Stochastik für Wirtschaftswissenschaftler

Vorlesungsskript
Wintersemester 2013/14

von Markus Kunze

Inhaltsverzeichnis

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 identischen (oder zumindest ähnlichen) Voraussetzungen nicht immer zum selben Ergebnis führt. Man geht hierbei davon aus, dass alle *möglichen* Ergebnisse des Vorgangs bekannt sind. Diese werden in der Grundmenge Ω zusammengefasst, d.h. Ω besteht aus allen möglichen Versuchsausgängen. Die Elemente von Ω bezeichnet man auch als *Elementarereignisse*.

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

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

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

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

(e) Ein Pfeil wird auf eine Dartscheibe geworfen. Um den Ausgang dieses Zufallsexperiments zu beschreiben legen wir ein Kartesisches Koordinatensystem in den Mittelpunkt der Dartscheibe und wählen die Einheiten so, dass der Radius der Dartscheibe 1 ist. Dann können wir $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ wählen, wobei (x, y) das Elementarereigniss "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.

Häufig ist man bei einem Zufallsexperiment nicht an dem tatsächlichen Ausgang interessiert sonder nur daran, ob das Ergebnis zu einer vorgegebenen Menge von Ergebnissen interessiert. Im Beispiel 1.1.1(e) interessiert man sich etwa lediglich dafür, wieviele Punkte

man für den Wurf erhält, ob man also beispielsweise in die Region für "18 Punkte" getroffen hat; wo genau man diese Region getroffen hat ist zweitrangig.

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

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) Das Ereignis $A = \{1, 2, 3\}$ ("Innerhalb der ersten drei Würfe fällt eine Sechs") ist beim Mensch-Ärgere-Dich-Nicht von Interesse.

(d) $A = (86400, \infty)$ ist das Ereignis "Die Glühbirne brennt länger als einen Tag".

(e) Ist der Radius des Bull's Eye $r \in (0, 1)$, so bezeichnet $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$ das Ereigniss "Bull's Eye!".

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

Beispiel 1.1.3. Beim Roulette ist die Grundmenge $\Omega = \{0, 1, 2, \ldots, 37\}$.

Sei $A = \{1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36\}$ das Ereignis "rouge". $B = (A^c) \setminus \{0\}$ ist das Ereignis "noir". $A \cup B$ ist das Ereignis "Es fällt nicht Null". $A \cap B$ ist das unmögliche Ereignis. A und B sind also unvereinbar.

Ist $C = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35\}$ die Menge der ungeraden Zahlen (das Ereignis "impair") so ist $A \cap C$ das Ereignis "Eine rote, ungerade Zahl fällt".

Beispiel 1.1.4. Eine Münze wird drei Mal geworfen. Wir wählen

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

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

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

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

Das Ereignis "Es fällt mindestens zweimal Kopf" lässt sich schreiben als

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

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

Leider ist dies keine einwandfreie Definition, da die relative Häufigkeit selbst zufallsbehaftet ist, also vom Zufallsexperiment abhängt. 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 1.1.5. Ein *endlicher Wahrscheinlichkeitsraum* ist ein Paar $(\Omega, \mathbb{P})$, bestehend aus einer endlichen Menge Ω und einer Abbildung $\mathbb{P} : \mathscr{P}(\Omega) \to [0,1]$, sodass

(i) $\mathbb{P}(\Omega) = 1$ (Normiertheit)

(ii) Sind $A, B \in \mathscr{P}(\Omega)$ 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 1.1.6. *Es sei $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum, A, B Ereignisse.*

(1) Ist $A \subset B$, so ist $\mathbb{P}(A) \leq \mathbb{P}(B)$. Weiter gilt $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. Insbesondere ist $\mathbb{P}(B^c) = 1 - \mathbb{P}(B)$ und $\mathbb{P}(\emptyset) = 0$.

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

Seien nun $A_1, \ldots, A_n$ Ereignisse.

(3) Es ist

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) \leq \sum_{j=1}^{d} \mathbb{P}(A_j)\,.$$

(4) Sind $A_1, \ldots, A_n$ paarweise unvereinbar, so ist

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) = \sum_{j=1}^{d} \mathbb{P}(A_j)\,.$$

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

(3) folgt induktiv aus (1) und (4) folgt induktiv aus Eigenschaft (ii) eines Wahrscheinlichkeitsmaßes.

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

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

Es folgt aus der endlichen Additivität eines Wahrscheinlichkeitsmaßes, Teil (4) von Propoisition 1.1.6, dass ein Wahrscheinlichkeitsmaß bereits durch die Werte auf den Elementarereignissen eindeutig bestimmt ist. In der Tat, ist $\Omega = \{\omega_1, \ldots, \omega_n\}$ und ist $p_j := \mathbb{P}(\{\omega_j\})$ so ist

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

Beachte, dass notwendigerweise $p_1, + \ldots + p_n = 1$ sein muss. Umgekehrt liefert aber auch jede Wahl von $p_1, \ldots, p_n$ mit Summe 1 ein Wahrscheinlichkeitsmaß.

Beispiel 1.1.7. 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

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

wählen. Mit dieser Wahl ergibt sich $\mathbb{P}(\{2, 4, 6\}) = p_2 + p_4 + p_6 = 1/2$.

Beachte, dass unsere Definition nicht zwingend verlangt, dass alle Elementarereignisse gleich wahrscheinlich sind. In der Tat könnte für einen gezinkten Würfel $p_6 = 0,25$ während $p_1, \ldots, p_5 = 0,15$ 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 1.1.8. Die Ecken eines Würfels werden gleichmäßig abgeschliffen, sodass der Würfel auch auf jeder der acht Ecken liegen bleiben kann. Dabei ist die Wahrscheinlichkeit einer jeden Ecke nur 1/4 so groß wie die Wahrscheinlichkeit einer jeden Seite. Wir wählen als Grundmenge $\Omega = \{s_1, \ldots, s_6, e_1, \ldots, e_8\}$ wobei s_j das Elementarereignis bezeichnet auf der Seite mit der Zahl j liegen zu bleiben. Die Ecken wurden von 1 bis 8 durchnummeriert und e_j bezeichnet das Elementarereignis auf der Ecke mit der Zahl j liegen zu bleiben.

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

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

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 *Laplace'schen Wahrscheinlichkeitsraum.* Ein Zufallsexperiment welches durch einen Laplaceschen Wahrscheinlichkeitsraum beschrieben wird nennt man häufig *Laplace'sches (Zufalls-)Experiment.*

Weil sich die Wahrscheinlichkeiten der Elementarereignisee 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 wir dieses Experiment als Laplace Experiment modellieren können. Hier ist man auf statistische Methoden angewiesen um ein geeignetes Modell zu finden.

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

Beispiel 1.2.2. 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 Grundram $\Omega = \{KK, KZ, ZZ\}$ 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 (denn die Münzen beeinflussen sich gegenseitig nicht). 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 $F_1, \ldots, F_k$ zu belegen und hat man n_1 Möglichkeiten Fach F_1 zu belegen, n_2 Möglichkeiten Fach F_2 zu belegen, ..., n_k Möglichkeiten Fach F_k zu belegen, so hat man insgesamt $n_1 \cdot \ldots \cdot n_k$ Möglichkeiten die k Fächer zu belegen.*

Beispiel 1.2.4. Thorsten hat ein Bewerbungsgespräch und stellt dazu ein passendes Outfit zusammen. Neben typischen Studentenklamotten (die er seinem zukünfitgen Arbeitgeber lieber ersparen will) findet er 3 Hemden, 4 Krawatten und 2 Anzüge in seinem Kleiderschrank. Er kann damit (von modischen Überlegungen abgesehen) $3 \cdot 4 \cdot 2 = 24$ verschiedene Outfits zusammenstellen.

Beispiel 1.2.5. 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. Insgesammt 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 $n \cdot (n-1) \cdot \ldots \cdot (n-(k-1))$ Möglichkeiten.

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

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

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

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

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

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

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

Ziehen mit Zurücklegen ohne Beachtung der Reihenfolge:
Dieses Art des Urnenmodells liegt etwa im Falle des 100-maligen Würfelns vor, wenn anschliessend nur mitgeteilt wird, wie oft jede Zahl gewürfelt wurde. Es spielt in der Praxis kaum eine Rolle, wir diskutieren es daher nicht weiter.

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 $(365)_n$ Elemente.

Folglich ist

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

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". Die Menge A besteht aus denjenigen 6-elementigen Teilmengen der Zahlen von 1 bis 49, die genau 4 der gezogenen 6 Zahlen (und 2 der nichtgezogenen 43 übrigen Zahlen) erhalten. Es gibt $\binom{6}{4}$ Möglichkeiten 4 der gezogenen Zahlen auszuwählen und zu jeder dieser Möglichkeiten gibt es $\binom{43}{2}$ Möglichkeiten diese mit nichtgezogenen Zahlen zu einem vollständigen "Tipp" aufzufüllen. Es ist also

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

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

Allgemeiner betrachtet man eine Urne mit n Kugeln, die 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 gelten dass $n_1 + n_2 + \cdots + n_r = n$ ist. Nun interessiert man sich fr das Ereignis beim Ziehen von k Kugeln genau k_j Kugeln der Farbe j zu ziehen, wobei $k_1 + \cdots + k_r = k$ ist. Es gibt

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

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

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

Beispiel 1.2.8. Wie hoch ist die Wahrscheinlichkeit beim Poker ein "Full House", d.h. zieht man zufällig 5 Karten aus Spiel mit 52 Karten, wie hoch ist die Wahrscheinlichkeit, 3 Karten einer Sorte und 2 Karten einer anderen Sorte zu ziehen.

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

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

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

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

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

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

1.3 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 Abwägung von Wahrscheinlichkeiten berücksichtigen. Beispielsweise wird er, wenn er selbst 2 Asse 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, das ein gewisses Ereignis B (z.B. "ich selbst habe 2 Asse") eingetreten ist. Im Falle eines Laplace Wahrscheinlichkeitsraums kann man diese Idee verwirklichen, indem man den Grundraum einschränkt: Man berücksichtig 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 $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum und B ein Ereignis mit positiver Wahrscheinlichkeit. Die *bedingte Wahrscheinlichkeit* $\mathbb{P}(A|B)$ *von* A *gegeben* B ist definiert als

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

Der Name bedingte *Wahrscheinlichkeit* ist gerechtfertigt:

Lemma 1.3.2. *Sei* $(\Omega, \mathbb{P})$ *ein endlicher Wahrscheinlichkeitsraum und* $B \subset \Omega$ *mit* $\mathbb{P}(B) > 0$. *Dann ist* $\mathbb{P}(\,\cdot\,|B)$ *ein Wahrscheinlichkeitsmaß auf* Ω *und auch ein Wahrscheinlichkeitsmaß auf* B.

Beweis. Weil $\emptyset \subset A\cap B \subset B$ ist folgt aus Proposition 1.1.6 (a), dass $0 = \mathbb{P}(\emptyset) \leq \mathbb{P}(A\cap B) \leq \mathbb{P}(B)$ und somit $0 \leq \mathbb{P}(A|B) \leq 1$ für alle $A \subset \Omega$ ist. Weil $\Omega \cap B = B$ gilt trivialerweise $\mathbb{P}(\Omega|B) = 1$. 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 $\mathbb{P}((A_1 \cup A_2) \cap B)) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B)$. Teilt man durch $\mathbb{P}(B)$, so folgt die Additivität von $\mathbb{P}(\,\cdot\,|B)$. □

Insbesondere gelten also die Aussagen von Proposition 1.1.6 für bedingte Wahrscheinlichkeiten.

Beispiel 1.3.3. Beim Skat werden 32 Karten wie folgt aufgeteilt: 10 Karten für jeden der drei Spieler und 2 Karten 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 und die übrigen zwei Karten kommen in den Skat).

Wir berechnen 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 zwei Asse hat.* 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

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

Der Sager erhält also in etwa 80 % der Fälle mindestens ein Ass.

Das Ereignis B, der Geber hat zwei Asse, tritt in $\binom{4}{2}\binom{28}{8}\binom{22}{10}\binom{12}{10}$ Fällen auf (zunächst geben wir dem Geber 2 Asse und 8 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 Asse, 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 Asse und 8 nicht-Asse, der Sager bekommt 10 von den verbliebenen 20 nicht-Assen, der Hörer 10 von den verbliebenen 10 nicht-Assen plus 2 Asse). 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 $\mathbb{P}(A|B) \approx 0,7143$.

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

Proposition 1.3.4. *(Multiplikationsregel) Es seien $A_1, \ldots, A_n$ Ereignisse in einem endlichen Wahrscheinlichkeitsraum derart, dass $\mathbb{P}(A_1 \cap \ldots \cap A_n) > 0$. Dann ist*

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

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 3 mal wiederholt. Es sei A_j das Ereignis, dass im j-ten Versuch eine rote Kugel gezogen wurde.

Mit der Multiplikationsregel ist

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

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

Nun kommen wir zu zwei wichtigen Aussagen über bedingte Wahrscheinlichkeiten. In der Formulierung verwenden wir den Begriff der *Partition.* Eine Partition einer Menge Ω ist eine Folge $\Omega_1, \ldots, \Omega_n$ derart, dass (i) die Ω_j paarweise disjunkt sind, also $\Omega_i \cap \Omega_j = \emptyset$ für $i \neq j$ und (ii) die Vereinigung aller Ω_j gerade Ω ist, also $\bigcup_{j=1}^{n} \Omega_j = \Omega$.

Satz 1.3.6. *Es sei $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum und $\Omega_1, \ldots, \Omega_n$ eine Partition von Ω.*

(1) (Satz von der totalen Wahrscheinlichkeit) Für $A \subset \Omega$ 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) (Satz von Bayes) Für $A \subset \Omega$ mit $\mathbb{P}(A) > 0$ gilt

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

Beweis. (1) Weil Ω die disjunkte Vereinigung der Ω_j ist, ist A die disjunkte Vereinigung der $A \cap \Omega_j$. Wegen der 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) folgt aus (1) und der Gleichheit $\mathbb{P}(A \cap \Omega_j) = \mathbb{P}(A|\Omega_j)\mathbb{P}(\Omega_j)$. □

Beispiel 1.3.7. 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

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

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

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

Beispiel 1.3.8. Ein Test auf eine bestimmte Kranheit erkennt mit 99% Wahrscheinlichkeit eine erkrankte Person als krank (= "Sensitivität"). Die Wahrscheinlichkeit, dass eine nicht erkrankte Person als gesund erkannt wird betrage 95% (= "Spezifität"). Es ist bekannt, dass 1% der Bevölkerung an dieser Krankheit leiden. Wie groß ist die Wahrscheinlichkeit an der Krankheit zu leiden, wenn der Test zu dem Ergebnis kommt, dass man krank ist?

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

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

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

Beispiel 1.3.9. Bei einer Spielshow kommt folgende Situation vor. Der Kandidat sieht drei (verschlossene) Tore. Hinter einem Tor befindet sich der Hauptpreis, ein Auto. Hinter den beiden anderen Toren befindet sich eine Ziege. Der Kandidat wählt ein Tor aus. Danach öffnet der Moderator ein Tor, hinter dem sich eine Ziege verbirgt. Er bietet sodann dem Kandidaten an, doch noch das Tor zu wechseln. Sollte der Kandidat das Tor wechseln oder bei seiner urspünglichen Wahl bleiben?

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

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

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

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

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

Definition 1.3.10. Es sei $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum $A, B \subset \Omega)$ Ereignisse. Dann heißen A und B *unabhängig*, wenn $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Allgemeiner heißt eine Familie $A_1, \ldots, A_n$ von Ereignisses *unabhängig*, falls für alle $k = 1, \ldots, n$ und alle Indizes $1 \leq i_1 < \cdots < i_k \leq n$ stets

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

gilt.

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

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

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:

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

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

1.4 Wiederholung von Zufallsexperimenten

Wir erinnern daran, dass Wahrscheinlichkeiten in gewissem Sinne relative Häufigkeiten repräsentieren sollen. Relative Häufigkeiten wiederum kann man bilden da wir 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 $(\Omega_1, \mathbb{P}_1), \ldots, (\Omega_n, \mathbb{P}_n)$ für gewisse Zufallsexperimente. Im Falle der Wiederholung haben wir also $(\Omega_i, \mathbb{P}_i) = (\Omega, \mathbb{P})$ für einen festen Wahrscheinlichkeitsraum. Führen wir diese Experimente nacheinander und unabängig voneinander aus, so bietet sich als Grundmenge für das zusammengesetzte Experiment das kartesische Produkt

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

an. Jedes Versuchsergebnis $\omega \in \bar{\Omega}$ ist also ein n-Tupel $(\omega_1, \ldots, \omega_n)$ 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

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

Wir nennen $(\bar{\Omega}, \bar{\mathbb{P}})$ das *Produkt der Wahrscheinlichkeitsräume* $(\Omega_1, \mathbb{P}_1), \ldots, (\Omega_n, \mathbb{P}_n)$. Weiter schreiben wir $\bar{\Omega} = \Omega_1 \times \cdots \times \Omega_n$ und $\bar{\mathbb{P}} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n$.

Beispiel 1.4.1. Es sei $\Omega_1 = \{K, Z\}$ und $\mathbb{P}(\{K\}) = \mathbb{P}(\{Z\}) = \frac{1}{2}$. Dann beschreibt $(\Omega_1, \mathbb{P}_1)$ das Werfen einer fairen Münze. Es sei weiter $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$ und $\mathbb{P}(\{j\}) = \frac{1}{6}$, sodass $(\Omega_2, \mathbb{P}_2)$ 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 Laplacscher 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 $j \in \{1, 2, \ldots, 6\}$ 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 kann man sich für das Ereignis "Im i-ten Experiment tritt Ereignis A_i ein". Dieses Ereignis wird in unserem zusammengesetzen Modell beschrieben durch die Menge

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

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

Wir rechnen dies nach 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

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

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 $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum und betrachte $\bar{\Omega} := \Omega^n$ und $\bar{\mathbb{P}} = \mathbb{P}^{\otimes n}$.

Beispiel 1.4.2. Wir betrachten ein Ereignis $A \subset \Omega$ 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 $E_0 = A^c \times \cdots \times A^c$ und daher $\bar{\mathbb{P}}(E_0) = \mathbb{P}(A^c) \cdot \ldots \cdot \mathbb{P}(A^c) = (1-p)^n$. Für $k = 1$ zerlegen wir das Ereignis E_1 wie folgt:

$$E_1 = A \times A^c \times \ldots \times A^c \cup A^c \times A \times A^c \times \cdots \times A^c \cup \ldots \cup A^c \times \cdots \times A^c \times A.$$

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 Anzal der Erfolge, so erhählt man eine Zahl in der Menge $\{0, 1, \dots, n\}$. Dabei ist die Wahrscheinlichkeit $\tilde{\mathbb{P}}(\{k\})$ dass man genau k-mal Erfolg hatte gerade $\binom{n}{k} p^k (1-p)^{n-k}$. Beachte, dass wegen des Binomialsatzes

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

ist, $\tilde{\mathbb{P}}$ also in der Tat ein Wahrscheinlichkeitsmaß ist. Dieses Maß nennt man *Binomialverteilung* mit Parametern n und p, oder auch $\mathsf{b}_{n,p}$ Verteilung.

Beispiel 1.4.3. 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 natürliche Grundmenge $\Omega = \mathbb{N}$ nicht endlich ist. Allerding scheint es natürlich, die Wahrscheinlichkeit des Elementarereignisses k ("Im k-ten Versuch hat man zum ersten Mal Erfolg") als $p(1-p)^{k-1}$ anzusetzen, 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 erhählt man

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

Also summieren sich auch hier die Wahrscheinlichkeiten der Elementarereignisse zu 1 auf und wir sind versucht ein Wahrscheinlichkeitsmaß auf $\mathbb{N}$ via

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

zu definieren. Dies ergibt in der Tat Sinn und wir modifizieren unser Definition eines Wahrscheinlichkeitsraumes entsprechend.

Definition 1.4.4. Ein *diskreter Wahrscheinlichkeitsraum* ist ein Paar $(\Omega, \mathbb{P})$, bestehend aus einer höchstens abzählbaren Menge Ω und einer Abbildung $\mathbb{P} : \mathscr{P}(\Omega) \to [0, 1]$, derart, dass

(i) $\mathbb{P}(\Omega) = 1$ (Normiertheit)

(ii) Ist $(A_k)_{k \in \mathbb{N}}$ eine Folge paarweise disjunkter Teilmengen von Ω, so ist

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

Diese Eigenschaft heißt *σ-Additivität*.

Bemerkung 1.4.5. (a) Eine Menge heißt *höchstens abzählbar*, wenn sie entweder endlich, oder abzählbar unendlich ist. Ist die Menge Ω endlich, so ist die σ-Additivität äquivalent zur endlichen Additivität die in Definition 1.1.5 gefordert wird. Mit Ω ist nänlich auch $\mathscr{P}(\Omega)$ endlich sodass wenn (A_k) eine Folge paarweise disjunkter Mengen ist, alle bis auf endlich viele Mengen leer sein müssen.

(b) Ist $(\Omega, \mathbb{P})$ 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 erhählt man aus der σ-Additivität, dass

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

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.

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. die Körpergröße oder das Einkommen der Person.

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

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".

Es ist $X : \Omega \to \mathbb{R}$ 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 $(\Omega, \mathbb{P})$ wird nur noch $(X(\Omega), \mathbb{P}_X)$ betrachtet. Ein ähnliches Vorgehen hatten wir bereits bei der Diskussion von Beispiel 1.1.1(c) in Beispiel 1.4.3 gesehen. Es gibt keinen diskreten Wahrscheinlichkeitsraum, der das Zufallsexperiment beschreibt einen Würfel unendlich oft zu werfen. Wenn wir uns jedoch lediglich dafür interessieren, wie lange es dauert bis zum ersten Mal eine Sechs gewürfelt wird, so können wir für diese "Zufallsvariable" sehr wohl eine sinnvolle Verteilung definieren. Das ist genau, was wir in Beispiel 1.4.3 gemacht haben.

Betrachten wir nochmals Beispiel 1.4.2. 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 $X : \bar{\Omega} \to \mathbb{R}$ 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. Die Notation $\{\omega \in \Omega : X(\omega) = x\}$ ist relativ umständlich. Oft schreibt man stattdessen $\{X = x\}$. 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 $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ 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 undendlich viele Summandan, 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 $\sum_{x\in X(\Omega)}|x|\mathbb{P}(X=x)=\infty$.

(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 (welcher zu bestimmen ist). Anschliessend zieht er eine Karte aus einem Kartenspiel (französisches Blatt, d.h. 32 Karten). Zieht er ein Ass, so werden ihm €10 ausgezahlt. Zieht er eine Bildkarte (Bube, Dame, König) so erhält er €2 ausgezahlt. Ansonsten erhählt er nichts. Für welchen Wert von E ist diese Spiel *fair*, d.h. der erwartete Gewinn ist €0.

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

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

Demnach ist das Spiel genau dann fair, wenn der Einsatz €2 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 $(\Omega,\mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X:\Omega\to\mathbb{R}$ 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, ist $X(\omega) = \{x_j\}_{j\in J}$ wobei die Indexmenge J entweder endlich oder abzählbar unendlich ist. Es sei A_j das Ereignis $\{X = x_j\}$. Weil Ω abzählbar ist ist $A_j = \{\omega_{j,k}\}_{k\in K_j}$, wobei K_j eine endliche oder abzählbar unendliche Indexmenge ist.

Nun ist

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

Hier haben wir beim zweiten Gleichheitszeichen die σ-additivität von $\mathbb{P}$ verwendet, beim dritten die Gleichheit $x_j = X(\omega_{jk})$ und bei der letzten die Tatsache das mit der Summation über alle j und k ganz Ω ausgeschöpft wird. Bei der letzten Gleichheit verwenden wir den "großen Umordnungssatz der Analysis" der es erlaubt die Summanden einer absolut konvergenten Reihe beliebig zu Gruppieren und aufzusummieren. □

Nun folgt sofort:

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

(1) Ist $0 \le Y \le X$ und hat X endlichen Erwartungswert, so hat auch Y endlichen Erwartungswert und es gilt $\mathbb{E}Y \le \mathbb{E}X$.

(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

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

nach Voraussetzung.

(2) Beachte, dass

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

ist. Aufgrund der Rechenregeln für Reihen mit positiven Einträgen ist

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

Nach Teil (1) hat also $\alpha X + \beta Y$ endliche Erwartung. Der Rest der Behauptung folgt aus den Rechenregeln für absolut konvergente Reihen. □

Wir betrachten nun ein weiteres Beispiel.

Beispiel 1.5.9. 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. Dazu benützt man die *Varianz*

Definition 1.5.10. Es sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit endlichem Erwartungswert μ. Ist die Reihe

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

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

Beispiel 1.5.11. Wir berechnen die Varianz der Zufallsvariablen X_n aus Beispiel 1.5.9. Weil $\mathbb{E}X_n \equiv 0$ ist, folgt

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

Somit hat X_n Varianz n^2 und Streuung n.

Beispiel 1.5.12. Wir berechnen die Varianz des Spieles in Beispiel 1.5.6 bei beliebigem Einsatz E. Es ist $\mathbb{E}X = 2 - E$. Daher

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

Beachte, dass die Varianz nichts weiteres ist als der Erwartungswert der Zufallsvariablen $g(X)$ wobei die Funktion $g : \mathbb{R} \to \mathbb{R}$ gegeben ist durch $g(t) = (t-\mu)^2$ wo $\mu = \mathbb{E}X$. Beachte, dass $g(t) = t^2 - 2\mu t + \mu^2$ ist. Aus dieser Gleichheit und der Linearität des Integrals (Proposition 1.5.8(2)) folgt

$$\mathsf{Var}X = \mathbb{E}\big(X^2 - 2\mu X - \mu^2\big) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2 .$$

Beachte weiterhin, dass X endliche Varianz hat genau dann, wenn X^2 endlichen Erwartungswert hat. Das folgt aus Proposition 1.5.8, denn $(X-\mu)^2 = X^2 + Y$ wobei $Y = \mu^2 - 2\mu X$ eine Zufallsvariable mit endlichem Erwartungswert ist. Hat also $(X-\mu)^2$ endlichen Erwartungswert, so auch $X^2 = (X-\mu)^2 - Y$. Hat umgekehrt X^2 endlichen Erwartungswert, so auch $X^2 + Y$. Somit gilt

Lemma 1.5.13. *Seit $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable mit endlichem Erwartungswert. Dann hat X genau dann endliche Varianz, wenn $\mathbb{E}(X^2) < \infty$. In diesem Fall ist $\mathsf{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.*

Manchmal ist es auch nützlich eine Zufallsvariable X mit anderen Funktionen g zu verknüpfen. Wir geben einige Beispiele:

Beispiel 1.5.14. Eine *europäische Kaufoption* (engl. *european call option*) beinhaltet das Recht (aber nicht die Pflicht) ein bestimmtes Gut (Basiswert oder underlying) an einem in der Zukunft liegenden Zeitpunkt (Ausübungszeitpunkt) zu einem heute festgelegten Preis (strike price) zu kaufen.

Wir betrachten eine europäische Kaufoption für eine Aktie, deren Kurs zum Ausübungszeitpunkt durch eine Zufallsvariable S gegeben ist. Der strike price sei K. Die Auszahlung aus dem Ausüben der Option (engl. payoff) ergibt sich dann wie folgt:

Ist der Wert der Aktie größer als K, so beträgt die Auszahlung $S - K$, denn wir können ja eine Aktie zum Preis von K kaufen und diese sofort zum Preis von S am Markt wieder verkaufen. Ist hingegen der Wert der Aktie kleiner als K, so ist die Option wertlos, denn wir können die Aktie ja billiger am Markt kaufen (d.h. wir üben die Option nicht aus).

Definieren wir $x^+ = \max\{x, 0\}$ so ist die Auszahlung gegeben als $(S - K)^+$, d.h. als Verknüpfung des Aktienkurses mit der sog. payoff funktion $g(t) := (t - K)^+$.

Beispiel 1.5.15. Eine *Indikatorfunktion* ist eine Funktion die nur die Werte 1 und 0 annimmt. Ist M eine Menge und $A \subset M$, 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.

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

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

Weiterhin von Bedeutung sind Verknüpfungen mit den Funktionen $g(t) := |t|^r$.

Definition 1.5.16. Sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable und $r \in \mathbb{N}$. Wir sagen dass X ein *endliches Momente r-ter Ordnung besitzt*, oder dass X *endliche r-te Momente besitzt* falls $|X|^r$ endlichen Erwartungswert besitzt. In diesem Fall heißt $\mathbb{E}X^r$ *Moment r-ter Ordnung.*

Bemerkung 1.5.17. Ist $k \leq r$, so ist

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

Folglich ist $|X|^k \leq \gamma(1 + |X|^r)$. Folglich 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 auftauchen.

Die Binomialverteilung

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

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

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

Eine Binomialverteilung zählt die Anzahl der Erfolge in n Versuchen wobei die Erfolgswahrscheinlichkeit p beträgt, siehe Beispiel 1.4.2.

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 noch die Varianz einer binomialverteilten Zufallsvariablen. Wir wissen, dass $\mathsf{Var}X = \mathbb{E}X^2 - (np)^2$ 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

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

Die hypergeometrische Verteilung

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

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

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

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

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

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

sein. Äquivalent hierzu ist

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}. \tag{1.2}$$

Ruft man sich die Interpretation der hypergeometrischen Verteilung ins Gedächtnis, so ist die Gleichheit (1.2) klar:

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

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

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

wobei wir die Gleicheit (1.2) im letzten Schritt benutzt haben. 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.

Dies zeigt, dass sich eine hypergeometrische Verteilung annährend wie eine Binomialverteilung verhält. Allerdings zieht man bei der Binomialverteilung quasi "aus einem unendlichen Vorrat" (oder auch mit Zurücklegen) zieht, nachdem man eine schwarze Kugel gezogen hat, ist die Wahrscheinlichkeit wieder eine schwarze Kugel zu ziehen unverändert. Bei der hypergeometrischen Verteilung ist dies anders. Dieser Unterschied zeigt sich bereits bei der Varianz. Ist $X \sim \mathsf{hg}_{m,n,k}$ so sieht man ähnlich wie oben, dass

$$\mathbb{E}(X(X-1)) = k(k-1)\frac{m(m-1)}{(m+n)(m+n-1)}.$$

Somit

$$\begin{aligned}\mathsf{Var}X &= \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}X - (\mathbb{E}X)^2 \\ &= k(k-1)\frac{m(m-1)}{(m+n)(m+n-1)} + k\frac{m}{m+n} - k^2\frac{m^2}{(m+n)^2} \\ &= k\frac{m}{m+n}\cdot\frac{(k-1)(m-1)(m+n) + (m+n)(m+n-1) - km(m+n-1)}{(m+n)(m+n-1)} \\ &= k\frac{m}{m+n}\cdot\frac{(m+n-k)n}{(m+n)(m+n-1)} \\ &= k\frac{m}{m+n}\Big(1-\frac{m}{m+n}\Big)\frac{m+n-k}{m+n-1}.\end{aligned}$$

Vergleicht man dies mit der Varianz einer binomialverteilten Zuvallsvariablen mit Parameter k und Erfolgswahrscheinlichkeit $p = \frac{m}{m+n}$, 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

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

Eine geometrische Verteilung beschreibt beim Wiederholen eines Zufallsexperimentes mit Erfolgswahrscheinlichkeit p die Anzahl der Versuche bis zum ersten Erfolg, siehe Beispiel 1.4.3.

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

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

Ist also $X \sim \mathsf{geom}_p$, 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

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

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

Die Poisson Verteilung

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

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

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

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

$$\begin{aligned}
\mathsf{b}_{n,p_n}(\{k\}) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \frac{1}{k!} n(n-1)\cdot \ldots \cdot (n-k-1) \Big(\frac{\lambda_n}{n}\Big)^k \Big(1-\frac{\lambda_n}{n}\Big)^{n-k} \\
&= \frac{1}{k!}\frac{n}{n}\cdot\frac{n-1}{n}\cdot \ldots \cdot \frac{n-k-1}{n} \lambda_n^k \Big(1-\frac{\lambda_n}{n}\Big)^{n} \cdot \Big(1-\frac{\lambda_n}{n}\Big)^{-k} \\
&\to \frac{1}{k!}\cdot 1\cdot 1\cdot \ldots \cdot 1 \cdot \lambda^k e^{-\lambda}(1-0)^{-k},
\end{aligned}$$

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

Proposition 1.6.2. *Ist p_n eine Folge in $(0,1)$ mit $np_n \to \lambda > 0$, so gilt*

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

für jedes $k = 0, 1, 2, \ldots$.

Bemerkung 1.6.3. In den Anwendungen hat man in der Regel keine Folge p_n von Erfolgswahrscheinlichkeiten in Binomialexperimenten gegeben. Man interpretiert Proposition 1.6.2 daher wie folgt:

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

Wir verdeutlichen dies an einem Beispiel.

Beispiel 1.6.4. Radioaktiven Zerfall kann man wie folgt modellieren:

Jeder einzelne Atomkern hat (innerhalb einer gewissen Zeit, etwa 10 Sekunden) eine gewisse Wahrscheinlichkeit p zu zerfallen oder auch nicht. Man kann jedoch nicht einzelne Atomkerne betrachten. Stattdessen betrachtet man eine bestimmte Menge eines Radioaktiven Materials und zählt die Zerfälle in einer Sekunde. Diese Anzahl ist dann binomialverteilt mit Parametern n (=Anzahl der Atomkerne) und p. Typischerweise ist n sehr groß, (die Einheit mol der Stoffmenge ist die Anzahl der Atome in 12 Gramm eines bestimmten Kohlenstoffisotops. 1 mol sind ungefähr $6\cdot 10^{23}$ = viel) die Anzahl der Zerfälle die man in Versuchen beobachtet ist aber vergleichsweise klein (Größenordnung 5 Zerfälle je 10 Sekunden).

Demnach ist es plausibel, eine Poissonverteilung zu unterstellen. Nun muß jedoch der Parameter λ "geschätzt" werden. Damit werden wir uns im nächsten Kapitel beschäftigen.

Wir diskutieren ein weiteres Beispiel.

Beispiel 1.6.5. Befinden sich n Personen in einem Raum so ist die Zufallsvariable X, die die Anzahl der Personen, die heute Geburtstag haben, Binomialverteilt mit Parametern n und $p = 1/365$. Wir betrachten die Situation wo $n = 97$. Damit ist $X \sim \mathsf{b}_{97,\frac{1}{365}}$. Wir können diese Verteilung mit der Poisson Verteilung zum Parameter $\lambda = 97/365$ approximieren. Die folgende Tabelle gibt einige vergleichende Werte von Binomial und Poisson Verteilung an.

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

Wir berechnen nun Erwartungswert und Varianz einer Poission Verteilten Zufallsvariablen. Ist $X \sim \mathsf{Pois}_\lambda$, 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 $\mathsf{Var}X = \lambda^2 + \lambda - \lambda^2 = \lambda$.

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

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

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

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

auch die *Zähldichte* des Vektors X.

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. Beipielsweise gilt

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

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

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

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

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

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

Beispiel 1.7.3. Es sei $c \in [0, \frac{1}{2}]$. Die Verteilung des Vektors (X, Y) sei gegeben durch

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

In diesem Beispiel hängen die Randverteilungen nicht von c ab: X und Y nehmen die Werte 0 und 1 jeweils mit Wahrscheinlichkeit $1/2$ an. 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 $f : \mathbb{R}^n \to \mathbb{R}$ gegeben, so ist $f(X)$ eine Zufallsvariable. Einige interessante Kenngrößen von Zufallsvektoren sind als Erwartungwert 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. Somit ist folgende Definition sinnvoll:

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

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

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

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

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

Beispiel 1.7.5. Wie Berachten wiederum Beispiel 1.7.3. Es ist $\mathbb{E}X = \mathbb{E}Y = \frac{1}{2}$ und $\mathsf{Var}X = \mathsf{Var}Y = \frac{1}{4}$. Weiter ist

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

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

Ist die Korrelation negativ (also $c < 1/4$) so ist es genau andersherum.

Beispiel 1.7.6. Wir betrachten nochmals Beispiel 1.7.2. Es ist

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

Weiterhin ist

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

und

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

Für die Kovarianz von R und W finden wir

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

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 $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und X, Y Zufallsvariablen mit endlichen zweiten Momenten.*

(1) Es gilt die Cauchy-Schwarz'sche Ungleichung:

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

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

(3) Es ist

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

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

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 $\mathbb{E}XY \leq \frac{1}{2t}\mathbb{E}X^2$ für alle $t > 0$ und somit $\mathbb{E}XY \leq 0$. Weiterhin ist $\mathbb{E}XY \geq \frac{1}{2t}\mathbb{E}X^2$ für alle $t < 0$ und somit $\mathbb{E}XY \geq 0$. Insgesamt ist $\mathbb{E}XY = 0$ und daher gilt die Ungleichung.

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

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

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). □

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 $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Zufallsvariablen $X_1 1, \ldots, X_n$ heißen *unabhängig*, falls für alle Teilmengen $A_1, \ldots, A_n \subset \mathbb{R}$ stets

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

gilt. Eine Folge $X_1, X_2, X_3, \ldots$ heißt unabhängig, falls $X_1, \ldots, X_n$ für alle $n \in \mathbb{N}$ unabhängig sind.

Bemerkung 1.7.9. (a) Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann unabängig, wenn für alle Wahlen von $A_1, \ldots, A_n \subset \mathbb{R}$ die Ereignisse $\{X_1 \in A_1\}, \ldots, \{X_n \in A_n\}$ unabhängig sind. Wir müssen hier keine Teilfamilien (wie in Definition 1.3.10) betrachten, weil wir gewisse $A_j = \mathbb{R}$ wählen können, sodass $\{X_j \in A_j\} = \Omega$ ist.

(b) Es sei $\{x_j : j \in J\}$ die Menge aller Werte die von allen Zufallsvariablen $X_1, \ldots X_n$ angenommen werden können. Es ist leicht zu sehen, dass $X_1, \ldots, X_n$ unabhängig sind genau dann, wenn

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

für alle Wahlen von $x_{j_1}, \ldots, x_{j_n}$. Mit anderen Worten, es genügt einelementige Mengen zu betrachten. Anschaulich kann man dies an einem Tableau wie in Beispiel 1.7.2 beschreiben. Zwei Zufallsvariablen sind unabhängig genau dann, wenn sich die Zähldichte des Vektors als Produkt der Zähldichten der Randverteilungen ergibt.

Beispiel 1.7.10. Die Zufallsvariablen R und W in Beispiel 1.7.2 sind nicht unabhängig. Beispielsweise ist

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

Beispiel 1.7.11. 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

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

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 1.7.12. *Sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum, X, Y unabhängige Zufallsvariablen mit endlichem zweiten Moment. Dann sind X und Y unkorreliert.*

Beweis. Sei $\mu = \mathbb{E}X, \nu = \mathbb{E}Y$. Weil $(X - \mu)(Y - \nu) = XY - \mu Y - \nu X + \mu\nu$ ist, liefert die Linearität des Erwartungswertes $\mathsf{Cov}(X, Y) = \mathbb{E}(XY) - \mu\nu$. Es genügt also zu zeigen, dass $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$ ist. Dies folgt aber sofort aus

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

□

Das folgende Beispiel zeigt, dass die Umkehrung von Lemma 1.7.12 nicht gilt.

Beispiel 1.7.13. 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

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

Allerdings ist

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

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. Wir illustrieren dies an einigen Beispielen.

Proposition 1.7.14. *Es sei $(\Omega, \mathbb{P})$ 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

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

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. □

Wir diskutieren einige Beispiele:

Beispiel 1.7.15. Es seien X, Y unabhängige Zufallsvariablen mit $X \sim \mathsf{Pois}_{\lambda_1}$ und $X \sim \mathsf{Pois}_{\lambda_2}$. Dann ist $X + Y \sim \mathsf{Pois}_{\lambda_1+\lambda_2}$. In der Tat, folgt aus Proposition 1.7.14, dass

$$\begin{aligned}\mathbb{P}(X + Y = n) &= \sum_{k=0}^{n} \mathbb{P}(X = n)\mathbb{P}(Y = n - k) = \sum_{j=0}^{n} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda^2)}{n!}.\end{aligned}$$

Beispiel 1.7.16. Es seien X, Y unabhängige Zufallsvariablen mit $X \sim \mathsf{b}_{n_1,p}$ und $Y \sim \mathsf{b}_{n_2,p}$. Dann ist $X + Y \sim \mathsf{b}_{n_1+n_2,p}$. In der Tat folgt aus Proposition 1.7.14, dass (wir definieren $\binom{a}{b} = 0$ für $b > a$)

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

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 ist gerade Formel (1.2)

1.8 Das Gesetz der großen Zahlen

Wir hatten Wahrscheinlichkeiten so definiert, dass sie gewisse Eigenschaften von *relativen Häufigkeiten* wiederspiegeln. 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: $X_1, X_2, X_3, \ldots$. Dabei ist $X_n = 1$, wenn A im n-ten Versuch eingetreten ist, sonst ist $X = 0$. Wenn wir das Experiment unabhängig wiederholen, sind die Zufallsvariablen X_j unabhängig. Definieren wir $S_n := X_1 + \ldots + X_n$, 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 zunäst Erwartungswert und Varianz einer binomialverteilten Zufallsvariable auf andere Art und weise.

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

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

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). Somit

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

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

Satz 1.8.2. *(Tschebyscheff Ungleichung)*
Es sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable mit endlicher Varianz. Dann gilt für $\varepsilon > 0$

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

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

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

was äquivalent zur Behauptung ist. □

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

Beispiel 1.8.3. Sei X eine Zufallsvariable mit endlichen Momenten zweiter Ordnung. 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:

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

Werfen wir beispielsweise eine Münze 1000 mal, so ist die Anzahl X der "Köpfe" binomial verteilt mit Parametern $n = 1000$ und $p = \frac{1}{2}$. Demnach ist $\mathbb{E}X = np = 500$ und $\mathsf{Var}X = np(1-p) = 250$, 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 werden mit Wahrscheinlichkeit mindestens 88% zwischen 450 und 550 Köpfe geworfen.

Satz 1.8.4. *(Schwaches Gesetz der großen Zahlen)*
Es sei X_n 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

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n} X_n - \mu\Big| \geq \varepsilon\Big) = 0 \tag{1.3}$$

Beweis. Wir schreiben $S_n := n^{-1}(X_1 + \ldots + X_n)$ Wegen der Linearität des Erwartungswerts $\mathbb{E}S_n = \mu$. Wegen der Unabängigkeit der Zufallsvariablen folgt aus Proposition 1.7.7(4), dass

$$\mathsf{Var}S_n = \frac{1}{n^2}\sum_{k=1}^{n} \mathsf{Var}X_n = \frac{\sigma^2}{n}\,.$$

Nun folgt aus der Tschebyscheff Ungleichung

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

für $n \to \infty$. □

Bemerkung 1.8.5. (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 mindestens zwei verschiedene Werte 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 (1.3) sagt man S_n konvergiert *in Wahrscheinlichkeit* gegen μ. Allgemeiner sagt man eine Folge Y_n von Zufallsvariablen konvergiert in Wahrscheinlichkeit gegen Z, falls $\mathbb{P}(|Y_n - Z| \geq \varepsilon) \to 0$ für alle $\varepsilon > 0$.

Beispiel 1.8.6. Im Falle, dass $X = \mathbb{1}_A$ ein Indikator ist (dann ist $\mathbb{E}X = \mathbb{P}(A)$ und $\mathsf{Var}X = \mathbb{P}(A)(1 - \mathbb{P}(A))$ betrachtet man eine Folge unabängiger Zufallsvariablen X_n die die gleiche Verteilung wie X hat. ("das Zufallsexperiment wird unendlich oft wiederholt"). In diesem Fall ist S_n gerade die relative Häufigkeit von A. Das schwache Gesetz der großen Zahlen besagt, dass $S_n \to \mathbb{P}(A)$ in Wahrscheinlichkeit.

Beispiel 1.8.7. 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. Zum klären dieser Frage verwenden wir die Abschätzung aus Satz 1.8.4. Sei hierzu

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

Dann ist X_n Bernoulli verteilt mit Erfolgswahrscheinlichkeit $p = 0,5$. Insbesondere ist $\mathbb{E}X_n \equiv 0,5$ und $\mathsf{Var}(X_n) \equiv 0,25$. Aus der Tschebyscheff Ungleichung folgt

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

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

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

Man muss also mindestens 50.000 mal werfen.

Ohne Beweis geben wir noch an:

Satz 1.8.8. *(Starkes Gesetz der großen Zahlen)*
Es sei X_n eine Folge unabhängiger und identisch verteilter (d.h. alle Zufallsvariablen haben die gleiche Verteilung) Zufallsvariablen. Weiter sei $\mathbb{E}X_1 = \mu$. Dann ist

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

Mit anderen Worten, die Menge derer ω sodass $n^{-1}(X_1(\omega)+\ldots+X_n(\omega))$ gegen μ konvergiert, hat Wahrscheinlichkeit 1.

Kapitel 2

Schätzen von Parametern

Bisher haben wir eine mathematische Theorie entwickelt, die es uns erlaubt, gewisse zufällige Phänomene zu modellieren. In Beispiel 1.6.4 hatten wir erklärt, warum die Poisson Verteilung ein plausibles Modell für den radioaktiven Zerfall ist. Allerdings liefert uns unsere 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 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. 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" (etwa 1000 mal), zählt wie oft sie auf der flachen Seite landet (etwa k mal) und nimmt dann $k/1000$ als Nährungswert für p.

Dies ist ein typisches Beispiel eines *Schätzproblemes*: Man möchte einen Nährungswert 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

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

Definition 2.1.1. Es sei F eine Wahrscheinlichkeitsverteilung auf einer abzählbaren Menge M. Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Verteilung F.

Dann nennt man $X_1, \ldots, X_n$ eine *Zufallsstichprobe* vom Umfang n zur Verteilung F. Die Werte $x_1 = X_1(\omega), \ldots, x_n = X_n(\omega)$ heißen *Realisierung der Zufallsstichprobe.* Die Menge M^n aller potentiell möglichen Realisierungen einer Stichprobe nennt man *Stichprobenraum.*

Typischerweise ist die Verteilung F nicht 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 $X_1, \ldots, X_n$ eine Zufallsstichprobe zur Verteilung F. 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. $S := \sqrt{S^2}$ heißt Stichprobenstandardabweichung. Die konkreten, auf einer Realisierung der Stichprobe basierenden, Werte werden häufig mit kleinen Buchstaben bezeichnet:

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

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

Tag	1	2	3	4	5	6	7	8	9	10
Anzahl der Gäste	44	37	49	52	30	43	43	47	39	50

In diesem Fall sind weder die Verteilung F (von der wir zumindest unterstellen, dass es sie gibt) noch die Zufallsstichprobe $X_1, \ldots, X_{10}$ konkret bekannt. Die obige Tabelle gibt lediglich eine Realisierung $x_1, \ldots, x_{10}$ der Stichprobe an. Für diese Realisierung haben wir

$$\bar{x} = \frac{1}{10}\big(44 + 37 + 49 + 52 + 30 + 43 + 43 + 47 + 39 + 50\big) = 43,4$$

und

$$s^2 = \frac{1}{9}\big(0,6^2 + 6,4^2 + 5,6^2 + 8,6^2 + 13,4^2 + 0,4^2 + 0,4^2 + 3,6^2 + 4,4^2 + 6,6^2\big) = 44,71$$

was einer Standardabweichung von etwa $6,69$ entspricht.

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

(1) $\mathbb{E}\bar{X} = \mu$.

(2) $\mathsf{Var}\bar{X} = \sigma^2/n$

(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 $X_1, \ldots, X_n$ gilt

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

(3) Wir haben

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

Hier haben wir in der dritten Gleichheit die Definition von $\mathsf{Var}X$ 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. □

2.2 Schätzen von Parametern

Wir disktieren nun sogenannte *parametrische Modelle*. Hierbei ist die Verteilung F, die wir näher untersuchen wollen, zwar unbekannt, es ist aber bekannt (oder wir nehmen es zumindest an), dass F zu einer bestimmten Familie $\{F_\theta : \theta \in \Theta\}$ 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 (siehe etwa Beispiel 1.6.4), so kann man die Familie $F_\theta = \mathsf{Pois}_\theta$ betrachten. Hier verwenden wir $\Theta = (0, \infty) \subset \mathbb{R}$.

(b) Im Falle des Restaurant Besitzers aus Beispiel 2.1.3 erscheint es plausibel in erster Nährung anzunehmen, dass die zugrunde liegende Verteilung eine Binomialverteilung ist:

Es gibt eine gewisse Anzahl n potentieller Gäste, die unabhängig voneinander jeden Tag entscheiden, ob sie essen gehen oder nicht. Dies geschieht mit Wahrscheinlichkeit p.[1] Wir könnten also die Familie $\mathsf{b}_{n,p}$ verwenden, wobei $\theta = (n, p) \in \Theta = \mathbb{N} \times [0, 1]$ ist.

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

Häufig ist $g(\theta) = \theta$, es ist also der Parameter selbst zu schätzen. Es gibt aber auch andere Beispiele, man denke etwa an das Problem, die Varianz einer Binomialverteilung zu schätzen. Häufig unterscheiden wir nicht zwischen der Funktion $T : M^n \to \mathbb{R}^m$ und dem Schätzer $T(X_1, \ldots, X_n)$.

Wir geben einige Beispiele von Schätzern.

[1] Offensichtlich hat dieses Modell Schwächen, denn die Gäste kommen oft in Gruppen, entscheiden also nicht unabhängig voneinander. Auch ist nicht klar ob wir unser Stichprobe wirklich als Folge *unabhängiger* Zufallsvariablen auffassen können. Vielleicht treffen die Gäste ihre Entscheidung nicht unabhängig von der des Vortages (Wer will schon zweimal hintereinander im gleichen Restaurant essen?)

Beispiel 2.2.3. Im Beispiel 2.0.1 hatten wir das Problem betrachtet, den Parameter $p \in [0,1]$ einer Bernoulliverteilung $\mathsf{b}_{1,p}$ zu schätzen. Dazu sei $X_1, \ldots, X_n$ eine Zufallsstichprobe zur Verteilung $\mathsf{b}_{1,p}$ (wobei wir in Beispiel 2.0.1 den Wert $X_j = 1$ mit dem Elementarereignis F, 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 $T_1(X_1, \ldots, X_n) = \bar{X}$, betrachtet. Unsere Definition lässt aber auch andere Schätzer zu. Zum Beispiel:

$$T_2(X_1, \ldots, X_n) = \frac{1}{2}$$
$$T_3(X_1, \ldots, X_n) = \frac{1}{2}(X_1 + X_n)$$

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 $\{F_\theta : \theta \in \Theta\}$ und eine Zufallsstichprobe $X_1, \ldots, X_n$ 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 F_θ ist.

Definition 2.2.4. Es seien $X_1, \ldots, X_n, X_{n+1}, \ldots$ unabhängige Zufallsvariablen mit $X_k \sim F \in \{F_\theta : \theta \in \Theta\}$. Weiter sei $T_n := T_n(X_1, \ldots, X_n)$ 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 $\mathbb{E}_\theta T_n \to g(\theta)$ für $n \to \infty$.

(c) Die Folge T_n heißt *schwach konsistent*, falls für alle $\theta \in \Theta$

$$\mathbb{P}_\theta\big(|T_n - g(\theta)| \geq \varepsilon\big) \to 0$$

für $n \to \infty$, also $T_n \to g(\theta)$ in θ- Wahrscheinlichkeit

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. Die Folge der Schätzer T_1 ist schwach konsistent nach dem schwachen Gesetz der großen Zahlen. Beachten wir, dass T_3 nur die Werte 0, $\frac{1}{2}$ und 1 annimmt, so erhalten wir für $p = \frac{1}{2}$ und $\varepsilon = 1/4$

$$\mathbb{P}_{\frac{1}{2}}(|T_3 - p| \geq 1/4) = \mathbb{P}_\theta(X_1 = 0, X_n = 0 \text{ oder } X_1 = 1, X_n = 1) = \frac{1}{2} \not\to 0$$

sodass T_3 nicht schwach konsistent ist.

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

Momentenmethode

Es seien $X_1, \ldots, X_n$ eine Zufallsstichprobe zur Verteilung $F \in \{F_\theta : \theta \in \Theta\}$. Für $r \in \mathbb{N}$ ist $m_r(\theta) := \mathbb{E}_\theta X_1^r$ das *r-te Moment der Verteilung* F_θ. Wir können auch die empirschen r-ten Momente $\hat{m}_r := \frac{1}{n}\sum_{k=1}^n X_k^r$ betrachten. Es folgt aus dem Gesetz der großen Zahlen, dass $\hat{m}_r \to m_r(\theta)$ in θ-Wahrscheinlichkeit für $n \to \infty$.

Ist man nun an einem Parameter $\vartheta = g(\theta)$ interessiert, der sich als Funktion gewisser 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. das r-te Moment der Verteilung F_θ.

Beispiel 2.2.6. Interessiert man sich für den Erwartungswert einer Verteilung, so liefert die Momentenmethode den Schätzer $\bar{X}$. Schreibt man $\mathsf{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$, so erhält man mit der Momentenmethode als Schätzer für die Varianz

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

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

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

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

Beachte, dass diese Darstellung nicht eindeutig ist, denn es ist auch $\lambda = m_2(\lambda) - m_1(\lambda)^2$ sodass man mit dieser Darstellung aus der Momentenmethode $\hat{\lambda} = \frac{n-1}{n}S^2$ erhält.

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 F_N die Gleichverteilung auf den Zahlen $1, 2, \dots, N$, also $F_N(\{k\}) = N^{-1}$ für $k = 1, \dots, N$. Ist $X \sim F_N$ 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 21, 4 und 121 beobachtet).

Maximum-Likelihood Methode

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

Die beste Schätzung für einen Parameter θ ist diejenge,
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 $F \in \{F_\theta : \theta \in \Theta\}$. Ferner sei $f(x, \theta) = F_\theta(\{x\})$ die zugehörige Zähldichte. Dann heißt $L : \mathbb{R}^n \times \Theta \to [0, 1]$, gegeben durch

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

die zugehörige *Likelihood Funktion*. Es sei nun $\hat{\theta} : \mathbb{R}^n \to \Theta$ 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) Manchmal ist es einfacher statt der Likelihood funktion L die sogenannte log-Likelihood Funktion $\log L$ zu verwenden. Wegen der Monotonie von Logarithmus und Exponentialfunktion ist es äquivalnt die log-Likelihood funktion zu maximieren.

Beispiel 2.2.11. 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 ein Maximum, denn die Grenzwerte bei 0 und 1 sind jeweils $-\infty$). Es ist

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

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 2.2.12. Wir bestimmen einen Maximum Likelihood Schätzer für die Verteilungen $\{\mathsf{Pois}_\lambda : \lambda > 0\}$. 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 $\log L = n\bar{x}\log\lambda - n\lambda - \log(x_1! \cdot \ldots \cdot x_n!)$. Durch Differenzieren und Nullsetzten erhält man dass als Kandidaten für das Maximum $\lambda = \bar{x}$. Beachten wir noch dass

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

so ergibt sich, dass es sich hierbei in der Tat um die Maximumstelle handeln muss und der maximum likelihood Schätzer für λ ist gegeben durch $\hat{\lambda} = \bar{x}$.

2.3 Konfidenzintervalle

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

Beispiel 2.3.1. (Qualitätskontrolle)
Ein Zulieferer produziert gewisse elektronische Bauteile, die in der Automobilherstellung verwendet werden. Manche dieser Bauteile sind fehlerhaft und um die Anzahl der fehlerhaften Bauteile in einer Charge von 10.000 zu schätzen geht der Zulieferer wie folgt vor.

Er wählt zufällig 200 der 10.000 Bauteile aus und testet deren Funktionstätigkeit. Er stellt fest, dass von den 200 getesteten Bauteilen lediglich 3 defekt sind. Er vermutet daher, dass sich unter den 10.000 Bauteilen etwa 150 defekte Bauteile befinden.

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, wieviele defekte Bauteile sich in der Charge befinden. Vom rein logischen Standpunkt aus kann dies immer noch jede Zahl zwischen 3 und 9803 sein.

Allerdings legt das Gesetz der Großen Zahlen nahe, dass wenn n hinreichend groß ist, mit großer Wahrscheinlichkeit die Differenz $|\bar{X} - p|$ klein ist. Dies legt es nahe statt einer einzigen Zahl $\bar{X}$ ein Intervall $I = I(X_1,\ldots,X_n)$ anzugeben, in dem der wahre Parameter mit hoher Wahrscheinlichkeit liegt.

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

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

für alle $\theta\in\Theta$ gilt.

Bemerkung 2.3.3. Beachte: Das Konfidenzintervall ist zufällig, nicht der Parameter θ. α-Konfidenzintervall zu sein bedeutet, dass wenn die "wahre Verteilung" Parameter θ hat, so liegt $g(\theta)$ mit Wahrscheinlichkeit größer α im Interval.

Man kann Konfidenzintervalle beispielsweise mit der Tschebyscheffschen Ungleichung bestimmen. Wenn wir etwa eine Zufallsstichprobe zur Verteilung F 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

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

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 beschänkt) und wählen wir $\delta = \sqrt{\frac{\sigma^2}{n(1-\alpha)}}$, so ist

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

ein Konfidenzintervall zum Konfidenzniveau α.

Beispiel 2.3.4. In Beispiel 2.3.1 bestimmen wir ein 95%-Konfidenzintervall für die Bernoulliwahrscheinlichkeit p. Beachte, dass hier $\sigma^2 = p(1-p) \leq 1/4$. 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 geschätzen Wert ($\bar{x} = 0,015$) relativ groß ist. Wir erhalten also das (relativ lange) Konfidenzinterval $[0, 0.173]$. Will man das Konfidenzinterval verkleinern, so muss man den Stichprobenumfang erhöhen. Um Beispielsweise ein Konfidenzinterval das kürzer als $0,05$ ist wählt man n so, dass $2\delta = 0,05$, also

$$0,05 = 2\frac{1}{\sqrt{4n \cdot 0,05}} \quad \Leftrightarrow \quad n = \frac{1}{0,05^2 \cdot 0,05} = 8000.$$

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. Auch qualitativ gibt es hier einen Unterschied zu diskreten Wahrscheinlicheitsräumen, nämlich haben Elementarereignisse nicht mehr notwendigerweise positive Wahrscheinlichkeit. Wir diskutieren dies an einem Beispiel.

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

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

aller Folgen in $\{K, Z\}$ an. Das Elementarereignis $\omega = (\omega_1, \omega_2, \ldots)$ 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

$$A_1 = \{\omega : \omega_1 = K\} \subset \mathscr{P}(\Omega).$$

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

$$A_n = \{\omega : \omega_1 = \omega_2 = \ldots = \omega_n = K\}$$

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

Betrachten wir nun das Elementarereignis $A_\infty := \{(K, K, K, \ldots)\}$, dass in allen Würfen Kopf fällt, so ist $A_\infty \subset A_n$ 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 $\mathbb{P}(A_\infty) = 0$.

Indem man allgemeiner beliebige Ereignisse betrachtet, die lediglich endlich viele Stellen betreffen, kann man auf ähnliche Weise zeigen, dass alle Elementarereignisse Wahrscheinlichkeit 0 haben müssten.

Wir betrachten ein zweites Beispiel:

Beispiel 3.1.2. Wir ziehen zufällig eine Zahl aus dem Interval $I := (0,1)$. Hierbei sollen alle Zahlen gleichwahrscheinlich sein.

Zunächst ist unklar wie man diese Aussage interpretieren soll. Es gibt unendlich viele Zahlen in I, sodass sie nur dann "gleichwahrscheinlich" sein können, wenn alle Wahrscheinlichkeit 0 besitzen.

Dieser Wiederspruch löst sich jedoch auf, wenn man "wirkliche" Ereignisse statt Elementarereignissen betrachtet. Beispielsweise sollte die Wahrscheinlichkeit eine Zahl kleiner als $1/2$ zu ziehen genau so groß sein, wie die Wahrscheinlichkeit eine Zahl größer als $1/2$ zu ziehen. Allgemeiner sollte die Wahrscheinlichkeit, eine Zahl in einem Interval $J = (a,b)$ zu ziehen nur von der *Länge* $b-a$ des Intervals abhängen, nicht jedoch von dessen *Lage*. Da die Gesamtlänge von I gerade 1 ist, sollte die Wahrscheinlichkeit eine Zahl in (a,b) zu ziehen gerade $b-a$ sein.

Auch aus dieser Forderung ergibt sich, dass Elementarereignisse "Wahrscheinlichkeit 0" besitzen müssen. Ist nämlich $x_0 \in (0,1)$, so kann man für (genügend kleine) $\varepsilon > 0$ das Ereignis A_ε, eine Zahl in $(x-\varepsilon/2, x+\varepsilon/2)$ zu ziehen, betrachten. Nach obigem sollte $\mathbb{P}(A_\varepsilon) = \varepsilon$ sein. Weil aber $\{x_0\} \subset A_\varepsilon$ ist, müsste $\mathbb{P}(\{x_0\}) \leq \varepsilon$ für alle $\varepsilon > 0$, also $\mathbb{P}(\{x_0\}) = 0$ sein.

Beide Beispiele zeigen, dass es insbesondere bei nicht abzählbaren Wahrscheinlichkeitsräumen besser ist, Wahrscheinlichkeiten für "zusammengesetzte Ereignisse" statt für Elementarereignisse anzugeben. Weiterhin gibt es in beiden Beispiele gewisse Ereignisse, für die wir wissen welche Wahrscheinlichkeit sie haben (oder bei denen wir zumindest eine gute Vorstellung davon haben, welche Wahrscheinlichkeit sie haben sollten). Im ersten Beispiel sind dies Ereignisse, die nur endlich viele Würfe betreffen, im zweiten Beispiel die Ereignisse, eine Zahl aus einem bestimmten Interval zu ziehen.

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

Definition 3.1.3. Es sei Ω eine Menge. Eine *σ-Algebra* auf Ω ist eine Teilmenge $\Sigma \subset \mathscr{P}(\Omega)$, 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 3.1.4. Offensichtlich ist $\mathscr{P}(\Omega)$ eine σ-Algebra. Es ist die größte σ-Algebra auf Ω. Es gibt auch eine kleinste σ-Algebra auf Ω, nämlich $\Sigma = \{\emptyset, \Omega\}$.

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

Bemerkung 3.1.5. Weiterhin erfüllt eine σ-Algebra Σ auf Ω auch folgende Eigenschaften:

(1) Sind $A_1, \ldots, A_k \in \Sigma$, so auch $A_1 \cup \ldots \cup A_k$. Das folgt aus (iii) indem man $A_n = \emptyset$ für $n > k$ wählt.

(2) Sind $A_1, \ldots, A_k \in \Sigma$, so auch $A_1 \cap \ldots \cap A_k$. Das folgt aus (1), (ii) und deMorgan's Gesetz, nach dem $(A_1 \cap \ldots \cap A_k)^c = A_1^c \cup \ldots \cup A_k^c$ ist.

(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.

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

Lemma 3.1.6. *Ist $S \subset \mathscr{P}(\Omega)$, so gibt es eine kleinste σ-Algebra auf Ω, die S enthält. Diese 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 $\mathscr{B}(\mathbb{R})$ bezeichnet. Es ist also $\mathscr{B}(\mathbb{R}) = \sigma((a,b) : a,b \in \mathbb{R}, a < b)$.

Es ist nicht möglich $\mathscr{B}(\mathbb{R})$ genauer zu beschreiben, aber jede Menge, die aus offenen Intervallen durch (wiederholte) Komplementbildung und Vereinigung gebildet werden kann, liegt in $\mathscr{B}$. Es ist schwierig, zu zeigen, dass $\mathscr{B}(\mathbb{R}) \neq \mathscr{P}(\mathbb{R})$; der Beweis beruht auf dem sogenannten *Auswahlaxiom*.

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

Beispiel 3.1.8. Für $a \in \mathbb{R}$ ist $(-\infty, a) \in \mathscr{B}(\mathbb{R})$ und $(a,\infty) \in \mathscr{B}(\mathbb{R})$. 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 $\mathscr{B}$. Schliesslich sind auch abgeschlossene Intervalle in $\mathscr{B}(\mathbb{R})$ enthalten, denn $[a,b] = (-\infty,b] \cap [a,\infty)$.

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 $\mathbb{P} : \Sigma \to [0,1]$ *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

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

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 $\bigcup A_n$ enthählt. 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 1.1.6 gelten auch in beliebigen Wahrscheinlichkeitsräumen. Der Beweis bleibt unverändert.

Wir zeigen noch eine weitere Eigenschaft von Wahrscheinlichkeitsmaßen.

Lemma 3.1.11. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A_n eine Folge in Σ mit $A_1 \subset A_2 \subset A_3 \subset \ldots$ und $A := \bigcup_{n\in\mathbb{N}}$. (Wir sagen: A_n* wächst gegen *A und schreiben $A_n \uparrow A$. Dann ist*

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

Ist C_n eine Folge in Σ mit $C_1 \supset C_2 \supset \ldots$ 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*

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

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

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

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

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

□

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

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. Weiterhin dürfen wir hier nicht beliebige Mengen A verwenden, denn die Potenzmenge $\mathscr{P}(\mathbb{R})$ des Bildbereiches $\mathbb{R}$ ist ja in der Regel zu groß um darauf ein Wahrscheinlichkeitsmaß zu definieren.

Wir definieren daher:

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

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

die *Verteilung von X*.

Wir hatten bereits bemerkt, dass die Borel σ-Algebra $\mathscr{B}(\mathbb{R})$ sehr groß und unübersichtlich ist. Daher ist es schwierig, die Verteilung konkret anzugeben. In Anlehnung an die Beispiele in der Einleitung stellt sich daher 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 $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann heißt $F_X : \mathbb{R} \to [0,1]$, definiert durch

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

Verteilungsfunktion von X.

Beispiel 3.2.3. Ist $X \sim \mathsf{b}_{1,p}$ 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 3.2.4. Ist X eine Zufallsvariable die lediglich die Werte $x_1 < x_2 < \ldots < x_n$ annimmt mit $\mathbb{P}(X = x_k) = p_k$, so ist

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

Nimmt allgemeiner die Zufallsvariable X die abzählbar vielen 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.

Wir stellen einige Eigenschaften von Verteilungsfunktionen zusammen:

Zunächst ist F offensichtlich monoton wachsend (jedoch nicht notwendigerweise strikt). Ist nämlich $x \leq y$ so ist $\{X \leq x\} \subset \{X \leq y\}$ und daher, wegen der Monotonie des Wahrscheinlichkeitsmaßes, $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$. Insbesondere existieren die Grenzwerte $\lim_{x \to -\infty} F_X(x)$ und $\lim_{x \to \infty} F_X(x)$.

Es gilt $\lim_{x \to -\infty} F_X(x) = 0$ und $\lim_{x \to \infty} F_X(x) = 1$. Das folgt aus Lemma 3.1.11 und den Beziehungen $\{X \leq -n\} \downarrow \emptyset$ und $\{X \leq n\} \uparrow \Omega$.

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 $F_X(x_n) \to F_X(x)$. Also ist F rechtsseitig stetig.

Wir haben gezeigt:

Lemma 3.2.5. *Es sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$. Dann ist die Verteilungsfunktion F_X monoton wachsend, rechtsseitig stetig und erfüllt*

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

Zentral ist nun folgender Satz, den wir hier nicht beweisen.

Satz 3.2.6. *Es sei $F : \mathbb{R} \to \mathbb{R}$ eine monoton wachsende, rechtsseitig stetige Funktion mit $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to\infty} F(x) = 1$. Dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}_F$ auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit*

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

Theorem 3.2.6 sagt genau, dass die Verteilung einer Zufallsvariablen eindeutig durch ihre Verteilungsfunktion bestimmt ist. Eine wichtige Klasse von Verteilungen sind *absolutstetige Verteilungen.*

Definition 3.2.7. 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 3.2.8. Beispiel 3.2.4 zeigt, dass diskrete Zufallsvariablen unstetige Verteilungsfunktionen (genauer: Verteilungsfunktionen, die Treppenfunktionen sind) haben. Ein Sprung in der der Verteilungsfunktion an der Stelle x in Höhe p bedeutet, dass die Zufallsvariable den Wert x mit Wahrscheinlichkeit p annimmt.

Ist die Verteilungsfunktion stetig (insbesondere, bei abolutstetigen Verteilungsfunktionen), so nimmt die Zufallsvariable keinen Wert mit positiver Wahrscheinlichkeit an. In Gleichung 3.1 darf man also für absolutstetige Verteilungen auch schreiben

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

Bemerkung 3.2.9. 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, rechsseitig stetige Funktion mit $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to\infty} F(x) = 1$ ist. Somit ist F eine Verteilungsfunktion und f die Dichte dieser Verteilungsfunktion. Man nennt daher jede nichtnegative Funktion f mit $\int_{\mathbb{R}} f(t)\,dt = 1$ bereits *Dichte.*

In gewisser Weise sind absolutstetige Zufallsvariablen ähnlich zu diskreten Zufallsvariablen in Beispiel 3.2.4. Ist f_X die Zähldichte der diskreten Zufallsvariablen X so ist die Verteilungsfunktion F_X gegeben durch

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

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 3.2.10. 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

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

der Erwartungswert von X. Wir sagen, X habe endliche Varianz, falls

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

endlich ist.

Bemerkung 3.2.11. Ähnlich wie im diskreten Fall kann zeigen, dass X endliche Varianz hat, genau dann, wenn $\mathbb{E}X^2 = \int_{-\infty}^{\infty} t^2 f(t)\, dt$ endlich ist. In diesem Fall ist

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

Beispiel 3.2.12. Setzen wir

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

so ist f Dichte einer Verteilungsfunktion. Es ist nämlich

$$\int_{-\infty}^{\infty} f(t)\, dt = \int_{1}^{\infty} \frac{1}{t^2}\, dt = \Big[\frac{-1}{t}\Big|_1^{\infty} = 0 - (-1) = 1.$$

Ist f die Dichte der Verteilungsfunktion von X, so ist

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

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

X hat *keinen* endlichen Erwartungswert, es ist nämlich

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

Wir geben noch eine wichtige Formel für die Transformation von Dichten an.

Lemma 3.2.13. *Es sei X absolutsetetig mit Dichte f. Weiter seien $c, r \in \mathbb{R}$ mit $r \neq 0$. Dann ist $X + c$ absolutsetig mit Dichte g, wo $g(t) = f(t - c)$ ist. Weiter ist rX absolutstetig mit Dichte h, wo $h(t) = \frac{1}{|r|} f(t/r)$.*

Beweis. Es sei $Y = X + c$. Dann ist $F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(X \leq x - c) = F_X(x - c)$. Andererseits liefert die Substitution $t = s - c$

$$F_Y(x) = F_X(x - c) = \int_{-\infty}^{x-c} f(t)\, dt = \int_{-\infty}^{x} f(s - c)\, ds.$$

Dies zeigt, dass Y absolutstetig mit Dichte g ist. Sei nun $Z = rX$. Wir nehmen zunächst an, dass $r > 0$ ist. Dann ist $F_Z(x) = \mathbb{P}(rX \leq x) = \mathbb{P}(X \leq x/r) = F_X(x/r)$. Substituiert man nun $t = s/c$, so folgt

$$F_Z(x) = F_X(x/r) = \int_{-\infty}^{x/r} f(t)\,dt = \int_{-\infty}^{x} f(s/r)\,\frac{ds}{r}$$

was zeigt, dass Z absolutstetig mit Dichte h ist. Ist andererseits $r < 0$, so ist $F_Z(x) = \mathbb{P}(rX \leq x) = \mathbb{P}(X \geq x/r)$. Hier liefert die Substitution $t = s/r$

$$F_Z(x) = \int_{x/r}^{\infty} f(t)\,dt = \int_{-\infty}^{x} f(s/r)\,\frac{ds}{r}$$

was die Behauptung in diesem Falle ist. □

Gelegentlich ist es wichtig (beispielsweise beim Berechnen von Konfidenzintervallen), Zahlen x zu finden, sodass $\mathbb{P}(X \leq x) = F_X(x) = \alpha$ für ein Vorgegebenes α. Man will also die Funktion F_X umkehren. Zwar ist F_X monoton, aber im Allgemeinen nicht streng monoton, sodass x nicht eindeutig bestimmt sein muss. Man verwendet daher folgende allgemeine Inverse

Definition 3.2.14. Es sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Für $p \in (0,1)$ sei

$$F^{-1}(p) := \inf\{x : F(x) \geq p\}.$$

Dann heißt *Quantilfunktion* der Verteilungsfunktion oder *verallgemeinerte Inverse*. $F^{-1}(p)$ heißt p-Quantil der Verteilung.

Ist die Verteilungsfunktion streng monoton wachsend, so ist sie bijektiv und die Quantilfunktion ist mit der Umkehrfunktion identisch. Die Interpretation des p-Quantils ist wie folgt: Ist X eine Zufallsvariable mit gegebener Verteilung F und ist $x = F^{-1}(p)$ das p-Quantil der Verteilung, so nimmt die Zufallsvariable mit Wahrscheinlichkeit p einen Wert kleiner oder gleich x an.

3.3 Wichtige absolutstetige Verteilungen

Stetige Gleichverteilung

Wir sagen X ist *gleichverteilt auf dem Interval* (a,b) und schreiben $X \sim \mathsf{U}(a,b)$, 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.

Dies ist in gewisser Weise die Verallgemeinerung von Beispiel 3.1.2: Wird zufällig eine Zahl aus dem Interval (a,b) gezogen, so hat das Ergebnis Verteilung $\mathsf{U}(a,b)$. Ein weiteres 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 noch Erwartungswert und Varianz einer gleichverteilten Zufallsvariable.

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

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

Beweis. Es ist

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

Weiter ist für $\mu = (a+b)/2$ gerade

$$\mathsf{Var}X = \frac{1}{b-a}\int_a^b (t-\mu)^2\,dt = \frac{1}{b-a}\Big[\frac{(t-\mu)^3}{3}\Big|_a^b = \frac{1}{3(b-a)}\frac{1}{8}((b-a)^3-(a-b)^3) = \frac{(b-a)^2}{12}.$$

□

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 $X \sim \mathsf{exp}_\lambda$. Beachte, dass

$$\int_{-\infty}^{\infty} f_\lambda(t)\,dt = \int_0^\infty \lambda e^{-\lambda t}\,dt = \Big[-e^{-\lambda t}\Big|_0^\infty = 0-(-1) = 1$$

sodass f_λ in der Tat eine Dichte ist.

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

Wir berechnen wiederum Erwartungswert und Varianz.

Lemma 3.3.2. *Es sei $X \sim \mathsf{exp}_\lambda$. Dann ist*

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

Beweis. Mit partieller Integration erhalten wir

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

Mit zweifacher partieller Integration folgt

$$\mathbb{E}X^2 = \int_0^\infty t^2\lambda e^{-\lambda t}\,dt = \Big[-t^2e^{-\lambda t}\Big|_0^\infty - \int_0^\infty -2te^{-\lambda t}\,dt = \frac{2}{\lambda}\int_0^\infty t\lambda e^{-\lambda t}\,dt = \frac{2}{\lambda^2},$$

wobei wir obige Rechnung im letzten Schritt verwendet haben. Daher ist

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

□

Die Exponentialverteilung hat eine wichtige Eigenschaft, die man *Gedächtnislosigkeit* nennt. Ist nämlich $X \sim \mathsf{exp}_\lambda$, so ist

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

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 $A \subset B$ und daher $A \cap B = A$. Nun beachten wir, dass für $x \in (0,\infty)$

$$\mathbb{P}(X \geq x) = \int_x^\infty \lambda e^{-\lambda t}\, dt = \Big[-e^{-\lambda t}\Big|_x^\infty = e^{-\lambda x}.$$

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 $X \sim \mathsf{N}_{\mu,\sigma^2}$. Ist $\mu = 0$ und $\sigma^2 = 1$, so sagen wir, X ist *standardnormalverteilt*.

Bei der Normalverteilung ist es nicht einfach nachzurechen, dass die "Dichte f" wirklich eine Dichte ist. Das liegt daran, dass die Funktion $t \mapsto e^{-t^2}$ keine durch elementare Funktionen ausdrückbare Stammfunktion besitzt. Es gilt jedoch

Lemma 3.3.3. *(Gauß'sches Fehlerintegral)*
Es ist

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

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

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

Es folgt, dass

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

also ist f tatsächlich eine Dichte.

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

Es folgt sofort aus Lemma 3.2.13, dass wenn $X \sim \mathsf{N}_{\mu,\sigma^2}$ und $c \in \mathbb{R}$, $r > 0$, so ist $rX + c \sim \mathsf{N}_{r\mu+c,r^2\sigma^2}$. Es hat nämlich rX Dichte

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

also ist $rX \sim \mathsf{N}_{r\mu,r^2\sigma^2}$. Andererseits hat $X + c$ Dichte

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

sodass $X + c \sim \mathsf{N}_{\mu+c,\sigma^2}$.

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

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

Wie schon bemerkt kann man Φ nicht elementar ausdrücken. In Anwendungen verwendet man daher oft Tabellen mit Werten von Φ. Häufig sind dort nur Werte $\Phi(x)$ für $x \geq 0$ aufgefürt. Folgt aus der Symmetrie von φ, dass $\Phi(-x) = 1 - \Phi(x)$, sodass diese Information ausreicht.

Wir können nun Erwartungswert und Varianz einer normalverteilten Zufallsvariablen bestimmen.

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

Beweis. Es sei zunächst $X \sim \mathsf{N}_{0,1}$, also X standardnormalverteilt. Wegen $\varphi(-x) = \varphi(x)$ folgt mit der Substitution $t = -s$, dass

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

und daher $\mathbb{E}X = 0$. Mit partieller integration folgt nun

$$\mathsf{Var}X = \mathbb{E}X^2 = \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} = \Big[\frac{-te^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} = 1.$$

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

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

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

3.4 Zufallsvektoren

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

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

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

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

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

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

sodass

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

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

Bemerkung 3.4.2. (a) Ähnlich wie im Fall von Zufallsvariablen kann man zeigen, dass die Verteilungsfunktion ein Wahrscheinlichkeitsmaß auf der Borel σ-Algebra $\mathscr{B}(\mathbb{R}^n)$ von $\mathbb{R}^n$ (die σ-Algebra, die von den Rechecken $(a_1, b_1) \times \ldots \times (a_n, b_n)$ erzeugt wird) 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

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

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

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

die Dichte von X_j. 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 andere Wahrscheinlichkeiten berechnen. Für $A \in \mathscr{B}(\mathbb{R}^d)$ ist nämlich

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

(e) Schliesslich lassen sich mittels der Dichte auch andere Erwartungswerte definieren. Ist nämlich $g : \mathbb{R}^n \to \mathbb{R}$, so ist

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

sofern der Erwartungswert endlich ist. Insbesondere kann man auf diese Art die Kovarianz zweier Zufallsvariablen berechnen.

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

$$\int_{\mathbb{R}^2} f(x, y)\, dxdy = \int_0^1 \int_0^1 x + 2xy\, dxdy = \int_0^1 \left[\frac{1}{2}x^2 + x^2 y\right]_0^1 dy$$

$$= \int_0^1 \frac{1}{2} + y\,dy = \Big[\frac{1}{2}y + \frac{1}{2}y^2\Big|_0^1 = 1.$$

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

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

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

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

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

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

und

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

Somit ist

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

Folglich sind X und Y unkorreliert.

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

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

gilt.

Man kann nun folgenden Satz beweisen:

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

(i) $X_1, \ldots, X_n$ *sind unabängig.*

(ii) $F_X(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$.

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

(iii) $f(t_1, \ldots, t_n) = f_1(t_1) \cdot \ldots f_n(t_n)$.

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

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

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

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

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

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

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

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

Wir diskutieren nun noch die Verteilung der Summe zweier unabhängiger Zufallsvariablen. Sind X und Y unabhängig und absolutstetig mit Dichten f_X, respektive f_Y, so hat der Zufallsvektor (X, Y) gerade Dichte $(x, y) \mapsto f_X(x)f_Y(y)$. Wir interessieren uns für die Verteilung von $S := X + Y$. Es sei $B = \{(x, y) \in \mathbb{R}^2 : x + y \leq t\}$. Um die Verteilungsfunktion von S zu bestimmen müssen wir

$$\mathbb{P}(S \leq t) = \int_B f_X(x)f_Y(y) \, dxdy$$

berechnen. Wir substituieren $u = x + y$ und $y = v$ und erhalten

$$\int_B f_X(x) f_Y(y)\, dxdy = \int_{-\infty}^{t} \int_{-\infty}^{\infty} f_X(u-v) f_Y(v)\, dv\, du.$$

Das zeigt, dass S absolutstetig ist mit Dichte

$$(f_X * f_Y)(u) = \int_{-\infty}^{\infty} f_X(u-v) f_Y(v)\, dv.$$

$f_X * f_Y$ heißt *Faltung* von f_X und f_Y.

Als Anwendung zeigen wir folgendes Resultat:

Proposition 3.4.8. *Es seien X, Y unabhängig mit $X \sim \mathsf{N}_{\mu_1,\sigma_1^2}$ und $Y \sim \mathsf{N}_{\mu_2,\sigma_2^2}$. Dann ist $X + Y \sim \mathsf{N}_{\mu_1+\mu_2,\sigma_1^2+\sigma_2^2}$. Insbesondere ist die Summe unabhängiger, normalverteilter Zufallsvariablen normalverteilt.*

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

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

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

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

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

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

Somit ergibt sich

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

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

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

$$\mathbb{P}\Big(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\Big) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt = \Phi(x).$$

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

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

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

$$\mathbb{P}\Big(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\Big) \to \int_a^b e^{-\frac{t^2}{2}}\, \frac{dt}{\sqrt{2\pi}}.$$

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

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 Anzahl der Erfolge $\mathsf{b}_{600,\frac{1}{6}}$-verteilt und die gesuchte Wahrscheinlichkeit ist

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

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

Seien hierzu $X_1, \ldots, X_{600}$ 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

$$\begin{aligned}\mathbb{P}(90 \le S_{600} \le 100) &= \mathbb{P}\Big(\frac{90-100}{9,13} \le \frac{S_{600} - 600\mu}{\sqrt{600}\sigma} \le \frac{100-100}{9,13}\Big) \\ &\approx \Phi(0) - \Phi(-1.095) = 0,5 - (1 - \Phi(1,095)) \approx 0,36,\end{aligned}$$

also ungefähr 36 Prozent.

Bemerkung 3.5.4. Der zentrale Grenzwertsatz macht keine Angaben darüber, wann die Normalverteilung eine gute Annährung an die Verteilung eines standartisierten Mittels ist bzw. darüber, wie gut diese Annährung ist. Bei der Approximation der Binomialverteilung $\mathsf{b}_{n,p}$ hat sich als Faustregel etabliert, dass die Approximation gut ist, falls $np(1-p) \geq 9$.

Bemerkung 3.5.5. Bei der Approximation der Binomialverteilung durch die Normalverteilung ist folgendes zu beachten:

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

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

Verwenden wir die Stetigkeitskorrektur in Beispiel 3.5.3, so erhalten wir

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

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

Wir geben nun noch ein etwas anderes Anwendungsbeispiel.

Beispiel 3.5.6. Der Airbus A380 hat gewöhnlich 526 Sitzplätze. Aus Erfahrungen ist bekannt, dass ein verkauftes Ticket mit einer Wahrscheinlichkeit von 0,1 storniert wird. Wie viele Tickets kann man für einen Flug verkaufen, wenn dieser mit einer Wahrscheinlichkeit von höchstens 2% überbucht sein soll?

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

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

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

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

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

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 $X_1, \ldots, X_n$ zur Normalverteilung $\mathsf{N}_{\mu,\sigma^2}$ die Parameter μ und σ^2 zu schätzen. Wir verwenden hierzu die *Maximum Likelihood Methode*. Im Falle diskreter Wahrscheinlichkeiten, hatten wir als Likelihood Funktion L gerade das Produkt der Zähldichten verwendet: $L(x_1, \ldots, x_n; \theta) = f(x_1; \theta) \cdot \ldots \cdot f(x_n; \theta)$. 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 $X_1, \ldots, X_n$ eine Zufallsstichprobe zur Verteilungsfunktion $F \in \{F_\theta : \theta \in \Theta\}$. Weiter sei F_θ absolutstetig mit Dichte $f(\cdot, \theta)$. Dann heißt $L : \mathbb{R}^n \times \Theta \to [0\infty)$, definiert durch

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

die zugehörige *Likelihood Funktion*. Ist $\hat{\theta} : \mathbb{R}^n \to \Theta$ eine Funktion mit

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

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

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

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

Es ist einfacher, die log-Likelihood Funktion

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

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

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

Für die partielle Ableitung nach σ^2 ist

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

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

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

Die einzige kritische Stelle von ℓ ist demnach der Punkt

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

Eine genauere Untersuchung der log-Likelihood Funktion zeigt, dass dies in der Tat eine (globale!) Maximumstelle ist, d.h. dies sind die (in diesem Falle eindeutigen) Maximum Likelihood Schätzer. In der Praxis verwendet man als Schätzer für die Varianz σ^2 in der Regel den erwartungstreuen Schätzer s^2.

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

Konfidenzinterval für μ bei bekannter Varianz

Es sei also $X_1, \ldots, X_n$ eine Zufallsstichprobe zur Verteilung $\mathsf{N}_{\mu,\sigma_0^2}$, 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 $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$. Es folgt induktiv aus Proposition 3.4.8, dass $X_1 + \ldots + X_n \sim \mathsf{N}_{n\mu,n\sigma_0^2}$ und daher $\bar{X} \sim \mathsf{N}_{\mu,\sigma_0^2/n}$. Nun normalisiert man, indem man μ substrahiert und durch $\sqrt{\sigma_0^2/n}$ dividiert. Es folgt, dass $V := \sqrt{n}\frac{\bar{X}-\mu}{\sigma_0} \sim \mathsf{N}_{0,1}$. Beachte, dass diese Verteilung *nicht* mehr von dem unbekannten Parameter μ abhängt. Um ein Konfidenzinterval zum Konfidenzniveau α zu konstruieren geht man nun wie folgt vor:

Zunächst bestimmt man ein Interval, 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

$$\alpha = \mathbb{P}_\mu(|V| \le c) = \mathbb{P}_\mu\Big(-c \le \sqrt{n}\frac{\bar{X}-\mu}{\sigma_0} \le c\Big) \tag{3.3}$$

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

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.

Beispiel 3.6.2. Bei einer Stichprobe von 20 Brötchen wurden folgende Gewichte (in Gramm) gemessen (nach aufsteigendem Gewicht):

44,61	45,15	45,41	45,44	45,56	45,65	46,51	46,59	46,72	46,77
46,98	47,23	47,86	47,90	48,25	48,76	48,86	48,98	49,09	49,20

Wir nehmen an, dass das Gewicht normalverteilt ist mit unbekanntem Mittelwert und Varianz 2. Als Schätzer für den Mittelwert verwenden wir $\bar{X} = 47,076$. 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 dann

$$\frac{c\sigma_0}{\sqrt{n}} = \frac{1,96\sqrt{2}}{\sqrt{20}} = 0,620.$$

Daher ist $[46.456, 47.696]$ ein 95% Konfidenzinvervall für den Mittelwert μ.

Konfidenzinvervall 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 der Definition von V die (bekannte) Standartabweichung σ_0 durch S, die Wurzel aus dem Schätzer für die Varianz, zu ersetzen. Um ein Konfidenzinvervall zu bestimmen muss man nun die Verteilung der resultierenden Zufallsvariable kennen. Wir diskutieren dies in allgemeiner Situation.

Definition 3.6.3. 1. Es seien $X_1, \dots, X_r$ unabhängig und $\mathsf{N}_{0,1}$-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 $X \sim \mathsf{N}_{0,1}$ 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 3.6.4. (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 3.6.5. *Es seien $X_1, \dots, X_n$ unabhängig und $\mathsf{N}_{\mu,\sigma^2}$-verteilt. Dann sind $\bar{X}$ und S^2 unabhängig und es gilt*

$$\bar{X} \sim \mathsf{N}_{\mu, \frac{\sigma^2}{n}}, \quad \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2, \quad \sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}.$$

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

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

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

Somit ist

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

Wir haben also ein α-Konfidenzintervall für μ gefunden.

Beispiel 3.6.6. Wir betrachten wiederum Beispiel 3.6.2. Man rechnet leicht nach, dass $S^2 = 2,172$, also $S = 1,474$ ist. Um ein 95%-Konfidenzinvervall zu berechnen, bestimmen wir zunächst das 0,975-Quantil der t-Verteilung mit 19 Freiheitsgraden. Aus einer Tabelle entnehmen wir den Wert $2,093$. Somit ergibt sich

$$\frac{cS}{\sqrt{n}} = \frac{2,093 \cdot 1,474}{\sqrt{20}} \approx 0,687$$

Daher ist $[46.386, 47.765]$ ein 95%-Konfidenzinvervall für μ. Beachte, dass dieses invervall länger ist als das in Beispiel 3.6.2 berechnete, obwohl die dort angenommene Varianz 2 nahe bei der geschätzten Varianz 2,172 liegt. Selbst wenn man in Beispiel 3.6.2 als Varianz genau 2,172 annehmen würde, so würde man immer noch ein kürzeres Konfidenzintervall bekommen. Der Preis dafür, dass wir σ^2 als unbekannt annehmen (dürfen!) ist ein längeres Konfidenzinvervall.

Konfidenzintervall für σ^2

Wir wollen nun noch ein Konfidenzinvervall für die Varianz einer Stichprobe $X_1, \dots, X_n$ zur Normalverteilung $\mathsf{N}_{\mu,\sigma^2}$ bestimmen. Wir hatten bereits gesehen, dass $R := \frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$. 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^2_{n-1}-Verteilung bestimmen.

Um ein α-Konfidenzintervall zu bestimmen, gehen wie folgt vor. Wir bestimmen a und b, sodass $\chi^2_{n-1}([a,b]) = \alpha$. Dann ist

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

Wir haben also ein α-Konfidenzinvervall gefunden. Es bleibt noch a und b zu wählen. Hier ist zu beachten, dass (anders als bei der Standartnormalverteilung 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^2_{n-1}-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^2_{n-1}-Verteilung.

Beispiel 3.6.7. Wir betrachten wieder Beispiel 3.6.2 wo wir $S^2 = 2,172$ bei einem Stichprobenumfang von $n = 20$ beobachtet hatten. Um ein Konfidenzinvervall zum Niveau 0,95 zu bestimmen benötigen wir noch Quantile der χ^2_{19}-Verteilung. Das 2,5%-Quantil von χ^2_{19} ist gegeben durch $a = 8,91$. Andererseits ist das 97,5%-Quantil gegeben durch $b = 32,85$. Somit ergibt sich als 95%-Konfidenzinvervall für σ^2

$$\Big[\frac{19 \cdot 2,172}{32,85}, \frac{19 \cdot 2,172}{8,91}\Big] = [1.256, 4.632].$$

Kapitel 4

Statistische Tests

4.1 Grundbegriffe

Wir betrachten wieder ein parametrisches Modell $\{F_\theta : \theta \in \Theta\}$ und eine zugehörige Zufallsstichprobe $X_1, \ldots, X_n$. Wir wollen nun die Beobachtung der $X_1, \ldots, X_n$ 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. Konkret wird eine Hypothese durch die Angabe einer Teilmenge Θ_0 von Θ definiert. Man sagt dann 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 $x_1, \ldots, x_n$ der Stichprobe $X_1, \ldots, X_n$ festlegt, ob die Hypothese oder die Alternative gewählt wird. Man hat also eine Zerlegung des Stichprobenraumes M^n in disjunkte, nichtleere Teilmengen K_0 und K_1, sodass wir uns für die Hypothese entscheiden, wenn $(X_1, \ldots, X_n) \in K_0$ (wir sagen "*die Hypothese wird akzeptiert*") und wir uns für die Alternative entscheiden, wenn $(X_1, \ldots, X_n) \in K_1$ (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 die Verteilungen F_p auf der Menge $\{0, 1\}$, gegeben durch $F_p(\{0\}) = 1 - p$ und $F_p(\{1\}) = p$. Ist $X \sim F_p$ so interpretieren wir $X = 1$ als "flache Seite" und $X = 0$ als "mit der Spitze schräg nach unten". Wir wollen die Hypothese $p = \frac{1}{2}$ (dies ist eine einfache Hypothese) gegen die Alternative $p \neq \frac{1}{2}$ testen.

Ein möglicher Test ist wie folgt gegeben. Wir beobachten Realisierungen $X_1, \ldots, X_{1000}$ von F_p-verteilten Zufallsvariablen, bilden den Mittelwert $\bar{X}$ und akzeptieren die Hypothese wenn $\bar{X} = \frac{1}{2}$ ist und verwerfen 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).

Als Beispiel einer zusammengesetzten Hypothese erwähnen wir $H_0 : p \leq \frac{1}{2}$ was wir gegen die Alternative $H_1 : p > \frac{1}{2}$ testen könnten.

Offensichtlich sind nicht alle Tests in obigem Beispiel gleich gut. Bei Test T_2 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 $\binom{1000}{500}\frac{1}{2}^{500}\frac{1}{2}^{500} \approx 0$. 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

$$G(\theta) := \mathbb{P}_\theta((X_1, \ldots, X_n) \in K_1)$$

die *Gütefunktion* des Tests. Ist $\alpha \in (0, 1)$ und $G(\theta) \leq \alpha$ für alle $\theta \in \Theta_0$, so sagt man es handelt sich um einen *Test zum Signifikanzniveau* α. 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 Signifikanzniveau α ist also die Wahrscheinlichkeit einen Fehler erster Art zu begehen höchstens α. Der Test ist unverfälscht, falls für $\theta \in \Theta_1$ die Wahrscheinlichkeit, dass der Test die Hypothese verwirft mindestends α beträgt. Beachte, dass für $\theta \in \Theta_1$ ist $1 - G(\theta) = \mathbb{P}_\theta(X_1, \ldots, X_n \in K_0)$ die Wahrscheinlichkeit, einen Fehler zweiter Art zu begehen. Ist der Test unverfälscht, so ist die Wahrscheinlichkeit einen Fehler zweiter Art zu begehen höchstens $1 - \alpha$.

Beispiel 4.1.3. In Beispiel 4.1.1 betrachten wir wiederum die Tests T_1, T_2, T_3. Es sei G_j die Gütefunktion des Tests T_j. Dann gilt

$$G_1(p) = 1 - \binom{1000}{500} p^{500}(1-p)^{500}.$$

Wir haben bereits gesehen, dass die Wahrscheinlichkeit einen Fehler erster Art zu begehen sehr hoch ist, denn $G(\frac{1}{2}) \approx 1$. Es ist leicht zu sehen, dass G_1 in $\frac{1}{2}$ ein Minimum besitzt. Daher ist die Wahrscheinlichkeit, einen Fehler zweiter Art zu machen praktisch 0.

Beim Test T_3 ist es quasi umgekehrt. Die Gütefunktion ist gegeben durch $G_3(p) \equiv 0$. Daher ist die Wahrscheinlichkeit, einen Fehler erster Art zu machen 0. Andererseits ist jedoch die Wahrscheinlichkeit, einen Fehler zweiter Art zu machen 1.

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. Wir bestimmen die Wahrscheinlichkeit, einen Fehler erster Art zu machen, approximativ mit dem Gesetz der

großen Zahlen. Es ist

$$\begin{aligned}\mathbb{P}_{\frac{1}{2}}(0.47 \leq \bar{X} \leq 0.53)) &= \mathbb{P}_{\frac{1}{2}}(470 \leq S_n \leq 530) \\ &= \mathbb{P}_{\frac{1}{2}}(\frac{470-500}{\sqrt{250}} \leq \frac{S_n - 1000\frac{1}{2}}{\sqrt{\frac{1}{2}\cdot\frac{1}{2}\cdot 1000}} \leq \frac{530-500}{\sqrt{250}}) \\ &\approx \Phi(1,897) - \Phi(-1,897) = 2\Phi(1,897) - 1 \approx 0,9412\end{aligned}$$

Daher ist $G_2(\frac{1}{2}) \approx 1 - 0,9412 = 0,0588$. Es folgt, dass G_2 (zumindest approximativ) ein Test zum Signifikanzniveau 6% ist.

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 Einschänkungen 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).

Ein anderes Beispiel ist das Testen eines neuen Produktes. Will man testen, ob der Verbraucher dieses Produkt einem Vergleichsprodukt vorzieht, so ist es problematischer, wenn eine nichtvorhandene Präferenz als solche erkannt wird (denn dann investiert man in die Markteinführung eines Produktes, das der Verbraucher gar nicht will).

Diese Asymmetrie sollte man bei der Wahl von Hypothese und Alternative beachten: die Hypothese sollte so gewählt werden, dass der schwerwiegendere 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. Ähnlich sollte beim zweiten Beispiel die Hypothese "Der Verbraucher zieht das Vergleichsprodukt vor" gegen die Alternative "Der Verbraucher zieht das neue Produkt vor" getestet werden.

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.

In allen hier diskutieren Tests verwenden wir eine sogenannte *Teststatistik* T und entscheiden uns für oder gegen die Hypothese, abhängig davon, wo T liegt. Genauer bestimmen wir beim zweiseitigen Test ein Intervall $[a, b]$ und akzeptieren die Hypothese, falls $T \in [a, b]$. Um einen Test zum Signifikanzniveau α zu erhalten, muss (unter Nullhypothese) die Wahrscheinlichkeit, dass $T \in [a, b]$ liegt mindestens $1 - \alpha$ sein. Wenn wir also die Verteilung von T unter der Nullhypothese kennen, so kann man das Interval aus Quantilen der Verteilung bestimmen. Bei einseitigen Tests verwendet man entsprechend Intervalle der Form $(-\infty, b]$ oder $[a, \infty)$.

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 $T \sim \mathsf{N}_{0,1}$ resp. $T \sim t_{n-1}$.

Tests bei bekannter Varianz

Gegeben eine Stichprobe $X_1, \ldots, X_n$ zur Normalverteilung $\mathsf{N}_{\mu,\sigma_0^2}$ bei *bekannter Varianz* σ_0^2, verwenden wir die Teststatistik

$$T := \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma_0}.$$

Wir betrachten zunächst den zweiseitigen Test $H_0 : \mu = \mu_0$ gegen die Alternative $\mu \neq \mu_0$.

Unter der Nullhypothese $H_0 : \mu = \mu_0$ ist dann $T \sim \mathsf{N}_{0,1}$. Für einen Test zum Signifikanzniveau α sei $c := \Phi^{-1}(1 - \alpha/2)$. Dann ist

$$\mathbb{P}_{\mu_0}(|T| \leq c) = 2\Phi(c) - 1 = 1 - \alpha. \tag{4.1}$$

Und daher $\mathbb{P}_{\mu_0}(|T| \geq c) = \alpha$. Wenn wir also H_0 akzeptieren, wenn $|T| \leq c$ ist und sonst H_0 ablehnen, so ist dies, wegen (4.1) ein Test zum Signifikanznieau α.

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 Durchschittsgewicht von 197 g ergibt. Es soll nun zum Signifikanzniveau von 5% getestet werden, ob $\mu = 200$.

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.

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 setzen $c := \Phi^{-1}(1 - \alpha)$. Wir akzeptieren H_0 falls $T \leq c$.

Dies ist ein Test zum Signifikanzniveau α. In der Tat ist

$$\mathbb{P}_{\mu_0}(T > c) = 1 - \Phi(c) = 1 - (1 - \alpha) = \alpha.$$

Allerdings ist die Hypothese zusammengesetzt und wir müssen auch für $\mu \leq \mu_0$ nachweisen, dass der Fehler erster Art eine Wahrscheinlichkeit kleiner α hat. Beachte, dass

$$T = \sqrt{n}\frac{T - \mu}{\sigma_0} - \sqrt{n}\frac{\mu_0 - \mu}{\sigma_0} =: T_\mu - r_\mu.$$

Für $\mu \leq \mu_0$ ist $r_\mu \geq 0$. Daher ist für solche μ

$$\mathbb{P}_\mu(T > c) = \mathbb{P}_\mu(T_\mu > c + r_\mu) = 1 - \Phi(c + r_\mu) \leq 1 - \Phi(c) = \alpha. \tag{4.2}$$

Insgesamt haben wir gezeigt, dass es sich um einen Test zum Signifikanzniveau α handelt.

Beispiel 4.2.2. Es werden Briefumschläge für Luftpost produziert, die nicht mehr als 2 g wiegen dürfen. Eine Stichprobe von 20 Briefumschlägen ergibt ein Durchscnittsgewicht 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.

In Gleichung (4.2) haben wir bereits die Gütefunktion für den einseitigen Test bestimmt. Es ist nämlich

$$G(\mu) = \mathbb{P}_\mu(T > c) = 1 - \Phi(c + r_\mu) = 1 - \Phi\Big(c + \sqrt{n}\frac{\mu_0 - \mu}{\sigma_0}\Big).$$

Beachte, dass $G(\mu)$ monoton wachsend in μ ist mit $G(\mu_0) = \alpha$. Insbesondere folgt, dass der Test unverfälscht ist.

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. In einem Geschäft gibt ein Kunde durchschnittlich €70 aus. Um den Umsatz zu steigern, schaltet das Geschäft eine Anzeige in der lokalen Zeitung. Nun soll der Erfolg der Werbekampange überprüft werden. Hierzu werden an einem Tag der Geldbetrag eines jeden Kunden notiert. Die so erhobenen Daten werden als Zufallsstichprobe einer $\mathsf{N}_{\mu,\sigma^2}$-verteilten Zufallsvariable angesehen. Sodann wird die Hypothese $\mu \leq 70$ gegen die Alternative $\mu > 70$ getestet zu einem Signifikanzniveau von 2,5% getestet.

An dem fraglichen Tag hatte das Geschäft 28 Kunden die durchschnittlich €73 Euro bei einer Stichprobenstandartabweichung von €4 ausgaben. Somit ergibt sich für die Teststatistik $T = \sqrt{28}\frac{73-70}{4} = 3,97$. Das 97,5% Quantil der t_{27}-Verteilung ist 2,052, sodass die Hypothese verworfen wird. Somit war die Werbekampagne in der Tat erfolgreich.

Test für verbundene Stichprobem

Bei dieser Art Testprobleme haben wir eine Folge von *Paaren* von Beobachtungen, also zwei Stichproben $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_n$, die zwar in sich (also die Xe und die Ye untereinander), aber nicht notwendigerweise voneinander unabhängig 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 $X_1, \ldots, X_n$, die andere Methode $Y_1, \ldots, Y_n$) die es zu vergleichen gilt. Man spricht in diesem Falle von *Verbundenen Stichproben.*

Manchmal kann man hierbei einen der obigen Tests auf die Differenzen $X_1 - Y_1, \ldots, X_n - Y_n$ 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 $Z_1, \ldots, Z_{100}$ 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.