{\omega\}) \\ &= \alpha \sum_{\omega \in \Omega} (X(\omega) \mathbb{P}(\{\omega\})) + \beta \sum_{\omega \in \Omega} (Y(\omega) \mathbb{P}(\{\omega\})) \\ &= \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]. \end{aligned}$$

□

Bemerkung 1.5.9. Ist (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitraum, $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable und $g : X(\Omega) \rightarrow \mathbb{R}$ eine Zufallsvariable, so ist auch $g \circ X$ eine Zufallsvariable. Man schreibt meist $g(X)$. Also ist $\mathbb{E}[g(X)]$ definiert, falls $g(X)$ endlichen Erwartungswert hat. In diesem Fall gilt

$$\begin{aligned} \mathbb{E}[g(X)] &= \sum_{y \in g(X(\Omega))} y \mathbb{P}(y = g(X)) \\ &= \sum_{y \in g(X(\Omega))} y \cdot \left(\sum_{x \in X(\Omega), g(x)=y} \mathbb{P}(x = X) \right) \\ &= \sum_{y \in g(X(\Omega))} \sum_{x \in X(\Omega), g(x)=y} g(x) \mathbb{P}(x = X) \\ &= \sum_{x \in X(\Omega)} g(x) \mathbb{P}(x = X) \end{aligned}$$

Analog folgt, dass $g(X)$ endlichen Erwartungswert hat, falls $\sum_{x \in X(\Omega)} |g(x)| \mathbb{P}(x = X) < \infty$. Nach Lemma 1.5.7 gilt

$$\mathbb{E}[g(X)] = \sum_{\omega \in \Omega} g(X(\omega)) \mathbb{P}(\{\omega\}).$$

Beispiel 1.5.10. Eine *Indikatorfunktion* ist eine Funktion, die nur die Werte 1 und 0 annimmt. Ist M eine Menge und $A \subseteq M$, so schreiben wir $\mathbb{1}_A$ für die Funktion von M nach \mathbb{R} , die auf A den Wert 1 und auf A^c den Wert 0 annimmt, d.h.

$$\mathbb{1}_A : M \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{falls } x \notin A \end{cases}.$$

Sei nun X eine Zufallsvariable und $A \subseteq \mathbb{R}$. Dann hat $\mathbb{1}_A(X)$ endlichen Erwartungswert (denn $0 \leq \mathbb{1}_A(X) \leq \mathbf{1}$, wobei $\mathbf{1}$ die Zufallsvariable ist, die konstant 1 ist; diese hat endlichen Erwartungswert). Weiter ist

$$\mathbb{E}\mathbb{1}_A(X) = 1 \cdot \mathbb{P}(\mathbb{1}_A(X) = 1) + 0 \cdot \mathbb{P}(\mathbb{1}_A(X) = 0) = \mathbb{P}(X \in A).$$

Wir wollen nun neben den Erwartungswerten auch eine weitere Kenngröße von Zufallsvariablen, die sog. Varianz betrachten. Dies wird durch folgendes Beispiel motiviert.

Beispiel 1.5.11. Es sei $\Omega = \{-1, 1\}$ mit $\mathbb{P}(\{-1\}) = \mathbb{P}(\{1\}) = \frac{1}{2}$. Weiter sei für $n \in \mathbb{N}$ die Zufallsvariable X_n definiert durch $X_n(\omega) = n\omega$. Dann ist

$$\mathbb{E}X_n = (-n)\frac{1}{2} + n\frac{1}{2} = 0$$

Dieses Beispiel zeigt, dass der Erwartungswert zwar angibt, welchen Wert die Zufallsvariable im Schnitt annimmt, er aber keine Aussage darüber trifft, wie viel die Zufallsvariable von diesem Mittelwert abweicht. Dies macht die *Varianz*.

Definition 1.5.12. Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit endlichem Erwartungswert μ . Ist der Erwartungswert

$$\mathbb{E}[(X - \mu)^2]$$

endlich, so sagen wir X hat *endliche Varianz* und dieser Erwartungswert heißt *Varianz* von X und wird mit $\text{Var}X$ bezeichnet. Die Wurzel aus $\text{Var}X$ heißt *Streuung* von X oder *Standardabweichung* von X .

Man beachte, dass die Endlichkeit des Erwartungswertes

$$\mathbb{E}[(X - \mu)^2]$$

äquivalent zur absoluten Konvergenz der Reihe

$$\sum_{x \in X(\Omega)} (x - \mu)^2 \mathbb{P}(X = x)$$

ist. Sind diese Bedingungen erfüllt, so ist der Erwartungswert gleich dem Grenzwert der Reihe.

Beispiel 1.5.13. Wir berechnen die Varianz der Zufallsvariablen X_n aus Beispiel 1.5.11. Weil $\mathbb{E}X_n \equiv 0$ ist, folgt

$$\text{Var}X_n = (-n - 0)^2 \frac{1}{2} + (n - 0)^2 \frac{1}{2} = n^2.$$

Somit hat X_n Varianz n^2 und Standardabweichung n .

Beispiel 1.5.14. Wir berechnen die Varianz des Spieles in Beispiel 1.5.6 bei beliebigem Einsatz E . Es ist $\mathbb{E}X = 2 - E$. Daher

$$\begin{aligned}\text{Var}X &= (10 - E - (2 - E))^2 \frac{4}{32} + (2 - E - (2 - E))^2 \frac{12}{32} + (-E - (2 - E))^2 \frac{16}{32} \\ &= 64 \frac{1}{8} + 0 \frac{3}{8} + 4 \frac{1}{2} = 10\end{aligned}$$

Beachte weiterhin, dass X endliche Varianz hat genau dann, wenn X^2 endlichen Erwartungswert hat. Genauer gilt:

Lemma 1.5.15. Sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable.

- Hat X^2 endlichen Erwartungswert, so hat X endlichen Erwartungswert und endliche Varianz.
- Hat X endlichen Erwartungswert und endliche Varianz, so hat X^2 endlichen Erwartungswert.
- Sind diese Bedingungen erfüllt, so folgt $\text{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

Beweis. c) Mit $\mu := \mathbb{E}[X^2]$ folgt

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

a) Es gilt $|X| \leq 1 + X^2$. Somit folgt aus Proposition 1.5.8, dass $|X|$ endlichen Erwartungswert hat. Aber aus der Definition des Erwartungswertes folgt, dass X genau dann endlichen Erwartungswert hat, wenn $|X|$ endlichen Erwartungswert hat. Nun folgt aus

$$(X - \mu)^2 = X^2 - 2X\mu + \mu^2,$$

dass $(X - \mu)^2$ endlichen Erwartungswert hat, also dass X endliche Varianz hat.

b) Dies folgt aus

$$X^2 = (X - \mu)^2 + 2X\mu - \mu^2.$$

□

Proposition 1.5.16. Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$ und seien $a, b \in \mathbb{R}$. Dann gilt

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

Beweis. Wir setzen $\mu := \mathbb{E}X$ und $\tilde{\mu} := \mathbb{E}[aX + b]$. Wegen der Linearität des Erwartungswertes gilt dann $\tilde{\mu} = a\mu + b$. Also

$$\text{Var}(aX + b) = \mathbb{E}[(aX + b - \tilde{\mu})^2] = \mathbb{E}[a^2(X - \mu)^2] = a^2 \text{Var}X.$$

□

Definition 1.5.17. Sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum, $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable und $r \in \mathbb{N}$. Wir sagen, dass X ein *endliches Moment r -ter Ordnung besitzt*, oder dass X ein *endliches r -tes Moment besitzt*, falls $|X|^r$ endlichen Erwartungswert besitzt. In diesem Fall heißt $\mathbb{E}X^r$ *Moment r -ter Ordnung*.

Bemerkung 1.5.18. Ist $k \leq r$, so ist

$$\gamma := \sup_{t \in \mathbb{R}} \frac{|t|^k}{1 + |t|^r} < \infty$$

Folglich ist $|X|^k \leq \gamma(1 + |X|^r)$. Deshalb hat X ein endliches Moment k -ter Ordnung, falls es ein endliches Moment r -ter Ordnung besitzt.

1.6 Spezielle Verteilungen

Wir diskutieren nun einige spezielle Verteilungen von Zufallsvariablen mit abzählbarem Wertebereich, die in Anwendungen oft auftreten.

Die Binomialverteilung

In Beispiel 1.4.2 hatten wir die n -fache Wiederholung eines Bernoulli-Experiments mit Erfolgswahrscheinlichkeit p betrachtet. Wir sahen, dass die Anzahl X der Erfolge eine Zufallsvariable mit $X(\Omega) = \{0, 1, \dots, n\}$ und

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

ist. Wir nennen eine Zufallsvariable mit dieser Verteilung *binomialverteilt* mit Parametern n und p . Wir schreiben $X \sim \text{Bin}(n, p)$. Ist $n = 1$ so sagt man auch, die Zufallsvariable sei *Bernoulli-verteilt*.

Berechnen wir nun die Momente der Binomialverteilung, zunächst den Erwartungswert. Beachten wir, dass $k \binom{n}{k} = k \frac{n!}{k!(n-k)!} = n \frac{(n-1)!}{(k-1)!(n-1-(k-1)!)} = n \binom{n-1}{k-1}$ ist, so folgt

$$\begin{aligned} \mathbb{E}X &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} = np. \end{aligned}$$

Interpretation: Wiederholt man ein Experiment mit Erfolgswahrscheinlichkeit p n -mal, so hat man im Schnitt np Erfolge.

Wir berechnen nun die Varianz einer binomialverteilten Zufallsvariablen. Wir wissen, dass $\text{Var}X = \mathbb{E}X^2 - (np)^2$ ist, es genügt also, $\mathbb{E}X^2$ zu berechnen. Beachten wir noch, dass $\mathbb{E}X^2 = \mathbb{E}(X(X-1) + X) = np + \mathbb{E}(X(X-1))$ ist, so genügt es $\mathbb{E}(X(X-1))$ zu berechnen. Ähnlich wie oben erhalten wir:

$$\begin{aligned} \mathbb{E}X(X-1) &= \sum_{k=1}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^n \binom{n-2}{k-2} p^{k-2} (1-p)^{n-2-(k-2)} \\ &= n(n-1)p^2. \end{aligned}$$

Somit ist

$$\text{Var}X = \mathbb{E}X(X-1) + np - (np)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

Die hypergeometrische Verteilung

Man kann eine $\text{Bin}(n, p)$ -verteilte Zufallsvariable auch wie folgt erzeugen: Man zieht aus einer Urne, in der der Anteil schwarzer Kugeln p beträgt, n Kugeln. Dabei ziehen wir mit

Zurücklegen oder wir nehmen an, dass die Urne unendlich viele Kugeln enthält. Nun ist die Zahl der gezogenen schwarzen $Bin(n, p)$ -verteilt. Was aber passiert, wenn wir aus einer Urne mit endlich vielen Kugeln ohne Zurücklegen ziehen?

In einer Urne befinden sich m schwarze und n weiße Kugeln. Es werden k Kugeln ($k \leq m+n$) ohne Zurücklegen und ohne Beachten der Reihenfolge gezogen. Nummerieren wir die schwarzen Kugeln von 1 bis m und die weißen Kugeln von $m+1$ bis $n+m$ durch, so ist ein geeigneter Wahrscheinlichkeitsraum gegeben durch

$$\Omega = \{(\omega_1, \dots, \omega_{n+m}) \in \{0, 1\}^{n+m} \mid \sum_{i=1}^{n+m} \omega_i = k\},$$

wobei $(\omega_1, \dots, \omega_{n+m})$ das Ereignis bezeichnet, dass genau diejenigen Kugeln i mit $\omega_i = 1$ gezogen werden. Wir können annehmen, dass dieser Wahrscheinlichkeitsraum Laplacesch ist. Wie in Abschnitt 1.2 besprochen, hat dieser Wahrscheinlichkeitsraum $\binom{m+n}{k}$ Elemente. Für eine Zahl $j = 0, \dots, \min\{m, k\}$ wollen wir nun die Wahrscheinlichkeit berechnen, dass die Anzahl X der gezogenen schwarzen Kugeln genau gleich j ist. Hierzu überlegen wir

$$\begin{aligned} \#\{X = j\} &= \#\{(\omega_1, \dots, \omega_{n+m}) \in \Omega \mid \sum_{i=1}^m \omega_i = j\} \\ &= \#\{(\omega_1, \dots, \omega_{n+m}) \in \{0, 1\}^{n+m} \mid \sum_{i=1}^m \omega_i = j, \sum_{i=m+1}^{n+m} \omega_i = k - j\} \\ &= \#\left(\left\{(\omega_1, \dots, \omega_m) \in \{0, 1\}^m \mid \sum_{i=1}^m \omega_i = j\right\} \right. \\ &\quad \left. \times \left\{(\omega_{m+1}, \dots, \omega_{m+n}) \in \{0, 1\}^n \mid \sum_{i=m+1}^{m+n} \omega_i = k - j\right\}\right) \\ &= \#\left\{(\omega_1, \dots, \omega_m) \in \{0, 1\}^m \mid \sum_{i=1}^m \omega_i = j\right\} \\ &\quad \cdot \#\left\{(\omega_{m+1}, \dots, \omega_{m+n}) \in \{0, 1\}^n \mid \sum_{i=m+1}^{m+n} \omega_i = k - j\right\} \end{aligned}$$

Die Zahl der Möglichkeiten, aus allen $m+n$ Kugeln k Stück zu ziehen und dabei aus den m schwarzen Kugeln j Stück zu ziehen, ist also gleich der Anzahl der Möglichkeiten, aus den m schwarzen Kugeln j Stück zu ziehen, mal der Anzahl der Möglichkeiten, aus den n weißen Kugeln $k-j$ Stück zu ziehen. Wie in Abschnitt 1.2 besprochen, sind dies $\binom{m}{j}$ bzw. $\binom{n}{k-j}$ Möglichkeiten. Also gilt

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

für $j = 0, 1, \dots, \min\{k, m\}$. Eine Zufallsvariable X mit dieser Verteilung heißt *hypergeometrisch verteilt* mit Parametern m, n und k . Wir schreiben $X \sim Hg(m, n, k)$.

Wir berechnen nun den Erwartungswert einer hypergeometrisch verteilten Zufallsvariable. Wir verwenden wieder die Identität $\binom{a}{b} = \frac{a}{b} \binom{a-1}{b-1}$. Wir erhalten

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

$$= k \frac{m}{m+n} \sum_{l=0}^{k-1} \frac{\binom{m-1}{l} \binom{n}{k-1-l}}{\binom{m-1+n}{k-1}} = k \frac{m}{m+n}$$

wobei wir im letzten Schritt benutzt haben, dass $\frac{\binom{m-1}{l} \binom{n}{k-1-l}}{\binom{m-1+n}{k-1}}$ gerade die Zähl-dichte einer $Hg(m-1, n, k-1)$ ist. Diese Formel für den Erwartungswert lässt sich wie folgt interpretieren:

Zieht man eine Kugel, so ist die Wahrscheinlichkeit eine schwarze zu ziehen gerade $\frac{m}{m+n}$. Zieht man also k Kugeln, so sind im Schnitt gerade $k \frac{m}{m+n}$ schwarze darunter. Weiter bemerken wir, dass sich der selbe Wert wie bei einer $Bin(k, \frac{m}{m+n})$ -Verteilung ergibt.

Ist $X \sim Hg(m, n, k)$, so sieht man ähnlich wie oben, dass

$$\text{Var}X = k \frac{m}{m+n} \left(1 - \frac{m}{m+n}\right) \frac{m+n-k}{m+n-1}.$$

Vergleicht man dies mit der Varianz einer $Bin(k, \frac{m}{m+n})$ -Verteilung, so liegt der Unterschied gerade im letzten Faktor $\frac{m+n-k}{m+n-1}$. Dieser Faktor kann als Korrekturterm für das Ziehen aus einem endlichen Vorrat angesehen werden.

Die geometrische Verteilung

In Beispiel 1.4.8 hatten wir die Anzahl X der Versuche bis zum ersten Erfolg bei der Wiederholung eines Bernoulli-Experiments mit Erfolgswahrscheinlichkeit $p \in (0, 1)$ betrachtet. Wir haben gesehen, dass $X(\Omega) = \mathbb{N}$ ist und $\mathbb{P}(X = k) = p(1-p)^{k-1}$ ist. Eine Zufallsvariable mit dieser Verteilung heißt *geometrisch verteilt* mit Parameter p . Wir schreiben $X \sim \text{Geom}(p)$.

Zum Berechnen von Erwartungswert und Varianz einer geometrisch verteilten Zufallsvariable wiederholen wir, dass für $x \in (0, 1)$

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$$

ist. Durch Differenzieren erhält man

$$\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1} \quad \text{und} \quad \frac{2}{(1-x)^3} = \sum_{k=2}^{\infty} k(k-1)x^{k-2}.$$

Ist also $X \sim \text{Geom}(p)$, so ist

$$\mathbb{E}X = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

und

$$\begin{aligned} \mathbb{E}X(X-1) &= \sum_{k=2}^{\infty} k(k-1)p(1-p)^{k-1} \\ &= p(1-p) \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &= \frac{2p(1-p)}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}. \end{aligned}$$

Somit ergibt sich

$$\text{Var}X = \mathbb{E}(X(X-1)) + \frac{1}{p} - \frac{1}{p^2} = \frac{2-2p+p-1}{p^2} = \frac{1-p}{p^2}.$$

Beispiel 1.6.1. Im Schnitt benötigt man also 6 Versuche bis man eine Sechse auf einem fairen Würfel würfelt. Die Erfolgswahrscheinlichkeit in diesem Experiment beträgt nämlich $p = 1/6$ und der Erwartungswert einer $Geom(\frac{1}{6})$ -verteilten Zufallsvariable ist $(1/6)^{-1} = 6$.

Die Poisson-Verteilung

Betrachten wir nun die Anzahl von sog. "seltenen" Ereignissen, d.h. von Ereignissen, für die es viele Gelegenheiten gibt, die aber nur bei wenigen dieser Gelegenheiten eintreten, z.B. Fehler oder der radioaktive Zerfall von Atomen. Die Verteilung dieser Anzahlen kennen wir bereits, es ist die Binomialverteilung. Da aber die Anzahl der Gelegenheiten oftmals schwer zu bestimmen ist, wollen wir nun die Binomialverteilung durch eine Verteilung, die nur von einem Parameter abhängt, approximieren.

Wir betrachten die $Bin(n, p_n)$ -Verteilung, wobei wir annehmen, dass $\lambda_n := np_n \rightarrow \lambda$ für $n \rightarrow \infty$ konvergiert. Für festes k ist die Zähldichte gegeben durch

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

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

Somit haben wir gezeigt:

Proposition 1.6.2. 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 $k = 0, 1, 2, \dots$

Daher nennen wir eine Zufallsvariable X *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 $X \sim Pois(\lambda)$.

Wir diskutieren ein Beispiel zu Approximation der Binomial-Verteilung.

Beispiel 1.6.3. Befinden sich n Personen in einem Raum, so ist die Anzahl X der Personen in diesem Raum, die heute Geburtstag haben, eine binomialverteilte Zufallsvariable mit Parametern n und $p = 1/365$. Wir betrachten den Fall $n = 97$. Damit ist $X \sim Bin(97, \frac{1}{365})$. Wir können diese Verteilung mit der Poisson-Verteilung zum Parameter $\lambda = 97/365$ approximieren. Die folgende Tabelle gibt einige vergleichende Werte der Zähldichten von Binomial- und Poisson-Verteilung an.

k	0	5	50
$Bin(97, \frac{1}{365})$	0,7663	$7,7289 \cdot 10^{-6}$	$8,2558 \cdot 10^{-101}$
$Pois(\frac{97}{365})$	0,7666	$8,4683 \cdot 10^{-6}$	$4,2214 \cdot 10^{-94}$

Wir sehen, dass die Approximation im praktisch relevanten Bereich gut ist. Sind die Wahrscheinlichkeiten hingegen so klein, dass der Wert für die meisten Anwendungen ohnehin bedeutungslos ist, treten dramatische relative Fehler auf.

Wir berechnen nun Erwartungswert und Varianz einer Poisson-verteilten Zufallsvariablen. Ist $X \sim Pois(\lambda)$, so ist

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

Fast die gleiche Rechnung liefert $\mathbb{E}X(X-1) = \lambda^2$, sodass $\text{Var}X = \lambda^2 + \lambda - \lambda^2 = \lambda$.

1.7 Zufallsvektoren

Häufig sind auf einem (diskreten) Wahrscheinlichkeitsraum mehrere Zufallsvariablen definiert, die von Interesse sind. In diesem Fall möchte man diese gemeinsam betrachten und insbesondere auch untersuchen, ob man von der Kenntnis einer Zufallsvariablen Rückschlüsse auf die anderen Zufallsvariablen ziehen kann.

Definition 1.7.1. Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und X_1, \dots, X_n Zufallsvariablen. Die Funktion $X : \Omega \rightarrow \mathbb{R}^n$, $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$ heißt *Zufallsvektor*. Der Wertebereich $X(\Omega)$ ist enthalten im kartesischen Produkt $X_1(\Omega) \times \dots \times X_n(\Omega)$. Die *Verteilung* von X ist das Wahrscheinlichkeitsmaß \mathbb{P}_X auf $X_1(\Omega) \times \dots \times X_n(\Omega)$, gegeben durch

$$\mathbb{P}_X(\{(x_1, \dots, x_n)\}) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

Beachte, dass für gewisse Wahlen von x_1, \dots, x_n auch $\mathbb{P}_X(\{(x_1, \dots, x_n)\}) = 0$ sein kann. Man nennt

$$f_X(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

auch die *Zähldichte* des Vektors X .

Genau wie bei Zufallsvariablen erfüllen auch die Zähldichten von Zufallsvektoren

$$f(x_1, \dots, x_n) \geq 0 \quad \text{für } x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega_n) \quad (1.2)$$

$$\sum_{x_1, \dots, x_n} f(x_1, \dots, x_n) = 1. \quad (1.3)$$

Umgekehrt kann man zeigen, dass jede Funktion f , die (1.2) und (1.3) erfüllt, Zähldichte eines Zufallsvektors ist. Die Verteilung eines Zufallsvektors ist durch seine Zähldichte eindeutig festgelegt.

Beispiel 1.7.2. 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. Beispielsweise gilt

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

Analog erhält man die anderen Werte in folgender Tabelle.

	R	0	1	2	$W \downarrow$
W					
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	$\frac{1}{30}$
	$R \rightarrow$	$\frac{21}{60}$	$\frac{1}{2}$	$\frac{3}{20}$	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 und die letzte Zeile enthält die Zähldichte für R . Beispielsweise ist

$$\mathbb{P}(W = 0) = \frac{\binom{5}{0} \binom{20}{2}}{\binom{25}{2}} = \frac{20 \cdot 19}{25 \cdot 24} = 19/30$$

(hypergeometrisches Modell mit 5 weißen und 20 nicht-weißen Kugeln). Beachte, dass die Werte in dieser Spalte genau die Summe der Wahrscheinlichkeiten in der Zeile davor sind:

$$\mathbb{P}(W = 0) = \mathbb{P}(W = 0, R = 0) + \mathbb{P}(W = 0, R = 1) + \mathbb{P}(W = 0, R = 2) = \frac{3}{20} + \frac{1}{3} + \frac{3}{20} = \frac{19}{30}$$

wegen der Additivität des Wahrscheinlichkeitsmaßes und weil $\{W = 0\}$ die Vereinigung der disjunkten Mengen $\{W = 0, R = 0\}$, $\{W = 0, R = 1\}$ und $\{W = 0, R = 2\}$ ist.

Man nennt die Verteilungen von W und von R auch die *Randverteilungen* des Vektors (W, R) .

Beispiel 1.7.3. Die Verteilung eines Zufallsvektors (X, Y) sei gegeben durch

	Y	0	1	$X \downarrow$
X				
0		c	$\frac{1}{2} - c$	$\frac{1}{2}$
1		$\frac{1}{2} - c$	c	$\frac{1}{2}$
	$Y \rightarrow$	$\frac{1}{2}$	$\frac{1}{2}$	1

für $c \in [0, \frac{1}{2}]$. 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. Insbesondere bestimmen also die Randverteilungen eines Vektors die Verteilung des Vektors *nicht* eindeutig.

Hat man einen Zufallsvektor $X = (X_1, \dots, X_n)$ und eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ gegeben, so ist $f(X)$ eine Zufallsvariable. Einige interessante Kenngrößen von Zufallsvektoren sind als Erwartungswert solcher Zufallsvariablen definiert.

Seien X, Y Zufallsvariablen mit endlichen zweiten Momenten ($\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$). Weil $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$ ist, hat XY endlichen Erwartungswert. Deshalb können wir definieren:

Definition 1.7.4. Sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum, X, Y Zufallsvariablen mit endlichen zweiten Momenten. Weiter sei $\mu = \mathbb{E}X$ und $\nu = \mathbb{E}Y$. Dann heißt

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

die Kovarianz von X und Y . Sind zusätzlich $\text{Var}X > 0$ und $\text{Var}Y > 0$, so heißt

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

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

Beispiel 1.7.5. Wir betrachten wiederum Beispiel 1.7.3. 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 von Kovarianz und Korrelation geben. Sind diese Zahlen positiv ($c > 1/4$), so ist es wahrscheinlicher, dass X und Y in die gleiche Richtung von ihrem jeweiligen Erwartungswert abweichen - wenn X große Werte annimmt, dann ist es wahrscheinlicher, dass auch Y große Werte annimmt.

Sind Kovarianz und Korrelation negativ (also $c < 1/4$), so ist es genau andersherum.

Multipliziert man eine oder mehrere Komponenten eines Zufallsvektors mit einer deterministischen Zahl (betrachtet man also (aX, Y) oder (aX, aY) anstelle von (X, Y)), so skaliert sich die Kovarianz entsprechend, während die Korrelation unverändert bleibt. Deshalb misst die Korrelation auch die Stärke des beschriebenen Effekts, während bei der Kovarianz nur das Vorzeichen eine Aussage über die Richtung dieses Effekts macht und der exakte Zahlwert für die hier angegebene Interpretation bedeutungslos ist.

Beispiel 1.7.6. Wir betrachten nochmals Beispiel 1.7.2. Es ist

$$\mathbb{E}W = 0 \cdot \frac{19}{30} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{30} = \frac{2}{5} \quad \text{und} \quad \mathbb{E}R = 0 \cdot \frac{21}{60} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{3}{20} = \frac{4}{5}.$$

Weiterhin ist

$$\text{Var}W = \left(\frac{2}{5}\right)^2 \frac{19}{30} + \left(\frac{3}{5}\right)^2 \frac{1}{3} + \left(\frac{8}{5}\right)^2 \frac{1}{30} = \frac{23}{75}$$

und

$$\text{Var}R = \left(\frac{4}{5}\right)^2 \frac{21}{60} + \left(\frac{1}{5}\right)^2 \frac{1}{2} + \left(\frac{6}{5}\right)^2 \frac{3}{20} = \frac{69}{150}.$$

Für die Kovarianz von R und W finden wir

$$\begin{aligned} \text{Cov}(X, Y) &= \left(0 - \frac{2}{5}\right)\left(0 - \frac{4}{5}\right)\frac{3}{20} + \left(0 - \frac{2}{5}\right)\left(1 - \frac{4}{5}\right)\frac{1}{3} + \left(0 - \frac{2}{5}\right)\left(2 - \frac{4}{5}\right)\frac{3}{20} \\ &\quad + \left(1 - \frac{2}{5}\right)\left(0 - \frac{4}{5}\right)\frac{1}{6} + \left(1 - \frac{2}{5}\right)\left(1 - \frac{4}{5}\right)\frac{1}{6} + \left(1 - \frac{2}{5}\right)\left(2 - \frac{4}{5}\right)0 \\ &\quad + \left(2 - \frac{2}{5}\right)\left(0 - \frac{4}{5}\right)\frac{1}{30} + \left(2 - \frac{2}{5}\right)\left(1 - \frac{4}{5}\right)0 + \left(2 - \frac{2}{5}\right)\left(2 - \frac{4}{5}\right)0 \\ &= \frac{8}{25} \cdot \frac{3}{20} - \frac{2}{25} \cdot \frac{1}{3} - \frac{12}{25} \cdot \frac{3}{20} - \frac{12}{25} \cdot \frac{1}{6} + \frac{3}{25} \cdot \frac{1}{6} - \frac{32}{25} \cdot \frac{1}{30} \\ &= -\frac{23}{150} \end{aligned}$$

Auch hier trifft die Intuition aus dem vorherigen Beispiel zu: Wurden von einer Sorte Kugeln *überdurchschnittlich viele* gezogen, so sind von der anderen Sorte eher *unterdurchschnittlich viele* gezogen worden. Mit anderen Worten, tendenziell weichen die Zufallsvariablen mit unterschiedlichen Vorzeichen von ihren jeweiligen Erwartungswerten ab.

Die Korrelation ist gegeben durch

$$\rho(X, Y) = -\frac{23}{150} \sqrt{\frac{75 \cdot 150}{23 \cdot 69}} \approx -0,4082.$$

Wir stellen nun einige Eigenschaften von Varianz und Kovarianz zusammen:

Proposition 1.7.7. *Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und X, Y Zufallsvariablen mit endlichen zweiten Momenten.*

(1) *Es gilt die Cauchy-Schwarzsche Ungleichung:*

$$|\mathbb{E}(XY)| \leq (\mathbb{E}(X^2)\mathbb{E}(Y^2))^{\frac{1}{2}}$$

(2) *Sind $\text{Var}X, \text{Var}Y > 0$, so ist $\rho(X, Y) \in [-1, 1]$.*

(3) *Es ist*

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

(4) *Sind X und Y unkorreliert, so ist $\text{Var}(X + Y) = \text{Var}X + \text{Var}Y$.*

(5) *Es gilt*

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

Beweis. (1) Sei $Z_t := X - tY$. Dann ist $Z_t^2 = X^2 - 2tXY + t^2Y^2$, insbesondere hat Z_t endliche zweite Momente. Es ist

$$0 \leq \mathbb{E}Z_t^2 = \mathbb{E}X^2 - 2t\mathbb{E}XY + t^2\mathbb{E}Y^2.$$

Ist $\mathbb{E}Y^2 = 0$, so folgt $Y(\omega) = 0$ für alle $\omega \in \Omega$ mit $\mathbb{P}(\{\omega\}) > 0$. Daher ist $\mathbb{E}XY = 0$ und daher gilt die Ungleichung.

Sei nun $\mathbb{E}Y^2 > 0$ vorausgesetzt. Wir wählen $t = \frac{\mathbb{E}XY}{\mathbb{E}Y^2}$ und erhalten

$$0 \leq \mathbb{E}X^2 - 2\frac{\mathbb{E}XY}{\mathbb{E}Y^2}\mathbb{E}XY + \frac{[\mathbb{E}XY]^2}{[\mathbb{E}Y^2]^2}\mathbb{E}Y^2 = \mathbb{E}X^2 - \frac{[\mathbb{E}XY]^2}{\mathbb{E}Y^2}$$

was äquivalent zur behaupteten Ungleichung ist.

(2) Folgt sofort aus (1), angewandt auf $\tilde{X} = X - \mathbb{E}X$ und $\tilde{Y} := Y - \mathbb{E}Y$.

(3) Sei $\mu = \mathbb{E}X$ und $\nu = \mathbb{E}Y$. Dann ist $\mathbb{E}(X + Y) = \mu + \nu$ und weiter

$$(X + Y - (\mu + \nu))^2 = (X - \mu)^2 + 2(X - \mu)(Y - \nu) + (Y - \nu)^2.$$

Nimmt man Erwartungswerte, so folgt die Behauptung.

(4) Folgt sofort aus (3).

(5) Wegen der Linearität des Erwartungswerts gilt

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

□

Wir haben gesehen, dass die Korrelation ein Maß für die Wechselwirkung zweier Zufallsvariablen ist, wobei Korrelation 0 (also Unkorreliertheit) die geringste mögliche Wechselwirkung repräsentiert. Noch stärker ist der Begriff der *Unabhängigkeit*.

Definition 1.7.8. Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum.

1. Zufallsvariablen X_1, \dots, X_n heißen *unabhängig*, falls für alle Teilmengen $A_1, \dots, A_n \subseteq \mathbb{R}$ stets

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdot \dots \cdot \mathbb{P}(X_n \in A_n)$$

gilt.

2. Eine Folge X_1, X_2, X_3, \dots von unendlich vielen Zufallsvariablen heißt *unabhängig*, falls X_1, \dots, X_n für alle $n \in \mathbb{N}$ unabhängig sind.
3. Endlich oder unendlich viele Zufallsvariablen X_1, X_2, \dots heißen *paarweise unabhängig*, wenn X_i und X_j für $i \neq j$ unabhängig sind.

Bemerkung 1.7.9. (a) Die Zufallsvariablen X_1, \dots, X_n sind genau dann (paarweise) unabhängig, wenn für alle Wahlen von $A_1, \dots, A_n \subseteq \mathbb{R}$ die Ereignisse $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$ (paarweise) unabhängig sind.

(b) Ereignisse A_1, \dots, A_n sind genau dann unabhängig, wenn die Zufallsvariablen $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ unabhängig sind.

(c) In Definition 1.7.8 müssen wir anders als in Definition 1.3.10 keine Teilfamilien betrachten, weil wir gewisse $A_j = \mathbb{R}$ wählen können, sodass $\{X_j \in A_j\} = \Omega$ ist.

Um zu entscheiden, ob Zufallsvariablen X_1, \dots, X_n unabhängig sind, genügt es, einelementige Mengen A_1, \dots, A_n zu betrachten. Genauer gesagt gilt:

Proposition 1.7.10. Sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und X_1, \dots, X_n Zufallsvariablen auf diesem. Es sei $\{x_j : j \in J\}$ die Menge aller Werte, die von mindestens einer der Zufallsvariablen X_1, \dots, X_n angenommen werden können. Dann sind X_1, \dots, X_n genau dann unabhängig, wenn

$$\mathbb{P}(X_1 = x_{j_1}, \dots, X_n = x_{j_n}) = \mathbb{P}(X_1 = x_{j_1}) \cdot \dots \cdot \mathbb{P}(X_n = x_{j_n}) \quad (1.4)$$

für alle Wahlen von x_{j_1}, \dots, x_{j_n} , $j_1 \in J, \dots, j_n \in J$. Mit anderen Worten, Zufallsvariablen sind genau dann unabhängig, wenn sich die Zehldichte des Vektors als Produkt der Zehldichten der Randverteilungen ergibt.

In Fall $n = 2$ kann dies anschaulich an einem Tableau wie in Beispiel 1.7.2 beschreiben werden. Die Zufallsvariablen sind genau dann unabhängig, wenn sich jeder Eintrag des Tableaus als Produkt des Wertes am Ende seiner Zeile und des Wertes am Ende seiner Spalte ergibt.

Beweis. Sind die Zufallsvariablen X_1, \dots, X_n unabhängig, so ergibt sich (1.4), indem man $A_1 = \{x_{j_1}\}, \dots, A_n = \{x_{j_n}\}$ in Definition 1.7.8 setzt.

Nun gelte umgekehrt (1.4). Wir wollen zeigen, dass X_1, \dots, X_n unabhängig sind. Seien $A_1, \dots, A_n \subseteq \mathbb{R}$. Dann gilt

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \sum_{x_1 \in A_1} \dots \sum_{x_n \in A_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

$$\begin{aligned}
&= \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \\
&= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n).
\end{aligned}$$

Also sind X_1, \dots, X_n unabhängig. \square

Beispiel 1.7.11. Die Zufallsvariablen R und W in Beispiel 1.7.2 sind nicht unabhängig. Beispielsweise ist

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

Beispiel 1.7.12. Die Zufallsvariablen X und Y in Beispiel 1.7.3 sind genau dann unabhängig, wenn $c = 1/4$ ist. In der Tat, sind X und Y unabhängig, so muss

$$\mathbb{P}(X = 0, Y = 0) = c \stackrel{!}{=} \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 0),$$

also $c = 1/4$ sein. In diesem Fall gilt aber für alle Wahlen von $i, j \in \{0, 1\}$, dass

$$\mathbb{P}(X = i, Y = j) = \frac{1}{4} = \mathbb{P}(X = i) \mathbb{P}(Y = j).$$

Lemma 1.7.13. Sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum, X, Y unabhängige Zufallsvariablen mit endlichem zweiten Moment. Dann sind X und Y unkorreliert.

Beweis. Wegen $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$ (siehe Proposition 1.7.7(5)) genügt zu zeigen, dass $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$ ist. Dies folgt aber sofort aus

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{x \in X(\Omega), y \in Y(\Omega)} xy \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xy \mathbb{P}(X = x) \mathbb{P}(Y = y) \\
&= \left(\sum_{x \in X(\Omega)} x \mathbb{P}(X = x) \right) \left(\sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y) \right) \\
&= \mathbb{E}X \mathbb{E}Y.
\end{aligned}$$

\square

Das folgende Beispiel zeigt, dass die Umkehrung von Lemma 1.7.13 nicht gilt.

Beispiel 1.7.14. Es sei $\Omega = \{1, 2, 3, 4\}$ mit $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = 2/5$ und $\mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = 1/10$. Die Zufallsvariablen X und Y seien wie folgt definiert:

ω	1	2	3	4
$X(\omega)$	1	-1	2	-2
$Y(\omega)$	-1	1	2	-2

Dann ist $\mathbb{E}X = \mathbb{E}Y = 0$ und

$$\text{Cov}(X, Y) = \mathbb{E}XY - 0 \cdot 0 = -1 \cdot \frac{2}{5} - 1 \cdot \frac{2}{5} + 4 \cdot \frac{1}{10} + 4 \cdot \frac{1}{10} = 0.$$

Allerdings ist

$$\mathbb{P}(X = 1, Y = -1) = \frac{2}{5} \neq \frac{2}{5} \cdot \frac{2}{5} = \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = -1)$$

und somit sind X und Y nicht unabhängig.

Eine wichtige Anwendung der Unabhängigkeit ist es, dass sie es manchmal erlaubt, die Verteilung von Summen unabhängiger Zufallsvariablen zu bestimmen.

Proposition 1.7.15. *Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum, X, Y unabhängige Zufallsvariablen mit Werten in $\mathbb{N} \cup \{0\}$. Dann ist*

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k).$$

Beweis. Es ist

$$\{X + Y = n\} = \{X = 0, Y = n\} \cup \{X = 1, Y = (n - 1)\} \cup \dots \cup \{X = n, Y = 0\}.$$

Wegen der Additivität von \mathbb{P} folgt

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k)$$

wobei wir im letzten Schritt die Unabhängigkeit verwendet haben. \square

Wir diskutieren zwei Beispiele:

Beispiel 1.7.16. Es seien X, Y unabhängige Zufallsvariablen mit $X \sim Pois(\lambda_1)$ und $Y \sim Pois(\lambda_2)$. Dann ist $X + Y \sim Pois(\lambda_1 + \lambda_2)$. In der Tat, folgt aus Proposition 1.7.15, dass

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

Beispiel 1.7.17. Es seien X, Y unabhängige Zufallsvariablen mit $X \sim Bin(n_1, p)$ und $Y \sim Bin(n_2, p)$. Dann ist $X + Y \sim Bin(n_1 + n_2, p)$. In der Tat folgt aus Proposition 1.7.15, dass (wir definieren $\binom{a}{b} = 0$ für $b > a$)

$$\begin{aligned} \mathbb{P}(X + Y = z) &= \sum_{k=0}^z \mathbb{P}(X = k)\mathbb{P}(Y = z - k) \\ &= \sum_{k=0}^z \binom{n_1}{k} p^k (1-p)^{n_1-k} \binom{n_2}{z-k} p^{z-k} (1-p)^{n_2-(z-k)} \\ &= p^z (1-p)^{n_1+n_2-z} \sum_{k=0}^z \binom{n_1}{k} \binom{n_2}{z-k} \\ &= \binom{n_1 + n_2}{z} p^z (1-p)^{n_1+n_2-z}. \end{aligned}$$

Hier haben wir verwendet, dass $\sum_{k=0}^z \binom{n_1}{k} \binom{n_2}{z-k} = \binom{n_1+n_2}{z}$ ist. Dies folgt z.B. daraus, dass $\binom{n_1}{k} \binom{n_2}{z-k} / \binom{n_1+n_2}{z}$ in einem hypergeometrischen Modell die Zähl-dichte der Anzahl der gezogenen weißen Kugeln ist, wenn aus n_1 weißen und n_2 schwarzen Kugeln z Stück gezogen werden.

1.8 Das Gesetz der großen Zahlen

Wir wollten Wahrscheinlichkeiten so definieren, dass *relative Häufigkeiten* gegen Wahrscheinlichkeiten konvergieren. In diesem Abschnitt werden wir dies präzisieren und zeigen, dass es uns gelungen ist.

Betrachten wir ein Zufallsexperiment, so können wir für ein Ereignis A nach Durchführung des Experiments sagen, ob A eingetreten ist, oder nicht. Dies können wir auch durch eine Zufallsvariable ausdrücken. Wir setzen $X = \mathbb{1}_A$, d.h. $X = 1$, falls A eingetreten ist, sonst ist $X = 0$.

Wenn wir dieses Experiment wiederholen, so bekommen wir eine Folge von Zufallsvariablen: X_1, X_2, X_3, \dots . Dabei ist $X_n = 1$, wenn A im n -ten Versuch eingetreten ist, sonst ist $X_n = 0$. Wenn wir das Experiment unabhängig wiederholen, sind die Zufallsvariablen X_j unabhängig. Definieren wir $S_n := X_1 + \dots + X_n$, so ist S_n die Anzahl der Versuche, in denen A eingetreten ist. Die relative Häufigkeit der Versuche, in denen A eingetreten ist, ist $\frac{1}{n}S_n$.

Die oben beschriebene Situation hatten wir bereits diskutiert (beispielsweise hatten wir gesehen, dass S_n binomialverteilt mit Parametern n und $\mathbb{P}(A)$ ist). In diesem Abschnitt zeigen wir, dass die relative Häufigkeit in der Tat gegen die Wahrscheinlichkeit eines Ereignisses konvergiert. Wir zeigen sogar einen allgemeineren Satz, das *Gesetz der großen Zahlen*.

Um auf den Beweis vorzubereiten, berechnen wir zunächst Erwartungswert und Varianz einer binomialverteilten Zufallsvariable auf andere Art und Weise.

Beispiel 1.8.1. Es seien X_1, \dots, X_n unabhängige Zufallsvariablen mit $\mathbb{P}(X_j = 1) = p$ und $\mathbb{P}(X_j = 0) = 1 - p$ für $j = 1, \dots, n$. Dann ist die Summe $S_n = X_1 + \dots + X_n$ binomialverteilt mit Parametern n und p . Es gilt

$$\mathbb{E}S_n = \mathbb{E}X_1 + \dots + \mathbb{E}X_n = n\mathbb{E}X_1 = np,$$

denn der Erwartungswert ist linear (daher das erste Gleichheitszeichen) und der Erwartungswert von X_j hängt nicht von j ab (da alle Zufallsvariablen X_j die gleiche Verteilung haben). Wegen der Unabhängigkeit ist auch die Varianz additiv (siehe Proposition 1.7.7(4) in Verbindung mit Lemma 1.7.13). Somit

$$\text{Var}S_n = \text{Var}X_1 + \dots + \text{Var}X_n = n\text{Var}X_1 = np(1 - p).$$

Das wesentliche Hilfsmittel zum Beweis des Gesetzes der großen Zahlen ist folgendes Resultat:

Satz 1.8.2. (*Tschebyscheff-Ungleichung*)

Es sei (Ω, \mathbb{P}) ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable mit endlicher Varianz. Dann gilt für $\varepsilon > 0$

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

Beweis. Wir schreiben $\mu := \mathbb{E}X$ und setzen $Y = \varepsilon^2 \mathbb{1}_{\{|X - \mu| \geq \varepsilon\}}$, d.h. $Y = \varepsilon^2$ falls $|X - \mu| \geq \varepsilon$ und $Y = 0$ sonst. Damit ist $Y \leq |X - \mu|^2$. Wegen der Monotonie des Erwartungswertes folgt

$$\varepsilon^2 \mathbb{P}(|X - \mu| \geq \varepsilon) = \varepsilon^2 \mathbb{E} \mathbb{1}_{\{|X - \mu| \geq \varepsilon\}} = \mathbb{E}Y \leq \mathbb{E}|X - \mu|^2 = \text{Var}X,$$

was äquivalent zur Behauptung ist. □

Bemerkung 1.8.3. Die Tschebyscheff-Ungleichung erlaubt es uns abzuschätzen, wie weit eine Zufallsvariable von ihrem Erwartungswert abweicht.

Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$. Wir schreiben μ für den Erwartungswert von X und σ für die Standardabweichung von X . Dann lässt sich die Wahrscheinlichkeit, dass X von μ um mehr als k Standardabweichungen abweicht (daher der Name!) wie folgt abschätzen:

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{\text{Var}X}{k^2\sigma^2} = \frac{1}{k^2}.$$

Beispiel 1.8.4. Werfen wir beispielsweise eine Münze 1000 mal, so ist die Anzahl X der “Köpfe” binomialverteilt mit Parametern $n = 1000$ und $p = \frac{1}{2}$. Demnach ist $\mathbb{E}X = np = 500$ und $\text{Var}X = np(1-p) = 250$, sodass die Standardabweichung $\sigma = \sqrt{250} \approx 15,8$ beträgt. Somit ist

$$\mathbb{P}(X \notin [450, 550]) \leq \mathbb{P}(|X - 500| \geq 3\sigma) \leq \frac{1}{9} = 0,1111.$$

Somit ist die Wahrscheinlichkeit, dass zwischen 450 und 550 Köpfe geworfen werden, über 88% .

Beispiel 1.8.5. Wie oft muss man eine faire Münze mindestens werfen, damit mit einer Wahrscheinlichkeit von mindestens 95 % die relative Häufigkeit der Köpfe um höchstens 0,01 von der Wahrscheinlichkeit $p = 0,5$ abweicht?

Wir setzen

$$X_n = \begin{cases} 1, & \text{falls Kopf im } n\text{-ten Wurf} \\ 0, & \text{falls Zahl im } n\text{-ten Wurf.} \end{cases}$$

Dann ist X_n Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $p = 0,5$. Insbesondere ist $\mathbb{E}X_n = 0,5$ und $\text{Var}(X_n) = 0,25$ für alle $n \in \mathbb{N}$. Aus der Tschebyscheff-Ungleichung folgt

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - 0,5\right| \geq 0,01\right) \leq \frac{0,25}{0,01^2 n}.$$

Um $\mathbb{P}(|1/n \sum_{k=1}^n X_k - \mu| \leq 0,01) \geq 0,95$ sicher zu stellen genügt es, n so groß zu wählen, dass der letzte Bruch kleiner als 0,05 ist, also

$$n \geq \frac{0,25}{0,01^2 \cdot 0,05} = 50.000.$$

Es reicht also, die Münze 50.000 mal zu werfen.

Satz 1.8.6. (*Schwaches Gesetz der großen Zahlen*)

Es sei X_1, X_2, \dots eine Folge unabhängiger Zufallsvariablen mit gleichem Erwartungswert μ und gleicher Varianz σ^2 , die beide als endlich vorausgesetzt werden. Dann ist für alle $\varepsilon > 0$ stets

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| \geq \varepsilon\right) = 0. \quad (1.5)$$

Beweis. Wir schreiben $M_n := \frac{1}{n}(X_1 + \dots + X_n)$. Aus der Linearität des Erwartungswerts folgt $\mathbb{E}M_n = \mu$. Wegen der Unabhängigkeit der Zufallsvariablen gilt

$$\text{Var}M_n = \frac{1}{n^2} \sum_{k=1}^n \text{Var}X_k = \frac{\sigma^2}{n}.$$

Nun folgt aus der Tschebyscheff-Ungleichung

$$\mathbb{P}(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n} \rightarrow 0$$

für $n \rightarrow \infty$. □

Bemerkung 1.8.7. (a) Bisher haben wir lediglich diskrete Wahrscheinlichkeitsräume diskutiert. Wir haben auch bereits erwähnt, dass es nicht möglich ist, einen diskreten Wahrscheinlichkeitsraum zu konstruieren, der das unendliche Wiederholen eines Experimentes modelliert. Ähnlich ist es auch nicht möglich, einen diskreten Wahrscheinlichkeitsraum zu konstruieren, auf dem eine Folge unabhängiger Zufallsvariablen definiert ist, die jeweils mindestens zwei verschiedene Werte mit positiver Wahrscheinlichkeit, annehmen.

Es ist aber möglich einen (nicht diskreten) Wahrscheinlichkeitsraum zu konstruieren, auf dem eine solche Folge definiert werden kann. Auf einem solchen Raum ist obiger Beweis korrekt.

(b) Für die Konvergenz in (1.5) sagt man M_n konvergiert *in Wahrscheinlichkeit* gegen μ . Allgemeiner sagt man eine Folge Y_n von Zufallsvariablen konvergiert in Wahrscheinlichkeit gegen Z , falls $\mathbb{P}(|Y_n - Z| \geq \varepsilon) \rightarrow 0$ für alle $\varepsilon > 0$.

Beispiel 1.8.8. Im Falle, dass $X = \mathbb{1}_A$ ein Indikator ist (dann ist $\mathbb{E}X = \mathbb{P}(A)$ und $\text{Var}X = \mathbb{P}(A)(1 - \mathbb{P}(A))$) betrachtet man eine Folge unabhängiger Zufallsvariablen X_1, X_2, \dots , die die gleiche Verteilung wie X haben. (“Das Zufallsexperiment wird unendlich oft wiederholt”). In diesem Fall ist M_n gerade die relative Häufigkeit von A . Das schwache Gesetz der großen Zahlen besagt, dass $M_n \rightarrow \mathbb{P}(A)$ in Wahrscheinlichkeit.

Ohne Beweis geben wir noch an:

Satz 1.8.9. (*Starkes Gesetz der großen Zahlen*)

Es sei X_1, X_2, \dots eine Folge unabhängiger und identisch verteilter (d.h. alle Zufallsvariablen haben die gleiche Verteilung) Zufallsvariablen. Weiter sei $\mathbb{E}X_1 = \mu$. Dann ist

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mu\right) = 1.$$

Mit anderen Worten, die Menge der ω , für die $\frac{1}{n}(X_1(\omega) + \dots + X_n(\omega))$ gegen μ konvergiert, hat Wahrscheinlichkeit 1.

Kapitel 2

Schätzer und Konfidenzintervalle

Bisher haben wir eine mathematische Theorie entwickelt, die es uns erlaubt, gewisse zufällige Phänomene zu modellieren. Zum Beispiel modellieren wir die Anzahl der radioaktiven Zerfälle in einer bestimmten Menge eines bestimmten Stoffs als Poisson-verteilt. Allerdings liefert unsere bisherige Theorie keinen Anhaltspunkt, wie der Parameter λ in der Poisson-Verteilung zu wählen ist, damit wir hiermit wirklich radioaktiven Zerfall beschreiben; genauer gesagt wird der Parameter λ wohl proportional zur Masse der beobachteten Probe sein und darüber hinaus vom radioaktiven Element, dessen Zerfall beschrieben werden soll, abhängen. In der Praxis wird man daher häufig den Parameter aus gewissen Beobachtungen “schätzen”. Hier ist ein weiteres Beispiel:

Beispiel 2.0.1. Betrachten wir wieder das Werfen einer Reißzwecke aus Beispiel 1.2.1(b). Wir hatten dort als Grundraum den Raum $\Omega = \{F, S\}$ gewählt, wobei F das Ergebnis “Die Reißzwecke landet auf der flachen Seite (mit der Spitze senkrecht nach oben)” und S das Ergebnis “Die Reißzwecke landet mit der Spitze schräg nach unten” bezeichnet. Wir haben ein mathematisches Modell für das Werfen einer Reißzwecke, sobald wir ein Wahrscheinlichkeitsmaß auf Ω spezifizieren. Wir müssen also $p = \mathbb{P}(\{F\})$ bestimmen (denn damit ist auch $\mathbb{P}(\{S\}) = 1 - p$ eindeutig festgelegt).

Um einen guten Wert für p zu wählen, kann man wie folgt vorgehen: Man wirft eine Reißzwecke “oft” (z.B. 1000 mal), zählt wie oft sie auf der flachen Seite landet (z.B. k mal) und nimmt dann $k/1000$ als Näherungswert für p .

Dies ist ein typisches Beispiel eines *Schätzproblems*: Man möchte einen Näherungswert für einen Parameter einer Verteilung bestimmen. In diesem Kapitel werden wir solche Probleme genauer untersuchen. Insbesondere wollen wir allgemeine Prinzipien zur Konstruktion von Schätzern kennenlernen und die Güte einiger Schätzer beurteilen.

2.1 Zufallsstichproben und Schätzer

Die Beobachtungen, derer man sich in der Statistik bedient, werden mathematisch wie folgt modelliert.

Definition 2.1.1. Es sei P eine Wahrscheinlichkeitsverteilung auf einer abzählbaren Menge M . Es seien X_1, \dots, X_n unabhängige Zufallsvariablen mit Verteilung P . Dann nennt man X_1, \dots, X_n eine *Zufallsstichprobe* vom Umfang n zur Verteilung P . Die Werte $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ heißen *Realisierung der Zufallsstichprobe*. Die Menge M^n aller potentiell möglichen Realisierungen einer Stichprobe nennt man *Stichprobenraum*.

Ein Wort zur Verwendung des Symbols P für eine Verteilung. Da jede Verteilung ein Wahrscheinlichkeitsmaß ist, verwenden einige Autoren das gleiche Symbol für Wahrscheinlichkeitsmaße und Verteilungen. Allerdings ist für eine Zufallsvariable X ihre Verteilung, die wir auch mit \mathbb{P}_X bezeichnet hatten, in der Regel nicht gleich dem Wahrscheinlichkeitsmaß \mathbb{P} des Wahrscheinlichkeitsraums, auf dem X definiert ist. Deshalb verwenden wir mit P ein ähnliches, nicht jedoch das gleiche Symbol.

Typischerweise ist die Verteilung P nicht (vollständig) bekannt und es sollen aus einer Realisierung der Zufallsstichprobe Rückschlüsse auf die Verteilung gezogen werden. Als ersten Schritt kann man einige Kenngrößen der Stichprobe berechnen.

Definition 2.1.2. Es sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung P . Dann heißt

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$$

das Stichprobenmittel und

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

die Stichprobenvarianz. Weiter heißt $S := \sqrt{S^2}$ Stichprobenstandardabweichung. Die konkreten, auf einer Realisierung der Stichprobe basierenden, Werte werden häufig mit kleinen Buchstaben bezeichnet:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k \quad \text{und} \quad s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Beispiel 2.1.3. Der Besitzer eines Restaurants zählt jeden Tag die Anzahl seiner Gäste. Er erhält folgende Tabelle:

Tag	1	2	3	4
Anzahl der Gäste	44	37	49	52

In diesem Fall sind weder die Verteilung P (von der wir zumindest unterstellen, dass es sie gibt) noch die Zufallsstichprobe X_1, \dots, X_4 konkret bekannt. Die obige Tabelle gibt lediglich eine Realisierung x_1, \dots, x_4 der Stichprobe an. Für diese Realisierung haben wir

$$\bar{x} = \frac{1}{4} (44 + 37 + 49 + 52) = 45,5$$

und

$$s^2 = \frac{1}{3} (1,5^2 + 8,5^2 + 3,5^2 + 6,5^2) = \frac{516}{12} = 43$$

was einer Standardabweichung von etwa 6,56 entspricht.

Lemma 2.1.4. Es sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung P . Wir nehmen an, dass $\mu = \mathbb{E}X_1$ und $\sigma^2 = \text{Var}X_1$ existieren. Dann gilt

- (1) $\mathbb{E}\bar{X} = \mu$.
- (2) $\text{Var}\bar{X} = \sigma^2/n$
- (3) $\mathbb{E}S^2 = \sigma^2$.

Beweis. (1) Es ist

$$\mathbb{E}\bar{X} = \frac{1}{n} \sum_{k=1}^n \mathbb{E}X_k = \frac{1}{n} \sum_{k=1}^n \mu = \mu.$$

(2) Aufgrund der Unabhängigkeit der X_1, \dots, X_n gilt

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

(3) Wir haben

$$\begin{aligned} \mathbb{E}S^2 &= \frac{1}{n-1} \sum_{k=1}^n \mathbb{E}(X_k - \bar{X})^2 \\ &= \frac{1}{n-1} \sum_{k=1}^n \mathbb{E}((X_k - \mu)^2 + 2(X_k - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2) \\ &= \frac{1}{n-1} \left(n\sigma^2 + \frac{2}{n} \sum_{k=1}^n \sum_{j=1}^n \mathbb{E}(X_k - \mu)(\mu - X_j) + \sigma^2 \right) \\ &= \frac{1}{n-1} \left(n\sigma^2 - \frac{2}{n} \sum_{k=1}^n \mathbb{E}(X_k - \mu)^2 + \sigma^2 \right) \\ &= \frac{1}{n-1} (n\sigma^2 - 2\sigma^2 + \sigma^2) \\ &= \sigma^2. \end{aligned}$$

Hier haben wir in der dritten Gleichheit die Definition von $\text{Var}X$ eingesetzt und verwendet, dass die Varianz von $\bar{X} = \sigma^2/n$ ist. In der vierten Gleichheit haben wir verwendet, dass $(X_k - \mu)$ und $(X_j - \mu)$ für $k \neq j$ unabhängig, also unkorreliert, sind. \square

2.2 Parametrische Modelle

Wir diskutieren nun sogenannte *parametrische Modelle*. Hierbei ist die Verteilung P , die wir näher untersuchen wollen, zwar unbekannt, es ist aber bekannt (oder wir nehmen es zumindest an), dass P zu einer bestimmten Familie $\{P_\theta : \theta \in \Theta\}$ von Verteilungen gehört, wobei die Parametermenge Θ für gewöhnlich eine geeignete Teilmenge von \mathbb{R}^d ist.

Beispiel 2.2.1. (a) Wenn wir annehmen, dass eine gewisse Größe Poisson-verteilt ist, so kann man $P_\theta = \text{Pois}(\theta)$ für $\theta \in \Theta = (0, \infty) \subset \mathbb{R}$ betrachten.

(b) Im Falle des Restaurant Besitzers aus Beispiel 2.1.3 ist es schwierig eine geeignete Familie von Verteilungen zu finden. Angesichts fehlender Alternativen scheint die Annahme, dass die zugrunde liegende Verteilung eine Binomialverteilung ist, vertretbar: Es gibt eine gewisse Anzahl n potentieller Gäste, die unabhängig voneinander jeden Tag entscheiden, ob sie essen gehen oder nicht. Jeder Gast besucht an jedem Tag mit Wahrscheinlichkeit p das Restaurant. Wir können also die Familie $\{\text{Bin}(n, p) \mid (n, p) \in \Theta\}$ verwenden, wobei $\Theta = \mathbb{N} \times [0, 1]$ ist.

In dieser Situation will man nun entweder den Parameter θ oder die Zahl $g(\theta)$ für eine Funktion $g : \Theta \rightarrow \mathbb{R}^l$ schätzen. In Beispiel 2.2.1(b) könnte zum Beispiel der Erwartungswert

$g_1(n, p) = np$ oder auch $g_2(n, p) = p$ interessieren. In der allgemeinen Theorie betrachten wir nur die Schätzung von $g(\theta)$, denn die Schätzung von θ ist hierin als der Spezialfall, dass g die Identität ($g(\theta) = \theta$ für alle $\theta \in \Theta$) ist, enthalten.

Definition 2.2.2. Es sei eine Zufallsstichprobe X_1, \dots, X_n zu einer Verteilung $P \in \{P_\theta : \theta \in \Theta\}$ gegeben, wobei $\Theta \subset \mathbb{R}^d$. Weiter sei $g : \Theta \rightarrow \mathbb{R}^l$ eine Funktion. Sei schließlich eine Abbildung $T : M^n \rightarrow \mathbb{R}^l$, also vom Stichprobenraum M^n in den Raum \mathbb{R}^l , der den Wertebereich von g umfasst, gegeben. Dann heißt $T(X_1, \dots, X_n)$ *Schätzer* für $g(\theta)$.

Häufig unterscheiden wir nicht zwischen der Funktion $T : M^n \rightarrow \mathbb{R}^m$ und dem Schätzer $T(X_1, \dots, X_n)$.

Wir geben einige Beispiele von Schätzern.

Beispiel 2.2.3. Im Beispiel 2.0.1 hatten wir das Problem betrachtet, den Parameter $p \in [0, 1]$ einer Bernoulli-Verteilung, also einer Binomialverteilung $Bin(1, p)$ mit $n = 1$, zu schätzen. Dazu sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung $Bin(1, p)$ (wobei wir in Beispiel 2.0.1 den Wert $X_j = 1$ mit dem Elementarereignis P , und den Wert $X_j = 0$ mit dem Elementarereignis S identifizieren wollen).

Wir hatten in Beispiel 2.0.1 bereits den Schätzer T_1 , gegeben durch $T_1(X_1, \dots, X_n) = \bar{X}$, betrachtet. Unsere Definition lässt aber auch andere Schätzer zu. Zum Beispiel:

$$\begin{aligned} T_2(X_1, \dots, X_n) &= \frac{2}{5} \\ T_3(X_1, \dots, X_n) &= \frac{1}{2}(X_1 + X_n) \end{aligned}$$

Das Beispiel zeigt, dass wir weitere Kriterien benötigen, um die Güte eines Schätzers zu beurteilen. Es scheint klar, dass T_1 der "beste" Schätzer für p ist. Der Schätzer T_2 berücksichtigt die Stichprobe überhaupt nicht. Der Schätzer T_3 berücksichtigt zwar die Stichprobe, jedoch nicht alle verfügbaren Informationen.

Um weitere Eigenschaften eines Schätzers zu definieren führen wir folgende Notation ein. Gegeben eine parametrisierte Familie $\{P_\theta : \theta \in \Theta\}$ und eine Zufallsstichprobe X_1, \dots, X_n schreiben wir $\mathbb{P}_\theta(A)$ respektive $\mathbb{E}_\theta Y$ für die Wahrscheinlichkeit des Ereignisses A respektive den Erwartungswert der Zufallsvariablen Y unter der Annahme, dass die zugrunde liegende Verteilung der X_j gerade P_θ ist.

Definition 2.2.4. Es seien $X_1, \dots, X_n, X_{n+1}, \dots$ unabhängige Zufallsvariablen mit $X_k \sim P \in \{P_\theta : \theta \in \Theta\}$. Weiter sei $T_n := T_n(X_1, \dots, X_n)$ eine Folge von Schätzern für $g(\theta) \in \mathbb{R}$.

- (a) Der Schätzer T_n heißt *erwartungstreu*, falls $\mathbb{E}_\theta T_n = g(\theta)$ für alle $\theta \in \Theta$.
- (b) Die Folge T_n heißt *asymptotisch erwartungstreu*, falls für alle $\theta \in \Theta$ stets $\mathbb{E}_\theta T_n \rightarrow g(\theta)$ für $n \rightarrow \infty$.
- (c) Die Folge T_n heißt *schwach konsistent*, falls $T_n \rightarrow g(\theta)$ in θ -Wahrscheinlichkeit für $n \rightarrow \infty$ für alle $\theta \in \Theta$, d.h. für alle $\epsilon > 0$ gilt

$$\mathbb{P}_\theta(|T_n - g(\theta)| \geq \epsilon) \rightarrow 0.$$

Beispiel 2.2.5. Wir betrachten wieder die Schätzer T_1, T_2 und T_3 für $\theta = p$ aus Beispiel 2.2.3. Dann sind T_1 und T_3 erwartungstreu, T_2 jedoch nicht. Es gilt nämlich

$$\mathbb{E}_p T_1(X_1, \dots, X_n) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}_p X_k = p$$

$$\begin{aligned}\mathbb{E}_p T_2(X_1, \dots, X_n) &= \frac{2}{5} \neq p \quad \text{für alle } p \in [0, 1] \setminus \{\frac{2}{5}\} \\ \mathbb{E}_p T_1(X_1, \dots, X_n) &= \frac{1}{2}(\mathbb{E}_p X_1 + \mathbb{E}_p X_k) = \frac{1}{2}(p + p) = p\end{aligned}$$

Die Folge der Schätzer T_1 ist schwach konsistent, da

$$T_1(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} \mathbb{E} X_1 = p$$

in p -Wahrscheinlichkeit nach dem schwachen Gesetz der großen Zahlen. Beachten wir, dass T_3 die Werte 0 , $\frac{1}{2}$ und 1 mit Wahrscheinlichkeit $\frac{1}{4}$, $\frac{1}{2}$ bzw. $\frac{1}{4}$ annimmt, so erhalten wir für $p = \frac{1}{2}$ und $\varepsilon = 1/4$

$$\mathbb{P}_{\frac{1}{2}}(|T_3 - \frac{1}{2}| \geq 1/4) = \mathbb{P}_{\frac{1}{2}}(X_1 = 0, X_n = 0 \text{ oder } X_1 = 1, X_n = 1) = \frac{1}{2} \neq 0$$

sodass T_3 nicht schwach konsistent ist. Ebenso ist T_2 nicht schwach konsistent.

Wir stellen nun zwei allgemeine Verfahren zur Konstruktion von Schätzern vor.

Momentenmethode

Es sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung $P \in \{P_\theta : \theta \in \Theta\}$. Für $r \in \mathbb{N}$ ist $m_r(\theta) := \mathbb{E}_\theta X_1^r$ das r -te Moment der Verteilung P_θ . Wir können auch die empirischen r -ten Momente $\hat{m}_r := \frac{1}{n} \sum_{k=1}^n X_k^r$ betrachten. Es folgt aus dem schwachen Gesetz der großen Zahlen, dass $\hat{m}_r \rightarrow m_r(\theta)$ in θ -Wahrscheinlichkeit für $n \rightarrow \infty$.

Ist man nun an einem Parameter $\vartheta = g(\theta) \in \mathbb{R}^l$ interessiert, der sich als Funktion der ersten l Momente ausdrücken lässt, etwa $\vartheta = f(m_1(\theta), \dots, m_l(\theta))$, so liegt es nahe, als Schätzer für ϑ gerade

$$\hat{\vartheta} = f(\hat{m}_1, \dots, \hat{m}_l)$$

zu verwenden.

Beispiel 2.2.6. Interessiert man sich für den Erwartungswert einer Verteilung, so liefert die Momentenmethode den Schätzer \bar{X} .

Nun wollen wir die Varianz schätzen. Es gilt $\text{Var} X = f(\mathbb{E} X, \mathbb{E} X^2)$ mit $f(x_1, x_2) = x_2 - x_1^2$. Daher erhält man mit der Momentenmethode als Schätzer für die Varianz

$$\begin{aligned}f\left(\frac{1}{n} \sum_{k=1}^n X_k, \frac{1}{n} \sum_{k=1}^n X_k^2\right) &= \frac{1}{n} \sum_{k=1}^n X_k^2 - \left(\frac{1}{n} \sum_{k=1}^n X_k\right)^2 \\ &= \frac{1}{n} \sum_{k=1}^n X_k^2 - \frac{1}{n} \sum_{k=1}^n X_k \bar{X} \\ &= \frac{1}{n} \sum_{k=1}^n X_k (X_k - \bar{X}) \\ &= \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 = \frac{n-1}{n} S^2.\end{aligned}$$

Hier haben wir verwendet, dass $\sum_{k=1}^n \bar{X}(X_k - \bar{X}) = \bar{X} \cdot (\sum_{k=1}^n X_k - n\bar{X}) = 0$ ist. Beachte, dass dieser Schätzer *nicht* erwartungstreu ist, denn aus Lemma 2.1.4 folgt $\mathbb{E} \frac{n-1}{n} S_n^2 = \frac{n-1}{n} \sigma^2$.

Wir geben noch einige Beispiele in denen der Parameter θ selbst geschätzt werden soll:

Beispiel 2.2.7. Ist $P_\lambda = \text{Pois}(\lambda)$, so ist $\lambda = m_1(\lambda)$. Somit liefert die Momentenmethode $\hat{\lambda} = \bar{X}$.

Beispiel 2.2.8. (Taxiproblem)

In einer großen Stadt gibt es N Taxis, die – gut zu erkennen – außen die Nummern $1, \dots, N$ tragen. Es stellt sich die Frage, wie man N schätzen kann, wenn man die Nummern x_1, \dots, x_n von n vorbeifahrenden Taxis notiert.

Hierzu sei U_N die Gleichverteilung auf den Zahlen $1, 2, \dots, N$, also $U_N(\{k\}) = N^{-1}$ für $k = 1, \dots, N$. Ist $X \sim U_N$, so ist

$$\mathbb{E}X = \sum_{k=1}^N k \frac{1}{N} = \frac{1}{N} \frac{N(N+1)}{2} = \frac{N+1}{2},$$

also $N = 2m_1(N) - 1$. Somit liefert die Momentenmethode als Schätzer für den Parameter N gerade $\hat{N} = 2\bar{X} - 1$.

Hat man also die Nummern 242, 681, 44 und 512 notiert, so liefert die Momentenmethode $\hat{N} = 2 \cdot 369,75 - 1 = 738,5$. Beachte, dass es passieren kann, dass \hat{N} kleiner als die größte beobachtete Zahl ist (etwa wenn man die Taxis mit den Nummern 22, 4 und 121 beobachtet; dann ist $\hat{N} = 97$).

Maximum-Likelihood-Methode

Die grundlegende Idee bei der Konstruktion von Schätzern mit der Maximum-Likelihood-Methode ist folgende:

Die beste Schätzung für einen Parameter θ ist diejenige,
bei der die beobachtete Stichprobe die höchste Wahrscheinlichkeit hat.

Formal geht man wie folgt vor:

Definition 2.2.9. Es sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung $P \in \{P_\theta : \theta \in \Theta\}$. Ferner sei $f(x, \theta) = P_\theta(\{x\})$ die zugehörige Zähldichte. Dann heißt $L : \mathbb{R}^n \times \Theta \rightarrow [0, 1]$, gegeben durch

$$L(x_1, \dots, x_n; \theta) = f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta)$$

die zugehörige *Likelihood-Funktion*. Es sei nun $\hat{\theta} : \mathbb{R}^n \rightarrow \Theta$ eine Funktion mit

$$L(x_1, \dots, x_n; \theta) \leq L(x_1, \dots, x_n; \hat{\theta}(x_1, \dots, x_n))$$

für alle x_1, \dots, x_n und alle $\theta \in \Theta$. Dann heißt $\hat{\theta}(X_1, \dots, X_n)$ *Maximum-Likelihood-Schätzer* für θ .

Bemerkung 2.2.10. (a) Weder existiert ein Maximum-Likelihood-Schätzer immer, noch ist er eindeutig bestimmt. In vielen Beispielen gibt es jedoch einen eindeutigen Maximum-Likelihood-Schätzer und in der Regel handelt es sich hierbei auch um einen “guten” Schätzer.

(b) Oft ist es einfacher statt der Likelihood-Funktion L die sogenannte log-Likelihood-Funktion $\log L$ zu maximieren. Wegen der Monotonie von Logarithmus und Exponentialfunktion nimmt diese an der selben Stelle wie L ihre Maxima an.

Um Maximum-Likelihood-Schätzer zu bestimmen, muss man also Funktionen maximieren. Die folgende Proposition hilft dabei:

Proposition 2.2.11. Sei $f : (a, b) \rightarrow \mathbb{R}$ eine stetige Funktion mit $-\infty \leq a < b \leq \infty$. Falls es eine Stelle $x_0 \in (a, b)$ gibt mit

$$f(x_0) \geq \max \left\{ \lim_{x \rightarrow a} f(x), \lim_{x \rightarrow b} f(x) \right\},$$

dann hat f eine Maximalstelle in $x_1 \in (a, b)$.

Die Voraussetzungen der Proposition sind insbesondere dann erfüllt, wenn

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow b} f(x) = -\infty.$$

Sind die Voraussetzungen der Proposition erfüllt, ist f darüber hinaus differenzierbar und hat die Ableitung von f nur eine Nullstelle, dann muss diese Nullstelle also bereits die Maximalstelle von f sein.

Beispiel 2.2.12. Wir bestimmen zunächst einen Maximum-Likelihood-Schätzer für die Erfolgswahrscheinlichkeit in einem Bernoulli-Experiment. In diesem Fall ist der Parameter $\theta = p \in [0, 1]$ und für $x \in \{0, 1\}$ ist $f(x, \theta) = p^x(1-p)^{1-x}$. Es ergibt sich für die Likelihood-Funktion

$$L(x_1, \dots, x_n, p) = \prod_{k=1}^n p^{x_k} (1-p)^{1-x_k}.$$

Hier ist es vorteilhaft, zur log-Likelihood-Funktion überzugehen. Wir erhalten

$$\log L(x_1, \dots, x_n, p) = \sum_{k=1}^n x_k \log(p) + (1-x_k) \log(1-p) = n\bar{x} \log(p) + n(1-\bar{x}) \log(1-p).$$

Um das Maximum zu bestimmen berechnen wir die Nullstellen der Ableitung (Beachte, $\log L$ hat nach Proposition 2.2.11 ein Maximum, denn die Grenzwerte bei 0 und 1 sind jeweils $-\infty$).

Es ist

$$0 \stackrel{!}{=} \frac{d}{dp} \log L(x_1, \dots, x_n, p) = \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p} \Leftrightarrow p = \bar{x}.$$

Demnach ist der Maximum-Likelihood-Schätzer $\hat{p} = \bar{x}$. In diesem Fall stimmt also der Maximum-Likelihood-Schätzer mit dem Schätzer, den man aus der Momentenmethode erhält, überein.

Beispiel 2.2.13. Wir bestimmen einen Maximum-Likelihood-Schätzer für die Verteilungen $\{Pois(\lambda) : \lambda > 0\}$. In diesem Falle ist

$$L(x_1, \dots, x_n, \lambda) = e^{-n\lambda} \frac{\lambda^{\sum_{k=1}^n x_k}}{x_1! \cdot \dots \cdot x_n!}$$

und daher $\log L = n\bar{x} \log \lambda - n\lambda - \log(x_1! \cdot \dots \cdot x_n!)$. Durch Differenzieren und Nullsetzen erhält man dass als Kandidaten für das Maximum $\lambda = \bar{x}$. Beachten wir noch dass

$$\lim_{\lambda \rightarrow 0} \log L(x_1, \dots, x_n, \lambda) = -\infty = \lim_{\lambda \rightarrow \infty} \log L(x_1, \dots, x_n, \lambda)$$

so ergibt sich nach Proposition 2.2.11, dass es sich hierbei in der Tat um die Maximumstelle handeln muss. Also ist Maximum-Likelihood-Schätzer für λ durch $\hat{\lambda} = \bar{x}$ gegeben.

2.3 Konfidenzintervalle

Interessieren wir uns für einen Parameter einer Verteilung, so liefert ein Schätzer eine einzige Zahl, die den uns unbekanntem Parameter approximieren soll. Allerdings ist dieser eine Wert für sich alleine genommen nicht aussagekräftig.

Beispiel 2.3.1. (Qualitätskontrolle)

Eine Maschine produziert Bauteile, die manchmal defekt sind. Der zuständige Werkstatt-leiter will überprüfen, wie hoch der Anteil der defekten Teile am Output einer Maschine ist. Hierzu nimmt er an, dass die Bauteile unabhängig voneinander defekt oder funktionsfähig sind. Nun entnimmt er eine Stichprobe von 200 Bauteilen und testet diese. Dabei stellt er fest, dass 3 hiervon defekt sind. Er vermutet daher, dass 1,5% der Bauteile defekt sind.

In obigem Beispiel war die Erfolgswahrscheinlichkeit p in einem Bernoulli-Experiment zu schätzen und wir haben hierzu den Schätzer \bar{X} verwendet. Allerdings können wir nicht mit Sicherheit sagen, wie hoch der Anteil defekter Bauteile ist. So kann dieses Experiment niemals einen Schätzwert zwischen 1,5% und 2% liefern, auch wenn der wahre Wert natürlich dazwischen liegen kann. Theoretisch wäre es auch denkbar, dass die Wahrscheinlichkeit, dass ein Bauteil defekt ist, 50% beträgt und nur zufällig so wenige defekte Bauteile in der Stichprobe sind. Die ist aber wenig realistisch, denn nach dem Gesetz der Großen Zahlen ist die Differenz $|\bar{X} - p|$ mit großer Wahrscheinlichkeit klein, wenn n hinreichend groß ist.

Deshalb wollen wir nun statt einer einzigen Zahl \bar{X} ein Intervall $I = I(X_1, \dots, X_n)$ angeben, in dem der wahre Parameter mit hoher Wahrscheinlichkeit liegt.

Definition 2.3.2. Es sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung $P \in \{P_\theta, \theta \in \Theta\}$. Weiter sei $g : \Theta \rightarrow \mathbb{R}$ und $\alpha \in (0, 1)$. Ein *Konfidenzintervall zum Konfidenzniveau α* oder *α -Konfidenzintervall* für $g(\theta)$ ist ein zufälliges Intervall $I = [a(X_1, \dots, X_n), b(X_1, \dots, X_n)]$, wobei $a, b : M^n \rightarrow \mathbb{R}$ mit $a(x_1, \dots, x_n) \leq b(x_1, \dots, x_n)$ für alle $(x_1, \dots, x_n) \in M^n$, falls

$$\mathbb{P}_\theta(g(\theta) \in I) \geq \alpha$$

für alle $\theta \in \Theta$ gilt.

Bemerkung 2.3.3. Beachte: Das Konfidenzintervall ist zufällig, nicht der Parameter θ .

Dass ein zufälliges Intervall ein α -Konfidenzintervall ist, bedeutet, dass $g(\theta)$ mit Wahrscheinlichkeit größer α im Intervall liegt, wenn die "wahre Verteilung" Parameter θ hat.

Man kann Konfidenzintervalle beispielsweise mit der Tschebyscheffschen Ungleichung bestimmen. Wenn wir etwa eine Zufallsstichprobe zur Verteilung P betrachten, die Erwartungswert μ und Varianz σ^2 hat, und verwenden wir \bar{X} als Schätzer für μ , so ist nach Lemma 2.1.4

$$\mathbb{E}\bar{X} = \mu \quad \text{und} \quad \text{Var}\bar{X} = \frac{\sigma^2}{n}.$$

Nach der Tschebyscheffschen Ungleichung gilt

$$\mathbb{P}(|\bar{X} - \mu| \geq \delta) \leq \frac{\sigma^2}{n\delta^2}.$$

Ist also σ^2 bekannt (oder zumindest beschränkt) und wählen wir $\delta = \sqrt{\frac{\sigma^2}{n(1-\alpha)}}$, so ist

$$[\bar{X} - \delta, \bar{X} + \delta]$$

ein Konfidenzintervall zum Konfidenzniveau α , denn

$$\mathbb{P}(\mu \in [\bar{X} - \delta, \bar{X} + \delta]) \geq \mathbb{P}(\mu \in (\bar{X} - \delta, \bar{X} + \delta)) \geq 1 - \frac{\sigma^2}{(1-\alpha)} = \alpha.$$

Beispiel 2.3.4. In Beispiel 2.3.1 bestimmen wir ein 95%-Konfidenzintervall für die Bernoulli-wahrscheinlichkeit p . Beachte, dass hier $\sigma^2 = p(1-p) = 1/4 - (p - 1/2)^2 \leq 1/4$. Weiter ist der Stichprobenumfang $n = 200$. Somit können wir

$$\delta = \sqrt{\frac{1}{4 \cdot 200 \cdot 0,05}} \approx 0,1581$$

wählen. Beachte, dass dies im Vergleich zum geschätzten Wert ($\bar{x} = 0,015$) relativ groß ist. Wir erhalten also das (relativ lange) Konfidenzintervall $[-0.143, 0.173]$ für $p = \mu$. Da $\mu \in [0, 1]$, folgt

$$\mathbb{P}(\mu \in [-0.143, 0.173]) = \mathbb{P}(\mu \in [0, 0.173])$$

und $[0, 0.173]$ ist ebenfalls 95%-Konfidenzintervall. Will man das Konfidenzintervall weiter verkleinern, so kann man z.B. den Stichprobenumfang erhöhen.

Der Grund, warum mit der Tschebyscheff Ungleichung relativ lange Konfidenzintervalle entstehen, liegt in der Allgemeinheit der Ungleichung. Wenn man spezielle Eigenschaften der Verteilung des Schätzers berücksichtigt, so erhält man kürzere Konfidenzintervalle. Wir kommen später darauf zurück.

Kapitel 3

Allgemeine Wahrscheinlichkeitsräume

3.1 Einleitung

Wir hatten schon bemerkt, dass der Begriff des diskreten Wahrscheinlichkeitsraums nicht ausreicht, um das unendliche Wiederholen eines Zufallsexperiments zu modellieren. Der Grund dafür ist, dass die Menge aller möglichen Ausgänge nicht mehr abzählbar ist. Ein weiterer Unterschied zu diskreten Wahrscheinlichkeitsräumen ist, dass es nicht mehr notwendigerweise Elementarereignisse positiver Wahrscheinlichkeit geben muss. Wir diskutieren dies an einem Beispiel.

Beispiel 3.1.1. Werfen wir unendlich oft eine faire Münze, so bietet sich als Grundraum der Raum

$$\Omega := \{\omega = (\omega_1, \omega_2, \omega_3, \dots) : \omega_n \in \{K, Z\} \forall n \in \mathbb{N}\}$$

aller Folgen in $\{K, Z\}$ an. Das Elementarereignis $\omega = (\omega_1, \omega_2, \dots)$ bedeutet hierbei gerade, dass im ersten Wurf ω_1 , im zweiten Wurf ω_2 usw. geworfen wurde. Beachte, dass Ω nicht abzählbar ist.

Es sei

$$A_1 = \{\omega \in \Omega : \omega_1 = K\} \subseteq \mathcal{P}(\Omega).$$

Dann bezeichnet A das Ereignis “Im ersten Wurf Kopf” und sollte gerade Wahrscheinlichkeit $\frac{1}{2}$ haben. Ist allgemeiner

$$A_n = \{\omega \in \Omega : \omega_1 = \omega_2 = \dots = \omega_n = K\}$$

das Ereignis, dass in den ersten n Würfeln Kopf gefallen ist, so sollte A_n Wahrscheinlichkeit 2^{-n} haben.

Betrachten wir nun das Elementarereignis $A_\infty := \{(K, K, K, \dots)\}$, dass in allen Würfeln Kopf fällt, so ist $A_\infty \subseteq A_n$ für alle $n \in \mathbb{N}$. Aufgrund der Monotonie der Wahrscheinlichkeit sollte $\mathbb{P}(A_\infty) \leq \mathbb{P}(A_n) = 2^{-n}$ für alle $n \in \mathbb{N}$ gelten. Es folgt also $\mathbb{P}(A_\infty) = 0$.

Analog folgt, dass alle Elementarereignisse Wahrscheinlichkeit 0 haben müssen.

Ein weiteres Beispiel ergibt sich bei Zufallsvariablen die stetige Merkmale, wie Temperatur oder Zeit, beschreiben. Hier möchte man auf, dass die Zufallsvariablen keinen Wert mit positiver Wahrscheinlichkeit annimmt. In einem Wahrscheinlichkeitsraum, auf dem eine solche Zufallsvariable definiert ist, müssen alle Elementarereignisse Wahrscheinlichkeit 0 haben.

Ein weiterer Unterschied zwischen allgemeinen und diskreten Wahrscheinlichkeitsräumen betrifft die Additivität: Haben alle Elementarereignisse Wahrscheinlichkeit 0, so muss diese eingeschränkt werden, damit

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = 0$$

nicht für alle Ereignisse $A \subseteq \Omega$ gilt.

Es bleibt die Frage, ob es in dieser Situation stets ein Wahrscheinlichkeitsmaß (im Sinne einer σ -additiven Abbildung von $\mathcal{P}(\Omega)$ nach $[0, 1]$) gibt. Leider ist dies nicht immer der Fall. Es zeigt sich, dass die Potenzmenge $\mathcal{P}(\Omega)$ in der Regel zu groß ist.

Deshalb werden wir nicht mehr jede Teilmenge von Ω als Ereignis betrachten. Wir nehmen aber an, dass die Menge aller Ereignisse eine sogenannte σ -Algebra bildet.

Definition 3.1.2. Es sei Ω eine Menge. Eine σ -Algebra auf Ω ist ein Mengensystem $\Sigma \subseteq \mathcal{P}(\Omega)$, sodass

- (i) $\emptyset \in \Sigma$.
- (ii) Ist $A \in \Sigma$, so ist auch $A^c = \Omega \setminus A \in \Sigma$.
- (iii) Ist $A_n \in \Sigma$ für $n \in \mathbb{N}$, so ist auch $\bigcup_{n \in \mathbb{N}} A_n \in \Sigma$.

Beispiel 3.1.3. a) Offensichtlich ist $\mathcal{P}(\Omega)$ eine σ -Algebra auf Ω .

b) Eine weitere σ -Algebra auf Ω ist $\{\emptyset, \Omega\}$.

c) Ist Ω eine Menge mit mindestens zwei Elementen und $\emptyset \neq A \subseteq \Omega$ mit $A \neq \Omega$, so ist $\{\emptyset, A, A^c, \Omega\}$ eine σ -Algebra.

Bemerkung 3.1.4. Weiterhin erfüllt eine σ -Algebra Σ auf Ω auch folgende Eigenschaften:

- (1) Sind $A_1, \dots, A_k \in \Sigma$, so auch $A_1 \cup \dots \cup A_k$. Das folgt aus (iii), indem man $A_n = \emptyset$ für $n > k$ wählt.
- (2) Sind $A_1, \dots, A_k \in \Sigma$, so auch $A_1 \cap \dots \cap A_k$. Das folgt aus (1), (ii) und deMorgan's Gesetz (siehe Proposition 3.1.5) unten.
- (3) Genau so sieht man, dass Σ mit der Folge $(A_n)_{n \in \mathbb{N}}$ auch deren Durchschnitt $\bigcap_{n \in \mathbb{N}} A_n$ enthält.

Proposition 3.1.5. Seien $A_i, i \in I$, endlich oder unendlich viele Teilmengen einer Grundmenge Ω . Dann ist

$$\left(\bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c \quad \text{und} \quad \left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c.$$

In der Regel ist es schwer, eine σ -Algebra konkret anzugeben. Daher ist folgendes Resultat wichtig:

Lemma 3.1.6. Ist $S \subseteq \mathcal{P}(\Omega)$, so gibt es eine kleinste σ -Algebra auf Ω , die S enthält. Präziser: Es gibt eine σ -Algebra Σ_0 auf Ω , so dass $S \subseteq \Sigma_0$ gilt und für jede σ -Algebra Σ auf Ω mit $S \subseteq \Sigma$ auch $\Sigma_0 \subseteq \Sigma$ gilt.

Diese σ -Algebra Σ_0 bezeichnet man mit $\sigma(S)$ und nennt sie die von S erzeugte σ -Algebra.

Definition 3.1.7. Die von den offenen Intervallen in \mathbb{R} erzeugte σ -Algebra heißt *Borel σ -Algebra* und wird mit $\mathcal{B}(\mathbb{R})$ bezeichnet. Es ist also $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) : a, b \in \mathbb{R}, a < b\})$.

Es ist nicht möglich $\mathcal{B}(\mathbb{R})$ genauer zu beschreiben, aber alle praktisch relevanten Teilmengen von \mathbb{R} liegen in $\mathcal{B}(\mathbb{R})$. Es gilt $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$, aber dies ist schwer zu zeigen.

Wir diskutieren dies nicht näher und geben stattdessen Beispiele für Mengen in $\mathcal{B}(\mathbb{R})$.

Beispiel 3.1.8. Für $a \in \mathbb{R}$ ist $(-\infty, a) \in \mathcal{B}(\mathbb{R})$ und $(a, \infty) \in \mathcal{B}(\mathbb{R})$. Es ist nämlich $(-\infty, a) = \bigcup_{n \in \mathbb{N}} (a - n, a)$ und $(a, \infty) = \bigcup_{n \in \mathbb{N}} (a, a + n)$. Somit liegen für $a \in \mathbb{N}$ auch die Komplemente $[a, \infty) = (-\infty, a)^c$ und $(-\infty, a] = (a, \infty)^c$ in \mathcal{B} . Schließlich sind auch kompakte Intervalle in $\mathcal{B}(\mathbb{R})$ enthalten, denn $[a, b] = (-\infty, b] \cap [a, \infty)$.

Vereinbarung: In Übungs- und Klausuraufgaben zur Vorlesung für “Stochastik für Wirtschaftswissenschaftler” braucht für Teilmengen von \mathbb{R} nicht nachgewiesen zu werden, dass diese in der Borel- σ -Algebra liegen.

Wir kommen nun zur zentralen Definition:

Definition 3.1.9. Ein *Messraum* ist ein Paar (Ω, Σ) , bestehend aus einer Menge Ω und einer σ -Algebra Σ auf Ω . Ist (Ω, Σ) ein Messraum, so heißt eine Abbildung $\mathbb{P} : \Sigma \rightarrow [0, 1]$ *Wahrscheinlichkeitsmaß* auf (Ω, Σ) , falls

1. $\mathbb{P}(\Omega) = 1$ (Normiertheit)
2. Ist $(A_n)_{n \in \mathbb{N}}$ eine Folge paarweise disjunkter Mengen in Σ , so ist

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Ist (Ω, Σ) ein Messraum und \mathbb{P} ein Wahrscheinlichkeitsmaß darauf, so nennt man das Tripel $(\Omega, \Sigma, \mathbb{P})$ einen *Wahrscheinlichkeitsraum*.

Bemerkung 3.1.10. (a) Beachte, dass Σ mit der Folge A_n auch deren Vereinigung $\bigcup A_n$ enthält. Daher ist $\mathbb{P}(\bigcup A_n)$ in (ii) wohldefiniert.

- (b) Jeder diskrete Wahrscheinlichkeitsraum ist ein Wahrscheinlichkeitsraum mit der Potenzmenge als σ -Algebra.
- (c) Die Eigenschaften von Wahrscheinlichkeitsmaßen in Proposition 1.1.5 gelten auch in beliebigen Wahrscheinlichkeitsräumen. Der Beweis bleibt unverändert.

Auch andere Konzepte (wie beispielsweise die Begriffe “Unabhängigkeit” und “bedingte Wahrscheinlichkeit”) übertragen sich ohne Änderungen auf allgemeine Wahrscheinlichkeitsräume.

Lemma 3.1.11. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A_n eine Folge in Σ mit $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ 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 $C_1 \supseteq C_2 \supseteq \dots$ 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. Übung!

□

3.2 Zufallsvariablen und ihre Verteilungen

Gegeben einen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ sind wir versucht, wiederum jede Abbildung X von Ω nach \mathbb{R} Zufallsvariable zu nennen. Dabei gibt es jedoch folgendes Problem:

Wenn wir die Verteilung \mathbb{P}_X von X durch $\mathbb{P}_X(A) := \mathbb{P}(X \in A)$ definieren, so muss $\{X \in A\}$ in der σ -Algebra Σ liegen.

Wir definieren daher:

Definition 3.2.1. Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ heißt *Zufallsvariable*, falls für $A \in \mathcal{B}(\mathbb{R})$ stets $\{\omega \in \Omega : X(\omega) \in A\} \in \Sigma$ liegt. In diesem Fall heißt das Wahrscheinlichkeitsmaß \mathbb{P}_X auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, gegeben durch

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A),$$

die *Verteilung von X* .

Wir hatten bereits bemerkt, dass die Borel σ -Algebra $\mathcal{B}(\mathbb{R})$ sehr groß und unübersichtlich ist. Daher ist es schwierig, die Verteilung konkret anzugeben. Es ist auch nicht mehr richtig, dass Verteilungen durch ihre Werte auf einelementigen Mengen bestimmt sind. Es stellt sich also die Frage, ob die Verteilung einer Zufallsvariablen bereits durch die Werte auf gewissen Mengen eindeutig bestimmt ist.

Dies ist in der Tat der Fall. Wir definieren:

Definition 3.2.2. Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann heißt $F_X : \mathbb{R} \rightarrow [0, 1]$, definiert durch

$$F_X(x) := \mathbb{P}(X \leq x)$$

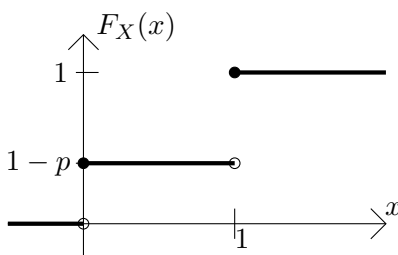
Verteilungsfunktion von X .

Proposition 3.2.3. Seien X und Y zwei Zufallsvariablen mit der selben Verteilungsfunktion, $F_X(t) = F_Y(t)$ für alle $t \in \mathbb{R}$. Dann haben X und Y auch die selbe Verteilung, $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$ für beliebige $A \in \mathcal{B}(\mathbb{R})$.

Beispiel 3.2.4. Ist $X \sim \text{Bin}(1, p)$ Bernoulli-verteilt, so ist

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases}$$

Denn für $x < 0$ ist $\{X \leq x\} = \emptyset$, also $\mathbb{P}(X \leq x) = 0$. Für $0 \leq x < 1$ ist $\{X \leq x\} = \{X = 0\}$ und daher $\mathbb{P}(X \leq x) = \mathbb{P}(X = 0) = 1 - p$. Schließlich ist für $x \geq 1$ gerade $\{X \leq x\} = \Omega$ und daher $F_X(x) = \mathbb{P}(\Omega) = 1$.



Ähnliche Überlegungen zeigen folgendes:

Beispiel 3.2.5. Ist X eine Zufallsvariable die lediglich die Werte $x_1 < x_2 < \dots < x_n$ annimmt mit $\mathbb{P}(X = x_k) = p_k$, so ist

$$F_X(x) = \begin{cases} 0 & x < x_1 \\ \sum_{j=1}^k p_j & x_k \leq x < x_{k+1} \\ 1 & x \geq x_n \end{cases}$$

Das Schaubild der Verteilungsfunktion sieht ähnlich wie oben aus, hat aber nun n Sprungstellen.

Nimmt allgemeiner die Zufallsvariable X abzählbar viele Werte $x_k, k \in \mathbb{N}$, an mit $\mathbb{P}(X = x_k) = p_k$, so ist

$$F_X(x) = \sum_{k: x_k \leq x} p_k$$

In diesem Fall sagt man X habe *diskrete Verteilung* oder manchmal X sei diskret.

Die Verteilungsfunktion F_X hat in diesem Fall unendlich viele Sprungstellen – es handelt sich jedoch weiterhin um eine Funktion, die nicht stetig wächst, sondern Sprungstellen hat und dazwischen konstant ist.

Ein Sprung in einer Verteilungsfunktion an der Stelle x in Höhe p bedeutet, dass die Zufallsvariable den Wert x mit Wahrscheinlichkeit p annimmt.

Wir stellen einige Eigenschaften von Verteilungsfunktionen zusammen:

Eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ heißt *rechtsseitig stetig*, falls für jede monoton fallende, konvergente Folge $(x_n)_{n \in \mathbb{N}}$ gilt $\lim_{n \rightarrow \infty} g(x_n) = g(\lim_{n \rightarrow \infty} x_n)$.

Lemma 3.2.6. *Es sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$. Dann ist die Verteilungsfunktion F_X monoton wachsend, rechtsseitig stetig und erfüllt*

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

Beweis. Zunächst ist F offensichtlich monoton wachsend (jedoch nicht notwendigerweise strikt). Ist nämlich $x \leq y$ so ist $\{X \leq x\} \subseteq \{X \leq y\}$ und daher, wegen der Monotonie des Wahrscheinlichkeitsmaßes, $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$.

Ist nun x_n eine fallende Folge, die gegen x konvergiert, so ist $\{X \leq x_n\} \downarrow \{X \leq x\}$. Es folgt aus Lemma 3.1.11 dass $F_X(x_n) \rightarrow F_X(x)$. Also ist F rechtsseitig stetig.

Es gilt $\lim_{x \rightarrow -\infty} F_X(x) = 0$ und $\lim_{x \rightarrow \infty} F_X(x) = 1$. Das folgt aus Lemma 3.1.11 und den Beziehungen $\{X \leq -x_n\} \downarrow \emptyset$ und $\{X \leq x_n\} \uparrow \Omega$ für jede Folge $(x_n)_{n \in \mathbb{N}}$, die gegen ∞ konvergiert. \square

Es gilt auch die Umkehrung dieses Lemmas. Da der Beweis deutlich schwieriger ist, übergehen wir ihn.

Satz 3.2.7. *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 genau ein Wahrscheinlichkeitsmaß \mathbb{P}_F auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit*

$$\mathbb{P}_F((-\infty, x]) = F(x) \tag{3.1}$$

Eine wichtige Klasse von Verteilungen sind *absolutstetige Verteilungen*.

Definition 3.2.8. 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 3.2.9. Ist die Verteilungsfunktion stetig (insbesondere, bei absolutstetigen Verteilungsfunktionen), so nimmt die Zufallsvariable keinen Wert mit positiver Wahrscheinlichkeit an. In diesem Fall gilt also

$$\mathbb{P}(X \in (a, b]) = \mathbb{P}(X \in (a, b)) = \mathbb{P}(X \in [a, b)) = \mathbb{P}(X \in [a, b]) = F(b) - F(a) = \int_a^b f(t) dt.$$

Bemerkung 3.2.10. 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.

In gewisser Weise sind die Dichten von absolutstetigen Zufallsvariablen ähnlich zu den Zähldichten von diskreten Zufallsvariablen in Beispiel 3.2.5. Ist f_X die Zähldichte der diskreten Zufallsvariablen X , so ist die Verteilungsfunktion F_X gegeben durch

$$F_X(x) = \sum_{t \leq x} f_X(t),$$

wobei zu beachten ist, dass f_X ja nur an höchstens abzählbar vielen Stellen verschieden von 0 ist. Bei absolutstetigen Zufallsvariablen hat man stattdessen eine Dichte f , die man bis x "aufintegriert" um die Verteilungsfunktion zu erhalten.

Diese Analogie verwenden wir auch bei der Definition von Erwartungswert, Varianz, etc. von absolutstetigen Zufallsvariablen.

Definition 3.2.11. Es sei X eine absolutstetige Zufallsvariable und f die Dichte der Verteilung von X . Wir sagen, X hat *endlichen Erwartungswert*, falls

$$\int_{-\infty}^{\infty} |t|f(t) dt < \infty.$$

In diesem Fall heißt

$$\mathbb{E}X := \int_{-\infty}^{\infty} tf(t) dt$$

der Erwartungswert von X .

Wie im diskreten Fall gilt für eine Zufallsvariable X mit Dichte f und eine Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$, dass

$$\mathbb{E}g(X) = \int_{-\infty}^{\infty} f(t)g(t) dt,$$

falls $g(X)$ endlichen Erwartungswert hat.

Wir sagen, X habe endliche Varianz, falls

$$\text{Var}X := \mathbb{E}[(X - \mathbb{E}X)^2] = \int_{-\infty}^{\infty} (t - \mathbb{E}X)^2 f(t) dt$$

endlich ist.

Bemerkung 3.2.12. Ähnlich wie im diskreten Fall kann man zeigen, dass X genau dann endliche Varianz hat, wenn $\mathbb{E}X^2 = \int_{-\infty}^{\infty} t^2 f(t) dt$ endlich ist. In diesem Fall ist

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

Beispiel 3.2.13. Setzen wir

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

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 Zufallsvariablen 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.

X hat *keinen* endlichen Erwartungswert, es ist nämlich

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

Gelegentlich ist es wichtig (beispielsweise beim Berechnen von Konfidenzintervallen), Zahlen x zu finden, sodass $\mathbb{P}(X \leq x) = F_X(x) = p$ für ein vorgegebenes $p \in (0, 1)$. Wir behandeln diese Frage nur in dem Spezialfall, dass es ein Intervall I gibt, so dass die Verteilungsfunktion F_X aufgefasst als Funktion $I \rightarrow (0, 1)$ bijektiv ist. Dies ist z.B. dann erfüllt, wenn die Verteilungsfunktion auf \mathbb{R} stetig und streng monoton wachsend ist; dann ist $I = \mathbb{R}$. Dann hat $x := (F_X)^{-1}(p)$ die gewünschte Eigenschaft und heißt *p-Quantil*.

3.3 Wichtige absolutstetige Verteilungen

Stetige Gleichverteilung

Wir sagen X ist *gleichverteilt auf dem Intervall* (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 $(0, 2\pi)$.

Wir berechnen Erwartungswert und Varianz einer gleichverteilten Zufallsvariable.

Lemma 3.3.1. *Ist $X \sim U(a, b)$, so ist*

$$\mathbb{E}X = \frac{a+b}{2} \quad \text{und} \quad \text{Var}X = \frac{(b-a)^2}{12}.$$

Beweis. Es ist

$$\mathbb{E}X = \frac{1}{b-a} \int_a^b t \, dt = \frac{1}{b-a} \left[\frac{t^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Weiter ist für $\mu = (a+b)/2$ gerade

$$\begin{aligned} \text{Var}X &= \frac{1}{b-a} \int_a^b (t-\mu)^2 \, dt = \frac{1}{b-a} \left[\frac{(t-\mu)^3}{3} \right]_a^b \\ &= \frac{1}{3(b-a)} \frac{1}{8} ((b-a)^3 - (a-b)^3) = \frac{(b-a)^2}{12}. \end{aligned}$$

□

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 $X \sim \text{Exp}(\lambda)$. 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 tritt bei sogenannten Wartezeitproblemen auf. Wichtige Beispiele sind: Die Lebensdauer von Glühbirnen oder die Wartezeit auf den nächsten Anruf.

Wir berechnen wiederum Erwartungswert und Varianz.

Lemma 3.3.2. *Es sei $X \sim \text{Exp}(\lambda)$. Dann ist*

$$\mathbb{E}X = \frac{1}{\lambda} \quad \text{und} \quad \text{Var}X = \frac{1}{\lambda^2}.$$

Beweis. Mit partieller Integration erhalten wir

$$\mathbb{E}X = \int_0^{\infty} t \lambda e^{-\lambda t} \, dt = \left[-te^{-\lambda t} \right]_0^{\infty} - \int_0^{\infty} -e^{-\lambda t} \, dt = 0 - \left[\frac{e^{-\lambda t}}{\lambda} \right]_0^{\infty} = \frac{1}{\lambda}.$$

Mit zweifacher partieller Integration folgt

$$\mathbb{E}X^2 = \frac{2}{\lambda^2},$$

wobei die Details Übung sind. Daher ist

$$\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

□

Die Exponentialverteilung hat eine wichtige Eigenschaft, die man *Gedächtnislosigkeit* nennt. Ist nämlich $X \sim \text{Exp}(\lambda)$, so ist

$$\mathbb{P}(X \geq a + b | X \geq a) = \mathbb{P}(X \geq b). \quad (3.2)$$

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 (3.2) 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 $A \subseteq B$ und daher $A \cap B = A$. Für $x \in (0, \infty)$ gilt

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

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 $X \sim \mathcal{N}(\mu, \sigma^2)$. Ist $\mu = 0$ und $\sigma^2 = 1$, so sagen wir, X ist *standard-normalverteilt*.

Da der Nachweis von $\int_{-\infty}^{\infty} f(t) dt = 1$ sehr schwierig ist, übergehen wir ihn.

Die Bedeutung der Normalverteilung entsteht vor allem durch den Zentralen Grenzwertsatz (den wir später behandeln), demzufolge viele Zufallsvariablen zumindest "annähernd" normalverteilt sind.

Proposition 3.3.3. *Ist $X \sim \mathcal{N}(\mu, \sigma^2)$ und sind $a, b \in \mathbb{R}$, so ist $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.*

Beweis. Es gilt

$$\begin{aligned} \mathbb{P}(aX + b \leq y) &= \mathbb{P}\left(X \leq \frac{y-b}{a}\right) \\ &= \int_{-\infty}^{(y-b)/a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{((u-b)/a-\mu)^2}{2\sigma^2}} \frac{1}{a} du \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi a^2 \sigma^2}} e^{-\frac{(u-b-a\mu)^2}{2a^2\sigma^2}} du. \end{aligned}$$

Also hat $aX + b$ die Verteilungsfunktion einer $\mathcal{N}(a\mu + b, a^2\sigma^2)$ -Verteilung. \square

Insbesondere folgt, dass $\frac{X-\mu}{\sigma}$ standardnormalverteilt ist, wenn $X \sim \mathcal{N}(\mu, \sigma^2)$. Somit kann man normalverteilte Zufallsvariablen durch Transformation immer in standardnormalverteilte Zufallsvariablen überführen. Daher erhalten Dichte und Verteilungsfunktion der Standardnormalverteilung besondere Bezeichner. Es sei

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad \text{und} \quad \Phi(x) = \int_{-\infty}^x \varphi(t) dt.$$

Man kann Φ nicht elementar ausdrücken. In Anwendungen verwendet man daher oft Tabellen mit Werten von Φ . Häufig sind dort nur Werte $\Phi(x)$ für $x \geq 0$ aufgeführt. Aus der Symmetrie von φ , d.h. $\varphi(-t) = \varphi(t)$, folgt dass

$$\Phi(-x) = \int_{-\infty}^{-x} \varphi(t) dt = \int_x^{\infty} \varphi(-t) dt = \int_{-\infty}^{\infty} \varphi(t) dt - \int_{-\infty}^x \varphi(t) dt = 1 - \Phi(x),$$

sodass diese Information ausreicht.

Wir können nun Erwartungswert und Varianz einer normalverteilten Zufallsvariablen bestimmen.

Lemma 3.3.4. *Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $\mathbb{E}X = \mu$ und $\text{Var}X = \sigma^2$. Man sagt daher auch, X sei normalverteilt mit Erwartungswert μ und Varianz σ^2 .*

Beweis. Es sei zunächst $X \sim \mathcal{N}(0, 1)$. Aus der Symmetrie von φ folgt mit der Substitution $t = -s$, dass

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

Mit partieller Integration folgt nun

$$\text{Var}X = \mathbb{E}X^2 = \int_{-\infty}^{\infty} t \cdot \frac{t \cdot e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = \left[\frac{-te^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{-e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = 1.$$

Sei nun $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann ist $Y := \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. Daher ist

$$0 = \mathbb{E}Y = \sigma^{-1} \mathbb{E}(X - \mu) = \sigma^{-1} ((\mathbb{E}X) - \mu)$$

und daher $\mathbb{E}X = \mu$. Weiter ist $\text{Var}X = \text{Var}(X - \mu) = \mathbb{E}((\sigma Y)^2) = \sigma^2 \mathbb{E}Y^2 = \sigma^2$. \square

3.4 Zufallsvektoren

Definition 3.4.1. Es seien X_1, \dots, X_n Zufallsvariablen, die auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ definiert sind. Dann heißt die Abbildung $X : \Omega \rightarrow \mathbb{R}^n$, gegeben durch

$$X(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

Zufallsvektor. Die Funktion $F : \mathbb{R}^n \rightarrow [0, 1]$ gegeben durch

$$F(x_1, \dots, x_n) := \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

heißt *Verteilungsfunktion* des Zufallsvektors X . Die Verteilungsfunktion (oder auch der Zufallsvektor X selbst) heißt *absolutstetig*, falls es eine Funktion $f : \mathbb{R}^n \rightarrow [0, \infty)$ gibt mit

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \dots dt_n = 1$$

sodass

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \dots dt_n.$$

In diesem Fall heißt f *Dichte* von F (oder von X).

Bemerkung 3.4.2. (a) Wir nennen die von den Rechtecken $(a_1, b_1) \times \dots \times (a_n, b_n)$ mit $a_i, b_i \in \mathbb{R}$, $a_i < b_i$, erzeugte σ -Algebra auf \mathbb{R}^n Borel- σ -Algebra $\mathcal{B}(\mathbb{R}^n)$. Wiederum enthält die Borel- σ -Algebra aller praktisch relevanten Teilmengen, weshalb wir in Übungs- und Klausuraufgaben darauf verzichten, bei Teilmengen des \mathbb{R}^n zu überprüfen, ob diese in der Borel- σ -Algebra liegen. Ähnlich wie im Fall von Zufallsvariablen kann man zeigen, dass die Verteilungsfunktion ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}^n)$ bestimmt. Dieses Maß ist gerade die Verteilung des Vektors X .

(b) Aus der Verteilungsfunktion des Vektors können auch die Verteilungsfunktionen der einzelnen Komponenten bestimmt werden. Mit Lemma 3.1.11 folgt nämlich

$$\begin{aligned} F_{X_j}(x) &= \mathbb{P}(X_j \leq x) = \lim_{N \rightarrow \infty} \mathbb{P}(X_1 \leq N, \dots, X_{j-1} \leq N, X_j \leq x, X_{j+1} \leq N, \dots, X_n \leq N) \\ &= \lim_{N \rightarrow \infty} F_X(N, \dots, N, x, N, \dots, N). \end{aligned}$$

In dieser Situation nennt man manchmal den letzten Grenzwert *Randverteilungsfunktion*.

(c) Auf ähnliche Weise lassen sich auch die Dichten der einzelnen Komponenten aus der Dichte des Vektors berechnen. Es ist nämlich

$$f_j(s) = \int_{\mathbb{R}^{n-1}} f(t_1, \dots, t_{j-1}, s, t_{j+1}, \dots, t_n) dt_1 \dots dt_n,$$

wobei f_j die Dichte von X_j bezeichnet. Diese wird also aus der "gemeinsamen Dichte" durch ausintegrieren der anderen Variablen berechnet. Man sagt die f_j seien die *Randdichten*.

(d) Mittels der Dichte lassen sich auch Wahrscheinlichkeiten berechnen. Sei nämlich X ein Zufallsvektor mit Dichte f . Dann gilt für $A \in \mathcal{B}(\mathbb{R}^n)$, dass

$$\mathbb{P}(X \in A) = \int_A f(t_1, \dots, t_n) dt_1 \dots dt_n.$$

(e) Für eine Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ und einen Zufallsvektor X mit Dichte f ist

$$\mathbb{E}g(X_1, \dots, X_n) = \int_{\mathbb{R}^n} g(t_1, \dots, t_n) f(t_1, \dots, t_n) dt_1 \dots dt_n$$

sofern der Erwartungswert endlich ist. Insbesondere kann man auf diese Art die Kovarianz zweier Zufallsvariablen berechnen, die wie im diskreten Fall definiert ist.

Beispiel 3.4.3. 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 $f(x, y) \geq 0$ für alle $x, y \in \mathbb{R}$ und

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

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

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

Schließlich berechnen wir noch die Kovarianz von X und Y . Hierzu benötigen wir zunächst die Erwartungswerte. Es ist

$$\mathbb{E}X = \int_{\mathbb{R}} t f_X(t) dt = \int_0^1 t \cdot 2t dt = \left[\frac{2}{3} t^3 \Big|_0^1 \right] = \frac{2}{3}$$

und

$$\mathbb{E}Y = \int_{\mathbb{R}} t f_Y(t) dt = \int_0^1 t \left(\frac{1}{2} + t\right) dt = \left[\frac{t^2}{4} + \frac{t^3}{3} \Big|_0^1 \right] = \frac{7}{12}.$$

Somit ist

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}\left(X - \frac{2}{3}\right)\left(Y - \frac{7}{12}\right) \\ &= \int_0^1 \int_0^1 \left(x - \frac{2}{3}\right)\left(y - \frac{7}{12}\right)(x + 2xy) dx dy \\ &= \int_0^1 \left(y - \frac{7}{12}\right) \int_0^1 x^2 + 2x^2y - \frac{2}{3}x - \frac{4}{3}xy dx dy \\ &= \int_0^1 \left(y - \frac{7}{12}\right) \left[\frac{x^3}{3} + \frac{2}{3}x^3y - \frac{x^2}{3} - \frac{2}{3}x^2y \Big|_0^1 \right] dy \\ &= 0. \end{aligned}$$

Folglich sind X und Y unkorreliert.

Definition 3.4.4. Zufallsvariablen X_1, \dots, X_n , die auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ definiert sind, heißen *unabhängig*, falls für alle $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ stets

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdot \dots \cdot \mathbb{P}(X_n \in A_n)$$

gilt.

Man kann nun folgenden Satz beweisen:

Satz 3.4.5. Seien X_1, \dots, X_n Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$. Weiter sei F_X die Verteilungsfunktion des Vektors (X_1, \dots, X_n) und F_{X_j} die Verteilungsfunktion der Zufallsvariable X_j für $j = 1, \dots, n$. Dann sind folgende Aussagen äquivalent:

(i) X_1, \dots, X_n sind unabhängig.

(ii) $F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$ für alle $x_1, \dots, x_n \in \mathbb{R}$.

Besitzt X eine Dichte f und X_j die Dichte f_j , so sind obige Aussagen äquivalent zu

(iii) $f(t_1, \dots, t_n) = f_1(t_1) \cdot \dots \cdot f_n(t_n)$, wobei f die Dichte von X und f_j die Dichte von X_j bezeichnet.

Beispiel 3.4.6. Die Zufallsvariablen X und Y aus Beispiel 3.4.3 sind unabhängig. Es ist nämlich

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

Wir betrachten ein weiteres Beispiel.

Beispiel 3.4.7. Es sei $D = \{(x, y) : 0 \leq y \leq 2x, 0 \leq x \leq 1\}$ und $f = \mathbb{1}_D$. Dann ist f eine Dichte. Es ist nämlich

$$\int_{\mathbb{R}^2} \mathbb{1}_D dx dy = \int_0^1 \int_0^{2x} dy dx = \int_0^1 2x dx = 1.$$

Ist (X, Y) ein Zufallsvektor mit Dichte f , so sagt man (X, Y) sei *gleichverteilt auf dem Dreieck D* .

Die Randdichten von f sind wie folgt gegeben:

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

und

$$f_Y(t) = \int_{\mathbb{R}} \mathbb{1}_D(x, t) dx = \mathbb{1}_{(0,2)}(t) \int_{t/2}^1 dx = \left(1 - \frac{t}{2}\right)\mathbb{1}_{(0,1)}(t).$$

Hat der Vektor (X, Y) Dichte f , so sind X und Y nicht unabhängig, es ist nämlich

$$\begin{aligned} f_X(x)f_Y(y) &= 2x\mathbb{1}_{(0,1)}(x)\left(1 - \frac{y}{2}\right)\mathbb{1}_{(0,2)}(y) \\ &= (2x - xy)\mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,2)}(y) \\ &\neq \mathbb{1}_D(x, y) = f_{X,Y}(x, y). \end{aligned}$$

3.5 Der Zentrale Grenzwertsatz

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

Satz 3.5.1. (Zentraler Grenzwertsatz) Es sei X_1, X_2, \dots eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit endlichen Momenten zweiter Ordnung und positiver Varianz. Wir setzen $\sigma^2 := \text{Var}X_1$ und $\mu := \mathbb{E}X_1$. Weiter sei $S_n := \sum_{k=1}^n X_k$. Dann ist

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

Wir bemerken: Es ist $\mathbb{E}S_n = n\mu$ und, wegen der Unabhängigkeit, $\text{Var}(S_n) = n\sigma^2$. Daher sind

$$\mathbb{E}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}}\right] = 0 \quad \text{und} \quad \text{Var}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = 1$$

unabhängig von n .

Alternative Interpretation: Ist $M_n := \frac{1}{n}S_n$ das Mittel der ersten n Zufallsvariablen, so ist $\frac{\sqrt{n}}{\sigma}(M_n - \mu) = \frac{S_n - n\mu}{\sigma\sqrt{n}}$.

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

Bemerkung 3.5.2. Beachte, dass $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$ ist. Es folgt aus dem Zentralen Grenzwertsatz, dass

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

Man darf hier $<$ - und \leq - Zeichen austauschen.

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

Beispiel 3.5.3. 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 Zahl der Erfolge $\text{Bin}(600, \frac{1}{6})$ -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

$$\begin{aligned} \mathbb{P}(90 \leq S_{600} \leq 100) &= \mathbb{P}\left(\frac{90 - 100}{9,13} \leq \frac{S_{600} - 600\mu}{\sqrt{600}\sigma} \leq \frac{100 - 100}{9,13}\right) \\ &\approx \Phi(0) - \Phi(-1,095) = 0,5 - (1 - \Phi(1,095)) \approx 0,363, \end{aligned}$$

also ungefähr 36 Prozent.

Bemerkung 3.5.4. 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. In Beispiel 3.5.3 ist somit

$$\mathbb{P}(90 \leq S_{600} \leq 100) = \mathbb{P}(89 < S_{600} < 101) = \mathbb{P}(89,5 \leq S_{600} \leq 100,5).$$

Berechnet man diese Wahrscheinlichkeiten näherungsweise mit dem zentralen Grenzwertsatz, so ergeben sich 0,363, 0,430 bzw. 0,397. Vergleichen wir diesen Wert mit dem exakten Wert 0,4025 (die Berechnung ist zwar aufwendig, aber nicht unmöglich), so sehen wir, dass der letztgenannte Wert am besten ist. Es entspricht auch der allgemeinen Erfahrung, dass, wenn die Wahrscheinlichkeit, dass eine diskrete Zufallsvariable in einem bestimmten Intervall liegt, mit Hilfe des zentralen Grenzwertsatzes approximiert wird, man am besten die Mittelwerte zwischen den beiden in Frage kommenden Zahlen als Intervallgrenzen einsetzt.

Wir geben nun noch ein etwas anderes Anwendungsbeispiel.

Beispiel 3.5.5. Ein Airbus A380 hat gewöhnlich 526 Sitzplätze. Aus Erfahrungen ist bekannt, dass ein verkauftes Ticket mit einer Wahrscheinlichkeit von 0,1 storniert wird. Wir nehmen an, dass die Fluggäste, die Entscheidung, ob sie stornieren oder nicht, unabhängig voneinander treffen. 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 $\text{Bin}(1, 0,9)$ -verteilter Zufallsvariablen und $S_n = \sum_{k=1}^n X_k$. Wir suchen die größte Zahl n , sodass $\mathbb{P}(S_n \geq 526) \leq 0,02$. Es ist

$$0,02 \stackrel{!}{=} \mathbb{P}(S_n \geq 526) = \mathbb{P}\left(\frac{S_n - n \cdot 0,9}{\sqrt{n \cdot 0,1 \cdot 0,9}} \geq \frac{526 - n \cdot 0,9}{\sqrt{n \cdot 0,09}}\right) \approx 1 - \Phi\left(\frac{526 - n \cdot 0,9}{\sqrt{n \cdot 0,09}}\right)$$

Wir benötigen das 0,98-Quantil der Normalverteilung. kann man in Tabellen nachschlagen. Es gilt $\Phi^{-1}(0,98) \approx 2,05$. Wir erhalten also

$$2,05 \approx \Phi^{-1}(0,98) \approx \frac{526 - n \cdot 0,9}{\sqrt{n \cdot 0,09}}.$$

Wir schreiben $m = \sqrt{n}$ und erhalten die Gleichung

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

Löst man diese quadratische Gleichung, so erhält man $m = 23,8$, also $n = 568,1$. Die Fluggesellschaft darf also höchstens 568 Tickets verkaufen.

3.6 Schätzung der Parameter in der Normalverteilung

In Kapitel 2 hatten wir bereits die Schätzung von Parametern gewisser Verteilungen diskutiert, dabei aber (weil wir absolutstetige Verteilungen noch nicht eingeführt hatten) die Normalverteilung außen vor gelassen. Allerdings ist diese Verteilung in Anwendungen von größtem Interesse. Daher wollen wir das Schätzen der Parameter der Normalverteilung hier nachholen.

Es gilt also aus einer Zufallsstichprobe X_1, \dots, X_n zur Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ die Parameter μ und σ^2 zu schätzen.

Wir nutzen zunächst die *Momentenmethode*. Für $Y \sim \mathcal{N}(\mu, \sigma^2)$ gilt $(\mu, \sigma^2) = (\mathbb{E}Y, \mathbb{E}Y^2 - (\mathbb{E}Y)^2)$. Somit ergibt sich als Momentenschätzer, wenn Realisierungen x_1, \dots, x_n beobachtet werden,

$$(\hat{\mu}, \widehat{\sigma^2}) = \left(\bar{x}, \frac{1}{n} \sum_{k=1}^n (x_k)^2 - (\bar{x})^2\right) = \left(\bar{x}, \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2\right).$$

Nu wenden wir uns der *Maximum-Likelihood-Methode* zu. Im Falle diskreter Wahrscheinlichkeiten hatten wir als Likelihood-Funktion L gerade das Produkt der Zähldichten verwendet: $L(x_1, \dots, x_n; \theta) = f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta)$. Im Falle von absolutstetigen Verteilungen mit Dichte $f(t, \theta)$ verwenden wir das Produkt der Dichten als Likelihood-Funktion.

Definition 3.6.1. Es sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilungsfunktion $P \in \{P_\theta : \theta \in \Theta\}$. Weiter sei P_θ absolutstetig mit Dichte $f(\cdot, \theta)$. Dann heißt $L : \mathbb{R}^n \times \Theta \rightarrow [0, \infty)$, definiert durch

$$L(x_1, \dots, x_n; \theta) := f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta)$$

Likelihood-Funktion. Ist $\hat{\theta} : \mathbb{R}^n \rightarrow \Theta$ eine Funktion mit

$$L(x_1, \dots, x_n; a) \leq L(x_1, \dots, x_n; \hat{\theta}(x_1, \dots, x_n))$$

für alle $a \in \Theta$, so heißt $\hat{\theta}(X_1, \dots, X_n)$ *Maximum-Likelihood-Schätzer* für θ .

Wir berechnen nun Maximum-Likelihood-Schätzer für die Parameter μ und σ^2 einer Normalverteilung. Es ist

$$L(x_1, \dots, x_n; \mu, \sigma^2) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_k - \mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left(-\frac{\sum_{k=1}^n (x_k - \mu)^2}{2\sigma^2}\right).$$

Es ist einfacher, die log-Likelihood Funktion

$$\ell(\mu, \sigma^2) := \log L(x_1, \dots, x_n; \mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2 - \frac{n}{2} \log(2\pi\sigma^2)$$

zu betrachten. Um einen Kandidaten für die Maximumstelle zu finden, berechnen wir die kritischen Punkte der Log-Likelihood-Funktion, also die Lösungen des Gleichungssystems $\nabla \ell = 0$. Für die partielle Ableitung nach μ finden wir

$$\frac{\partial \ell}{\partial \mu} = \left(-\frac{1}{2\sigma^2}\right) \sum_{k=1}^n 2(x_k - \mu) \stackrel{!}{=} 0 \quad \Leftrightarrow \quad 0 = \sum_{k=1}^n (x_k - \mu) = n\bar{x} - n\mu \quad \Leftrightarrow \quad \mu = \bar{x}$$

Für die partielle Ableitung nach σ^2 ist

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{2\sigma^4} \sum_{k=1}^n (x_k - \mu)^2 - \frac{n}{2} \frac{2\pi}{2\pi\sigma^2} = \frac{1}{2\sigma^2} \left(\frac{1}{\sigma^2} \sum_{k=1}^n (x_k - \mu)^2 - n\right) \stackrel{!}{=} 0$$

Wir wissen bereits, dass in einem kritischen Punkt $\mu = \bar{x}$ sein muss. Wir setzen das in obige Gleichung ein und erhalten

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \quad \Leftrightarrow \quad \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - \bar{x})^2 = n \quad \Leftrightarrow \quad \sigma^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Die einzige kritische Stelle von ℓ ist demnach der Punkt

$$(\mu, \sigma^2) = \left(\bar{x}, \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2\right).$$

Eine genauere Untersuchung der log-Likelihood Funktion zeigt, dass dies in der Tat eine (globale!) Maximumstelle ist, d.h. dies ist der (in diesem Falle eindeutige) Maximum-Likelihood-Schätzer. Für die Normalverteilung stimmen also Momenten- und Maximum-Likelihood-Schätzer überein.

Obwohl sich sowohl aus der Momenten- als auch aus der Maximum-Likelihood-Methode $\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$ als Schätzer für σ^2 ergibt, zieht man in der Praxis meist $\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$ vor, weil dieser Schätzer erwartungstreu ist.

Konfidenzintervall für μ bei bekannter Varianz

Als nächstes konstruieren wir Konfidenzintervalle für den Mittelwert μ . Dabei muss unterschieden werden, ob die Varianz σ^2 bekannt ist, oder nicht

Es sei also X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung $\mathcal{N}(\mu, \sigma_0^2)$, wobei σ_0^2 eine bekannte, feste, positive Zahl ist und μ ein unbekannter, reeller Parameter ist. Um μ zu schätzen verwenden wir den Schätzer $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$. Wir könnten nun mittels Tschebyscheff-Ungleichung ein Konfidenzintervall herleiten. Allerdings ergibt sich ein kürzeres Konfidenzintervall, wenn wir folgende Proposition verwenden.

Proposition 3.6.2. *Seien $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$, unabhängige Zufallsvariablen. Dann gilt $X_1 + \dots + X_n \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.*

Insbesondere ist $X_1 + \dots + X_n \sim \mathcal{N}(n\mu, n\sigma_0^2)$ und daher $\bar{X} \sim \mathcal{N}(\mu, \sigma_0^2/n)$. Es folgt, dass $V := \sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \sim \mathcal{N}(0, 1)$. Beachte, dass diese Verteilung *nicht* mehr von dem unbekanntem Parameter μ abhängt. Um ein Konfidenzintervall zum Konfidenzniveau α zu konstruieren, geht man nun wie folgt vor:

Zunächst bestimmt man ein Intervall, in dem V mit Wahrscheinlichkeit α liegt. Weil V symmetrisch verteilt ist, wählen wir auch ein symmetrisches Intervall, also ein $c > 0$ sodass

$$1 - \alpha = \mathbb{P}_\mu(|V| > c) = 2\mathbb{P}_\mu(V > c) = 2(1 - \Phi(c))$$

also $c = \Phi^{-1}(\frac{1+\alpha}{2})$. Somit ergibt sich

$$\alpha = \mathbb{P}_\mu(|V| \leq c) = \mathbb{P}_\mu\left(-c \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \leq c\right) \quad (3.3)$$

$$= \mathbb{P}_\mu\left(\mu \in \left[\bar{X} - \frac{c\sigma_0}{\sqrt{n}}, \bar{X} + \frac{c\sigma_0}{\sqrt{n}}\right]\right). \quad (3.4)$$

Somit ist $[\bar{X} - c\sigma_0/\sqrt{n}, \bar{X} + c\sigma_0/\sqrt{n}]$ mit $c = \Phi^{-1}(1 + \alpha/2)$ ein α -Konfidenzintervall für μ .

An diesem Beispiel kann man sehen, wie man allgemein vorgeht, um (unter Ausnutzung spezieller Eigenschaften der zu Grunde liegenden Verteilung) Konfidenzintervalle für einen Parameter $\vartheta \in \Theta$ zu konstruieren:

- 1) Wähle einen (Punkt-) Schätzer $\hat{\theta}$ für θ .
- 2) Bestimme die Verteilung von $\hat{\theta}$.
- 3) Wähle eine Transformation g_θ , so dass die Verteilung von $g_\theta(\hat{\theta})$ nicht von θ abhängt.
- 4) Wähle nun ein Intervall J mit $\mathbb{P}(g_\theta(\hat{\theta}) \in J) = \alpha$.
- 5) Löse die Bedingung $g_\theta(\hat{\theta}) \in J$ nach θ auf. Falls sich ein Intervall ergibt, ist dies das gesuchte Konfidenzintervall.

Insbesondere sind Konfidenzintervalle nicht eindeutig bestimmt.

Beispiel 3.6.3. Bei einer Stichprobe von 5 Brötchen wurden folgende Gewichte (in Gramm) gemessen (nach aufsteigendem Gewicht):

$$43 \quad 44 \quad 44 \quad 47 \quad 50$$

Wir nehmen an, dass das Gewicht normalverteilt ist mit unbekanntem Mittelwert und Standardabweichung 3. Wir bestimmen ein 95%-Konfidenzintervall für μ wie folgt:

Zunächst schlagen wir $c = \Phi^{-1}(\frac{1,95}{2}) = \Phi^{-1}(0,975) = 1,96$ in einer Tabelle der Verteilung der Standardnormalverteilung nach. Es ist nun

$$\frac{c\sigma_0}{\sqrt{n}} = \frac{1,96 \cdot 3}{\sqrt{5}} = 2,63 \quad \text{und} \quad \bar{X} = 45,6.$$

Daher ist $[42,97, 48,23]$ ein 95%-Konfidenzintervall für den Mittelwert μ .

Mit der Tschebyscheff-Ungleichung hätte sich

$$\left[\bar{X} - \sqrt{\frac{\sigma^2}{n \cdot (1-\alpha)}}, \bar{X} + \sqrt{\frac{\sigma^2}{n \cdot (1-\alpha)}} \right] = [39,6, 51,6]$$

ergeben.

Konfidenzintervall für μ bei unbekannter Varianz

Im allgemeinen kann man in Anwendungen *nicht* annehmen, dass die Varianz bekannt ist. Sie muss also ebenfalls geschätzt werden. Es ist naheliegend, in obigem Zugang die (nun nicht mehr bekannte) Standardabweichung σ_0 durch S , die Wurzel aus dem Schätzer S^2 für die Varianz, zu ersetzen. Um ein Konfidenzintervall zu bestimmen, muss man nun die Verteilung der resultierenden Zufallsvariable kennen. Wir diskutieren dies in allgemeiner Situation.

Definition 3.6.4. 1. Es seien X_1, \dots, X_r unabhängig und $\mathcal{N}(0,1)$ -verteilt. Dann heißt die Verteilung von $Y := X_1^2 + \dots + X_r^2$ χ^2 -Verteilung mit r Freiheitsgraden (sprich: chi-quadrat). Wir schreiben $Y \sim \chi_r^2$.

2. Es seien X, Y unabhängig mit $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi_r^2$. Dann heißt die Verteilung von $Z := \frac{X}{\sqrt{Y/r}}$ t -Verteilung mit r Freiheitsgraden. Wir schreiben $Z \sim t_r$.

Bemerkung 3.6.5. (a) Sowohl die χ^2 -Verteilungen als auch die t -Verteilungen sind absolutstetig. Die Dichten dieser Verteilungen können explizit angegeben werden. Für uns sind jedoch insbesondere die Verteilungsfunktion und deren Umkehrfunktion (d.h. die Quantilfunktion) von Bedeutung. Diese liegen in Tabellen vor.

(b) Die t -Verteilung ist symmetrisch.

Zur Bestimmung von Konfidenzintervallen wichtig ist folgendes Resultat, welches wir hier ohne Beweis angeben.

Satz 3.6.6. *Es seien X_1, \dots, X_n unabhängig und $\mathcal{N}(\mu, \sigma^2)$ -verteilt. Dann sind \bar{X} und S^2 unabhängig und es gilt*

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2, \quad \sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}.$$

Wir können also statt der Größe $V = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$, die von der Standardabweichung σ abhängt, die Größe $T = \sqrt{n} \frac{\bar{X} - \mu}{S}$, die nicht von der Standardabweichung abhängt, betrachten und haben wiederum eine Größe, deren Verteilung nicht von den Parametern μ und σ^2 abhängt. Somit können wir nun ein α -Konfidenzintervall für μ wie folgt bestimmen:

Es sei c ein $(1 + \alpha)/2$ -Quantil der t_{n-1} -Verteilung, sodass

$$\mathbb{P}_{\mu, \sigma^2}(|T| > c) = 2\mathbb{P}_{\mu, \sigma^2}(T > c) = 2\left(1 - \frac{1 + \alpha}{2}\right) = 1 - \alpha.$$

Somit ist

$$\mathbb{P}_{\mu, \sigma^2}\left(\bar{X} - \frac{cS}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cS}{\sqrt{n}}\right) = \mathbb{P}_{\mu, \sigma^2}(|T| \leq c) = \alpha.$$

Wir haben also ein α -Konfidenzintervall für μ gefunden.

Beispiel 3.6.7. Wir betrachten wiederum Beispiel 3.6.3. Man rechnet leicht nach, dass $S^2 = 8,3$, also $S = 2,88$ ist. Um ein 95%-Konfidenzintervall zu berechnen, bestimmen wir zunächst das 0,975-Quantil der t -Verteilung mit 4 Freiheitsgraden. Aus einer Tabelle entnehmen wir den Wert 2,78. Somit ergibt sich

$$\frac{cS}{\sqrt{n}} = \frac{2,78 \cdot 2,88}{\sqrt{5}} \approx 3,58$$

Daher ist $[42,02, 49,18]$ ein 95%-Konfidenzintervall für μ . Beachte, dass dieses Intervall länger ist als das in Beispiel 3.6.3 berechnete, obwohl die dort angenommene Standardabweichung 3 größer als die nun geschätzte Standardabweichung 2,88 ist. Der Preis dafür, dass wir σ^2 als unbekannt annehmen (dürfen!) ist ein längeres Konfidenzintervall.

Konfidenzintervall für σ^2

Wir wollen nun noch ein Konfidenzintervall für die Varianz einer Stichprobe X_1, \dots, X_n zur Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ bestimmen. Wir nehmen an, dass der Erwartungswert μ bekannt sei. Wir wissen bereits, dass $R := \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$. Insbesondere hängt die Verteilung dieser Größe nicht mehr von den Parametern μ und σ^2 ab. Wir können daher Konfidenzintervalle für σ^2 mittels Quantilen der χ_{n-1}^2 -Verteilung bestimmen.

Um ein α -Konfidenzintervall zu bestimmen, gehen wie folgt vor. Wir bestimmen a und b , sodass $\mathbb{P}(R \in [a, b]) = \alpha$. Dann ist

$$\alpha = \mathbb{P}_{\mu, \sigma^2}\left(\frac{n-1}{\sigma^2} S^2 \in [a, b]\right) = \mathbb{P}_{\mu, \sigma^2}\left(\frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right).$$

Wir haben also ein α -Konfidenzintervall gefunden. Es bleibt noch a und b zu wählen. Hier ist zu beachten, dass (anders als bei der Standardnormalverteilung und der t -Verteilung), die χ^2 -Verteilung *nicht* symmetrisch ist, genauer gesagt nimmt eine χ^2 -verteilte Zufallsvariable nur positive Werte an. Daher wählt man a und b in der Regel wie folgt:

Wir wählen a so, dass $R \in [0, a]$ Wahrscheinlichkeit $(1 - \alpha)/2$ hat, also ist a das $(1 - \alpha)/2$ -Quantil der χ_{n-1}^2 -Verteilung. Dann wählen wir b so, dass $R \in [b, \infty)$ ebenfalls Wahrscheinlichkeit $(1 - \alpha)/2$ hat, also ist b das $1 - (1 - \alpha)/2 = (1 + \alpha)/2$ -Quantil der χ_{n-1}^2 -Verteilung.

Beispiel 3.6.8. Wir betrachten wieder Beispiel 3.6.3, wo wir $S^2 = 8,3$ bei einem Stichprobenumfang von $n = 5$ beobachtet hatten. Um ein Konfidenzintervall zum Niveau 0,95 zu bestimmen benötigen wir noch Quantile der χ_4^2 -Verteilung. Das 2,5%-Quantil von χ_4^2 ist

gegeben durch $a = 8,91$. Andererseits ist das 97,5%-Quantil gegeben durch $b = 32,85$. Somit ergibt sich als 95%-Konfidenzintervall für σ^2

$$\left[\frac{4 \cdot 8,3}{32,85}, \frac{4 \cdot 8,3}{8,91} \right] = [2,98, 68,5].$$

Auf Grund des kleinen Stichprobenumfangs ist dies recht ungenau.

Kapitel 4

Statistische Tests

4.1 Grundbegriffe

Wir betrachten wieder ein parametrisches Modell $\{P_\theta : \theta \in \Theta\}$ und eine zugehörige Zufallsstichprobe X_1, \dots, X_n . Wir wollen nun die Beobachtung der X_1, \dots, X_n verwenden, um bestimmte Aussagen über die zugrundeliegende Verteilung (genauer: den zugrundeliegenden Parameter θ) zu testen. Wir verwenden folgende Terminologie:

Eine *Hypothese* ist eine Aussage über den Parameter θ . Modelliert wird eine Hypothese durch eine Teilmenge Θ_0 von Θ . Man sagt, die Hypothese *trifft zu*, falls $\theta \in \Theta_0$. Eine Hypothese heißt *einfach*, falls Θ_0 einelementig ist, ansonsten sagt man, die Hypothese ist *zusammengesetzt*. Die Negation der Hypothese, also die Aussage $\theta \notin \Theta_0$ – äquivalent, die Aussage $\theta \in \Theta_1 := \Theta \setminus \Theta_0$ – heißt *Alternative*. Manchmal nennt man die Hypothese auch *Nullhypothese* (und schreibt $H_0 : \theta \in \Theta_0$) und die Alternative *Alternativhypothese* und schreibt $H_1 : \theta \in \Theta_1$).

Ein *Test* (der Hypothese $\theta \in \Theta_0$ gegen die Alternative $\theta \in \Theta_1$) ist eine Entscheidungsregel, die für jede Realisierung x_1, \dots, x_n der Stichprobe X_1, \dots, X_n festlegt, ob die Hypothese oder die Alternative gewählt wird. Man hat also eine disjunkte Zerlegung des Stichprobenraumes M^n in Teilmengen K_0 und K_1 , sodass wir uns für die Hypothese entscheiden, wenn $(X_1, \dots, X_n) \in K_0$ (wir sagen “*die Hypothese wird akzeptiert*”) und wir uns für die Alternative entscheiden, wenn $(X_1, \dots, X_n) \in K_1$ (wir sagen “*die Hypothese wird verworfen*”).

K_0 heißt auch *Annahmebereich*, K_1 *kritischer Bereich*.

Beispiel 4.1.1. Wir betrachten erneut das Werfen einer Reißzwecke und möchten untersuchen, ob es gleich wahrscheinlich ist, auf der flachen Seite oder mit der Spitze schräg nach unten liegen zu bleiben. Als parametrisches Modell betrachten wir $X \sim \text{Bin}(1, p)$, wobei wir $X = 1$ als “flache Seite” und $X = 0$ als “mit der Spitze schräg nach unten” interpretieren. Wir wollen die Hypothese $p = \frac{1}{2}$ (dies ist eine einfache Hypothese) gegen die Alternative $p \neq \frac{1}{2}$ testen.

Wir beobachten Zufallsvariablen $X_1, \dots, X_{1000} \sim \text{Bin}(1, p)$ und bilden den Mittelwert \bar{X} .

Ein möglicher Test akzeptiert die Hypothese, wenn $\bar{X} = \frac{1}{2}$ ist, und verwirft sie sonst (Test T_1).

Eine andere Möglichkeit wäre es, die Hypothese zu akzeptieren, wenn $\bar{X} \in [0.47, 0.53]$ liegt, und sie sonst zu verwerfen (Test T_2).

Schließlich wäre es auch möglich, die Hypothese immer zu akzeptieren (Test T_3).

Offensichtlich sind nicht alle Tests in obigem Beispiel gleich gut. Bei Test T_1 ist das Problem, dass selbst wenn der wahre Parameter $p = \frac{1}{2}$ ist, nicht notwendigerweise $\bar{X} = \frac{1}{2}$ sein muss. Genauer gesagt besitzt dieses Ereignis (sofern $p = \frac{1}{2}$) gerade Wahrscheinlichkeit $\binom{1000}{500} \frac{1}{2}^{500} \frac{1}{2}^{500} \approx 0,025$. Beachte, dass falls der Stichprobenumfang n ungerade ist, $\bar{X} = \frac{1}{2}$ Wahrscheinlichkeit 0 besitzt. Offensichtlich ist, wenn die Hypothese zutrifft, die Wahrscheinlichkeit, dass der Test die Hypothese akzeptiert, bei Test T_2 größer. Bei Test T_3 ist sie sogar noch größer. Allerdings irrt Test T_3 immer, wenn die Alternative richtig gewesen wäre.

Um diese Phänomene genauer zu untersuchen, führen wir folgende Begriffe ein.

Definition 4.1.2. Verwirft ein Test die Hypothese, obwohl sie richtig gewesen wäre, so sagt man, es liegt ein *Fehler erster Art* vor. Akzeptiert ein Test die Hypothese, obwohl die Alternative richtig gewesen wäre, so sagt man, es liegt ein *Fehler zweiter Art* vor.

Hat man nun einen Test mit Annahmebereich K_0 und kritischem Bereich K_1 gegeben, so heißt

$$G(\theta) := \mathbb{P}_\theta((X_1, \dots, X_n) \in K_1)$$

die *Gütefunktion* des Tests. Die Zahl

$$\alpha := \sup\{G(\theta) : \theta \in \Theta_0\}$$

heißt *Umfang* des Tests. Ist zudem $G(\theta) \geq \alpha$ für $\theta \in \Theta_1$, so sagt man der Test sei *unverfälscht*.

Interpretation: Die Gütefunktion gibt an, mit welcher Wahrscheinlichkeit die Hypothese verworfen wird, wenn der wahre Parameter θ ist. Für $\theta \in \Theta_0$ ist also $G(\theta)$ gerade die Wahrscheinlichkeit, einen Fehler erster Art zu begehen. Bei einem Test zum Umfang α ist also die Wahrscheinlichkeit einen Fehler erster Art zu begehen höchstens α .

Beispiel 4.1.3. Wir betrachten wiederum die Tests T_3 und T_2 aus Beispiel 4.1.1. Es sei G_j die Gütefunktion des Tests T_j .

Die Gütefunktion von T_3 gegeben durch $G_3(p) \equiv 0$.

Für den Test T_2 ist die Gütefunktion gegeben durch

$$G_2(p) = 1 - \sum_{k=470}^{530} \binom{1000}{k} p^k (1-p)^{1000-k}.$$

Natürlich ist es wiederum schwer, diese Funktion explizit auszurechnen. Sie kann mit Hilfe des Zentralen Grenzwertsatzes approximiert werden.

Wir bemerken, dass Fehler erster Art und Fehler zweiter Art unterschiedlich behandelt werden. Bei einem Test zum Signifikanzniveau α ist die Wahrscheinlichkeit einen Fehler erster Art zu begehen begrenzt. Allerdings gibt es keine Einschränkungen hinsichtlich des Fehlers zweiter Art. In Anwendungen ist es jedoch häufig der Fall, dass ein Fehler schwerwiegender ist als der andere.

Man denke etwa daran, ein neues Medikament auf Nebenwirkungen zu testen: Es ist wesentlich schwerwiegender vorhandene Nebenwirkungen nicht zu entdecken als falschen Alarm zu schlagen (und nichtvorhandene Nebenwirkungen zu erkennen).

Diese Asymmetrie sollte man bei der Wahl von Hypothese und Alternative beachten: die Hypothese sollte so gewählt werden, dass der schwerwieendere Fehler der Fehler erster Art ist. Bei den Medikamenten sollte man also die Hypothese "Das Medikament hat Nebenwirkungen" gegen die Alternative "Das Medikament hat keine Nebenwirkungen" testen.

Problem: In der Regel existieren nur für eine der beiden Wahlen Tests.

4.2 Tests für den Erwartungswert einer Normalverteilung

Wir betrachten nun einige “Standardtests”, die den Erwartungswert μ einer Normalverteilung betreffen. Man unterscheidet hierbei *einseitige Tests*, bei denen die Hypothese $H_0 : \mu \leq \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$ (resp. die Hypothese $H_0 : \mu \geq \mu_0$ gegen die Alternative $H_1 : \mu < \mu_0$) getestet wird und *zweiseitige Tests*, bei denen die Hypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu \neq \mu_0$ getestet wird.

Eine weitere Unterscheidung ist, ob die Varianz bekannt oder unbekannt ist.

In allen hier diskutierten Tests verwenden wir eine sogenannte *Teststatistik* T und entscheiden uns für oder gegen die Hypothese, abhängig davon, wo T liegt.

Als Teststatistik treten $T := \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0}$ bei bekannter Varianz σ_0^2 und $T := \sqrt{n} \frac{\bar{X} - \mu_0}{s}$ bei unbekannter Varianz auf. Ist der wahre Parameter μ_0 (also unter Nullhypothese beim zweiseitigen Test), so ist $T \sim \mathcal{N}(0, 1)$ resp. $T \sim t_{n-1}$.

Tests bei bekannter Varianz

Gegeben eine Stichprobe X_1, \dots, X_n zur Normalverteilung $\mathcal{N}(\mu, \sigma_0^2)$ bei *bekannter Varianz* σ_0^2 , wollen wir Tests zu vorgegebenem Umfang $\alpha \in (0, 1)$ konstruieren. Wir betrachten zunächst den zweiseitigen Test für $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu \neq \mu_0$.

Wir akzeptieren H_0 , wenn $|T| \leq c$ für ein geeignetes $c > 0$ ist, und lehnen H_0 sonst ab.

Damit der Test Umfang α hat, muss $\mathbb{P}_{\mu_0}(|T| > c) = \alpha$ gelten. Unter der Nullhypothese $\mu = \mu_0$ ist $T \sim \mathcal{N}(0, 1)$. Daher wählen wir $c := \Phi^{-1}(1 - \alpha/2)$.

Beispiel 4.2.1. Eine Maschine füllt 200 g Packungen mit Müsli ab. Aus Erfahrungen ist bekannt, dass das tatsächliche Gewicht, das von der Maschine abgefüllt wird, normalverteilt mit Varianz 4 ist. Um zu überprüfen, ob die Maschine korrekt eingestellt ist, werden 10 Packungen nachgewogen, was ein Durchschnittsgewicht von 197 g ergibt. Es soll nun zum Signifikanzniveau von 5% getestet werden, ob $\mu = 200$ plausibel ist.

Lösung: Es ist $\Phi^{-1}(0,975) = 1,96$. Für die Teststatistik T ergibt sich $T = \sqrt{10} \frac{197 - 200}{2} = -4,743$. Die Hypothese wird also verworfen; die Maschine ist nicht richtig eingestellt.

Nun betrachten wir einen einseitigen Test, bei dem wir $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$ testen (Um $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$ zu testen, geht man analog vor).

Wir betrachten wieder $T = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0}$ und akzeptieren H_0 falls $T \leq c$ für ein (neues) geeignetes c . Dieses ist so zu wählen, dass

$$\sup\{G(\mu) \mid \mu \leq \mu_0\} = \alpha.$$

Wir berechnen die Gütefunktion in Abhängigkeit von c :

$$\begin{aligned} G(\mu) &= \mathbb{P}_\mu((X_1, \dots, X_n) \in K_1) = \mathbb{P}_\mu(\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0} > c) = \mathbb{P}_\mu(\sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} > c + \sqrt{n} \frac{\mu_0 - \mu}{\sigma_0}) \\ &= 1 - \Phi(c + \sqrt{n} \frac{\mu_0 - \mu}{\sigma_0}) \end{aligned}$$

Wir stellen fest, dass die Gütefunktion monoton wachsend ist. Somit

$$\alpha = \sup\{G(\mu) \mid \mu \leq \mu_0\} = G(\mu_0) = 1 - \Phi(c)$$

und wir müssen $c = \Phi^{-1}(1 - \alpha)$ wählen. Weiter ist der Test unverfälscht.

Beispiel 4.2.2. Es werden Briefumschläge für Luftpost produziert, die im Schnitt nicht mehr als 2 g wiegen dürfen. Eine Stichprobe von 20 Briefumschlägen ergibt ein Durchschnittsgewicht von 2,01 g. Ferner ist bekannt, dass das Gewicht normalverteilt mit Erwartungswert μ und Varianz $0,03^2$ ist. Es soll zu einem Signifikanzniveau von 1% getestet werden, ob $\mu \leq 2$.

Für die Teststatistik ergibt sich $T = \sqrt{20} \frac{2,01-2}{0,03} \approx 1,49$. Da $\Phi(0,99) = 2,33$ wird die Hypothese akzeptiert. Es ist also denkbar, dass das erwartete Gewicht unter $2g$ liegt.

Tests bei unbekannter Varianz

In Anwendungen ist in der Regel die Varianz σ^2 nicht bekannt. Daher verwendet man die Teststatistik $T := \sqrt{n} \frac{\bar{X} - \mu_0}{s}$. Man konstruiert nun Tests wie im Fall von bekannter Varianz, ersetzt aber die Quantile der Standardnormalverteilung durch entsprechende Quantile der t_{n-1} -Verteilung. Genauer konstruiert man Tests zum Signifikanzniveau α wie folgt:

- Beim zweiseitigen Test von $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu \neq \mu_0$ lehnt man die Hypothese ab, sofern $|T| \geq c$ ist, wobei c das $1 - \alpha/2$ -Quantil der t_{n-1} -Verteilung ist.
- Beim einseitigen Test von $H_0 : \mu \leq \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$ lehnt man die Hypothese ab, sofern $T > c$, wobei c das $1 - \alpha$ -Quantil der t_{n-1} -Verteilung ist.

Beispiel 4.2.3. Ein Diätkonzept, bei dem die Teilnehmer durchschnittlich $7,0kg$ abnehmen, wird überarbeitet. Um den Erfolg der Überarbeitung zu kontrollieren, werden die Gewichtsabnahmen von 28 Teilnehmern des neuen Konzepts erfasst. Die so erhobenen Daten werden als Zufallsstichprobe einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariable angesehen. Sodann wird die Hypothese $\mu \leq 7,0$ gegen die Alternative $\mu > 7,0$ zu einem Signifikanzniveau von 2,5% getestet.

Die durchschnittliche Gewichtsabnahme betrug $7,3kg$ bei einer Stichprobenstandardabweichung von $0,4kg$. Somit ergibt sich für die Teststatistik $T = \sqrt{28} \frac{7,3-7,0}{0,4} = 3,97$. Das 97,5% -Quantil der t_{27} -Verteilung ist 2,052, sodass die Hypothese verworfen wird. Somit war die Überarbeitung in der Tat erfolgreich.

Test für verbundene Stichproben

Bei dieser Art Testprobleme haben wir eine Folge von *Paaren* von Beobachtungen, also zwei Stichproben X_1, \dots, X_n und Y_1, \dots, Y_n , so dass die Paare $(X_1, Y_1), \dots, (X_n, Y_n)$ zwar voneinander unabhängig sind, die beiden Komponenten eines Paares jedoch i.d.R. abhängig sind. Beispielsweise kann man sich vorstellen, dass eine bestimmte Größe (etwa die Länge eines Werkstückes) mit zwei verschiedenen Methoden gemessen wird (die eine Methode liefert X_1, \dots, X_n , die andere Methode Y_1, \dots, Y_n), die es zu vergleichen gilt. Man spricht in diesem Falle von *Verbundenen Stichproben*.

Manchmal kann man hierbei einen der obigen Tests auf die Differenzen $X_1 - Y_1, \dots, X_n - Y_n$ anwenden. Wir geben ein Beispiel.

Beispiel 4.2.4. Um zu untersuchen, ob eine geplante Vereinfachung des Steuerrechts zu Mindereinnahmen des Staates führt, gehen wir wie folgt vor:

Wir wählen $n = 100$ Steuererklärungen des vergangenen Jahres zufällig aus und berechnen die Steuer nach dem geltenden Steuerrecht und nach dem vorgeschlagenen neuen Steuerrecht. Nennen wir für Steuererklärung k die Steuerschuld nach dem vorgeschlagenen Recht X_k und die Steuerschuld nach geltendem Recht Y_k . Wir bilden nun die Differenz $Z_k := X_k - Y_k$. Bei den untersuchten Steuererklärungen ergibt sich $\bar{Z} = 120$ bei einer Stichprobenstandardabweichung von 725.

Wir nehmen an, dass Z_1, \dots, Z_{100} normalverteilt sind und testen $H_0 : \mu \leq 0$ ("Mindereinnahmen") gegen die Alternative $H_1 : \mu > 0$. Beachte, dass wir hier wiederum die Hypothese und die Alternative so gewählt haben, dass der schwerwiegendere Irrtum (Es gibt

Mindereinnahmen für den Staat, aber unser Test sagt, es gibt keine) der Fehler erster Art ist.

Als Signifikanzniveau wählen wir $\alpha = 0,05$. Es ist dann $T = \sqrt{100} \frac{120-0}{725} \approx 1,655$. Weiterhin ist das 95%-Quantil der t_{99} -Verteilung gegeben durch $c = 1,660$. Weil $T \leq c$ ist, akzeptiert der Test die Nullhypothese. Mindereinnahmen sind also nicht auszuschließen.