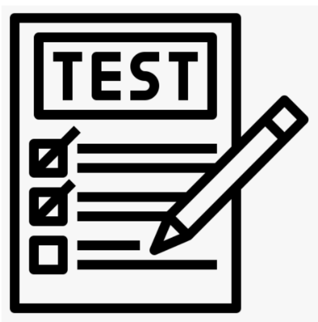
Stochastik für L2/L5[*]

(Lehramt an Haupt-, Real- und Förderschulen in Hessen)

Wintersemester 2025/26

Ralph Neininger

Hinweis <u>nur</u> für Prüfer:innen: In diesem Skript werden ausschließlich endliche Wahrscheinlichkeitsräume besprochen. Deshalb haben alle auftretenden reellen Zufallsvariablen Momente beliebiger Ordnung. Es treten somit im Gesetz großer Zahlen sowie im Zentralen Grenzwertsatz keine Integrierbarkeitsvoraussetzungen auf. Zudem sind diese Aussagen ohne Folgen unabhängiger Zufallsvariablen formuliert, um den Rahmen endlicher Wahrscheinlichkeitsräume nicht zu sprengen.



Hinweis für Studierende: Die in Kapitel 7 besprochene Probeklausur gibt einen Eindruck von Art und Umfang der Klausuraufgaben. Großteils sind Lösungen durch kurze Rechnungen zu ermitteln. Das mündliche Staatsexamen (Erste Staatsprüfung) dagegen ist von anderer Form und erfolgt nach Absprache mit der/dem Prüfer:in. Typischerweise werden dort Begriffe (Definitionen), Aussagen (Sätze), Zusammenhänge und Beweisideen besprochen. Empfehlenswert ist eine Vorbesprechung dazu.

[*]Laut Modulhandbuch L2/L5 Mathematik ist dies die Vorlesung „Elementare angewandte Mathematik“, laut Modulhandbuch L3 Informatik (Prüfungsordnung 2023) die Vorlesung „Angewandte Mathematik“. Mein Dank gebührt Herrn Adrian Traptschev für Unterstützung beim Erstellen von Aufgaben, Lösungen, R-Programmen und Graphiken sowie Frau Prof. Gaby Schneider für eine gründliche Durchsicht des Skripts.

Inhaltsverzeichnis

1 Vorbemerkungen 3

2 Endliche Wahrscheinlichkeitsräume 6
1 Ereignisse 6
2 Wahrscheinlichkeitsräume 7
3 Kombinatorik 14
4 Bedingte Wahrscheinlichkeiten und Unabhängigkeit 20

3 Einführung in R 29
5 Zur Syntax von R 29
6 Beispiele 33

4 Zufallsvariablen 36
7 Zufallsvariablen 36
8 Erwartungswert und Varianz 41

5 Summen unabhängiger Zufallsvariablen 46
9 Ein Gesetz großer Zahlen 46
10 Normalapproximation 49
11 Poissonapproximation* 54

6 Grundbegriffe schließender Statistik 56
12 Schätzen 57
13 Konfidenzintervalle 59
14 Testen 61

7 Probeklausur 65
15 Aufgaben 65
16 Musterlösungen 67

Literaturverzeichnis 70

1 Vorbemerkungen

Die Veranstaltung. Das vorliegende Skript ist Begleitmaterial einer zweistündigen Vorlesung „Elementare angewandte Mathematik“, die mit zweistündigen Übungen an der Goethe Universität Frankfurt a.M. für das Lehramt L2/L5 für Haupt-, Real- und Förderschulen angeboten wird. (Die Vorlesung wird ebenfalls von Studierenden des Lehramts L3 Informatik gehört.) Ab dem Wintersemester 2025/26 ist diese Vorlesung inhaltlich auf Stochastik fokussiert (im Gegenssatz zum Angebot vor dem Wintersemester 2025/26, dem das Skript von Prof. Kersting zugrunde lag, das auch Aspekte der Diskreten Mathematik, der Zahlentheorie sowie algorithmische Aspekte umfasst). Im vorliegenden Skript finden sich auch Bezüge zur Didaktik, die wie folgt markiert sind:

> **Bemerkung zur Didaktik/Schule:** In diesen Bemerkungen finden sich einige Hinweise auf didaktische Aspekte der Stochastik, Verbindungen zur Lehrveranstaltung „Didaktik der Stochastik“ sowie Hinweise auf die Kerncurricula der Sekundarstufe I des Hessischen Kultusministeriums, siehe
> `https://kultus.hessen.de/sites/kultus.hessen.de/files/2021-07/kerncurriculum_mathematik_realschule.pdf`
> `https://kultus.hessen.de/sites/kultus.hessen.de/files/2021-07/leitfaden_mathematik_sekundarstufe_i.pdf`
> Eine fachdidaktische Behandlung der Stochastik ist Gegenstand der Veranstaltung „Didaktik der Stochastik“.
> In den Kerncurricula werden „Kompetenzbereiche des Faches“ sowie „Inhaltliche Konzepte des Faches“ formuliert. Die Kompetenzbereiche betreffen Aspekte wie Problemlösen, Modellieren, Argumentieren oder mit symbolischen und formalen Elementen umzugehen. Sie sind unabhängig von den speziellen mathematischen Inhalten relevant.
> Als inhaltliche Konzepte der Kerncurricula für die Sekundarstufe I werden „Zahl und Operation“, „Raum und Form“, „Größen und Messen“, „Funktionaler Zusammenhang“ sowie „Daten und Zufall“ identifiziert. Wenngleich diese Konzepte offenbar auch mehrfach verschiedene Gebiete der Mathematik, etwa Algebra, Analysis, Geometrie oder Stochastik, betreffen, so sind sie dennoch in den jeweils entsprechenden Gebieten besonders dominant.
> Im vorliegenden Skript werden Begriffe und Zusammenhänge besprochen, die den Bereich „Daten und Zufall“ betreffen.

Teil der Vorlesung ist die freie Programmiersprache R

`https://www.r-project.org/`

für Simulationen, statistische Berechnungen und Graphiken. Die Studierenden des Lehramts sollen damit ein Hilfsmittel zur grafischen Darstellung von Daten und zur statistischen Datenanalyse kennenlernen und zudem in die Lage versetzt werden, Simulationen zur Illustration stochastischer Phänomene und Gesetzmäßigkeiten für ihren Unterricht vorbereiten zu können.

Der Zufall. Die Fragen, was denn Zufall sei, und ob Zufall existiere, sind im Grunde ungeklärt. Denkbar ist, dass uns wahre (deterministische) Mechanismen hinter Phänomenen unbekannt sind und diese uns nur deshalb zufällig erscheinen. Denkbar ist dagegen ebenso, dass der Zufall gegebener Bestandteil der Natur ist. Diese Fragen brauchen wir aber nicht zu klären. Selbst wenn wir etwa die Augenzahl beim Würfeln aus einer hinreichend genauen physikalischen Beschreibung des Wurfs deterministisch im Voraus bestimmen könnten, spricht vieles dafür anzunehmen, dass der Vorgang vom Zufall gesteuert werde. Diese weiter gefasste Vorstellung des Zufalls würde dann die beiden oben beschriebenen gegensätzlichen Standpunkte umfassen.
In der Stochastik werden Wahrscheinlichkeiten, Ereignisse und Zufallsvariable als mathematische Begriffe definiert und untersucht. Dies erlaubt, Gesetzmäßigkeiten des Zufalls aufzudecken und zu beweisen. Scheinbar willkürliche Phänomene folgen beweisbaren Gesetzen, z.B. bei zahlreichem Wurf einer Münze wird etwa in der Hälfte der Fälle „Zahl“ beobachtet. Was hierbei „etwa“ bedeutet, wird z.B. in den Gesetzen der großen Zahlen mathematisch spezifiziert und zum Teil quantifiziert.

Stochastische Modellierung. Die mathematische Sprache der Stochastik ermöglicht, Aspekte der Realität durch mathematische Modelle idealisiert zu beschreiben. Diese Modelle können sodann untersucht werden, um zu Vorhersagen zu kommen, die dann wieder mit der Realität verglichen werden können. Als einfaches Beispiel werfen wir drei Würfel gleichzeitig und ermitteln die Gesamtaugenzahl. Ist 11 ebenso wahrscheinlich wie 12? Mögliche Würfelkonstellationen sind:

$$\left.\begin{array}{ll} \text{„11“:} & 641,\ 632,\ 551,\ 542,\ 533,\ 443 \\ \text{„12“:} & 651,\ 642,\ 633,\ 552,\ 543,\ 444 \end{array}\right\} \text{jeweils 6 Möglichkeiten}$$

Glücksspieler des 17. Jahrhunderts wussten aus Erfahrung, dass 11 häufiger ist als 12, was sie durch empirische Daten belegten. Um dem nachzugehen, liegt es nahe, ein mathematisches (idealisiertes) Modell zu entwerfen, in dem sich die beiden relevanten Wahrscheinlichkeiten berechnen lassen. Wir kommen darauf am Ende von Kapitel 3 zurück.

Allgemein kommt man zu folgender stochastischen Betrachtungsweise:

(i) Modellbildung: Man präzisiert, welche Versuchsausgänge betrachtet werden sollen.

(ii) Man nimmt an, dass zu jedem Ereignis A eine Wahrscheinlichkeit $\mathbb{P}(A) \in [0, 1]$ gehört, die man für „einfache“ Ereignisse festlegt.

(iii) Man versucht auf der Grundlage konsistenter Rechenregeln aus Wahrscheinlichkeiten für einfache Ereignisse die Wahrscheinlichkeiten komplizierter Ereignisse zu bestimmen oder zu approximieren.

(iv) Man überprüft die Ergebnisse an der Wirklichkeit. Dies macht gegebenenfalls eine Korrektur am Modell notwendig.

Wahrscheinlichkeitstheorie und Statistik. Die Stochastik teilt sich in die zwei Teilgebiete Wahrscheinlichkeitstheorie und Statistik. Grob gesagt geht man in der Wahrscheinlichkeitstheorie davon aus, dass die ein Phänomen steuernden Wahrscheinlichkeiten bekannt sind (etwa im Rahmen einer Modellierierung) und untersucht davon ausgehend das Modell weiter. In der (schließenden) Statistik sind diese dagegen unbekannt. Dafür liegen Daten vor, aus denen Schlussfolgerungen gezogen werden können, etwa auf solche steuernde Wahrscheinlichkeiten. Die beiden Teilgebiete Wahrscheinlichkeitstheorie und Statistik haben Bezüge in das jeweils andere Gebiet. Kapitel 6 stellt Grundlagen der schließenden Statistik (auch Inferenzstatistik genannt) bereit.

> **Bemerkung zur Didaktik/Schule:** Der Unterschied zwischen Wahrscheinlichkeitstheorie und Statistik findet sich im oben genannten PDF Dokument des Kerncurriculums für die Sekundarstufe I etwa auf Seite 29 unten, wo das Inhaltsfeld „Daten und Zufall“ näher beschrieben wird. Dort finden sich in der oberen Spalte „statistische Erhebungen und ihre Auswertung“ sämtlich Themen der Statistik, während in der unteren Spalte „Umgang mit dem Zufall“ Themen der Wahrscheinlichkeitstheorie genannt sind.

2 Endliche Wahrscheinlichkeitsräume

1 Ereignisse

Mengen. Es werden hier einige Grundbegriffe zu Mengen wiederholt: Eine *Menge* ist eine Zusammenfassung einer Anzahl einzelner Objekte. Diese Objekte werden *Elemente* der Menge genannt. Es gibt verschiedene Möglichkeiten, Mengen darzustellen, z.B. durch explizite Angabe ihrer Elemente, etwa $\Omega = \{A, C, G, T\}$ bedeutet, dass die Menge Ω aus den Elementen A, C, G und T besteht. Z.B. Zahlenräume wie $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ und $\mathbb{R}$ sind Mengen. Eine Menge kann ein Element höchstens einmal enthalten, d.h. $\{1, 2, 3\}$ und $\{1, 1, 2, 3\}$ bezeichnen dieselbe Menge, die aus den Elementen $1, 2$ und 3 besteht. Die *leere Menge* $\emptyset$ ist die Menge, die kein Element enthält.
Eine *Teilmenge* A einer Menge Ω, geschrieben $A \subset \Omega$, ist eine Menge, die nur Elemente der Menge Ω enthalten kann (sie kann z.B. alle Elemente von Ω enthalten, d.h. $A = \Omega$ oder auch gar keine Elemente enthalten, d.h. $A = \emptyset$). Für eine Menge $\Omega \neq \emptyset$ bezeichnet $\mathcal{P}(\Omega)$ die *Potenzmenge* von Ω, d.h. die Menge aller Teilmengen von Ω. Z.B. für $\Omega = \{1, 2\}$ ist $\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.
Nehmen wir an, Ω habe 4 Elemente (oder n Elemente). Wieviele Elemente hat dann $\mathcal{P}(\Omega)$?
Das *kartesische Produkt* $A \times B$ für Mengen A und B ist

$$A \times B := \{(a, b) \,|\, a \in A, b \in B\},$$

also die Menge aller Paare (a, b) mit Komponenten $a \in A$ und $b \in B$. Entsprechend wird $A_1 \times \cdots \times A_n$ für Mengen $A_1, \ldots, A_n$ erklärt. Man erinnere sich, dass etwa (s_1, s_2, s_3) ein geordnetes Tupel (hier Tripel) bezeichnet, bei dem die Reihenfolge der Einträge wesentlich ist. Es ist also etwa $(3, 2, 3) \neq (3, 3, 2)$.
Für zwei Mengen A, B bezeichnen $A \cup B$, $A \cap B$, $A \setminus B$ deren *Vereinigung*, *Durchschnitt* bzw. (mengentheoretische) *Differenz*. Liegt Ω fest, so wird für $A \subset \Omega$ mit $A^c = \Omega \setminus A$ das *Komplement* von A in Ω bezeichnet.
Für Abbildungen $f : A \to B$ ist das *Urbild* $f^{-1}(C) = \{a \in A \,|\, f(a) \in C\}$ für alle $C \subset B$ erklärt. Das Urbild vertauscht mit den oben notierten Verknüpfungen von Mengen, d.h. es gilt stets $f^{-1}(C_1 \cup C_2) = f^{-1}(C_1) \cup f^{-1}(C_2)$, $f^{-1}(C_1 \cap C_2) = f^{-1}(C_1) \cap f^{-1}(C_2)$, $f^{-1}(C_1 \setminus C_2) = f^{-1}(C_1) \setminus f^{-1}(C_2)$ und $f^{-1}(C^c) = f^{-1}(C)^c$ für $C, C_1, C_2 \subset B$.

Ereignisse. Um Zufallsexperimente mathematisch zu modellieren, fassen wir im ersten Schritt die möglichen Ausgänge des Experiments zu einer Menge Ω zusammen, die in diesem Kontext auch *Ergebnismenge* genannt wird. Wird z.B. ein Würfel geworfen, so kann man $\Omega = \{1, 2, 3, 4, 5, 6\}$ wählen, beim Werfen einer Münze mit den Seiten Kopf und Zahl etwa $\Omega = \{K, Z\}$. Wird ein Glücksrad gedreht, das fünf Sektoren mit den Farben rot, blau, gelb, orange und schwarz hat, so kann man $\Omega = \{r, b, g, o, s\}$ wählen. Wird ein

Würfel dreimal hintereinander geworfen, so kann

$$\begin{aligned}\Omega &= \{(s_1, s_2, s_3) \mid s_i \in \{1, \dots, 6\} \text{ für } i = 1, \dots, 3\} \\ &= \{1, \dots, 6\} \times \cdots \times \{1, \dots, 6\} \\ &= \{1, \dots, 6\}^3\end{aligned} \tag{1}$$

gewählt werden. Dies ist ein *mehrstufiges Zufallsexperiment*. Werden Zufallsexperimente hintereinander ausgeführt, wie hier dreimal den Würfel hintereinander zu werfen, so kann man die Ergebnismenge des Gesamtexperiments, also des mehrstufiges Zufallsexperiments, offenbar stets als kartesisches Produkt der Ergebnismengen der einzelnen Experimente schreiben.

Stellt man sich den Ausgang einer physikalischen Messung einer reellen Größe, etwa einer Temperatur oder einer Geschwindigkeit (ohne Einheiten), als mit vom Zufall behaftet vor, so kann $\Omega = \mathbb{R}$ gewählt werden. Werden n solche Messwerte ermittelt, etwa die Temperatur am Morgen jedes Tages eines Monats, so kann $\Omega = \mathbb{R}^n$ gewählt werden (mit $n = 31$).

Man betrachtet nicht nur einzelne mögliche Ausgänge eines Zufallsexperiments, sondern auch komplizierter Aussagen über das Experiment, z.B. dass beim Werfen eines Würfels eine gerade Zahl geworfen wird. Um dies zu beschreiben, fasst man die einzelnen Ausgänge dafür zu einer Menge zusammen, also dass eine gerade Zahl geworfen wird, wird durch die Menge $\{2, 4, 6\}$ beschrieben. Man nennt so eine Menge ein *Ereignis*. Ereignisse sind also immer Teilmengen der Menge aller möglichen Ausgänge Ω. Etwa das Ereignis, dass beim dreimaligen Werfen des Würfels insgesamt die Augensumme 13 fällt, ist das Ereignis A, das durch

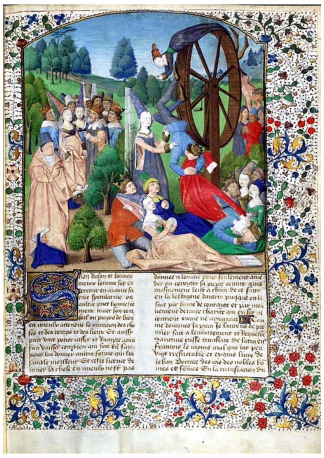
$$A = \{(s_1, s_2, s_3) \in \{1, \dots, 6\}^3 \mid s_1 + s_2 + s_3 = 13\}$$

beschrieben wird. Dass bei der Messung einer Temperatur ein Wert zwischen 0°C und 20°C ermittelt wird, ist das Ereignis, das (ohne Einheiten) gegeben ist durch das Intervall $[0, 20] \subset \mathbb{R}$.

Hat man alle möglichen Ausgänge eines Experiments durch eine Menge Ω modelliert, so nennt man $A = \Omega$ das *sichere Ereignis* (denn es tritt sicher ein, egal wie das Zufallsexperiment ausgeht) und $A = \emptyset$ das *unmögliche Ereignis*. Besteht ein Ereignis aus genau einem Element, also $A = \{\omega\}$ für ein $\omega \in \Omega$, so nennt man A ein *Elementarereignis*.

2 Wahrscheinlichkeitsräume

Im zweiten Schritt der Modellierung eines Zufallsexperiments werden den Ereignissen *Wahrscheinlichkeiten* zugeordnet. Dies ist in manchen Situationen offensichtlich. Etwa bei einem fairen Würfel würde man jeder der möglichen Augenzahlen $1, \dots, 6$ aus Symmetriegründen dieselbe Wahrscheinlichkeit zuordnen. In anderen Situationen, etwa dem in der Didaktik gerne betrachteten Riemen-Quader (ein Quader, dessen Seitenlängen nicht alle gleich sind, der zum Würfeln verwendet wird) oder beim Wurf einer Reißzwecke, ist dies nicht offensichtlich. Wir kommen auf diesen Modellierungsaspekt im Folgenden öfter zurück.



Quelle: Wikipedia, public domain

Wir betrachten ein Glücksrad, z.B. mit acht Sektoren (links abgebildet ist das „Rad des Schicksals" (lat. *rota fortunae*). Die Wahrscheinlichkeiten (momentan noch nicht mathematisch definiert), dass nach dem Drehen ein bestimmter Sektor auftritt, brauchen nicht für alle Sektoren gleich zu sein. Es ist übliche Konvention, dem Ereignis, dass irgendein Sektor nach Drehen auftritt (also dem sicheren Ereignis), Wahrscheinlichkeit 100%, also 1, zuzuordnen. Betrachten wir nun zwei verschiedene Sektoren und nehmen an, dass diese z.B. mit Wahrscheinlichkeiten von 20% bzw. 15% auftreten. Dann wird man nicht bestreiten, dass die Wahrscheinlichkeit, dass nach Drehen einer dieser beiden Sektoren auftritt (egal welcher), 35% beträgt. Es stellt sich heraus, dass das Rechnen mit Wahrscheinlichkeiten allein auf diese beiden Eigenschaften gegründet werden kann, die wir in der folgenden Definition als Axiome annehmen werden.

Definition 2.1. Ein *endlicher Wahrscheinlichkeitsraum* $(\Omega, \mathbb{P})$ besteht aus einer endlichen Menge $\Omega \neq \emptyset$ und einer Abbildung $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ mit

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

(ii) $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$
für alle disjunkten Mengen $A_1, A_2 \subset \Omega$, also mit $A_1 \cap A_2 = \emptyset$.

Die Abbildung $\mathbb{P}$ heißt *Wahrscheinlichkeitsmaß*, Eigenschaft (ii) nennt man *Additivität.*

Bemerkung 1. Das Wahscheinlichkeitmaß ordnet also jedem Ereignis A eine Wahrscheinlichkeit $\mathbb{P}(A)$, also eine Zahl aus $[0, 1]$, zu. Dabei steht $\mathbb{P}$ für *probability*, statt $\mathbb{P}$ wird auch WS geschrieben. Für das sichere Ereignis wird in (i) die Wahrscheinlichkeit 1 vereinbart.

Aus den Eigenschaften (i) und (ii) in Definition 2.1 können alle gängigen Rechenregeln für Wahrscheinlichkeiten hergeleitet werden. Wir gehen dies im einzelnen durch:

Lemma 2.2. Sei $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum. Dann gelten

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

(b) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ für alle $A \subset \Omega$.

(c) $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ für $A, B \subset \Omega$.

(d) $\mathbb{P}(A) \leq \mathbb{P}(B)$, falls $A \subset B$ für $A, B \subset \Omega$. (*Monotonie*)

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

(f) $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ für paarweise disjunkte Mengen $A_1, \ldots, A_n \subset \Omega$.

(g) $\mathbb{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$ für beliebige $A_1, \ldots, A_n \subset \Omega$. (Sub-Additivität)

Beweis.

(a) Wähle $A_1 = A_2 = \emptyset$. Damit sind A_1 und A_2 disjunkt. Wir können die endliche Additivität anwenden und erhalten

$$\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset) \stackrel{\text{(ii)}}{=} \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) = 2\mathbb{P}(\emptyset).$$

Es folgt also $\mathbb{P}(\emptyset) = 0$, denn die Annahme $\mathbb{P}(\emptyset) > 0$ führt zum Widerspruch.

(b) Die Mengen A, A^c sind disjunkt, und es gilt $\Omega = A \cup A^c$. Damit folgt

$$1 \stackrel{\text{(i)}}{=} \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) \stackrel{\text{(ii)}}{=} \mathbb{P}(A) + \mathbb{P}(A^c), \quad \text{also} \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

(c) Wir schreiben $B = (A \cap B) \cup (B \setminus A)$ als disjunkte Zerlegung. Damit folgt

$$\mathbb{P}(B) \stackrel{\text{(ii)}}{=} \mathbb{P}(B \cap A) + \mathbb{P}(B \setminus A), \quad \text{also} \quad \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(d) Da $A \subset B$, ist $B = A \cup (B \setminus A)$ eine disjunkte Zerlegung. Es folgt

$$\mathbb{P}(B) \stackrel{\text{(ii)}}{=} \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A), \quad \text{da } \mathbb{P}(B \setminus A) \geq 0.$$

(e) Wir schreiben $A \cup B = A \cup (B \setminus A)$ als disjunkte Zerlegung. Somit folgt

$$\mathbb{P}(A \cup B) \stackrel{\text{(ii)}}{=} \mathbb{P}(A) + \mathbb{P}(B \setminus A) \stackrel{\text{(c)}}{=} \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(B \cap A).$$

(f) Die Aussage kann mit vollständiger Induktion über n bewiesen werden. Für $n = 1$ ist die Aussage trivial, für $n = 2$ ist die Aussage wahr nach (ii). Für den Induktionsschritt sei die Aussage wahr für ein $n \geq 2$. Dann gilt für paarweise disjunkte Mengen $A_1, \ldots, A_{n+1} \subset \Omega$, dass $B_1 := A_1 \cup \cdots \cup A_n$ und $B_2 := A_{n+1}$ disjunkt sind. Es folgt

$$\begin{aligned} \mathbb{P}(A_1 \cup \cdots \cup A_{n+1}) &= \mathbb{P}(B_1 \cup B_2) \\ &\stackrel{\text{(ii)}}{=} \mathbb{P}(B_1) + \mathbb{P}(B_2) \\ &\stackrel{\text{IV}}{=} (\mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)) + \mathbb{P}(A_{n+1}), \end{aligned}$$

wobei für IV die Induktionsvoraussetzung auf B_1 angewandt wurde.

(g) Seien $B_1 = \emptyset$ und $B_i := \bigcup_{j=1}^{i-1} A_j$ für $i = 1, \ldots, n$. Es gilt dann

$$\begin{aligned} \bigcup_{i=1}^n A_i &= A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \cdots \cup (A_n \setminus (A_1 \cup \ldots \cup A_{n-1})) \\ &= \bigcup_{i=1}^n (A_i \setminus B_i), \end{aligned} \tag{2}$$

wobei in (2) nun eine Vereinigung paarweise disjunkter Mengen steht. Es folgt

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \overset{(f)}{=} \sum_{i=1}^{n} \mathbb{P}(A_i \setminus B_i) \overset{(d)}{\leq} \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Damit sind alle Aussagen bewiesen. ∎

Bemerkung zur Didaktik/Schule: Die eben bewiesenen Rechenregeln lassen sich durch *Venn-Diagramme* (Mengendiagramme) veranschaulichen.

Bemerkung 2. Sei $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum. Wir schreiben explizit $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. Dann ist $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ bereits durch die Werte $p_i = \mathbb{P}(\{\omega_i\})$ für $i = 1, \ldots, n$ vollständig festgelegt. Denn für jedes $A \in \mathcal{P}(\Omega)$ ist $A = \bigcup_{\omega_i \in A} \{\omega_i\}$ eine Darstellung als Vereinigung paarweise disjunkter Mengen. Damit gilt

$$\mathbb{P}(A) \overset{(f)}{=} \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{\omega_i \in A} p_i = \sum_{i}^{n} \mathbf{1}_A(\omega_i) p_i. \tag{3}$$

Hierbei bezeichnet $\mathbf{1}_A$ die Indikatorfunktion von A, gegeben durch

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

Umgekehrt kann auch für beliebige $p_i \in [0,1]$ mit $\sum_i p_i = 1$ mittels (3) ein Wahrscheinlichkeitsmaß definiert werden.

Weiterführende Bemerkung: Die folgende Bemerkung geht über den Rahmen der Vorlesung hinaus, braucht nicht verstanden zu werden und kann übergangen werden.

Bemerkung 3. Für abzählbar unendliche Mengen Ω, wie $\mathbb{N}$, $\mathbb{Z}$ oder $\mathbb{Q}$, können Wahrscheinlichkeitsräume ähnlich definiert werden wie in Definition 2.1. Wir müssen lediglich die Additivität aus (ii) dort ersetzen durch die allgemeinere σ-Additivität

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

für jede Folge $(A_i)_{i\in\mathbb{N}}$ paarweise disjunkter Mengen $A_i \subset \Omega$. Man sieht dann leicht ein, dass die Eigenschaften aus Lemma 2.2 ebenso gelten.
Für überabzählbare Ω, wie $\mathbb{R}$, $\mathbb{R}^n$ oder $\mathbb{C}$ können Wahrscheinlichkeitsräume ebenfalls wie im eben beschriebenen Fall abzählbar unendlicher Mengen definiert werden. Man kann im Allgemeinen dann nicht mehr beliebige Teilmengen von Ω als Ereignisse betrachten, jedoch hinreichend allgemeine Klassen von Teilmengen von Ω. Etwa für $\Omega = \mathbb{R}$ kann man alle Intervalle sowie Mengen, die durch abzählbare Vereinigungen und Schnitte von Intervallen gebildet werden, als Ereignisse betrachten, was in der Regel in der Praxis ausreichend ist. Diesen Aspekt brauchen wir im Weiteren nicht im Auge zu behalten. Die Rechenregeln aus Lemma 2.2 gelten ebenso.

Bezeichnung 1. Wir haben einige technische Begriffe der Stochastik eingeführt, die wir im Folgenden mit einer Bedeutung versehen wollen.

(i) Ω heißt *Grundraum, Ergebnismenge, Stichprobenraum* oder *Ergebnisraum.*

(ii) Teilmengen von Ω heißen *Ereignisse.*

(iii) $\mathbb{P}$ heißt *Wahrscheinlichkeitsmaß* (oder *Wahrscheinlichkeitsverteilung*), $\mathbb{P}(A)$ bezeichnet die *Wahrscheinlichkeit des Ereignisses* A.

Im Rest dieses Abschnitts werden Beispiele für Wahrscheinlichkeitsräume diskutiert:

Laplace-Modelle (Gleichverteilung)

Räume $\Omega = \{1, \dots, n\}$ mit $\mathbb{P}(\{i\}) = \frac{1}{n} =: p_i$, also $p_1 = \cdots = p_n$, heißen *Laplace-Modelle* und $\mathbb{P}$ dann *Gleichverteilung.* Hier haben alle Elementarereignisse also dieselbe Wahrscheinlichkeit. Häufig begründen Symmetrieannahmen im Rahmen einer Modellierung die *Laplace-Annahme*, also die Annahme eines Laplace-Modells, also der Gleichverteilung als Wahrscheinlichkeitsmaß. Es gilt dann

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{|A|}{|\Omega|}.$$

Dabei bezeichnet $|A|$ die *Kardinalität* von $A \subset \Omega$, also die Anzahl der Elemente von A. Man spricht bei der Berechnung von Wahrscheinlichkeiten in Laplace-Modellen von „Günstige durch Mögliche“.

Beispiel 2.3. (Würfeln mit einem fairen Würfel) Sei $\Omega = \{1, \dots, 6\}$, wobei $\mathbb{P}(\{i\}) =: p_i$ und $p_1 = p_2 = \cdots = p_6$. Wegen $1 = \mathbb{P}(\Omega) = \sum_{i=1}^{6} p_i = 6p_1$ folgt $p_i = \frac{1}{|\Omega|} = 1/6$.

Beispiel 2.4. (Geburtstagsproblem) Es befinden sich m Personen in einem Raum, wir interessieren uns für das Ereignis $A =$“Mindestens zwei Anwesende haben am selben Tag Geburtstag“. Zur Modellierung sei

$$\Omega = \{\omega = (\omega_1, \dots, \omega_m) \,|\, \omega_i \in \{1, \dots, 365\} \text{ für } i = 1, \dots, m\}.$$

Damit ist $A = \{\omega \in \Omega \,|\, \exists\, 1 \leq i < j \leq m : \omega_i = \omega_j\}$. Wir machen die idealisierte Modellannahme, dass die (m-Tupel der) Geburtstage gleichverteilt seien und ignorieren den 29. Februar als möglichen Geburtstag. Damit liegt ein Laplace-Modell vor. Es gilt

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = 1 - \frac{|A^c|}{|\Omega|},$$

wobei

$$|\Omega| = 365^m, \quad A^c = \{\omega \in \Omega \,|\, \forall\, 1 \leq i < j \leq m : \omega_i \neq \omega_j\}.$$

Das Ereignis A^c entspricht „es gibt keine zwei mit demselben Geburtstag“. Die Anzahl der Möglichkeiten in A^c nimmt mit wachsendem $m \leq 365$ wesentlich langsamer zu als die Mächtigkeit $|\Omega| = 365^m$. Deshalb strebt $\mathbb{P}(A)$ rasch gegen 1:

$$|A^c| = \begin{cases} \prod_{i=1}^m (366 - i), & m \leq 365 \\ 0, & m > 365. \end{cases}$$

m	20	23	40	150
$\mathbb{P}(A)$	0.411	0.507	0.891	$1 - 10^{-15}$

Einpunktverteilung (Dirac-Maß)

Sei Ω eine beliebige endliche Menge und $\omega_0 \in \Omega$. Wir definieren

$$\delta_{\omega_0}(A) := \mathbf{1}_A(\omega_0) = \begin{cases} 1, & \text{falls } \omega_0 \in A, \\ 0, & \text{falls } \omega_0 \notin A, \end{cases}$$

für alle $A \subset \Omega$. Dann ist $(\Omega, \delta_{\omega_0})$ ein endlicher Wahrscheinlichkeitsraum. Das Wahrscheinlichkeitsmaß δ_{ω_0} heißt *Dirac-Maß* (in ω_0).

Poisson-Verteilung*

Man kann statt mit endlichem Ω ähnlich mit $\Omega = \mathbb{N}_0$ arbeiten also mit Experimenten, die grundsätzlich beliebig viele (unendlich viele) Ausgänge zulassen. Wir gehen auf technische Aspekte, dies mathematisch zu formalisieren, kurz in Bemerkung 3 ein. Ein wichtiges Wahrscheinlichkleitsmaß auf $\mathbb{N}_0$ ist die Poisson-Verteilung Π_λ mit Parameter $\lambda > 0$.
Es sei $\Omega = \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ und

$$\Pi_\lambda(\{n\}) := e^{-\lambda} \frac{\lambda^n}{n!}, \quad n \in \mathbb{N}_0. \tag{4}$$

Es ist damit ein Wahrscheinlichkeitsmaß Π_λ festgelegt, vgl. Bemerkung 2, denn es gilt

$$\Pi_\lambda(\Omega) = \sum_{n=0}^{\infty} \Pi_\lambda(\{n\}) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1.$$

Poisson-Verteilungen werden zur Modellierung der Anzahl seltener Phänomene pro Zeiteinheit verwendet. Beispiele sind die Anzahl fehlerhafter Teile in einer großen Produktion, dir Emission von α-Teilchen beim radioaktiven Zerfall oder die Anzahl der Druckfehler in einem Buch. Weshalb dabei jeweils zur Modellierung die Poisson-Verteilungen geeignet ist, werden wir später in Abschnitt 11 sehen.

Weitere Verteilungen

Zahlreiche weitere Wahrscheinlichkeitsmaße finden sich in Kapitel 3, z.B. die aus der Schule bekannte Binomialverteilung, die geometrische und die hypergeometrische Verteilung sind Wahrscheinlichkeitsmaße.

> **Bemerkung zur Didaktik/Schule:** Diese Bemerkung stellt eine Verbindung zu Themen der Veranstaltung „Didaktik der Stochastik" her. Sie ist knapp gefasst und für das vorliegende Skript nicht wesentlich und kann übergangen werden.
>
> Der „axiomatische Wahrscheinlichkeitsbegriff" aus Definition 2.1 ist mathematisch sauber, tragfähig und modern. Er erfordert von den Lernenden allerdings ein gewisses Maß an Erfahrung und mathematischer Reife und unterstützt (absichtlich, da axiomatisch) weniger den Aufbau von stochastischen Grundvorstellungen. Es stellen sich deshalb hier im Hinblick auf die Situation in der Schule didaktische Fragen. Dazu einige Stichworte zu anderen Zugängen, Wahrscheinlichkeiten zu definieren und zu interpretieren. (Die in diesem Abschnitt abgeleiteten Rechenregeln und Sachverhalte sind stets gültig.)
> Der „Laplacesche Wahrscheinlichkeitsbegriff" betrifft die oben diskutierten Laplace-Modelle, in denen die Elementarereignisse alle gleichwahrscheinlich sind. Man kann dann Wahrscheinlichkeiten stets als „Günstige durch Mögliche" erklären, was technisch oft auf kombinatorische Fragen führt, wie sie im nächsten Abschnitt behandelt werden. Der Laplacesche Wahrscheinlichkeitsbegriff ist auf Situationen beschränkt, die von speziellen Symmetrien leben (z.B. Würfel oder Münze), um die Gleichverteilung jeweils zu rechtfertigen.
> Der „frequentistische Wahrscheinlichkeitsbegriff" versteht Wahrscheinlichkeiten von Ereignisse als Grenzwerte relativer Häufigkeiten, mit denen das Ereignis in wiederholten, voneinander unabhängigen Zufallsexperimenten eintritt. Die sogenannten „Frequentisten" stellen deshalb die Gesetze der großen Zahlen, die wir später besprechen, an den Anfang, um Wahrscheinlichkeiten zu definieren.
> Beim „Bayesschen Wahrscheinlichkeitsbegriff" fließen subjektive, persönliche Einschätzungen eines Sachverhaltes in die Interpretation von Wahrscheinlichkeiten ein. Die „Bayesianer" verstehen Wahrscheinlichkeiten als Grad der eigenen Überzeugung.
>
> Eine ausführliche Darstellung der Entwicklung des Wahrscheinlichkeitsbegriffs für Studierende des Lehramts findet sich in [1, Abschnitt 3.1]. Zudem interessant sind Hinweise in [2, 6].
>
> Vorstellungen zum Wahrscheinlichkeitsbegriff finden sich in den Schwerpunktsetzungen des Inhaltsfelds „Daten und Zufall" der Jahrgangsstufen 5/6. Das Rechnen mit Wahrscheinlichkeiten wird in den Jahrgangsstufen 7–10 unter „zwei- und mehrstufige Zufallsexperimente" behandelt.

3 Kombinatorik

Wir betrachten zunächst vier Abzählprobleme.

A1 Wie viele 10-stellige Dualzahlen gibt es?

A2 Auf wie viele Arten können 3 verschiedene Autos auf 8 Parkplätzen parken?

A3 Wie viele mögliche Ergebnisse gibt es beim Lotto „6 aus 49“ (ohne Superzahl)?

A4 Auf wieviele Arten können 10 gleiche 1-Euro-Münzen auf 3 Kinder verteilt werden?

Es sei im Folgenden stets $M = \{1, \dots, n\}$.

Modell I: **Stichprobe der Länge k aus M in Reihenfolge mit Zurücklegen.**

$$\Omega_I = M^k = M \times \cdots \times M = \{(\omega_1, \dots \omega_k) \mid \omega_i \in M \text{ für } i = 1, \dots, k\}.$$

Satz 3.1. Es gilt $|\Omega_I| = n^k$.

Beispiel 3.2. Zum Abzählproblem A1 vom Anfang dieses Abschnitts: Wir generieren die 10 stelligen Dualzahlen, indem wir ihre Ziffern (Bits) aus der Menge $\{0, 1\}$ ziehen. Betrachten wir die beiden 10-stelligen Dualzahlen 0111010110 und 1011010110. Offenbar macht es einen Unterschied, in welcher Reichenfolge die Bits gezogen wurden, wir ziehen also unter Beachtung der Reihenfolge (also eine Stichprobe in Reihenfolge). Zudem können Bits mehrfach in der Dualzahl vorkommen. Wir müssen also ein gezogenes Bit notieren und wieder in M zurücklegen, bevor wir erneut ziehen. Es handelt sich also um eine Stichprobe mit Zurücklegen. Damit liegt tatsächlich Modell I vor mit $M = \{0, 1\}$, also $n = 2$ und $k = 10$. Es ergeben sich $2^{10} = 1024$ verschiedene 10-stellige Dualzahlen.

Modell II: **Stichprobe der Länge k aus M in Reihenfolge ohne Zurücklegen** ($k \leq n$).

$$\Omega_{II} = \left\{(\omega_1, \dots, \omega_k) \in M^k \,\middle|\, \omega_i \neq \omega_j \text{ für } i \neq j\right\}.$$

Satz 3.3. Es gilt $|\Omega_{II}| = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$.

Ist $k = n$, so befinden wir uns im Spezialfall $\Omega = \mathcal{S}_n$, der Menge aller Permutationen von M, auch *symmetrische Gruppe* von M genannt. Es folgt, dass $|\mathcal{S}_n| = n!$ gilt.

Beispiel 3.4. Zum Abzählproblem A2 vom Anfang dieses Abschnitts: Zu jedem der drei Autos muss ein Parkplatz zugeordnet werden. Wir ziehen für jedes Auto also aus der Menge $M = \{1, \dots, 8\}$ der möglichen Parkplätze. Da auf jedem Parkplatz nur ein Auto geparkt werden kann, ziehen wir ohne Zurücklegen, d.h. nachdem etwa Parkplatz 3 gezogen wurde, steht er nicht weiter zur Verfügung. Es handelt sich also um eine Stichprobe ohne Zurücklegen. Da es einen Unterschied macht, wie die Autos am Ende auf den drei gezogenen Parkplätzen zu stehen kommen (sie wollen ja ihr Auto wiederfinden und merken sich z.B. die Parkplatznummer), ist die Reihenfolge des Ziehens relevant, also eine Stichprobe in Reihenfolge. Es liegt also Modell II vor mit $M = \{1, \dots, 8\}$, also $n = 8$ und $k = 3$. Es ergeben sich $8!/5! = 8 \cdot 7 \cdot 6 = 336$ Möglichkeiten.

Modell III: **Stichprobe der Länge k aus M ohne Reihenfolge ohne Zurücklegen.**

$$\Omega_{\mathrm{III}} = \left\{ \{\omega_1, \dots, \omega_k\} \,\middle|\, \omega_i \in M, \omega_i \neq \omega_j \text{ für alle } 1 \le i < j \le n \right\}.$$

Satz 3.5. Es gilt $|\Omega_{\mathrm{III}}| = \frac{n!}{k!(n-k)!} =: \binom{n}{k}$.

Beweis.

Betrachte zunächst $\Omega_{\mathrm{II}} = \left\{ (\omega_1, \dots, \omega_k) \in M^k \,|\, \omega_i \neq \omega_j \text{ für } i \neq j \right\}$ und die Äquivalenzrelation $\sim$ auf Ω_{II} : $(\omega_1, \dots, \omega_k) \sim (\omega_1', \dots, \omega_k')$, falls eine Permutation π von $\{1, \dots, k\}$ existiert mit $\omega_i = \omega'_{\pi(i)}$ für $i = 1, \dots, k$. Offenbar können wir die Quotientenmenge $\Omega_{\mathrm{II}}/\sim$ mit Ω_{III} identifizieren. Jede Äquivalenzklasse hat $k!$ Elemente. Ein Repräsentant ist etwa jeweils $(\omega_1, \dots, \omega_k) \in \Omega_{\mathrm{II}}$ mit $\omega_1 < \omega_2 < \dots < \omega_k$. Damit folgt $|\Omega_{\mathrm{III}}| = |\Omega_{\mathrm{II}}|/k! = n!/(k!(n-k)!)$. ∎

Aus dem Beweis folgt, dass man statt Ω_{III} alternativ auch

$$\Omega'_{\mathrm{III}} = \left\{ (\omega_1, \dots, \omega_k) \in M^k \,\middle|\, \omega_1 < \omega_2 < \dots < \omega_k \right\}$$

wählen kann.

Beispiel 3.6. Zum Abzählproblem A3 vom Anfang dieses Abschnitts: Offenbar können wir beim Lotto die Menge der Kugeln, aus denen gezogen wird, mit $M = \{1, \dots, 49\}$ modellieren. Es ist also $n = 49$. Beim Ziehen der Kugeln werden diese nicht zurückgelegt, können also nur einmal gezogen werden. Es handelt sich also um eine Stichprobe ohne Zurücklegen. Spielt die Reihenfolge des Ziehens eine Rolle? Nein, es werden ja 6 Zahlen auf dem Lottoschein angekreuzt und man gewinnt unabhängig davon, in welcher Reihenfolge die angekreuzten Zahlen gezogen werden. Also ist die Stichprobe ohne Reihenfolge. Es liegt somit Modell III vor mit $n = 49$ und $k = 6$. Damit ergeben sich $\binom{49}{6} = 13.983.816$ Möglichkeiten.

Modell IV: **Stichprobe der Länge k aus M ohne Reihenfolge mit Zurücklegen.**

$$\Omega_{\mathrm{IV}} = \left\{ (\omega_1, \dots, \omega_k) \in M^k \,\middle|\, \omega_1 \le \omega_2 \le \dots \le \omega_k \right\}.$$

Satz 3.7. Es gilt $|\Omega_{\text{IV}}| = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$.

Beweis.
Wir betrachten $M^* = \{1, \ldots, n+k-1\}$ und

$$\Omega^*_{\text{III}} = \left\{ (\omega^*_1, \ldots, \omega^*_k) \in (M^*)^k \,\middle|\, \omega^*_1 < \omega^*_2 < \cdots < \omega^*_k \right\}$$

sowie die Abbildung $f : \Omega_{\text{IV}} \to \Omega^*_{\text{III}}$ mit

$$(\omega_1, \ldots, \omega_k) \mapsto (\omega_1, \omega_2 + 1, \ldots, \omega_k + k - 1).$$

Man sieht leicht, dass f bijektiv ist. Damit gilt $|\Omega_{\text{IV}}| = |\Omega^*_{\text{III}}|$. Nach Modell III gilt $|\Omega^*_{\text{III}}| = \binom{n+k-1}{k}$. ∎

Beispiel 3.8. Zum Abzählproblem A4 vom Anfang dieses Abschnitts: Die zehn Münzen sollen an drei Kinder verteilt werden, d.h. wir müssen zu jeder der Münzen jeweils ein Kind auswählen (ziehen), das die Münze erhält. Wir modellieren die Menge der Kinder deshalb mit $M = \{1, 2, 3\}$ und ziehen $k = 10$ Mal. Da bei jeder der 10 Münzen alle drei Kinder verfügbar sind, ist des eine Stichprobe mit Zurücklegen. Am Ende interessiert uns lediglich, wieviele Münzen Kind 1, Kind 2 und Kind 3 haben. Die Reihenfolge, in der die Münzen verteilt wurden, um zu dieser finalen Aufteilung der Münzen zu kommen, ist unerheblich. Deshalb ist es eine Stichprobe ohne Reihenfolge. Es liegt also Modell 4 mit $n = 3$ und $k = 10$ vor. Es gibt deshalb $\binom{n+k-1}{k} = \binom{12}{10} = \binom{12}{2} = 66$ Möglichkeiten.

Definition 3.9. Die Größen $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ für $k = 0, 1, \ldots, n$ heißen *Binomialkoeffizienten*. Wir setzen $\binom{n}{k} := 0$ für $k < 0$ oder $k > n$.

Bemerkung zur Didaktik/Schule: Die folgende Bemerkung ist für die Vorlesung nicht wesentlich und kann übergangen werden.

Bemerkung 4. Es gibt eine alternative Klassifikation der Modelle I-IV, bei der statt Ziehen aus einer Urne mit und ohne Zurücklegen Objekte auf Zellen verteilt und diese Möglichkeiten gezählt werden. Dies ist aus der statistischen Physik motiviert.
Die vier Modelle

k-maliges sukzessives Ziehen aus einer Urne mit n nummerierten Kugeln:

- mit/ohne Zurücklegen,
- mit/ohne Beachten der Reihenfolge des Ziehens,

lassen sich übertragen in

Besetzung von n Zellen durch k Objekte

- mit/ohne Mehrfachbesetzungen,
- unterscheidbare/ununterscheidbare Objekte.

Pauliprinzip: Mehrfachbesetzungen sind verboten.

Man findet eine Darstellung dazu in Kapitel 9 des Buchs [5]

Der Rest dieses Abschnitts besteht aus Anwendungen und Verallgemeinerungen der vier kombinatorischen Grundmodelle.

Korollar 3.10 (Binomischer Lehrsatz)**.** Für alle $x, y \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Beweis.

$$\begin{aligned}(x+y)^n =&(x+y)\cdots(x+y) = \sum_{A\subset\{1,\ldots,n\}} x^{|A|}y^{|A^c|}\\ =&\sum_{k=0}^{n}\sum_{A\subset\{1,\ldots,n\}:|A|=k} x^k y^{n-k} \overset{(3.5)}{=} \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.\end{aligned}$$

∎

Korollar 3.11. Für $n \in \mathbb{N}$ gelten

$$\sum_{k=0}^{n}\binom{n}{k} = 2^n, \quad \sum_{k=0}^{n}\binom{n}{k}(-1)^k = 0.$$

Beweis.
Es gilt $\sum_{k=0}^{n}\binom{n}{k} = \sum_{k=0}^{n}\binom{n}{k}1^k 1^{n-k} = (1+1)^n = 2^n$ und mit $x = -1, y = 1$ in Korollar 3.10 erhält man auch die zweite Summe. ∎

Beispiel 3.12. In einer Urne seien s schwarze und w weiße Kugeln, $n := s + w$. Es werden $k \leq n$ Kugeln ohne Zurücklegen gezogen. Was ist die Wahrscheinlichkeit, dass die Stichprobe genau ℓ schwarze und $k - \ell$ weiße Kugeln enthält?

Wir können dies wie folgt modellieren: Seien

$$A = \{1, \ldots, n\}, \quad A_s = \{1, \ldots, s\}, \quad A_w = A \setminus A_s = \{s+1, \ldots, n\}.$$

Ferner sei

$$\Omega = \left\{(\omega_1, \ldots, \omega_k) \in A^k \,\middle|\, \omega_1 < \omega_2 < \cdots < \omega_k\right\},$$

also $|\Omega| = \binom{n}{k}$ (Modell III). Es sei

$$\begin{aligned} B_\ell &= \text{„genau } \ell \text{ schwarze Kugeln unter } k \text{ gezogenen“} \\ &= \{\omega \in \Omega \mid \omega_i \in A_s \text{ für } i = 1, \ldots, \ell, \text{ und } \omega_i \in A_w \text{ für } i = \ell+1, \ldots, k\}. \end{aligned}$$

Es gilt $|B_\ell| = \binom{s}{\ell}\binom{w}{k-\ell}$. Das Laplace-Modell liefert

$$\mathbb{P}(B_\ell) = \frac{\binom{s}{\ell}\binom{w}{k-\ell}}{\binom{s+w}{k}} =: h(\ell; k, s+w, s).$$

Satz 3.13. Für die Parameter $s, w \in \mathbb{N}$ und $1 \leq k \leq s+w$ ist durch

$$p_\ell := h(\ell; k, s+w, s) = \frac{\binom{s}{\ell}\binom{w}{k-\ell}}{\binom{s+w}{k}}, \quad \ell = 0, \ldots, k \tag{5}$$

eine Wahrscheinlichkeitsverteilung auf $\{0, \ldots, k\}$ definiert. Sie heißt *hypergeometrische Verteilung.*

Beweis.
Seien Ω und B_ℓ wie oben. Die B_ℓ sind für $\ell = \max\{0, k-w\}, \ldots, \min\{k, s\}$ paarweise disjunkt mit $\Omega = \bigcup B_\ell$. Es folgt

$$\sum_{\ell=0 \vee (k-w)}^{k \wedge s} h(\ell; k, s+w, s) = 1.$$

Für $\ell < 0 \vee (k-w)$ oder $\ell > k \wedge s$ gilt $h(\ell; k, s+w, s) = 0$, da entsprechende Binomialkoeffizienten in (5) nach der Definition 3.9 Null sind. Wir haben also $\sum_{\ell=0}^{k} p_\ell = 1$. ∎

Beispiel 3.14 (Beispiele zur hypergeometrischen Verteilung).
(1) Die Wahrscheinlichkeit, genau ℓ Richtige im Lotto „6 aus 49“ zu haben: $s = 6$, $w = 43$, $k = 6$, also $h(\ell; 6, 49, 6)$.

(2) Bei einer Qualitätskontrolle von n Produktionsstücken seien s defekt, $w = n - s$ nicht defekt. Es werde eine Stichprobe der Größe k genommen. Die Wahrscheinlichkeit, dass genau ℓ Defekte unter der Stichprobe sind, ist $h(\ell, k, s+w, s)$.

(3) Die Wahrscheinlichkeit, dass Person A beim Skat 3 Asse erhält (sie erhält 10 von 32 Karten, in denen sich insgesamt 4 Asse befinden), ist

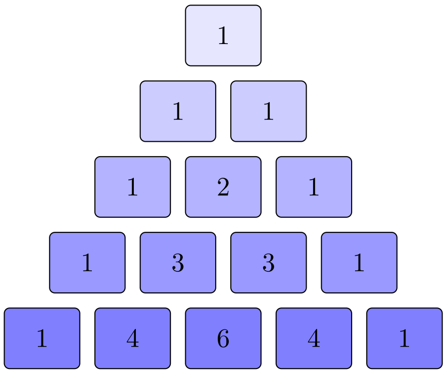
$$\frac{\binom{4}{3}\binom{28}{7}}{\binom{32}{10}} = \frac{66}{899}.$$

Bemerkung zur Didaktik/Schule: Für alle $n \geq 1$ und $1 \leq k \leq n$ gilt

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}. \tag{6}$$

Geben Sie einen analytischen Beweis durch Rechnung mit Fakultäten, siehe Definition 3.9. Geben Sie einen zweiten, kombinatorischen Beweis, der nutzt, dass $\binom{n}{k}$ die Anzahl k-elementiger Teilmengen einer n-elementigen Menge angibt. (*Hinweis dazu:* Sie können unterscheiden, ob eine k-elementige Teilmenge ein ausgezeichnetes Element einer n-elementigen Menge enthält oder nicht.) Welchen der beiden Beweise würden Sie in einem verständnisorientierten Unterricht bevorzugen?

Die Identität (6) begündet die Darstellung der Binomialkoeffizienten im *Pascalschen Dreieck* (die Zahlen ergeben sich jeweils als Summe der direkt rechts und links darüberliegenden Zahlen):



Ähnlich kann man folgende Identitäten kombinatorisch einsehen: Für alle $m \geq n \geq 1$ gilt

$$\binom{n+m}{m} = \sum_{k=0}^{n} \binom{n}{k}\binom{m}{m-k}, \tag{7}$$

$$\binom{m}{n} = \sum_{k=n}^{m} \binom{k-1}{n-1},$$

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

Indem Sie die Kardinalität der Potenzmenge $\mathcal{P}(\{1,\ldots,n\})$ auf zwei Arten zählen, können Sie kombinatorisch auch die Identität

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \quad n \in \mathbb{N}$$

aus Korollar 3.11 direkt einsehen.

Für mit fortgeschrittenen mathematischen Methoden bewiesene Identitäten sind verständnisorientierte Argumente nicht immer offensichtlich, z.B. gilt für alle $n \geq 1$ die Identität (versuchen Sie nicht, diese zu beweisen)

$$\sum_{k=0}^{n} \binom{n}{k}^2 \binom{3n+k}{2n} = \binom{3n}{n}^2.$$

4 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Beispiel 4.1. 1. Ein fairer Würfel werde geworfen. Modell: $\Omega = \{1, \dots, 6\}$, $\mathbb{P}(\{i\}) = 1/6$ für $i = 1, \dots, 6$. Ein Beobachter verrate, dass eine gerade Zahl geworfen wurde. Für die neue Situation gilt intuitiv

$$\widetilde{\mathbb{P}}(\{i\}) = \begin{cases} 0, & \text{falls } i \text{ ungerade} \\ \frac{1}{3}, & \text{falls } i \text{ gerade.} \end{cases}$$

2. Versicherungsproblem: Ein männlicher Bürger werde genau k Jahre alt mit Wahrscheinlichkeit p_k, $k \in \mathbb{N}$, mit $\sum_{k \geq 1} p_k = 1$. Gesucht: Wahrscheinlichkeit q_ℓ, dass er im ℓ-ten Lebensjahr stirbt, gegeben, dass er bereits das k-te Jahr erreicht hat. Dazu:

$$\begin{aligned} s_k &= \mathbb{P}\,(\text{wird mindestens } k \text{ Jahre alt}) \\ &= \mathbb{P}\left(\bigcup_{i=k}^{\infty} \{\text{wird genau } i \text{ Jahre alt}\}\right) = \sum_{i=k}^{\infty} p_i. \end{aligned}$$

Nun ist intuitiv (heuristisch über relative Häufigkeiten einsichtig)

$$q_\ell = \begin{cases} 0, & \text{für } \ell < k \\ p_\ell / s_k, & \text{für } \ell \geq k. \end{cases}$$

Allgemeines Konzept

Definition 4.2. Sei $(\Omega, \mathbb{P})$ ein endlicher Wahrscheinlichkeitsraum. Seien $A, B \subset \Omega$ mit $\mathbb{P}(B) > 0$. Dann heißt

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{8}$$

die *bedingte Wahrscheinlichkeit von* A unter (Bedingung) B.

> **Bemerkung zur Didaktik/Schule:** Wem die Definition der bedingten Wahrscheinlichkeit in (8) durch die Beispiele darüber nicht ausreichend motiviert erscheint, kann sie für den Fall von Laplace-Modellen expliziter motivieren. Man vergleiche dazu die Diskussion in [1, Seite 202] oder das Beispiel zum Ziehen zweier Kugeln aus einer Urne in [4, Seite 52], das dort auf Seite 53 unten dann nochmals aus dem Blickwinkel der formalen Definition (8) aufgegriffen wird.
> Das Rechnen mit bedingten Wahrscheinlichkeiten betrifft die Schwerpunktsetzungen zweistufige Zufallsexperimente, Baumdiagramme und Vierfeldertafeln sowie Pfadregeln des Inhaltsfelds „Daten und Zufall“ der Jahrgangsstufen 7/8 sowie für mehrstufige Zufallsexperimente in den Jahrgangsstufen 9/10.

Im Folgenden sei $(\Omega, \mathbb{P})$ stets ein endlicher Wahrscheinlichkeitsraum.

Satz 4.3 (Satz von der totalen Wahrscheinlichkeit)**.** Sei $B_1, B_2, \dots, B_n \subset \Omega$ eine Zerlegung von Ω, d.h. B_i sind paarweise disjunkt und $\bigcup_i B_i = \Omega$. Dann gilt für $A \subset \Omega$:

$$\mathbb{P}(A) = \sum_{\{i \,|\, \mathbb{P}(B_i)>0\}} \mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i).$$

Beweis.
Es ist $A = A \cap \Omega = A \cap \bigcup_i B_i = \bigcup_i (B_i \cap A)$, wobei die Mengen $B_i \cap A$ paarweise disjunkt sind für $i \in \mathbb{N}$. Es folgt

$$\mathbb{P}(A) = \sum_{\{i \,|\, \mathbb{P}(B_i)>0\}} \mathbb{P}(A \cap B_i) = \sum_{\{i \,|\, \mathbb{P}(B_i)>0\}} \mathbb{P}(A \,|\, B_i)\mathbb{P}(B_i).$$

∎

Satz 4.4 (Satz von Bayes (1763))**.** Sei $A \subset \Omega$ mit $\mathbb{P}(A) > 0$ und $B_1, B_2, \dots, B_n \subset \Omega$ eine Zerlegung von Ω mit $\mathbb{P}(B_i) > 0$ für $i = 1, 2, \dots, n$. Dann gilt

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

Beweis.
Es ist

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

wobei die letzte Gleichheit durch den Satz von der totalen Wahrscheinlichkeit (Satz 4.3) begündet wird. ∎

Der Satz von der totalen Wahrscheinlichkeit und der Satz von Bayes werden oft im Fall $n = 2$ benötigt. Dabei ist die Zerlegung von Ω gegeben durch $\Omega = B \cup B^c$. Wir nehmen für $A, B \subset \Omega$ an, dass $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(B^c) > 0$ gilt, was in Anwendungen (und Aufgaben) oft gegeben ist. Dann schreiben sich der Satz von der totalen Wahrscheinlichkeit und der Satz von Bayes von oben vereinfacht folgendermaßen:

Satz von der totalen Wahrscheinlichkeit:

$$\boxed{\mathbb{P}(A) = \mathbb{P}(A \,|\, B)\mathbb{P}(B) + \mathbb{P}(A \,|\, B^c)\mathbb{P}(B^c).}$$

Satz von Bayes:

$$\boxed{\mathbb{P}(B \,|\, A) = \frac{\mathbb{P}(A \,|\, B)\mathbb{P}(B)}{\mathbb{P}(A \,|\, B)\mathbb{P}(B) + \mathbb{P}(A \,|\, B^c)\mathbb{P}(B^c)}.} \tag{9}$$

Beispiel 4.5 (Test für eine seltene Krankheit)**.** Eine Krankheit trete insgesamt bei 0,5% der Bevölkerung auf.

$$\text{Ein Test führe bei } \begin{cases} 99\% \text{ der Kranken zur Reaktion,} \\ 2\% \text{ der Gesunden zur Reaktion.} \end{cases}$$

Man sagt auch, der Test sei positiv, wenn er zur Reaktion führt, andernfalls negativ. Gesucht ist die Wahrscheinlichkeit, dass eine Person, bei der der Test positiv ist, tatsächlich die Krankheit hat.
Formalisierung: Sei X eine zufällig ausgewählte Person und $B = \{X \text{ hat die Krankheit}\}$, d.h. $\mathbb{P}(B) = 0{,}005$, auch „Prävalenz“ genannt. Ferner sei $A = \{\text{bei } X \text{ ist der Test positiv}\}$, d.h. $\mathbb{P}(A \,|\, B) = 0{,}99$, auch „Sensitivität“ genannt, und $\mathbb{P}(A \,|\, B^c) = 0{,}02$. (Die „Spezifität“ ist damit $\mathbb{P}(A^c \,|\, B^c) = 0{,}98$.) Die gesuchte Wahrscheinlichkeit ist $\mathbb{P}(B \,|\, A)$. Nach dem Satz von Bayes mit der disjunkten Zerlegung $\Omega = B \cup B^c$ gilt nach (9), dass

$$\mathbb{P}(B \,|\, A) = \frac{\mathbb{P}(A \,|\, B)\mathbb{P}(B)}{\mathbb{P}(A \,|\, B)\mathbb{P}(B) + \mathbb{P}(A \,|\, B^c)\mathbb{P}(B^c)} = \frac{99}{497} \approx 0{,}2.$$

Von allen Personen, bei denen der Test positiv ist, sind also nur etwa 20% tatsächlich krank.

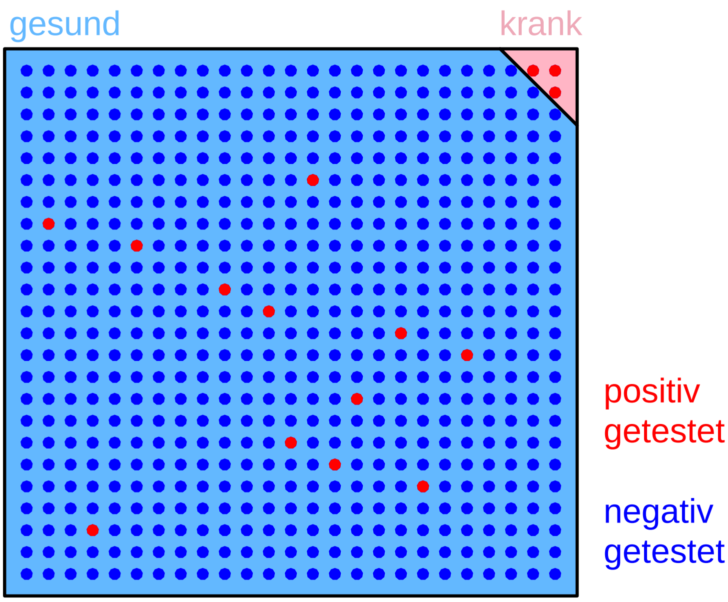


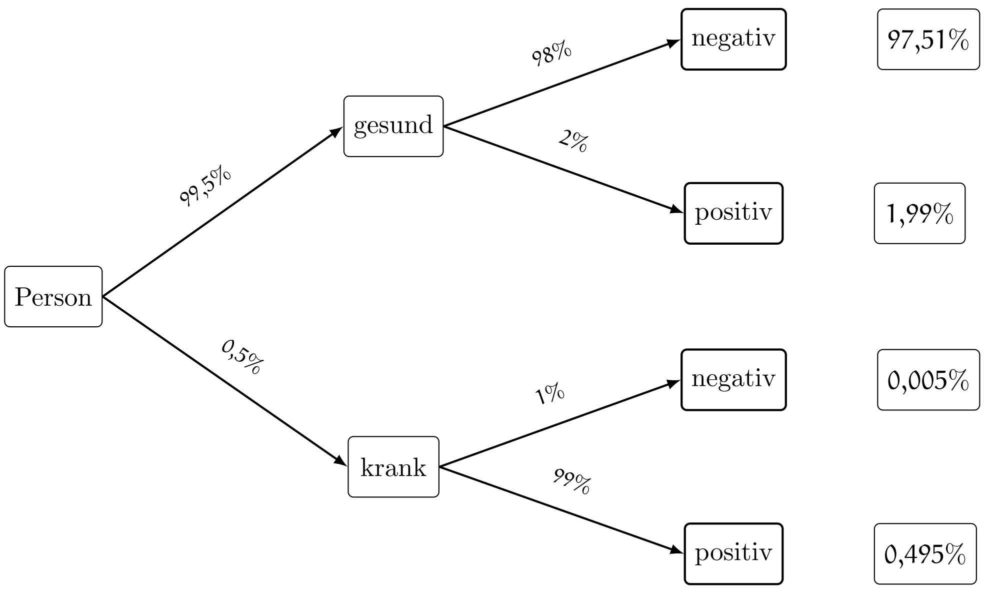
Abbildung 1: Beispielhafte Veranschaulichung der Situation aus Beispiel 4.5.

Bemerkung zur Didaktik/Schule. Bei Anwendungen des Satzes von Bayes (auch Bayessche Regel genannt) kommt es öfter zu Ergebnissen, die zunächst wenig plausibel erscheinen mögen. Wie kann es in Beispiel 4.5 sein, dass ein Verfahren, das sowohl bei den Kranken als auch bei den Gesunden zuverlässig ist, doch im Falle eines positiven Tests mehr Unklarheit als Klarheit schafft?
Das liegt hier hauptsächlich daran, dass die Gruppen der Gesunden und Kranken so unterschiedlich groß sind. Abbildung 1 veranschaulicht diese Situation. Die Kranken sind durch die Kreise im pinken Bereich (Dreieck) dargestellt, die Gesunden durch die restlichen Kreise im hellblauen Bereich. Ein rot gefärbter Kreis bedeutet, dass

der Test positiv ist, blau gefärbt, dass der Test negativ ist. Die Verhältnisse im Bild entsprechen in etwa den Zahlen in Beispiel 4.5. Die Gesunden haben nur selten einen positiven Test, die Kranken (fast) alle. Da nun aber die Gruppe der Gesunden so groß ist gebenüber den Kranken, gibt es unter den Gesunden, absolut gesehen, doch wesentlich mehr positive Testergebnisse als unter den Kranken.
Wenn nun nach der Wahrscheinlichkeit gefragt wird, dass eine Person krank ist, gegeben der Test ist positiv, so bedeutet dies, dass unter den positiv getesteten Personen (also unter allen roten Kreisen) eine Person gleichverteilt gezogen wird und die Wahrscheinlichkeit gefragt ist, dass diese Person krank ist. Es wird nun offenkundig, dass unter allen roten Kreisen die der Gesunden in großer Überzahl sind, was dazu führt, dass die gezogene positiv getestete Person wesentlich häufiger gesund als krank ist.

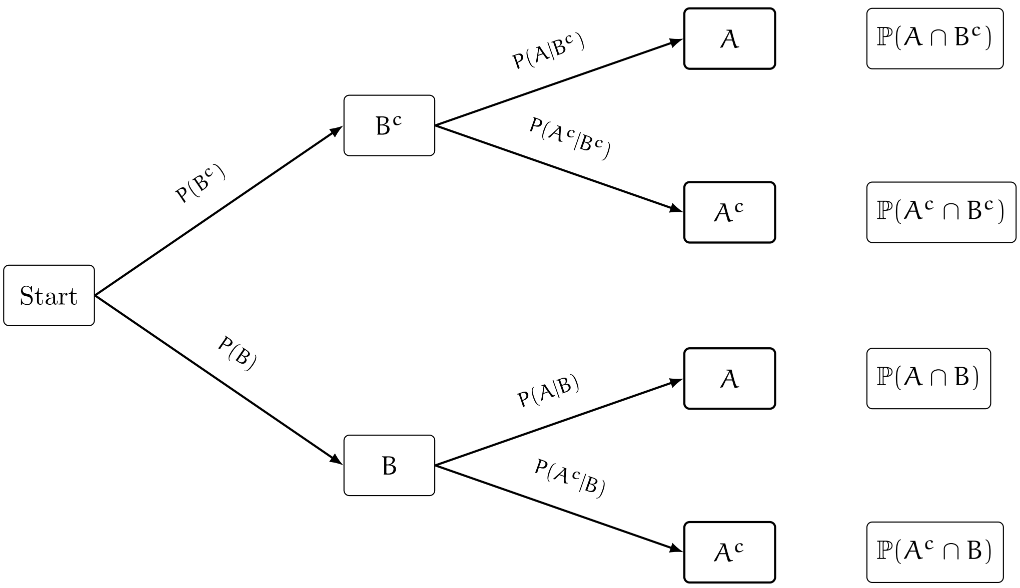
Übung: Ändern Sie nun in Beispiel 4.5 die Prävalenz von $0{,}5\%$ auf 50%. Berechnen Sie damit $\mathbb{P}(B \mid A)$ und machen Sie sich klar, wie eine entsprechende Graphik aussehen könnte.

Verbreitete Methoden, Wahrscheinlichkeiten (und Häufigkeiten) im Kontext des Satzes von Bayes graphisch darzustellen, sind Vierfeldertafeln und Baumdiagramme. Für unser Beispiel sieht das entsprechende Baumdiagramm wie folgt aus:



Entlang der Pfade eines Baumdiagramms werden die Wahrscheinlichkeiten multipliziert. Um z.B. die Wahrscheinlichkeit, dass eine Person gesund ist und positiv getestet wird (am Start links erst den Ast nach oben, von dort dann den Ast nach unten), zu ermitteln, müssen am ersten Ast die Wahrscheinlichkeit für gesund, am folgenden Ast dann die bedingte Wahrscheinlichkeit für positiven Test gegeben, dass die Person gesund ist, eingetragen werden. Wollen wir nun die gesuchte Wahrscheinlichkeit $\mathbb{P}(B \mid A)$ ermitteln, so finden wir den Zähler $\mathbb{P}(A \cap B)$ direkt im Baum, den Nenner $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$

durch Addition zweier Einträge.
Abstrakt mit allgemeinen Ereignissen A und B sieht ein Baumdiagramm also so aus:



Eine andere Möglichkeit, die relvanten Wahrscheinlichkeiten darzustellen, ist die Vierfeldertafel. Für allgemeine Ereignisse A und B sieht die Vierfeldertafel wie folgt aus:

	B	B^c	Total
A	$P(A \cap B)$	$P(A \cap B^c)$	$P(A)$
A^c	$P(A^c \cap B)$	$P(A^c \cap B^c)$	$P(A^c)$
Total	$P(B)$	$P(B^c)$	1

In unserem Beispiel erhalten wir also:

	krank	gesund	Total
positiv	0,495%	1,99%	2,485%
negativ	0,005%	97,51%	97,515%
Total	0,5%	99,5%	1

Falls keine Wahrscheinlichkeiten bekannt sind, sondern Daten erhoben werden, werden die entsprechenden Anzahlen in die Vierfeldertafel eingetragen. Man kann dann die gesuchten Wahrscheinlichkeiten entsprechend aus der Vierfeldertafel schätzen. Z.B. überträgt sich die Situation aus Abbildung 1 mit den 600 Personen wie folgt in die Vierfeldertafel:

	krank	gesund	Total
positiv	3	12	15
negativ	0	585	585
Total	3	597	600

Bemerkung zur Didaktik/Schule. Man findet eine Diskussion der Baumdiagramme und Vierfeldertafeln für die Schulpraxis etwa in [1, Tabelle 3.4 und Abb. 3.33]. Eine weitere Möglichkeit der Darstellung ist der Häufigkeitsdoppelbaum, der in der Vorlesung „Didaktik der Stochastik“ angesprochen wird, siehe dazu auch die Arbeit von Karin Binder, Stefan Krauss und Christoph Wassner „Der Häufigkeitsdoppelbaum als didaktisch hilfreiches Werkzeug von der Unterstufe bis zum Abitur“ in der Zeitschrift *Stochastik in der Schule*, 38(1):2–17, 2018.

Definition 4.6. (a) Zwei Ereignisse $A, B \subset \Omega$ heißen (stochastisch) *unabhängig*, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ gilt.

(b) $A_1, \dots, A_n \subset \Omega$ heißen unabhängig, falls für alle $2 \leq k \leq n$ und $1 \leq j_1 \leq \dots \leq j_k \leq n$ gilt

$$\mathbb{P}\left(\bigcap_{i=1}^{k} A_{j_i}\right) = \prod_{i=1}^{k} \mathbb{P}(A_{j_i}) \qquad \text{„Produktformel“.}$$

(c) $A_1, \dots, A_n \subset \Omega$ heißen *paarweise unabhängig*, falls für alle $1 \leq i < j \leq n$ gilt: A_i und A_j sind unabhängig.

Lemma 4.7. Sei $B \subset \Omega$ mit $\mathbb{P}(B) > 0$. Dann gilt:

$$A, B \text{ sind unabhängig } \Leftrightarrow \mathbb{P}(A \mid B) = \mathbb{P}(A).$$

Beweis.
„$\Rightarrow$“: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = (\mathbb{P}(A)\mathbb{P}(B))/\mathbb{P}(B) = \mathbb{P}(A)$.
„$\Leftarrow$“: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B)$. ∎

Beispiel 4.8 (Zweimaliges Würfeln mit fairem Würfel)**.** Es ist

$$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\} = \{(\omega_1, \omega_2) \mid \omega_i \in \{1, \dots, 6\} \text{ für } i = 1, 2\}.$$

Im Laplace-Modell gehen wir davon aus, dass die Wahrscheinlichkeit eines Ausgangs beim zweifachen Wurf gegeben ist durch $\mathbb{P}(\{(\omega_1, \omega_2)\}) = 1/36$ für alle $(\omega_1, \omega_2) \in \Omega$. Seien $A =$“beim ersten Wurf < 4“ und $B =$“beim zweiten Wurf ≥ 3“.

$$\begin{aligned} A &= \{(\omega_1, \omega_2) \in \Omega : \omega_1 < 4\}, \quad |A| = 18, \quad \mathbb{P}(A) = \frac{1}{2} \\ B &= \{(\omega_1, \omega_2) \in \Omega : \omega_2 \geq 3\}, \quad |B| = 24, \quad \mathbb{P}(B) = \frac{2}{3}. \\ A \cap B &= \{(\omega_1, \omega_2) \in \Omega : \omega_1 < 4, \omega_2 \geq 3\}, \quad |A \cap B| = 12, \quad \mathbb{P}(A \cap B) = \frac{1}{3}. \end{aligned}$$

Es folgt $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, also sind A und B unabhängig.

Beispiel 4.9 (Bernoulli-Experimente)**.** Ein Zufallsexperiment mit zwei möglichen Ausgängen, also etwa ein Münzwurf, heißt Bernoulli-Experiment. Wir wollen die n-fache Wiederholung modellieren. Sei $\Omega_i = \{0, 1\}$ und $\mathbb{P}_i(\{1\}) = p = 1 - \mathbb{P}(\{0\})$. Man bezeichnet den Ausgang „1“ als Erfolg und nennt $p \in [0, 1]$ die *Erfolgswahrscheinlichkeit.* Ein Modell für die n-fache unabhängige Wiederholung ist:

$$\Omega = \{0, 1\}^n,\ \mathbb{P}(\{\omega\}) = p^k(1-p)^{n-k} \text{ für } \omega = (\omega_1, \ldots, \omega_n) \text{ mit } \sum_{i=1}^{n} \omega_i = k.$$

Mit k wird dabei die Anzahl der Erfolge gezählt. Man liest an den Wahrscheinlichkeiten ab, dass für $p = \frac{1}{2}$ ein Laplace-Modell vorliegt, für $p \neq \frac{1}{2}$ nicht.
Man kann nun leicht wie in Beispiel 4.8 nachrechnen, dass Ereignisse, die sich auf unterschiedliche Wiederholungen beziehen, unabhängig sind. Wählen Sie z.B. $n = 4$ und prüfen Sie die Unabhängigkeit der beiden Ereignisse, dass in den ersten beiden Wiederholungen genau ein Erfolg eintritt und dass in den letzten beiden Wiederholungen mindestens ein Erfolg eintritt. Man spricht deshalb von n unabhängigen Wiederholungen eines Bernoulli-Experiments.

Zwei grundlegende Fragen dazu sind die Folgenden:

- Wie lassen sich Wahrscheinlichkeiten für die Anzahl der Erfolge beschreiben?
- Wie lässt sich die Wartezeit auf den ersten Erfolg beschreiben?

Um die Anzahl der Erfolge zu untersuchen, betrachten wir die Ereignisse

$$E_k = \left\{ \omega \in \Omega : \sum_{i=1}^{n} \omega_i = k \right\} = \text{„genau } k \text{ Erfolge in } n \text{ Experimenten“}. \tag{10}$$

Obwohl hier (bis auf den Spezialfall $p = \frac{1}{2}$) kein Laplace-Modell vorliegt, können wir die Wahrscheinlichkeit geschlossen darstellen. Zunächst beachten wir, dass nach Modell III aus Abschnitt 2 offenbar $|E_k| = \binom{n}{k}$ gilt. Damit folgt

$$\mathbb{P}(E_k) = \sum_{\omega \in E_k} \mathbb{P}(\{\omega\}) = \sum_{\omega \in E_k} p^k(1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}. \tag{11}$$

Satz 4.10. Für die Parameter $n \in \mathbb{N}$ und $p \in [0, 1]$ ist durch

$$b_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$$

eine Wahrscheinlichkeitsmaß auf $\{0, \ldots, n\}$ definiert. Es heißt *Binomialverteilung* mit Parameter $n \in \mathbb{N}$ und $p \in [0, 1]$. Statt $b_{n,p}$ wird auch $B(n, p)$ geschrieben.

Beweis.
Da die Ereignisse E_k in (10) für $k = 0, \ldots, n$ eine disjunkte Zerlegung von Ω bilden, müssen sich die Zahlen in (11) zu 1 summieren. Dies liefert die Behauptung, vgl. Bemerkung 2. Alternativ kann man dies aus dem Binomischen Lehrsatz (Korollar 3.10) schließen. ∎

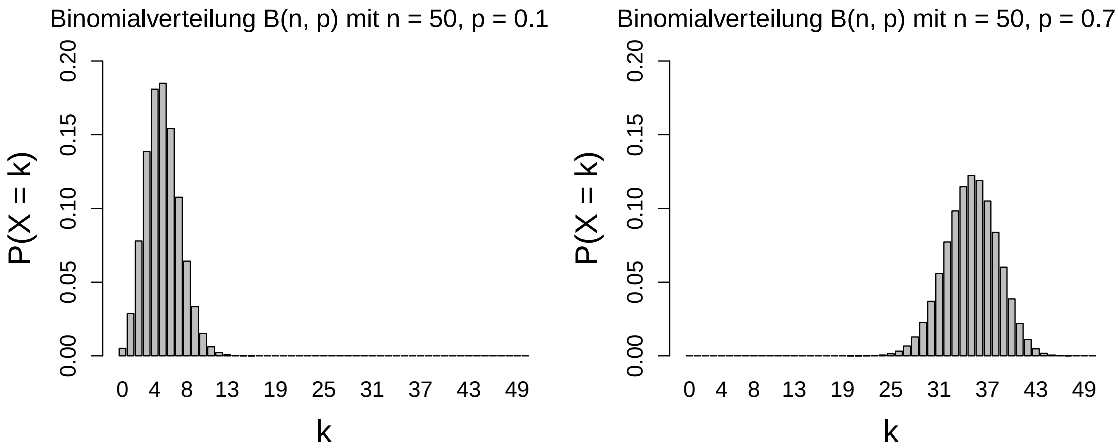


Abbildung 2: Gezeigt sind die Gewichte $b_{n,p}(\{k\})$ für $n = 50$ und $p = 0{,}1$ (links) sowie $p = 0{,}7$ (rechts), wobei auf der x-Achse $k = 0, \ldots, 50$ aufgetragen ist. Um derartige Diagramme mit R zu erzeugen, siehe Bemerkung 5.

Die Werte $b_{n,p}(\{k\})$, auch Gewichte oder Punktmassen der Binomialverteilung genannt, sind in Abbildung 2 beispielhaft dargestellt.

Korollar 4.11. Die Wahrscheinlichkeiten für die Anzahlen der Erfolge in n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ sind durch die Binomialverteilung $b_{n,p}$ beschrieben.

Wir betrachten nun die Wahrscheinlichkeit, im k-ten Teilexperiment erstmals einen Erfolg zu haben. Dies ist das Ereignis

$$\begin{aligned} F_k &= \{\omega \in \Omega : \omega_1 = \cdots = \omega_{k-1} = 0, \omega_k = 1\} \\ &= \{0\} \times \cdots \times \{0\} \times \{1\} \times \Omega_{k+1} \times \cdots \times \Omega_n, \end{aligned} \tag{12}$$

also ergibt sich

$$\mathbb{P}(F_k) = p(1-p)^{k-1}. \tag{13}$$

Man beachte, dass auch n Misserfolge möglich sind, weshalb die Zahlen in (13) für $k = 1, \ldots, n$ kein Wahrscheinlichkeitsmaß festlegen können. Für $k \in \mathbb{N}$ liefert (13) allerdings ein Wahrscheinlichkeitsmaß.

Satz 4.12. Zum Parameter $p \in (0, 1]$ ist durch

$$g_p(\{k\}) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

eine W-Verteilung auf $\mathbb{N}$ erklärt. Sie heißt *geometrische Verteilung* zum Parameter p.

Korollar 4.13. In einer Folge von unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in (0, 1]$ sind die Wahrscheinlichkeiten für den Index (Zeitpunkt), bei dem erstmals ein Erfolg eintritt, durch die geometrische Verteilung zum Parameter p beschrieben.

Bemerkung zur Didaktik/Schule: Die n-fachen unabhängigen Wiederholungen eines Bernoulli-Experiments sind Beispiele für mehrstufige Zufallsexperimente, die in den Jahrgangsstufen 7–10 behandelt werden.

3 Einführung in R

Die freie Programmiersprache R kann für stochastische Simulationen, statistische Auswertungen von Daten sowie für Graphiken zur Datenvisualisierung verwendet werden. Als Open Source Programm kann R unter

`https://www.r-project.org/`

frei für Windows, macOS und Linux heruntergeladen und installiert werden. Neben Anwendungen zu Stochastik in der Schule wird R in Studium und Forschung neben der mathematischen Statistik in vielen angewandten Bereichen verwendet, die empirisch arbeiten, etwa in den Wirtschaftswissenschaften, Medizin, Psychologie oder den Sozialwissenschaften.

In dieser Vorlesung wird eine knappe Einführung in die Syntax von R soweit gegeben, um mit einfachen zufälligen Simulationen einige stochastische Phänomene und Ergebnisse der Vorlesung zu illustrieren. Dies soll angehende Lehrkräfte in die Lage versetzen, Simulationen für den Unterricht vorzubereiten. Um schnell zu einem passenden Programm zu kommen, bieten sich KI-Systeme wie ChatGPT (Stand 2025) an, die oft zu einer gegebenen Problemstellung einen groben Code ausgeben, der dann noch angepasst und gegebenenfalls korrigiert werden kann. Die Darstellung von Daten (oder auch statistische Datenanalyse) ist eine weitere Stärke von R, die in der Vorlesung „Didaktik der Stochastik" mit besprochen werden kann.
Mit „R Studio" steht zudem eine integrierte Entwicklungsumgebung und grafische Benutzeroberfläche für R zur Verügung:

`https://www.rstudio.com/`

5 Zur Syntax von R

Wir fassen einige Grundlagen der Syntax von R zusammen.

Rechenoperationen und Standardfunktionen

In R Programmen werden Kommentare mit der Raute eingeleitet:

```
# Dies ist ein Kommentar.
```

Elementare Rechenoperationen werden wie folgt in R ausgeführt: Es wird + für Addition, − für Subtraktion, ∗ für Multiplikation, / für Division, und ˆ für Potenz verwendet. Insbesondere kann man R also auch als Taschenrechner nutzen, etwa

```
13 + 113
10 - 7
3 * 5
50 / 2                        # Achtung, 50:2 liefert nicht die Division
```

```
3 * 2.5             # Dezimalzahlen mit Punkten schreiben, nicht mit Komma
5 ^ 2
3 + 2*2                                             # Es gilt Punkt vor Strich
(3+2)*2
2*pi                                           # pi steht für die Kreiszahl π
```

Standardfunktionen sind in R verfügbar, etwa `log()`, `exp()`, `sin()`, `cos()`, `tan()`,..., also etwa

```
exp(1)
log(2)
3*cos(0)
abs(-123)                                   # abs() liefert den Absolutbetrag
sqrt(16)                                   # sqrt() liefert die Quadratwurzel
factorial(4)               # factorial() liefert die Fakultät, also hier 4!
choose(4,2)                      # choose(n,k) bedeutet (n k), also hier (4 2)
```

R unterscheidet Groß- und Kleinschreibung. Hilfe erhält man mit

```
help.start()
```

Zudem ist die `?`-Anweisung nützlich, um eine Dokumentation zu einer bestimmten Funktion aufzurufen, z.B.

```
?factorial
```

Es ist oft nützlich, ein Programm oder eine Sequenz von Anweisungen in einem separaten Editor zu schreiben und dann mit copy & paste in die Konsole von R zu kopieren, um sie ausführen zu lassen.

Variablen und Zuweisungen

Variablen und Zuweisungen von Werten zu Variablen spielen eine zentrale Rolle.

```
a <- 10
```

weist der Variablen a, die nicht zuvor spezifiziert werden muss, den Wert 10 zu. Dabei steht `<-` also für die Zuweisung. Den Wert von a gibt man nun wie folgt aus:

```
a
```

Wird a nun erneut ein Wert zugewiesen, so wird der alte Wert überschrieben.

```
a <- 2              # Die Variablen a wird nun mit dem Wert 2 überschrieben
3 * a                     # wir können mit Variablen rechnen wie mit Zahlen
a <- 3 * a                     # Dies verdreifacht den aktuellen Wert von a
10 * b                     # Dies ist fehlerhaft, da b noch keinen Wert hat
```

Vektoren

Das Arbeiten mit Vektoren ist zentraler Bestandteil von R, womit sich R von vielen anderen Programmiersprachen unterscheidet. Vektoren können auf verschiedene Arten erzeugt werden. Zum einen durch Auflisten der Komponenten des Vektors:

```
vektor1 <- c(1, 4, 5, 10, 56, 103)
vektor1
vektor2 <- c("Stelle 1","Stelle 2","Stelle 3","Stelle 4")
vektor2
```

Um auf Komponenten eines Vektors zuzugreifen, werden eckige Klammern verwendet, etwa liefert

```
vektor2[2]
```

die zweite Komponente des Vektors `vektor2`. Eine weitere Möglichkeit, Vektoren zu erzeugen, die später bei Schleifen eine Rolle spielt, ist `a:b`, was einen Vektor von Zahlen von a bis b in Einerschritten erzeugt, also etwa

```
1:10
10:1
```

Vektoren mit identischen Komponenten erhalten wir mit `rep(wert,anzahl)`, also liefert z.B.

```
rep(3,12)
```

einen Vektor der Länge 12, dessen Komponenten alle den Wert 3 haben.

Wesentliches Element von R ist, dass Rechenoperationen mit Vektoren möglich sind und dabei komponentenweise ausgeführt werden, z.B.

```
4 * c(5,7,1,3)
1:10 + 5
c(2,4,6) + c(7,4,1)
```

Ebenso wirken Funktionen auf Vektoren komponentenweise, also etwa

```
sqrt(c(1,4,9,16,25))
exp(1:10)
```

Zudem gibt es spezielle Funktionen für Vektoren:

```
vektor3 <- c(5, 7, 5, 8, 3)
length(vektor3)        # gibt die Anzahl der Komponenten von vektor3 aus
unique(vektor3)       # reduziert auf paarweise verschiedene Komponenten
min(vektor3)                # Gibt den minimalen Eintrag von vektor3 aus
max(vektor3)                # Gibt den maximalen Eintrag von vektor3 aus
sum(vektor3)             # Gibt die Summe aller Einträge von vektor3 aus
cumsum(vektor3)            # Gibt die kumulierende Summe von vektor3 aus
sort(vektor3)             # Gibt vektor3 in aufsteigender Reihenfolge an
mean(vektor3)                  # berechnet den arithmetischen Mittelwert
```

Schleifen, Wahrheitswerte und „if"-Abfragen

Zur wiederholten Ausführung derselben Anweisung können `for`-Schleifen verwendet werden. Zum Beispiel schreibt

```
for (i in 1:10){print(i)}
```

die Zahlen von 1 bis 10. Man erinnere sich, dass `1:10` für den Vektor $(1, 2, \ldots, 10)$ steht.

Logische Abfragen liefern einen der Wahrheitswerte `(TRUE)` oder `(FALSE)`. Zum Beispiel

```
10 < 4                                         # analog mit >, <=, >=
3 = 7                              # das entspricht nicht der Syntax von R
3 == 7                        # das doppelte = liefert die gewünschte Abfrage
4 != 5                                                   # != steht für ≠
```

Soll nun eine Anweisung nur ausgeführt werden, falls eine Bedingung erfüllt ist, also den Wahrheitswert `(TRUE)` hat, d.h. wahr ist, können „if"-Abfragen verwendet werden, z.B.

```
x <- 5
if (x > 4)print("x ist größer als 4")
```

und entsprechend

```
x <- 3
if (x > 4)print("x ist größer als 4")
```

Simulation in R

In R ist die Simulation zahlreicher Zufallsexperimente bereits implementiert.[1] Wollen wir aus einer endliche Menge M ohne Zurücklegen n Elemente ziehen (mit $n \leq |M|$), so schreiben wir die Elemente von M als Komponenten eines Vektors `vek`. Man kann dann n Elemente ohne Zurücklegen zufällig (uniform) aus M auswählen mit

```
sample(vek,n)
```

Dies liefert einen Vektor der Länge n. Wollen wir etwa eine Ziehung beim Lotto 6 aus 49 simulieren, so erreichen wir das mit

```
sample(1:49,6)
```

bzw. wollen wir die gezogenen Zahlen in aufsteigender Reihenfolge, wie sie üblicherweise bekanntgegeben werden, so

```
sort(sample(1:49,6))
```

Wollen wir stattdessen n mal mit Zurücklegen ziehen (nun kann auch $n > |M|$ gelten), muss zusätzlich `replace=TRUE` gesetzt werden. Wollen wir etwa den 10-fachen Wurf mit einem fairen Würfen simulieren, so erreichen wir das mit

```
sample(1:6,10,replace=TRUE)
```

6 Beispiele

Geburtstagsproblem. Wir kommen auf das Geburtstagsproblem aus Beispiele 2.4 zurück und wollen mit einer Simulation die Wahrscheinlichkeit dafür untersuchen, dass unter 23 Personen mindestens zwei am gleichen Tag des Jahres Geburtstag haben (unter der idealisierten Annahme, dass das Tupel der 23 Geburtstage gleichverteilt ist). Das folgende R-Programm simuliert das Szenario:

```
sim <- 10000                                    # Anzahl an Simulationen
m <- 23                                                # Anzahl Personen
erg <- 0                                     # Zähler für alle verschieden
for (i in 1:sim){
    X <- sample(1:365, m, replace = TRUE)     # ziehen der 23 Geburtstage
    if (length(unique(X))==m){         # prüfen auf verschiedene Einträge
```

[1] Das Erzeugen von Zufall kann dabei nicht exakt erfolgen, jedoch gibt es diverse sogenannte Pseudozufallszahlengeneratoren (typischerweise zahlentheoretischer Natur), die Approximationen von Realisierungen idealisiert modellierter Zufallsexperimente liefern.

```
        erg <- erg + 1                          # falls ja, erhöhe erg um 1
        }
    }
print(erg/sim)                                     # Ausgabe in der Konsole
```

Ausgegeben wird also der relative Anteil der 10.000 Simulationen, bei denen alle $m = 23$ Personen an verschiedenen Tagen des Jahres Geburtstag haben. Gemäß unserer Berechnungen in Beispiel 2 sollte dieser Wert nahe bei $1 - 0{,}507 = 49{,}3\%$ liegen. Man kann nun die Anzahl m der Personen sowie Anzahl `sim` der Wiederholungen variieren, um andere Wahrscheinlichkeiten mit verschiedenen Genauigkeiten zu simulieren. Wie genau die Simulationen für wachsende n an den tatsächlichen Wahrscheinlichkeiten liegen, wird in Abschnitt 9 in einem Gesetz großer Zahlen quantifiziert.

Ein Problem des Chevalier de Méré. Wir kommen nun auf das Beispiel aus den Vorbemerkungen zu diesem Skript zurück: Was ist beim Werfen von drei fairen Würfeln wahrscheinlicher: Gesamtaugenzahl 11 oder 12? Es handelt sich um eine Frage das Chevalier de Méré aus dem 17. Jahrhundert. Bevor wir das Problem mathematisch exakt mit den Mitteln aus Kapitel 2 lösen, verschaffen wir uns einen Überblick mittels einer Simulation mit R. Dazu simulieren wir den Wurf der drei Würfel oft und schauen, ob 11 oder 12 häufiger auftritt. Ein passendes R Programm könnte so aussehen:

```
sim <- 10000                                            # Stichprobengröße
count11 <- 0                                 # Zähler für Gesamtsumme 11
count12 <- 0                                 # Zähler für Gesamtsumme 12
for (i in 1:sim) {
    rolls <- sample(1:6, size = 3, replace = TRUE)        # 3 mal Werfen
    if (sum(rolls) == 11) count11 <- count11 + 1
    if (sum(rolls) == 12) count12 <- count12 + 1
    }
count11                                # gibt Anzahl für Gesamtsumme 11 aus
count12                                # gibt Anzahl für Gesamtsumme 12 aus
```

Dieses Programm gibt die Anzahlen der Würfe mit Gesamtaugenzahl 11 bzw. 12 aus. Man kann somit sehen, was in der Simulation häufiger auftrat. Indem man n in der ersten Zeile erhöht, bekommt man ein klarere Abgrenzung der beiden Fälle.

Um nun die Frage das Chevalier de Méré mathematisch exakt zu lösen, betrachte man das Laplacemodell mit Grundraum $\Omega = \{1, \dots, 6\}^3$, also mit $\mathbb{P}(\{\omega\}) = \frac{1}{216}$ für alle $\omega \in \Omega$. Die relevanten Ereignisse sind

$$\begin{aligned} A_{11} &= \{(\omega_1, \omega_2, \omega_3) \in \Omega \,|\, \omega_1 + \omega_2 + \omega_3 = 11\}, \\ A_{12} &= \{(\omega_1, \omega_2, \omega_3) \in \Omega \,|\, \omega_1 + \omega_2 + \omega_3 = 12\}. \end{aligned}$$

Da ein Laplace-Modell vorliegt, gilt

$$\mathbb{P}(A_{11}) = \frac{|A_{11}|}{216}, \qquad \mathbb{P}(A_{12}) = \frac{|A_{12}|}{216}.$$

Es reicht also, die Kardinalitäten von A_{11} und A_{12} zu ermitteln. Die in den Vorbemerkungen dieses Skripts suggerierte gleiche Anzahl von Würfelkonstellationen

$$\left.\begin{array}{ll} \text{„11“:} & \text{641, 632, 551, 542, 533, 443} \\ \text{„12“:} & \text{651, 642, 633, 552, 543, 444} \end{array}\right\} \text{jeweils 6 Möglichkeiten}$$

ist irreführend, denn z.B. zu 641 gibt es 6 verschiedene $\omega \in \Omega$, die die Konstellation 641 liefern (nämlich die sechs Permutationen der Zahlen $6, 4, 1$), während es etwa für 551 nur 3 verschiedene $\omega \in \Omega$ gibt (denn die 1 kann an erster, zweiter oder dritter Position stehen) und zu 444 nur ein $\omega \in \Omega$, nämlich $\omega = (4, 4, 4)$. Damit erhalten wir

$$|A_{11}| = 6 + 6 + 3 + 6 + 3 + 3 = 27, \quad |A_{12}| = 6 + 6 + 3 + 3 + 6 + 1 = 25.$$

Es ergibt sich also, dass die Augensumme 11 wahrscheinlicher ist als 12 mit den Wahrscheinlichkeiten

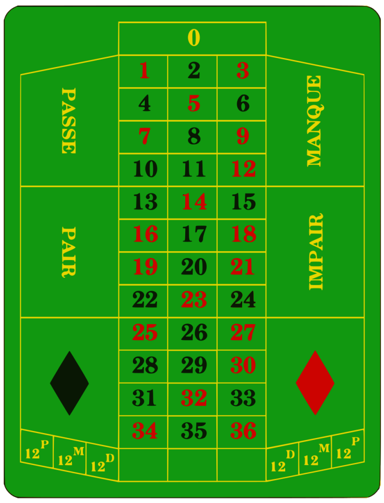
$$\mathbb{P}(A_{11}) = \frac{27}{216} = 12{,}5\%, \qquad \mathbb{P}(A_{12}) = \frac{25}{216} \approx 11{,}6\%.$$

4 Zufallsvariablen

Wir kommen zum Begriff der Zufallsvariable über ein Beispiel.

7 Zufallsvariablen

Quelle: https://openclipart.org, public domain



Quelle: Wikipedia, public domain

Beim Roulette wird eine der Zahlen $0, \dots, 36$ zufällig mit einer Kugel fair bestimmt. Setzt man auf rot oder schwarz bzw. auf gerade oder ungerade, wird im Gewinnfall (bei Null gewinnt stets die Spielbank) jeweils der doppelte Einsatz ausgezahlt, setzt man auf eine Zahl, das Sechsunddreißigfache des Einsatzes.

Eine Person setze je eine Einsatzmarke auf rot, gerade und auf die 1. Wir wollen den Gewinn bzw. Verlust X beschreiben, den die Person nach Ausspielung hat. Wir können die Zahl, die durch die Kugel bestimmt wird, als gleichwahrscheinlich annehmen. Wir modellieren das Zufallsexperiment also durch das Laplace-Modell $(\Omega, \mathbb{P})$, wobei $\Omega = \{0, \dots, 36\}$ und $\mathbb{P}(\{\omega\}) = \frac{1}{37}$ für alle $\omega \in \Omega$. Fällt die Kugel weder auf 1, noch auf gerade oder rot, so gewinnt die Person nichts und verliert den Einsatz von 3, d.h. hat Verlust von 3, was wir als $X = -3$ schreiben. Fällt die Kugel auf 1, so gewinnt die Person 36 für die 1, zusätzlich 2 für rot und verliert den Einsatz von 3 (Einsatzmarken), d.h. es ist $X = 36+2-3 = 35$. Für die anderen roten gerade Zahlen ist die Bilanz $X = 2+2-3 = 1$, für die schwarzen, geraden Zahlen und die roten ungeraden Zahlen (ungleich 1) ist $X = 2-3 = -1$. Für die restlichen Zahlen ist die Bilanz für die Person $X = -3$. Es hängt X also vom Ausgang $\omega \in \Omega$ ab, ist also eine Funktion von ω.

Beispiel 7.1. Zusammenfassend haben wir also $X : \Omega \to \mathbb{Z}$ (oder auch $X : \Omega \to \mathbb{R}$) mit

$$X(\omega) = \begin{cases} 35, & \text{falls } \omega = 1, \\ 1, & \text{falls } \omega \in \{12, 14, 16, 18, 30, 32, 34, 36\}, \\ -1, & \text{falls } \omega \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28\}, \\ -3, & \text{sonst.} \end{cases}$$

Der Gewinn bzw. Verlust der Person in Beispiel 7.1 ist also eine Funktion von einem endlichen Wahrscheinlichkeitsraum. Derartige Funktionen oder Abbildungen nennen wir Zufallsvariablen.

Definition 7.2. Sei $(\Omega, \mathbb{P})$ ein endlicher W-Raum und Ω' eine beliebige Menge. Jede Abbbildung

$$X : \Omega \to \Omega'$$

heißt Ω'-wertige *Zufallsvariable* (ZVe). Falls $\Omega' = \mathbb{R}$, so heißt X *reellwertige ZVe* (oder einfach nur ZVe), falls $\Omega' = \mathbb{R}^d$, so heißt X *Zufallsvektor.* Für $\omega \in \Omega$ heißt $X(\omega)$ *Realisierung* der Zufallsvariablen X.

Im Folgenden sei stets ein endlicher W-Raum $(\Omega, \mathbb{P})$ gegeben.

Satz 7.3. Sei $X : \Omega \to \Omega'$ eine Ω'-wertige ZVe. Dann ist durch

$$\mathbb{P}_X : \mathcal{P}(\Omega') \to [0, 1], \quad A' \mapsto \mathbb{P}\left(X^{-1}(A')\right)$$

ein Wahrscheinlichkeitsmaß auf Ω' definiert. $\mathbb{P}_X$ heißt *Verteilung der ZVe X.*

Beweis.
Wir müssen die definierenden Eigenschaften aus Definition 2.1 prüfen: $\mathbb{P}_X$ bildet offenbar in $[0, 1]$ ab. Wir haben

$$\mathbb{P}_X(\Omega') = \mathbb{P}\left(X^{-1}(\Omega')\right) = \mathbb{P}(\Omega) = 1.$$

Seien $A_1, \ldots, A_n \subset \Omega'$ paarweise disjunkte Mengen. Dann folgt

$$\mathbb{P}_X\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{i=1}^n A_i\right)\right) = \mathbb{P}\left(\bigcup_{i=1}^n X^{-1}(A_i)\right) = \sum_{i=1}^n \mathbb{P}\left(X^{-1}(A_i)\right) = \sum_{i=1}^n \mathbb{P}_X(A_i).$$

Es folgt die endliche Additivität. ∎

Notationen 1. Die folgenden Kurzschreibweisen sind gebräuchlich:
$X^{-1}(A) = \{\omega \in \Omega \,|\, X(\omega) \in A\} =: \{X \in A\}$.
Für reelle ZVe X: Sei $A = (-\infty, x]$, dann $X^{-1}(A) =: \{X \leq x\}$.

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

Beispiel 7.4. Für die Zufallsvariable X aus Beispiel 7.1 ist die Verteilung $\mathbb{P}_X$ das Wahrscheinlichkeitsmaß auf $\mathbb{Z}$ mit

$$\mathbb{P}_X(\{35\}) = \frac{1}{37}, \quad \mathbb{P}_X(\{1\}) = \frac{8}{37}, \quad \mathbb{P}_X(\{-1\}) = \frac{19}{37}, \quad \mathbb{P}_X(\{-3\}) = \frac{9}{37}$$

und $\mathbb{P}_X(\{k\}) = 0$ für alle $k \in \mathbb{Z} \setminus \{-3, -1, 1, 35\}$. Die Verteilung gibt also die Wahrscheinlichkeiten dafür an, dass die einzelnen Werte auftreten.

Beispiel 7.5 (n-maliger Münzwurf)**.** Wir betrachten den n-maligen Münzwurf mit Erfolgswahrscheinlichkeit $p \in [0, 1]$. Wir modellieren diesen mit

$$\Omega = \{0, 1\}^n, \quad \mathbb{P}\left(\{(\omega_1, \dots, \omega_n)\}\right) = p^k(1-p)^{n-k} \text{ mit } k = \sum_{i=1}^n \omega_i.$$

Die Zufallsvariable

$$X : \Omega \to \mathbb{N}_0, \quad \omega \mapsto \sum_{i=1}^n \omega_i$$

beschreibt die Anzahl der Erfolge, vgl. Beispiel 4.9. Es ist $X(\Omega) = \{0, \dots, n\}$. Wie lautet die Verteilung von X? Sei wieder $E_k := \{(\omega_1, \dots, \omega_n) \in \Omega : \sum_{i=1}^n \omega_i = k\}$. Dann gilt

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

Die Anzahl der Erfolge bei n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ ist binomial $b_{n,p}$ verteilt (d.h. die Verteilung $\mathbb{P}_X$ von X ist die Binomialverteilung mit Parametern n und p, vgl. Korollar 4.11.)

Beispiel 7.6 (k-maliges Ziehen ohne Zurücklegen aus einer Urne mit s schwarzen und w weißen Kugeln)**.** Betrachte eine ZVe $X =$“Anzahl gezogener schwarzer Kugeln“. Seien $A = \{1, \dots, n\}$, $A_s = \{1, \dots, s\}$, $A_w = A \setminus A_s$, $\Omega = \left\{(\omega_1, \dots, \omega_k) \in A^k \,|\, \omega_1 < \dots < \omega_k\right\}$ (vgl. Beispiel 3.12). Damit haben wir

$$X : \Omega \to \{0, \dots, k\}, \quad (\omega_1, \dots, \omega_k) \mapsto \sum_{i=1}^k \mathbf{1}_{A_s}(\omega_i).$$

Wie in Abschnitt 2 gezeigt, liefert das Laplace-Modell auf Ω:

$$\mathbb{P}_X(\{\ell\}) = \mathbb{P}\left(X^{-1}(\ell)\right) = \mathbb{P}(B_\ell) = \frac{\binom{s}{\ell}\binom{w}{k-\ell}}{\binom{s+w}{k}}.$$

Die Anzahl X der schwarzen Kugeln ist folglich hypergeometrisch verteilt (zu entsprechenden Parametern).

Ebenso: Bei einer Folge von unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ ist die Wartezeit bis zum ersten Erfolg eine Zufallsvariable, die geometrisch g_p verteilt ist.

Bemerkung 5. Für den Umgang mit Zufallsvariablen sowie deren Verteilungen sind die gängigen Verteilungen in R verfügbar. Die Bezeichnungen sind wie folgt:

Verteilung	Kürzel	Parameter
Binomialverteilung	`binom`	`size, prob`
geometrische Verteilung	`geom`	`prob`
hypergeometrische Verteilung	`hyper`	`m,n,k`

Will man nun Realisierungen von Zufallsvariablen mit einer vorgegebenen Verteilung generieren, so stellt man dem Kürzel ein `r` voran (für *random*). Z.B. erzeugt die Anweisung

```
rgeom(n=10, prob=0.2)
```

die Realisierungen von 10 geometrisch zum Parameter $0{,}2$ verteilten Zufallsvariablen (die im Sinne der folgenden Definitionen stochastisch unabhängig sind). Will man Punktmassen ausgeben, so stellt man ein `d` voran (für *distribution*). Z.B. ergibt

```
dgeom(5, prob=0.2)
```

die Wahrscheinlichkeit $\mathbb{P}(X = 5)$ für eine zum Parameter $p = 0{,}2$ geometrisch verteilte Zufallsvariable X. Stellt man ein `p` voran, z.B.

```
pgeom(7, prob=0.2)
```

so erhält man $\mathbb{P}(X \leq 7)$. Man nennt

$$x \mapsto \mathbb{P}(X \leq x) \tag{14}$$

die *Verteilungsfunktion* von X.
Damit lassen sich z.B. Diagramme wie in Abbildung 2 wie folgt generieren:

```
n <- 50                                    # Parameter n wird gesetzt
p <- 0.7                                   # Parameter p wird gesetzt
x <- 0:n                           # Vektor x = (0,...,n) wird gesetzt
weights <- dbinom(x, size = n, prob = p)
barplot(                       # barplot erstellt das Säulendiagramm
    heights <- weights,
    names.arg = x,
    xlab = "k",
    ylab = "P(X = k)",
    main = paste("Binomialverteilung B(n, p) mit n =", n, ", p =", p)
    )
```

Ist man wegen der Parameter der Verteilungen in R unsicher, so erhält man z.B. mittels `?dbinom` die Dokumentation der entsprechenden Funktion.

> **Bemerkung zur Didaktik/Schule:** Zufallsvariable (dort Zufallsgröße genannt) und Wahrscheinlichkeitsverteilung sind zwar erst Gegenstand des Kerncurriculums zur Oberstufe in der Qualifikationsphase Q3.3, sind aber eng verknüpft mit den Themen der Sekundarstufe 1, weshalb ein Verständnis für den Unterricht dort wesentlich ist. Die hier betrachteten grundlegenden Verteilungen im Kontext von unabhängigen Bernoulli-Experimenten (geometrische Verteilung, hypergeometrische Verteilung, Bi-

nomialverteilung) sind ebenfalls Bestandteil. Im Kontext von unabhängigen Bernoulli-Experimenten ist dort von „Bernoulli-Ketten" die Rede, was insofern eine unglückliche Bezeichnung ist, da es zu Missverständnissen mit Bernoulli-Experimenten kommen kann, die nicht unabhängig sind. Solche abhängigen Bernoulli-Ketten treten in zahlreichen Anwendungen auf, werden hier aber nicht besprochen.

Definition 7.7. Seien $(\Omega, \mathbb{P})$ ein endlicher W-Raum und $X_i : \Omega \to \Omega_i$ ZVe für $i = 1, \dots, n$. Die Familie $\{X_1, \dots, X_n\}$ von ZVen heißt *(stochastisch) unabhängig*, falls für jede Wahl $A_i \subset \Omega_i$ die Familie von Ereignissen $\{\{X_i \in A_i\} \mid i = 1, \dots, n\}$ unabhängig ist.

Satz 7.8. Sei $(\Omega, \mathbb{P})$ ein endlicher W-Raum und $X_1, \dots, X_n$ ZVe auf Ω, $X_i : \Omega \to \Omega_i$. Dann sind äquivalent:

a) $X_1, \dots, X_n$ sind unabhängig.

b) Für alle $(x_1, \dots, x_n) \in \Omega_1 \times \cdots \times \Omega_n$ gilt

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i = x_i\}\right) = \mathbb{P}\left(X_1 = x_1, \dots, X_n = x_n\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i\right).$$

c) Für beliebige $A_i \subset \Omega_i$ gilt

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i \in A_i\right).$$

Beweis.
a) $\Rightarrow$ c): folgt aus den Definitionen 7.7 und 4.6 b).
c) $\Rightarrow$ b): Wähle $A_i = \{x_i\}$ für $i = 1, \dots, n$.
b) $\Rightarrow$ a): Seien $A_i \subset \Omega_i$ Mengen. Zu zeigen ist, dass $\{\{X_i \in A_i\} : i = 1, \dots, n\}$ Familie unabhängiger Mengen ist, d.h. für jede Teilfamilie die Produktformel gilt. Wir betrachten ohne Einschränkung die Teilfamilie $\{1, \dots, s\} \subset \{1, \dots, n\}$. Dann gilt

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{s} \{X_i \in A_i\}\right) &= \mathbb{P}\left(\bigcap_{i=1}^{s} \bigcup_{x_i \in A_i} \{X_i = x_i\}\right) = \mathbb{P}\left(\bigcup_{x_1 \in A_1} \cdots \bigcup_{x_s \in A_s} \left(\bigcap_{i=1}^{s} \{X_i = x_i\}\right)\right) \\
&= \sum_{x_1 \in A_1} \cdots \sum_{x_s \in A_s} \mathbb{P}\left(\bigcap_{i=1}^{s} \{X_i = x_i\}\right) = \sum_{x_1 \in A_1} \cdots \sum_{x_s \in A_s} \prod_{i=1}^{s} \mathbb{P}(X_i = x_i) \\
&= \left(\sum_{x_1 \in A_1} \mathbb{P}(X_1 = x_1)\right) \cdots \left(\sum_{x_s \in A_s} \mathbb{P}(X_s = x_s)\right) \\
&= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_s \in A_s).
\end{aligned}$$

Also gilt die Produktformel. ∎

Satz 7.9. Seien $X_1, \dots, X_n$ unabhängige ZVe, $X_i : \Omega \to \Omega_i$. Seien $f_i : \Omega_i \to \Gamma_i$ Funktionen für $i = 1, \dots, n$. Dann sind die ZVe $Y_1, \dots, Y_n$ mit $Y_i = f_i \circ X_i$ unabhängig.

Beweis.
Seien $A_i \subset \Gamma_i$ für $i = 1, \dots, n$ beliebig. Dann ist

$$\{Y_i \in A_i\} = \{f_i \circ X_i \in A_i\} = \left\{X_i \in f_i^{-1}(A_i)\right\}.$$

Da $X_1, \dots, X_n$ unabhängig sind, ist $\{\{X_i \in f_i^{-1}(A_i)\} : i = 1, \dots, n\}$ eine unabhängige Familie von Ereignissen. Folglich ist $\{\{Y_i \in A_i\} : i = 1, \dots, n\}$ eine unabhängige Familie. Es folgt die Unabhängigkeit von $Y_1, \dots, Y_n$. ∎

8 Erwartungswert und Varianz

In diesem Abschnitt sei stets $(\Omega, \mathbb{P})$ mit $\Omega = \{\omega_1, \dots, \omega_n\}$ ein endlicher Wahrscheinlichkeitsraum, auf dem alle auftretenden Zufallsvariablen definiert sind.

Definition 8.1. Sei X eine reellwertige ZVe. Die *Erwartung* (Erwartungswert, EW) von X ist gegeben durch

$$\mathbb{E}[X] = \sum_{i=1}^{n} X(\omega_i)\mathbb{P}(\{\omega_i\}).$$

Beispiel 8.2. Wir werfen zwei Würfel und wollen die Erwartung für die Gesamtaugenzahl berechnen: Zur Modellierung setzen wir

$$\Omega = \{1, \dots, 6\}^2 = \{(\omega_1, \omega_2) | \omega_i \in \{1, \dots, 6\}, i = 1, 2\},$$

und $\mathbb{P}(\{(\omega_1, \omega_2)\}) = 1/36$ für alle $(\omega_1, \omega_2) \in \Omega$. Die Zufallsvariable, die die Gesamtaugenzahl beschreibt, ist $X : \Omega \to \mathbb{N}$ gegeben durch $X(\omega_1, \omega_2) = \omega_1 + \omega_2$. Um $\mathbb{E}[X]$ gemäß Definition zu berechnen, müssen wir die 36 Elemente von Ω durchzählen. Die Reihenfolge ist beliebig. Wir können also wie folgt rechnen:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{(\omega_1,\omega_2)\in\Omega} (\omega_1 + \omega_2)\frac{1}{36} = \frac{1}{36}\sum_{i=1}^{6}\sum_{j=1}^{6}(i + j) \\ &= \frac{1}{36}\sum_{i=1}^{6}(6i + 21) = \frac{1}{36}(6 \cdot 21 + 6 \cdot 21) = \frac{12 \cdot 21}{36} = 7. \end{aligned}$$

Lemma 8.3. Sei $\{x_1, \dots, x_m\}$ der Wertebereich von X. Dann gilt

$$\mathbb{E}[X] = \sum_{i=1}^{m} x_i\, \mathbb{P}(X = x_i).$$

Beweis.

Es gilt

$$\mathbb{E}[X] = \sum_{i=1}^{n} X(\omega)\,\mathbb{P}(\{\omega\}) = \sum_{i=1}^{m} \sum_{\{\omega\,|\,X(\omega)=x_i\}} X(\omega)\,\mathbb{P}(\{\omega\}) = \sum_{i=1}^{m} x_i\,\mathbb{P}(X = x_i).$$

Dies liefert die Behauptung. ∎

Bemerkung 6. Der Erwartungswert von X hängt also nur von der Verteilung $\mathbb{P}_X$ der Zufallsvariable X ab. Derartige Größen heißen *Verteilungsgrößen*.

Beispiel 8.4. Wir wiederholen viermal ein Bernoulliexperiment mit Erfolgswahrscheinlichkeit $p = \frac{1}{8}$. Die Zufallsvariable Y bezeichne die Anzahl der Erfolge. Wollen wir $\mathbb{E}[Y]$, also die erwartete Anzahl der Erfolge berechnen, können wir auf die Definition zurückgreifen. Da der kanonische Grundraum 16 Elemente hat, führt dies auf eine Summe mit 16 Summanden. Wir kennen allerdings die Verteilung von Y, denn Y ist binomial $b_{4,1/8}$ verteilt. Also können wir $\mathbb{E}[Y]$ auch mit Lemma 8.3 berechnen. Der Wertebereich von Y ist hier $\{0, 1, 2, 3, 4\}$. Es ist damit

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{i=0}^{4} i \cdot b_{4,1/8}(\{i\}) = \sum_{i=0}^{4} i \cdot \binom{4}{i} \left(\frac{1}{8}\right)^{i} \left(\frac{7}{8}\right)^{(4-i)} \\
&= \frac{0 + 1 \cdot \binom{4}{1} 7^3 + 2 \cdot \binom{4}{2} 7^2 + 3 \cdot \binom{4}{3} 7^1 + 4 \cdot 1}{8^4} \\
&= \frac{4 \cdot 343 + 12 \cdot 49 + 12 \cdot 7 + 4}{8^4} = \frac{1}{2}.
\end{aligned}$$

Einen dritten, systematischeren Weg, $\mathbb{E}[Y]$ zu berechnen, werden wir in Beispiel 8.7 sehen.

Beispiel 8.5. Für die Zufallsvariable X in Beispiel 7.1 erhalten wir über ihre Verteilung (siehe Beispiel 7.4)

$$\mathbb{E}[X] = 35 \cdot \frac{1}{37} + 1 \cdot \frac{8}{37} - 1 \cdot \frac{19}{37} - 3 \cdot \frac{9}{37} = -\frac{3}{37}.$$

Satz 8.6. Seien X, Y reellwertige ZVe. Dann gelten:

(i) Für $\lambda \in \mathbb{R}$ gilt $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$.

(ii) Es gilt $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

(iii) Sind X, Y unabhängig, so gilt $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(iv) Falls $X \geq 0$, so gilt $\mathbb{E}[X] \geq 0$.

(v) Falls $X \geq Y$ (punktweise), so gilt $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

(vi) Es ist $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$ für alle $A \subset \Omega$.

Beweis.
Der Beweis dieser Eigenschaften lässt sich direkt durch Rückgriff auf die Definition und Nachrechnen führen und wird deshalb hier nicht ausgeführt. Wichtig ist, die Rechenregeln für den Erwartungswert zu verstehen und anwenden zu können. ∎

Beispiel 8.7. Sei X binomial $b_{n,p}$ verteilt mit $n \in \mathbb{N}$ und $p \in [0,1]$. Dann gilt $\mathbb{E}[X] = np$.

Beweis.
Sei X binomial $b_{n,p}$ verteilt. Es ist X verteilt wie die Anzahl der Erfolge bei n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p. Also gilt $X = \sum_{i=1}^{n} \mathbf{1}_{A_i}$, wobei $A_i = \{\text{Erfolg im } i\text{-ten Teilexperiment}\}$, also $\mathbb{P}(A_i) = p$ für $i = 1, \dots, n$. Es folgt

$$\mathbb{E}[X] \overset{8.6(ii)}{=} \sum_{i=1}^{n} \mathbb{E}[\mathbf{1}_{A_i}] \overset{8.6(vi)}{=} \sum_{i=1}^{n} \mathbb{P}(A_i) = np.$$

∎

Bemerkung zur Didaktik/Schule: Häufig wird für den Erwartungswert einer binomialverteilten Zufallsvariablen in Beispiel 8.7 folgender (didaktisch verfehlter) rechnerischer Beweis gegeben: Es gilt nach Lemma 8.3

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} np \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
&= np(p + (1-p))^{n-1} \qquad (15) \\
&= np,
\end{aligned}$$

wobei in (15) der Binomische Lehrsatz aus Korollar 3.10 angewandt wurde. Manchmal wird diese Rechnung in Schulbüchern nur für den Fall $n = 3$ präsentiert, da für allgemeines n der Binomische Lehrsatz nicht verfügbar ist. In jedem Falle fördert solch ein rechnerischer Beweis kaum stochastisches Denken. Dagegen sind die Struktur binomialverteilter Zufallsvariablen sowie grundlegende Eigenschaften des Erwartungswerts wie oben gezeigt imstande, die Formel $\mathbb{E}[X] = np$ erklärend zu beweisen.

Wir geben noch den Erwartungswert hypergeometrisch verteilter Zufallsvariablen (ohne Beweis) an:

Beispiel 8.8. Sei X hypergeometrisch verteilt zu Parametern n, s, w, k wie in Beispiel 7.6. Dann gilt

$$\mathbb{E}[X] = k\frac{s}{n}.$$

Der Erwartungswert kann als Maßzahl für den „Schwerpunkt“ bzw. den „mittleren Wert“ einer Verteilung aufgefasst werden. Wir betrachten zudem Maßzahlen für die Streuung um den Erwartungswert.

Definition 8.9. Seien X, Y reelle ZVe. Dann heißen:

- $\mathrm{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$ die *Varianz* von X,
- $\sigma_X := \sqrt{\mathrm{Var}(X)}$ die *Standardabweichung* von X,
- $\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$ die *Kovarianz* von X und Y,

X und Y heißen *unkorreliert*, falls $\mathrm{Cov}(X, Y) = 0$ gilt.

Satz 8.10. Seien $X, Y, X_1, \dots, X_n$ ZVe. Dann gilt für alle $a, b, c, d \in \mathbb{R}$:

(i) $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$,

(ii) $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$,

(iii) $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$,

(iv) $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$,

(v) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,

(vi) $\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$,

(vii) X, Y unabhängig $\Rightarrow$ X, Y unkorreliert.

Beweis.
(i)-(v) folgen direkt aus der Definition. (vi) Wegen (ii) können wir o.E. $\mathbb{E}[X_i] = 0$ für $i = 1, \dots, n$ annehmen. Dann folgt

$$\mathrm{Var}(X_1 + \cdots + X_n) \overset{(i)}{=} \mathbb{E}\left[(X_1 + \cdots + X_n)^2\right] = \sum_{i,j \le n} \mathbb{E}[X_i X_j] = \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j].$$

Wegen $\mathbb{E}[X_i] = 0$ ist dies $\sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$.
(vii) Seien X, Y unabhängig. Nach Satz 7.9 sind dann $(X - \mathbb{E}[X]), (Y - \mathbb{E}[Y])$ unabhängig. Also

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[X - \mathbb{E}[X]]\mathbb{E}[Y - \mathbb{E}[Y]] = 0 \cdot 0 = 0.$$

Somit sind X, Y unkorreliert. ∎

Satz 8.11 (Bienaymé)**.** Seien $X_1, \dots, X_n$ unabhängig. Dann gilt

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Beweis.
O.E. gelte $\mathbb{E}[X_i] = 0$ für $i = 1, \dots, n$. Nach Satz 8.10 (vi) gilt $\mathrm{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathbb{E}[X_i X_j]$. Nach Satz 8.10 (vii) sind X_i, X_j unkorreliert für $i \neq j$. Damit ist $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j] = 0$. Es folgt die Behauptung. ▮

Beispiel 8.12. Sei X binomial $b_{n,p}$ verteilt mit $n \in \mathbb{N}$ und $p \in [0, 1]$. Dann gilt

$$\mathrm{Var}(X) = np(1-p).$$

Beweis.
Wie im Beweis von Beispiel 8.7 haben wir $X = \sum_{i=1}^n \mathbf{1}_{A_i}$, wobei $\mathbb{P}(A_i) = p$ und $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ unabhängige ZVe sind. Es gilt $\mathrm{Var}(\mathbf{1}_{A_i}) = \mathbb{E}[\mathbf{1}_{A_i}^2] - \mathbb{E}[\mathbf{1}_{A_i}]^2 = p - p^2 = p(1-p)$. Nach Satz 8.11 liefert Unabhängigkeit der Indikatoren, dass $\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(\mathbf{1}_{A_i}) = np(1-p)$. ▮

Bemerkung zur Didaktik: Lage- und Streumaße werden in der „Didaktik der Stochastik“ im Kontext von Daten besprochen. Solche Daten stellen wir uns als Realisierungen von Zufallsviablen vor, also etwa Daten $x_1, \dots, x_n$ als Realisierungen $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ für (in der Regel) unabhängige Zufallsvariable $X_1, \dots, X_n$ und ein $\omega \in \Omega$. Die in der Didaktik besprochenen Mittelwerte (als Lagemaße) und Standardabweichungen (als Streumaße) kann man als Schätzer unserer Erwartungswerte und Varianzen verstehen, man vergleiche dazu Abschnitt 12. In der Didaktik werden weitere Maße für die „Mitte“ von Daten besprochen wie Median oder Modalwert. Als weiteres Lagemaß tritt der Begriff des *Quantils* auf. Für unsere Verteilungen von Zufallsvariablen aus Satz 7.3 ist $x_q \in \mathbb{R}$ ein q-Quantil (für $q \in (0, 1)$), falls

$$\mathbb{P}(X \leq x_q) \geq q \quad \text{und} \quad \mathbb{P}(X \geq x_q) \geq 1 - q. \tag{16}$$

Derartige Quantile treten später in (22) auf (für die Standardnormalverteilung).
Der Begriff der Kovarianz tritt in diesem Abschnitt nur als technische Größe bei Satz 8.10 (vi) sowie dem daraus folgenden Satz 8.11 von Bienaymé auf. Tatsächlich spielt die Kovarianz und die sich daraus ergebende Korrelation eine wichtige Rolle in der Datenanalyse, um Zusammenhänge zwischen Merkmalen zu beschreiben und zu untersuchen. Dies wird in der Vorlesung „Didaktik der Stochastik“ besprochen.

5 Summen unabhängiger Zufallsvariablen

In einem fairen Spiel zwischen zwei Personen werde vielfach unabhängig eine Münze geworfen. Bei Kopf erhält jeweils Person A einen vorgegebenen Einsatz $E > 0$ von Person B, bei Zahl erhält Person B den Einsatz E von Person A. Was kann man über den Gewinn von Person A asymptotisch sagen, wenn das Spiel sehr lange dauert?
Wir modellieren dies wie folgt: Es seien $X_1, \ldots, X_n$ unabhängige ZVen mit $\mathbb{P}(X_i = 1) = \frac{1}{2} = \mathbb{P}(X_i = -1)$. Das Ereignis $\{X_i = 1\}$ bedeute, dass Person A im i-ten Spiel gewinnt. Dann ist der Gewinn von Person A gegeben durch

$$S_n = E \cdot \sum_{i=1}^{n} X_i,$$

falls n Spiele gespielt werden. Es ist S_n also eine Summe unabhängiger ZVe. Summen unabhängiger Zufallsvariablen treten bei der stochastischen Modellierung in zahlreichen Situationen auf. Deshalb untersuchen wir das asymptotische Verhalten solcher Summen. Es zeigt sich dabei, dass der Zufall nichts völlig Willkürliches ist, sondern Gesetzen folgt, die wir im Rahmen mathematischer Modellierungen beweisen können.

9 Ein Gesetz großer Zahlen

In diesem Abschnitt untersuchen wir das Verhalten von S_n/n, wobei S_n eine Summe von n unabhängigen Zufallsvariablen ist. Wir erhalten ein sogenanntes *Gesetz der großen Zahlen*. Wir formulieren und interpretieren dies zunächst. Im Anschluss werden Methoden entwicklt (Chebyshev Ungleichung), um es zu beweisen.

Satz 9.1 (Schwaches Gesetz großer Zahlen)**.** Seien $X_1, \ldots, X_n$ unabhängige ZVe mit $\mathbb{E}[X_i] = \mu$ und $\mathrm{Var}(X_i) = \sigma^2$ für alle $i = 1, \ldots, n$. Bezeichne $S_n := \sum_{i=1}^n X_i$. Dann gilt für alle $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

Bemerkung 7. Sei $X_1, \ldots, X_n$ eine Folge von Bernoulli-Experimenten, die unabhängig mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ ausgeführt werden. Bezeichne

$$H_n(\omega) := \frac{1}{n} \sum_{i=1}^{n} X_i(\omega)$$

für $\omega \in \Omega$ die relative Anzahl von Erfolgen, siehe Abbildung 3. Es gilt $\mathrm{Var}(X_i) = p(1-p) = \sigma^2 \leq \frac{1}{4}$ für alle $p \in [0, 1]$. Das schwache Gesetz großer Zahlen liefert

$$\mathbb{P}(|H_n - p| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n} \leq \frac{1}{4\varepsilon^2 n}. \tag{17}$$

Für große n ist die Wahrscheinlichkeit, dass sich die relative Häufigkeit von der Erfolgswahrscheinlichkeit um mehr als ε unterscheidet, also klein. Dies begründet u.a. den frequentistischen Wahrscheinlichkeitsbegriff.

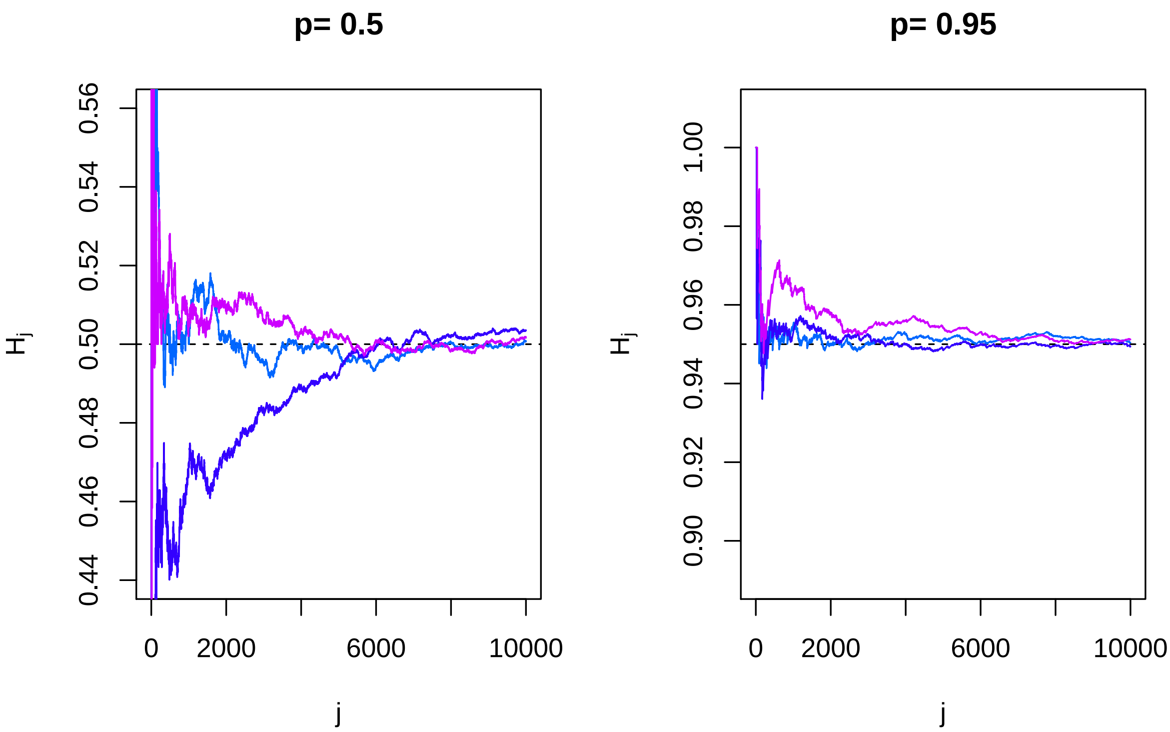


Abbildung 3: Gezeigt sind jeweils Simulationen $(H_j(\omega_i))_{j=1,\dots,10000}$ für $i = 1, 2, 3$, vgl. Bemerkung 7. Im linken Bild ist $p = \frac{1}{2}$, rechts $p = 0.95$. Zur Simulation wurde folgender R-Code verwendet:

```
n <- 10000
p <- 0.5                        # Hier entsprechend 0.95 ersetzen.
m <- 3
plot(0, 0, type="n", xlab="j", ylab=expression(paste(H[j])),
xlim=c(0,n), ylim=c(0,1))
abline(h=p, lty=2)
for (i in 1:m){
    X <- sample(c(1,0), prob=c(p,1-p), replace=TRUE, n)
    S <- cumsum(X)/(1:n)
    lines(S, col=rainbow(m, start=0.6, end=0.8)[i])
          }
```

Für den Beweis des schwachen Gesetzes großer Zahlen werden zwei Ungleichungen benötigt:

Satz 9.2 (Markovsche Ungleichungen)**.** Sei X eine nichtnegative Zufallsvariable und $\varepsilon > 0$, dann gelten

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}, \qquad \mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X^2]}{\varepsilon^2}. \tag{18}$$

Beweis.
Es sei

$$Y(\omega) := \begin{cases} \varepsilon, & \text{falls } X(\omega) \geq \varepsilon, \\ 0, & \text{falls } X(\omega) < \varepsilon. \end{cases}$$

Dann gilt $Y \leq X$ punktweise. Die Monotonie des Erwartungswerts aus Lemma 8.6 (v) liefert $\mathbb{E}[X] \geq \mathbb{E}[Y] = \varepsilon\mathbb{P}(X \geq \varepsilon)$. Dies ergibt die linke Ungleichung in (18). Die rechte Ungleichung in (18) folgt analog, indem man

$$\widehat{Y}(\omega) := \begin{cases} \varepsilon^2, & \text{falls } X(\omega) \geq \varepsilon, \\ 0, & \text{falls } X(\omega) < \varepsilon, \end{cases}$$

betrachtet und $\widehat{Y} \leq X^2$ bemerkt. ∎

Beispiel 9.3. Sei X binomial $b_{n,1/10}$ verteilt mit $n \in \mathbb{N}$, also X etwa die Anzahl der Erfolge bei einem entsprechenden n-fachen Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p = \frac{1}{10}$. Wir interessieren uns für die Wahrscheinlichkeit $\mathbb{P}(X \geq n/2)$, wobei wir annehmen, dass n gerade ist. Diese können wir explizit bestimmen als

$$\sum_{j=n/2}^{n} \binom{n}{j} \left(\frac{1}{10}\right)^j \left(\frac{9}{10}\right)^{n-j}.$$

Für große n ist diese Summe schwer auszuwerten, für unbestimmtes n ist die Größenordnung dieser Summe nicht unmittelbar klar. Hier helfen etwa die beiden Ungleichungen aus Satz 9.2: Es ist $\mathbb{E}[X] = n/10$ und $\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2 = n(1/10)(9/10) + n^2/100 \leq n^2/50$ für alle $n \geq 9$. Also erhalten wir die Abschätzungen

$$\mathbb{P}(X \geq n/2) \leq \frac{n/10}{n/2} = \frac{1}{5}, \quad \mathbb{P}(X \geq n/2) \leq \frac{n^2/50}{n^2/4} = \frac{2}{25}, \tag{19}$$

wobei die zweite Abschätzung, die offenbar schärfer ist, für alle $n \geq 9$ gilt. Diese Abschätzungen kann man für große n mit der Chebyshev Ungleichung aus dem folgenden Korollar 9.4 wesentlich verbessern, vgl. Beispiel 9.5.

Korollar 9.4 (Chebyshevsche Ungleichung)**.** Sei X eine reellwertige Zufallsvariable. Dann gilt für alle $\varepsilon > 0$

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Beweis.
Sei $Z := |X - \mathbb{E}[X]|$. Dann ist Z eine nichtnegative Zufallsvariable. Wir wenden die Markovsche Ungleichung in Form der rechten Ungleichung in (18) an und erhalten

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) = \mathbb{P}(Z \geq \varepsilon) \leq \frac{\mathbb{E}[Z^2]}{\varepsilon^2} = \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Dies ist die Behauptung. ∎

Beispiel 9.5. Wir kommen auf die Zufallsvariable X aus Beispiel 9.3 zurück sowie auf die Wahrscheinlichkeit $\mathbb{P}(X \geq n/2)$ mit n gerade. Wegen $\mathbb{E}[X] = n/10$ gilt

$$\mathbb{P}(X \geq n/2) = \mathbb{P}(X - E[X] \geq (4/10)n) \leq \mathbb{P}(|X - E[X]| \geq (4/10)n).$$

Wir können also die Chebyshev Ungleichung anwenden und erhalten

$$\mathbb{P}(X \geq n/2) \leq \mathbb{P}(|X - E[X]| \geq (4/10)n) \leq \frac{(9/100)n}{(16/100)n^2} = \frac{9}{16n}. \tag{20}$$

Vergleichen wir die Abschätzung (20) mit (19), so fällt auf, dass (20) mit wachsendem n immer schärfer wird, für große n also die Ungleichungen aus (19) verbessert. Für den folgenden Beweis des schwachen Gesetzes großer Zahlen werden wir deshalb auch die Chebyshev Ungleichung anwenden.

Beweis von Satz 9.1.
Bezeichne $X := \frac{1}{n}S_n$. Dann gilt $\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$ und

$$\mathrm{Var}(X) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) \leq \frac{M}{n},$$

wobei Satz 8.10 und der Satz von Bienaymé 8.11 verwendet werden. Die Chebyshevsche Ungleichung liefert

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - \mu\right| \geq \varepsilon\right) = \mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2} \leq \frac{M}{\varepsilon^2 n}.$$

Dies ist die Behauptung. ∎

10 Normalapproximation

Wir betrachten nochmals den Kontostand $S_n = \sum_{i=1}^n X_i$ von Person A (mit Einsatz $E = 1$) in Abschnitt 11, wobei $X_1, \dots, X_n$ unabhängig sind und wir hier $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = -1) = 1 - p =: q$ annehmen mit $p \in (0, 1)$. Mit $B_i := \frac{1}{2}(X_i + 1)$ sind $B_1, \dots, B_n$ die Ausgänge von n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p. Wir erhalten

$$S_n = -n + 2\sum_{i=1}^n B_i.$$

Wir wissen, dass $\sum_{i\leq n} B_i$ binomial $b_{n,p}$-verteilt ist. Damit gilt für $a < b$, dass

$$\mathbb{P}\left(S_n \in [a,b]\right) = b_{n,p}\left(\left[\frac{a+n}{2}, \frac{b+n}{2}\right]\right).$$

Die Verteilung $\mathbb{P}_{S_n}$ von S_n ist also vollständig beschrieben. Das schwache Gesetz großer Zahlen (auf $B_1, \ldots, B_n$ angewandt) liefert dann für jedes $\varepsilon > 0$, dass

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} B_i - p\right| \geq \varepsilon\right) \leq \frac{1}{4\varepsilon^2 n}$$

oder äquivalent, dass gilt

$$\mathbb{P}\left(\sum_{i=1}^{n} B_i \in (n(p-\varepsilon), n(p+\varepsilon))\right) \geq 1 - \frac{1}{4\varepsilon^2 n}.$$

Die Summe $\sum_{0\leq i\leq n} B_i$ liegt also „mit hoher Wahrscheinlichkeit“ (genauer: mit gegen 1 konvergierender Wahrscheinlichkeit) im Intervall $(n(p-\varepsilon), n(p+\varepsilon))$. Für die Binomialverteilung bedeutet dies für alle $\varepsilon > 0$:

$$b_{n,p}\left((n(p-\varepsilon), n(p+\varepsilon))\right) \geq 1 - \frac{1}{4\varepsilon^2 n}.$$

Die Binomialverteilung $b_{n,p}$ lässt sich in diesem Intervall genauer beschreiben. Wir betachten die Gewichte der Binomialverteilung. In Figur 4 (oben) sind diese Gewichte für $n = 100$ und $p = 1/3$ dargestellt. Für das Histogramm wird über dem Intervall $[k-\frac{1}{2}, k+\frac{1}{2}]$ ein Rechteck der Fläche $b_{n,p}(\{k\})$ markiert. Für große n wird das Histogramm dann sehr flach (da sich die Fläche aller Rechtecke stets zu 1 summiert). Man reskaliert deshalb wie folgt: Betrachte statt k nun

$$x(n,k) := \frac{k-np}{\sqrt{npq}} \quad \text{und trage über} \quad \left[x(n,k) - \frac{1}{2\sqrt{npq}}, x(n,k) + \frac{1}{2\sqrt{npq}}\right] \tag{21}$$

ein Rechteck der Fläche $b_{n,p}(\{k\})$ ein. Für k (mit gewissen, moderaten Einschränkungen und einer expliziten Rechnung, auf die wir hier nicht eingehen werden) nähert sich (genauer: konvergiert) für wachsendes n die Höhe dieses Rechtecks dem Wert $\varphi(x(n,k))$ an, wobei

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}.$$

Die Funktion φ heißt *Dichte der Standardnormalverteilung* und wird auch als *Gaußsche Glockenkurve* bezeichnet. Das reskalierte Histogramm nähert sich also in diesem Sinne der Dichte der Standardnormalverteilung an. Eine typische Situation sind Werte k, die von n abhängen, von der Form

$$k_n = np + x\sqrt{npq} + O(1)$$

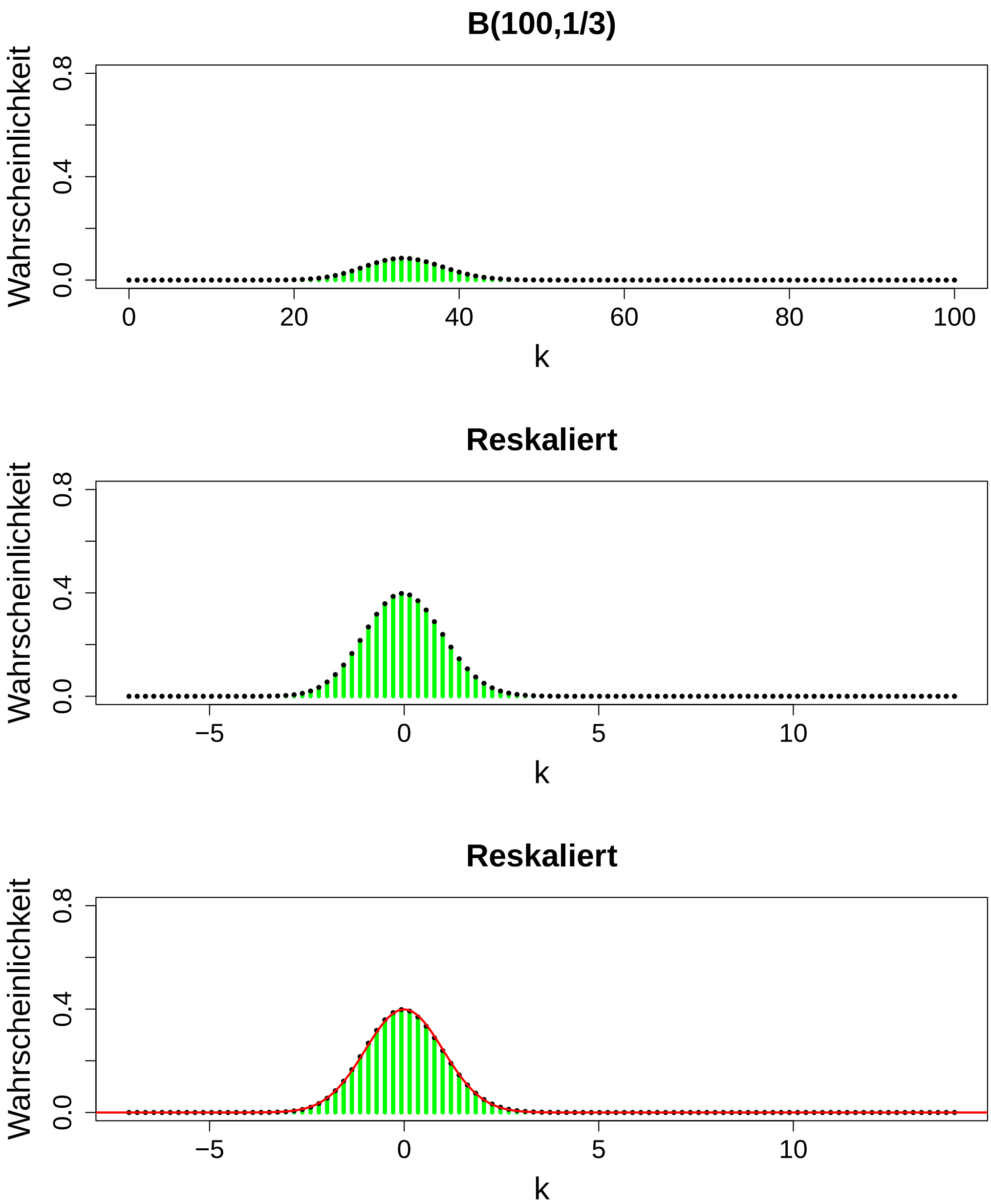


Abbildung 4: Gezeigt sind oben die Gewichte der Binomialverteilung mit $n = 100$ und $p = 1/3$. In der Mitte sind diese Gewichte reskaliert. Im unteren Bild die reskalierten Gewichte zusammen mit der Dichte φ der Standardnormalverteilung in rot.

mit $x \in \mathbb{R}$, wobei $O(1)$ (groß-O-Notation) einen von n abhängenden Term bezeichnet (also eine Folge), der beschränkt in n ist.
Um diesen Zusammenhang für Zufallsvariable umzuschreiben und zu erweitern, bezeichne S_n eine $b_{n,p}$-verteilte ZVe, also etwa die Anzahl der Erfolge bei n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit $p \in (0,1)$. Es ist dann

$$\begin{aligned}\mathbb{P}(S_n = k) &= \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} = \frac{k - np}{\sqrt{npq}}\right) = \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} = x(n,k)\right) \\ &= \mathbb{P}(S_n^* = x(n,k))\end{aligned}$$

im Sinne der folgenden Definition.

Definition 10.1. Für eine Zufallsvariable X mit $\mathrm{Var}(X) > 0$ heißt

$$X^* := \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}}$$

die *standardisierte Form* zu X.

Bemerkung 8. Es gilt stets $\mathbb{E}[X^*] = 0$ und $\mathrm{Var}(X^*) = 1$. Für die standardisierte ZVe S_n^* zu S_n gilt also nach obiger Diskussion

$$\mathbb{P}(S_n^* = x(n,k)) \approx \frac{1}{\sigma_n}\varphi(x(n,k)).$$

Im nächsten Schritt sollen nicht nur die Wahrscheinlichkeiten der „lokalen“ Ereignisse $\{S_n^* = x(n,k)\}$ approximiert werden, sondern auch Wahrscheinlichkeiten für Ereignisse der Form $\{a \leq S_n^* \leq b\}$ für $a < b$. Wir bezeichnen dazu mit

$$\Phi(x) := \int_{-\infty}^{x} \varphi(u)\,du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\,du$$

die *Verteilungsfunktion der Standardnormalverteilung*. Es ist kein expliziter Term bekannt, der $\Phi(x)$ beschreibt. Die Werte von Φ sind deshalb in vielen Programmen hinterlegt, z.B. erhält man in R den Wert $\Phi(x)$ mittels

```
pnorm(x).
```

Manchmal wird auch die Umkehrfunktion Φ^{-1} von Φ benötigt. Die Werte

$$\Phi^{-1}(q) \quad \text{für} \quad q \in (0,1) \tag{22}$$

sind die q-Quantile, vgl. (16), der Standardnormalverteilung, man erhält sie in R mittels

```
qnorm(q).
```

Satz 10.2 (Satz von de Moivre-Laplace)**.** Seien $0 < p < 1$, S_n eine $b_{n,p}$-verteilte Zufallsvariable sowie $a < b$. Dann gilt

$$\mathbb{P}\left(a \le S_n^* \le b\right) \approx \Phi(b) - \Phi(a).$$

(Es bedeutet $\approx$, dass der Grenzwert für $n \to \infty$ der linken Seite gerade die rechte Seite ist.)

Beispiel 10.3. Es werden 600 faire Würfel geworfen. Die Wahrscheinlichkeit, mindestens 90 Sechsen und höchstens 100 Sechsen zu werfen, wird gesucht. Exakt erhält man

$$b_{n,p}(\{90, \ldots, 100\}) = 0{,}4024\ldots$$

für $n = 600$ und $p = \frac{1}{6}$. Damit gilt $np = 100$, $\sigma_n = \sqrt{npq} = 9{,}13$. Die Approximation aus dem Satz von de Moivre-Laplace liefert

$$\mathbb{P}\left(90 \le S_n \le 100\right) = \mathbb{P}\left(\frac{90-100}{\sigma_n} \le S_n^* \le \frac{100-100}{\sigma_n}\right) \approx \Phi(0) - \Phi\left(-\frac{10}{9{,}13}\right) \approx 0{,}36,$$

wobei S_n die $b_{600,1/6}$-Verteilung hat.

Korrekturterme $\pm\frac{1}{2}$: Genauer kann mit den sogenannten Korrekturtermen $\pm\frac{1}{2}$ approximiert werden: Statt des Integrals über $[(90-100)/\sigma_n, (100-100)/\sigma_n]$ nehme man das Integral über $[(90-\frac{1}{2}-100)/\sigma_n, (100+\frac{1}{2}-100)/\sigma_n]$. Das entspricht der genauen Lage der Rechtecke mit den Grundseiten aus (21). Dies liefert

$$\mathbb{P}\left(90 \le S_n \le 100\right) \approx \Phi\left(\frac{0{,}5}{9{,}13}\right) - \Phi\left(-\frac{10{,}5}{9{,}13}\right) \approx 0{,}397.$$

Eine weitreichende Verallgemeinerung des Satzes 10.2 von de Moivre-Laplace ist der Zentrale Grenzwertsatz:

Satz 10.4 (Zentraler Grenzwertsatz)**.** Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariable, die alle dieselbe (beliebige) Verteilung besitzen mit $\mathrm{Var}(X_1) > 0$. Bezeichne $S_n = \sum_{1 \le i \le n} X_i$, sowie S_n^* die standardisierte Form zu S_n. Dann gilt für $a < b$, dass

$$\mathbb{P}(a \le S_n^* \le b) \approx \Phi(b) - \Phi(a).$$

(Wieder bedeutet $\approx$, dass der Grenzwert für $n \to \infty$ der linken Seite gerade die rechte Seite ist.)

Offenbar folgt der Satz 10.2 von de Moivre-Laplace aus dem Zentralen Grenzwertsatz für die Wahl $X_1, \ldots, X_n$ als Bernoulli-verteilte Zufallsvariablen mit Erfolgswahrscheinlichkeit $p \in (0, 1)$.

Bemerkung zum Kerncurriculum: Die Dichte der Normalverteilung (im Kerncurriculum als „Dichtefunktion“ bezeichnet) als Näherung der Binomialverteilung ist Themenfeld der Qualifikationsphase Q3.3 mit erhöhtem Niveau (Leistungskurs). Der Satz von de Moivre–Laplace leistet zwar die im Kerncurriculum gewünschte Näherung der Binomialverteilung durch die Normalverteilung, ist allerdings nur ein (enger, historisch relevanter) Spezialfall des Zentralen Grenzwertsatzes.
Im Kerncurriculum fehlt die für das stochastische Denken ebenso wichtige Näherung der Binomialverteilung (und verwandter Verteilungen) durch die Poissonverteilung, die im folgenden Abschnitt diskutiert wird. Oft kann für große n die Erfolgswahrscheinlichkeit p nicht als fest (also unabhängig von n) angesehen werden wie im Satz von de Moivre–Laplace. Der folgende Abschnitt ergänzt das Verständnis deshalb wesentlich im Fall $np \approx \lambda \in (0, \infty)$, wobei p dann offenbar von n abhängen wird.

11 Poissonapproximation*

Wir betrachten nun eine weitere Approximation der Binomialverteilung $b_{n,p}$. Oft ist n sehr groß und p sehr nahe bei Null. Man stellt sich dann das Produkt np als nahezu konstant vor. Wir nehmen der Einfachheit halber an, dass $np = \lambda$ ist mit einem $\lambda > 0$. Mit wachsendem n werde also stets $p = \frac{\lambda}{n}$ gesetzt. (Im vorigen Abschnitt war p dagegen im Satz von de Moivre-Laplace (Satz 10.2) konstant, also unabhängig von n.) Die Gewichte der Binomial-Verteilung lassen sich dann sehr gut durch die Gewichte der Poisson-Verteilung aus (4) approximieren:

Satz 11.1. Für $\lambda > 0$ sowie $p = p_n = \frac{\lambda}{n}$ gilt für alle $k \in \mathbb{N}_0$, dass

$$b_{n,p}(\{k\}) \approx \Pi_\lambda(\{k\}).$$

(Es bedeutet $\approx$, dass der Grenzwert für $n \to \infty$ der linken Seite gerade die rechte Seite ist.)

Dieses Ergebnis wird auch *Gesetz seltener Ereignisse* genannt, denn die Binomialverteilung beschreibt die Summe unabhängiger Bernoulli-verteilter Zufallsvariablen. Ist nun p klein, so treten wenige dieser Bernoulli-Experimente ein, also ist nur selten mit eintretenden Ereignissen zu rechnen. Satz 11.1, auch Poissonapproximation genannt, quantifiziert diesen Sachverhalt und gibt einen ersten Hinweis auf die Bedeutung der Poisson-Verteilung.
Der Beweis von Satz 11.1 verwendet eine Aussage der Analysis, die über den Rahmen der fachmathematischen Veranstaltungen des Studiengangs hinausgeht; er wird hier für Interessierte jedoch angegeben.

Beweis.
Mit der Bezeichnung $n^{\underline{k}} := n(n-1)\cdots(n-k+1)$ ist

$$\begin{aligned} b_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} &= \frac{(np)^k}{k!} \frac{n^{\underline{k}}}{n^k} \left(1 - \frac{np}{n}\right)^{-k} \left(1 - \frac{np}{n}\right)^n \\ &= \frac{\lambda^k}{k!} \frac{n^{\underline{k}}}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n. \end{aligned} \tag{23}$$

Für festes k ist leicht einzusehen, dass

$$\lim_{n\to\infty} \frac{n^{\underline{k}}}{n^k} = 1, \qquad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1.$$

Aus der Analysis ist der folgende Grenzwert bekannt:

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} = e^{-\lambda}.$$

Die Anwendung dieser drei Grenzwerte in (23) liefert die Behauptung. ∎

6 Grundbegriffe schließender Statistik

Die Stochastik teilt sich in zwei Teilgebiete ein:

$$\text{Stochastik} \begin{cases} \text{Wahrscheinlichkeitstheorie (Kap. 2,4,5)} \\ \text{Statistik} \end{cases}$$

In der Wahrscheinlichkeitstheorie nimmt man an, Wahrscheinlichkeitsmaße, die Zufallsexperimente „steuern", zu kennen, und möchte daraus Eigenschaften komplizierterer Wahrscheinlichkeiten und Größen herleiten, z.B. Gesetze großer Zahlen, Grenzwertsätze oder Poissonapproximation. In der Statistik dagegen sind Wahrscheinlichkeitsverteilungen, die Zufallsexperimente steuern, unbekannt. Stattdessen sind Daten verfügbar, die als Realisierungen der Zufallsexperimente gedeutet werden können. Die Statistik wird in zwei Gebiete unterteilen:

$$\text{Statistik} \begin{cases} \text{deskriptive Statistik (Darstellung von Daten: Tabellen, Graphiken)} \\ \text{schließende Statistik (Inferenzstatistik, induktive Statistik) (Kap. 6)} \end{cases}$$

Die deskriptive Statistik ist der übersichtlichen Darstellung von Daten gewidmet. Geeignete Darstellungsformen für Daten sind z.B. Häufigkeitstabellen, Säulen-, Balken-, Kreis- und Streifendiagramme, Dotplots, Boxplots, Liniendiagramme, Histogramme, Streu- und Flächendiagramme. Sie werden in der Vorlesung „Didaktik der Stochastik" besprochen.

In der schließenden Statistik (die auch in der „Didaktik der Stochastik" mit behandelt wird) kennt man das W-Maß $\mathbb{P}$, das ein Zufallsexperiment steuert, nicht und möchte aus Beobachtungen (Realisierungen) von Versuchsausgängen auf $\mathbb{P}$ oder zumindest Eigenschaften von $\mathbb{P}$ schließen. In diesem Kapitel werden sehr knapp einige Grundbegriffe typischer statistischer Aufgaben (Schätzen, Testen, Konfidenzintervalle) am Beispiel von wiederholten Bernoulliexperimenten mit unbekannter Erfolgswahrscheinlichkeit diskutiert.
Wir betrachten im Folgenden den Münzwurf (also ein Bernoulliexperiment mit Erfolgswahrscheinlichkeit $p \in [0, 1]$). Das Experiment werde n mal unabhängig ausgeführt. Im Gegensatz zu den Untersuchungen der Kapitel 2,4,5 ist nun p aber unbekannt. Stattdessen sind Daten $x_1, \ldots, x_n \in \{0, 1\}$ gegeben, also die Ausgänge der n Wiederholungen des Experiments. Diese können wir uns als Realisierungen von n unabhängigen Zufallsvariablen $X_1, \ldots, X_n$ vorstellen, die die Bernoulli-Verteilung mit Parameter p haben. Die Daten ergeben sich also für ein (uns unbekanntes $\omega \in \Omega$) als $x_i = X_i(\omega)$ für $i = 1, \ldots, n$. Wir wollen aus den Daten Rückschlüsse auf p ziehen. Die folgenden drei Problemstellungen sind typisch für die schließende Statistik:

a) Gib einen Schätzwert für p an — *Schätzproblem.*

b) Gib ein Intervall für p an — *Konfidenzintervall.*

c) Entscheide, ob eine Münze fair ist, d.h. ob $p = \frac{1}{2}$ oder $p \neq \frac{1}{2}$ — *Test.*

Man nennt die Menge der möglichen Ausgänge des Gesamtexperiments, also hier der n unabhängigen Wiederholungen des Bernoulli-Experiments, den *Stichprobenraum* S. Der Vektor der Daten ist also ein Element von S. Wir haben $S = \{0,1\}^n$. Die Erfolgswahrscheinlichkeit p ist uns unbekannt. Aufgrund der Struktur des Experiments als unabhängige Wiederholung von n Bernoulli-Experimenten mit (unbekannter) Erfolgswahrscheinlichkeit p können wir die Wahrscheinlichkeiten, einzelne Daten zu erhalten, aber in Abhängigkeit von p angeben: Für $x = (x_1, \dots, x_n) \in S$ ist

$$\mathbb{P}_p(\{x\}) = p^{\sum x_i}(1-p)^{n-\sum x_i}.$$

Statt $\mathbb{P}$ für das Wahrscheinlichkeitsmaß wie bisher schreiben wir nun $\mathbb{P}_p$, um anzudeuten, dass hier ein unbekannter Parameter p vorliegt. Für jedes $p \in [0,1]$ erhalten wir ein anderes Wahrscheinlichkeitsmaß $\mathbb{P}_p$, das dem Experiment zugrunde liegt.

12 Schätzen

Bei einem Schätzproblem soll der unbekannte Parameter, hier das p, aus den Daten geschätzt werden. Wir ordnen also den Versuchsausgängen einen Schätzwert für p zu. Jede solche Abbildung $T : S \to [0,1]$ heißt *Schätzer* (oder Punktschätzer, oder Schätzfunktion) für p. Dabei spielt zunächst keine Rolle, ob T intuitiv sinnvoll ist. Ziel ist es natürlich aber, Schätzer T zu finden, die gewünschte Eigenschaften haben.
Eine beliebte Methode, Schätzer zu konstruieren, ist, denjenigen Parameter p zu wählen, für den die beobachteten Daten größtmögliche Wahrscheinlichkeit haben. Man nennt den resultierenden Schätzer dann einen *Maximum-Likelihood-Schätzer* (ML-Schätzer). Um in unserem Beispiel den Maximum-Likelihood-Schätzer zu bestimmen, betrachten wir für $x \in S$ die *Likelihood-Funktion*

$$L_x : [0,1] \to [0,1], \; p \mapsto \mathbb{P}_p(\{x\}).$$

Falls L_x ein globales Maximum an einer Stelle $\hat{p}(x)$ annimmt, so ist $\hat{p}(x)$ Maximum-Likelihood-Schätzer von p.
Der ML-Schätzer schätzt also zu gegebener Beobachtung $x \in S$ denjenigen Parameter $\hat{\vartheta}(x)$, für den die beobachteten Daten die größte Wahrscheinlichkeit haben. Statt L_x zu maximieren, ist es oft einfacher, $\log L_x$ zu maximieren, hier also

$$\mathcal{L}_x(p) := \log L_x(p) = \left(\sum x_i\right) \log p + \left(n - \sum x_i\right) \log(1-p).$$

Da der Logarithmus monoton wachsend ist, ist L_x maximal genau dann, wenn $\mathcal{L}_x$ maximal ist. Wir betrachten die Ableitung $\mathcal{L}_x'$ (nach p) von $\mathcal{L}_x$, um Kandidaten zu finden. Es ist

$$\mathcal{L}_x'(p) = \frac{1}{p}\sum x_i - \frac{1}{1-p}\left(n - \sum x_i\right) \stackrel{!}{=} 0. \tag{24}$$

In der Nullstelle $\hat{p}(x) = \frac{1}{n}\sum x_i$ hat L_x ein globales Maximum. Es ist also

$$T(x) = \hat{p}(x) = \frac{1}{n}\sum_{1\le i\le n} x_i$$

ML-Schätzer für p. Die Gleichung (24),

$$\mathcal{L}_x'(p) = 0$$

heißt *ML-Gleichung*.
Erinnern wir uns, dass die Daten Realisierungen von Zufallsvariablen waren, also $x_i = X_i(\omega)$ für $i = 1, \ldots, n$, so ist klar, dass der Schätzer $T = \hat{p} = \frac{1}{n} \sum_{1 \le i \le n} X_i$ vom Zufall abhängt, also eine Zufallsvariable ist.

Als Beispiel für Gütekriterien für Schätzer betrachten wir die Erwartungstreue. Dass ein Schätzer erwartungstreu schätzt, bedeutet, dass im Mittel (also im Erwartungswert) der zu schätzende Parameter korrekt geschätzt wird. Hier tritt wieder das Problem auf, dass das Wahrscheinlichkeitsmaß $\mathbb{P}_p$, bezüglich dessen der Erwartungswert gebildet werden soll, unbekannt ist. Man fordert diese Eigenschaft deshalb für alle $\mathbb{P}_p$ gleichzeitig. Der Schätzer T heißt also *erwartungstreu*, falls

$$\mathbb{E}_p[T] = p \text{ für alle } p \in [0, 1],$$

wobei $\mathbb{E}_p$ den bezüglich $\mathbb{P}_p$ gebildeten Erwartungswert bezeichnet. Ist ein Schätzer nicht erwartungstreu, so nennt man ihn verzerrt (engl. *biased*).
Sind $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariable mit unbekanntem Erfolgsparameter p, so hat der ML-Schätzer die Form $T = \frac{1}{n} \sum_{i=1}^n X_i$. Wegen $\mathbb{E}_p[X_i] = p$ für alle $i = 1, \ldots, n$ erhalten wir

$$\mathbb{E}_p[T] = \mathbb{E}_p\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\mathbb{E}_p\left[\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_p[X_i] = \frac{1}{n}\sum_{i=1}^n p = p.$$

Dies gilt für alle $p \in [0, 1]$. In unserem Beispiel des unbekannten Erfolgsparameters beim Münzwurf ist der ML-Schätzer T also erwartungstreu.

Beispiel 12.1. Wir nehmen an, dass Autos in Frankfurt fortlaufend mit $1, 2 \ldots, N$ nummeriert seien, also insgesamt N Autos unterwegs seien, wobei N unbekannt ist. Wir nehmen die Nummern vorbeifahrender Autos auf und nehmen an, dass diese Nummern als Realisierungen unabhängiger, uniform auf $\{1, \ldots, N\}$ verteilter Zufallsvariblen $Y_1, \ldots, Y_n$ modelliert werden können, machen also die Laplace-Annahme der Gleichverteilung. Wir wollen aus den Daten die unbekannte Gesamtzahl N der Autos schätzen. Wir schreiben $\mathbb{P}_N$ für die entsprechende Verteilung, also die Gleichverteilung auf $\{1, \ldots, N\}^n$. Damit gilt für $(k_1, \ldots, k_n) \in \mathbb{N}^n$, dass

$$\mathbb{P}_N((k_1, \ldots, k_n)) = \begin{cases} \frac{1}{N^n}, & \text{falls } k_1, \ldots, k_n \le N, \\ 0, & \text{sonst.} \end{cases} \tag{25}$$

Um einen ML-Schätzer zu bestimmen, müssen wir diese Wahrscheinlichkeit in N maximieren, also das N finden, so dass $\mathbb{P}_N((k_1, \ldots, k_n))$ maximal wird. Dies ist offenbar für $N = \max\{k_1, \ldots, k_n\}$ der Fall. Der ML-Schätzer $\hat{N}$ für N ist also gegeben durch

$$\hat{N}(k_1, \ldots, k_n) = \max_{1 \le i \le n} k_i.$$

Ist der ML-Schätzer $\hat{N} = \max_{1 \le i \le n} Y_i$ erwartungstreu? Im Allgemeinen werden wir mit $\hat{N}$ den Wert N unterschätzen, denn es ist stets $\hat{N} \le N$ und $\hat{N} = N$ nur dann, falls eine der Beobachtungen N ist. Man kann $\mathbb{E}_N[\max\{Y_1, \dots, Y_n\}]$ auch mit etwas Aufwand ausrechnen und erhält

$$\mathbb{E}_N[\max\{Y_1, \dots, Y_n\}] = \sum_{i=1}^{N} \left(1 - \left(\frac{i-1}{N}\right)^n\right) < N.$$

Der ML-Schätzer $\hat{N}$ ist also verzerrt und unterschätzt N im Mittel.

> **Bemerkung zur Didaktik:** Ein weiteres bekanntes Beispiel eines verzerrten ML-Schätzers im Kontext der Schätzung der Varianz von Stichproben kann in der Vorlesung „Didaktik der Stochastik“ behandelt werden. Dort kann die Erwartungstreue durch einen Korrekturfaktor „geheilt“ werden.
> Erwartungstreue ist eine wünschenswerte Eigenschaft, allerdings existieren erwartungstreue Schätzer nicht immer und falls sie existieren, sind sie nicht unbedingt gute Schätzer. Auf weitere Gütekriterien, z.B. den mittleren quadratischen Fehler, vgl. etwa [7, Abschnitt 5.2], gehen wir hier nicht weiter ein.

13 Konfidenzintervalle

Wir kommen zurück zum Schätzen der Erfolgswahrscheinlichkeit beim Münzwurf. Wir betrachten wieder n unabhängige Wiederholungen des Bernoulli-Experiments mit Stichprobenraum $S = \{0,1\}^n$. Die Erfolgswahrscheinlichkeit p ist unbekannt, aber aufgrund der Struktur des Experiments haben wir für $x = (x_1, \dots, x_n) \in S$ und $p \in [0,1]$, dass

$$\mathbb{P}_p(\{x\}) = p^{\sum x_i}(1-p)^{n-\sum x_i}.$$

Ein Schätzer $\hat{p}$ für p gibt einen Hinweis auf die wahre Erfolgswahrscheinlichkeit, allerdings ist nicht klar, wie zuverlässig dieser Wert dann ist. Man möchte deshalb oft auch Abweichungen vom Schätzwert zulassen, etwa dazu ein Intervall um $\hat{p}$ angeben, um im Gegenzug Garantien zu erhalten, dass der wahre Wert mit hoher Wahrscheinlichkeit vom entsprechenden Intervall überdeckt wird. Um solche Wahrscheinlichkeiten abzuschätzen, tritt hier wieder (wie beim Begriff der Erwartungstreue) das Problem auf, dass wir nicht wissen, welches $\mathbb{P}_p$ das tatsächlich zugrundeliegende Wahrscheinlichkeitsmaß ist.

Definition 13.1. Ein *Konfidenzintervall* zum Niveau $\alpha \in (0,1)$ ist ein von den Daten $x \in S$ abhängendes Intervall $I(x) \subset [0,1]$, das den wahren (aber unbekannten) Parameter p mindestens mit Wahrscheinlichkeit $1-\alpha$ überdeckt. Gefordert wird also

$$\mathbb{P}_p(p \in I) \ge 1 - \alpha \quad \text{für alle } p \in [0,1]. \tag{26}$$

Man beachte, dass dabei I zufällig ist, da es von den Daten abhängt, die Realisierungen von Zufallsvariablen sind.

Hinweise zur Didaktik: Es kommt im Bereich der Konfidenzintervalle (und ähnlich bei Tests, die im folgenden Abschnitt besprochen werden) häufig zu Fehlvorstellungen, die auf Missverständnissen beruhen, welche Größen fest und welche zufällig sind. Aus den Daten $x \in S$, die wir als Realisierung einer Zufallsvariable X (mit Verteilung $\mathbb{P}_p$) modellieren, wird das Konfidenzintervall $I(x)$ konstruiert. Es ist also $I(X)$ ein zufälliges Intervall mit (fester) Realisierung $I(x)$. Das tatsächlich zugrunde liegende p ist dagegen fest, also deterministisch. Deshalb wird manchmal auch die Schreibweise $\{I(X) \ni p\}$ verwendet. Aussagen der Form „wir haben die Beobachtung $x \in S$ gemacht, der Wert p fällt also mit mindestens Wahrscheinlichkeit $1-\alpha$ in das Intervall $I(x)$" ergeben deshalb keinen Sinn. Es ist dagegen so, dass der wahre Wert p fest ist und das Intervall $I(X)$ mindestens mit Wahrscheinlichkeit $1-\alpha$ den Wert p überdeckt. Derartige Fehlvorstellungen werden etwa in einem fiktiven Dialog zwischen einem Statistiker und einem Anwender in [7, Seiten 61-64] diskutiert.

Der ML-Schätzer für p ist gegeben durch $\hat{p}(x) = \frac{1}{n}\sum_{i=1}^n x_i$. Wir betrachten deshalb Intervalle der Form $I(x) = (\hat{p}(x)-\varepsilon, \hat{p}(x)+\varepsilon)$ und versuchen, ε geeignet festzulegen. Die zu erfüllende Bedingung (26) ist damit

$$\mathbb{P}_p\left(|\hat{p}-p| \geq \varepsilon\right) \leq \alpha \quad \text{für alle } p \in [0,1].$$

Oder mit $X_1, \dots, X_n$ unabhängigen und Bernoulli $\mathrm{Ber}(p)$ verteilten Zufallsvariablen und $S_n = X_1 + \cdots + X_n$ ist dies gleichbedeutend mit

$$\mathbb{P}(|S_n/n - p| \geq \varepsilon) \leq \alpha. \tag{27}$$

Es gibt mehrere Möglichkeiten nun abzuschätzen bzw. zu approximieren. Um schnell zu einem (groben) Konfidenzintervall zu kommen, wenden wir die Chebyshevsche Ungleichung aus Korollar 9.4 an: Es ist

$$\mathbb{P}(|S_n/n - p| \geq \varepsilon) \leq \frac{\mathrm{Var}(S_n)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

für alle $p \in [0,1]$. Die Bedingung (26) ist also erfüllt, falls $4n\varepsilon^2\alpha \geq 1$, d.h.

$$I(x) = \left(\hat{p}(x) - \frac{1}{2\sqrt{\alpha n}}, \hat{p}(x) + \frac{1}{2\sqrt{\alpha n}}\right) \tag{28}$$

ist ein Konfidenzintervall für p zum Niveau α.

Man erkennt an der Länge des Konfidenzintervals in (28) einen Sachverhalt, der unter dem Begriff der $\sqrt{n}$-*Regeln* in der Stochastik bekannt ist: Möchte man zu gegebenem α ein schärferes Konfidenzintervall, z.B. von halber Länge, so braucht man viermal soviele Daten (also n muss durch $4n$ ersetzt werden), möchte man nur $\frac{1}{10}$ der Länge, so braucht man 100-mal soviele Daten.

Man kann für dieses Beispiel schärfere Konfidenzintervalle mit verschiedenen Methoden konstruieren, etwa mit dem Zentralen Grenzwertsatz, mit Beta-Quantilen oder Konzentrationsungleichungen, worauf hier nicht weiter eingegangen werden kann. Die $\sqrt{n}$-Regel bleibt dabei stets erhalten.

14 Testen

Wir betrachten wieder n unabhängige Wiederholungen des Bernoulli-Experiments mit Stichprobenraum $S = \{0,1\}^n$. Die Erfolgswahrscheinlichkeit p ist unbekannt, aber aufgrund der Struktur des Experiments haben wir für $x = (x_1, \dots, x_n) \in S$ und $p \in [0,1]$, dass

$$\mathbb{P}_p(\{x\}) = p^{\sum x_i}(1-p)^{n-\sum x_i}.$$

Beim Testen wird eine *Hypothese* (auch *Nullhypothese* genannt) über den wahren Parameter p betrachtet. In unserem Beispiel des Münzwurfs könnten wir etwa testen, ob die Münze fair ist. Man würde dann die Nullhypothese H_0 wählen, dass die Münze fair ist, also $H_0 = \{\frac{1}{2}\}$, und testen gegen die *Alternative* H_1, dass $p \neq 1/2$ ist, also $H_1 = [0,1]\setminus\{\frac{1}{2}\}$. Zudem wird ein *Niveau* α vorgegeben, etwa $\alpha = 0{,}05$ fest. Wir wollen nun nicht mehr p schätzen oder ein Konfidenzintervall dafür angeben, sondern entscheiden, ob wir es im Lichte der Stichprobe für plausibel halten, dass der wahre Parameter $p = \frac{1}{2}$ ist.

Definition 14.1. Ein *Test* ist eine Abbildung $T : S \to \{0,1\}$, wobei $T(x) = 1$ bedeutet, dass wir die Hypothese H_0 ablehnen, und $T(x) = 0$, dass wir H_0 nicht ablehnen. Die Menge $K = \{x \in S : T(x) = 1\}$ heißt *kritischer Bereich* von T.

Bemerkung 9. Man spricht von einem *Fehler 1. Art*, falls man die Hypothese fälschlich ablehnt, von einem *Fehler 2. Art*, falls man die Hypothese fälschlich annimmt.

Man möchte vor allem die Wahrscheinlichkeit für den Fehler 1. Art kontrollieren. Wieder ist das Problem, dass wir nicht wissen, welches der Wahrscheinlichkeitsmaße $\mathbb{P}_p$ das tatsächlich zugrunde Liegende ist. Wie bei der Erwartungstreue und bei den Konfidenzintervallen kontrollieren wir deshalb die Wahrscheinlichkeit für den Fehler 1. Art gleichzeitig für alle möglichen $p \in H_0$.

Definition 14.2. Ein Test hat *(Signifikanz-)Niveau* $\alpha \in [0,1]$, falls für alle $p \in H_0$ gilt, dass

$$\mathbb{P}_p(K) \leq \alpha.$$

Bemerkung 10. Bei einem Test mit Niveau α ist die Wahrscheinlichkeit für einen Fehler 1. Art also durch α beschränkt. Häufig liegt eine Asymmetrie bezüglich der Fehler 1. und 2. Art vor, z.B. soll getestet werden, ob eine gewisse Krankheit vorliegt, also H_0 oder nicht, also dann H_1, um gegebenenfalls eine Behandlung durchzuführen. Führt nun die Nichtbehandlung eines Kranken zu irreparablem Schaden, die Behandlung eines Gesunden nur zu materiellem Schaden, so muss der Fehler 1. Art kontrolliert werden. Typische Vorgehensweise: Man fixiert ein $\alpha \in [0,1]$ und sucht unter den Tests zum Niveau α, d.h. mit $\mathbb{P}_p(K) \leq \alpha$ für alle $p \in H_0$ denjenigen Test, der die Wahrscheinlichkeit für Fehler 2. Art minimiert.

In unserem Beispiel des Münzwurfs ist es plausibel, die Hypothese abzulehnen, falls $|\sum_{1\leq i\leq n} x_i - n/2| > c$ für einen kritischen Wert $c > 0$. Wir wählen also $c > 0$ minimal mit

$$\mathbb{P}_{\frac{1}{2}}\left(\left|\sum_{i=1}^n X_i - \frac{n}{2}\right| > c\right) = \sum_{k\,:\,|k-n/2|>c} \binom{n}{k}\left(\frac{1}{2}\right)^n \leq \alpha = 0{,}05,$$

wobei $X_1, \dots, X_n$ unabhängig sind mit $\mathbb{P}(X_i = 1) = 1/2 = \mathbb{P}(X_i = 0)$. Z.B. für $n = 100$ erhält man $c = 10$. Damit ist der kritische Bereich

$$K = \Big\{x \in S : \Big|\sum_{i \leq 100} x_i - 50\Big| > 10\Big\}.$$

Was passiert mit dem Fehler 2. Art? Falls etwa $p = 0{,}6$, so ist die Wahrscheinlichkeit für den Fehler 2. Art $0{,}538$. Die Daten reichen nicht aus für eine bessere Trennschärfe. Möchte man etwa $\mathbb{P}_{1/2}(K) \leq 0{,}05$ und $\mathbb{P}_{0{,}6}(K) \geq 0{,}9$, so muss n erhöht werden.

Ein weiterer wichtiger Begriff bei statistischen Tests ist der p-*Wert*. Der p-Wert ist eine Funktion der Daten: Für $x \in S$ gibt der p-Wert $\mathfrak{p}(x)$ an, wie wahrscheinlich eine „mindestens so extreme" Beobachtung wie x unter der Nullhypothese ist. In unserem Beispiel des Münzwurfs etwa wird man $\{y \in S : |\sum_i y_i - 50| \geq |\sum_i x_i - 50|\}$ als die Menge der Beobachtungen betrachten, die mindestens so extrem wie x sind, und den p-Wert deshalb als

$$\mathfrak{p}(x) := \mathbb{P}_{\frac{1}{2}}\Big(\Big\{y \in S : \Big|\sum_{i \leq 100} y_i - 50\Big| \geq \Big|\sum_{i \leq 100} x_i - 50\Big|\Big\}\Big)$$

erklären. Der Test lehnt also für die Beobachtung $x \in S$ genau dann ab, wenn $\mathfrak{p}(x) \leq \alpha$. Man sagt dann auch, dass die beobachtete Diskrepanz zur Nullhypothese zum Niveau α *signifikant* ist. Im Falle $\mathfrak{p}(x) > \alpha$ sagt man entsprechend, dass die Diskrepanz auf dem Niveau α nicht signifikant ist.
Es ist für den Anwender aussagekräftiger, einen p-Wert zu erhalten, als nur die Information über Annahme oder Ablehnung der Hypothese. Denn die Wahl des Signifikanzniveaus (typisch ist $\alpha = 0{,}05$) bleibt willkürlich, der p-Wert dagegen ist unabhängig vom Signifikanzniveau.

Hinweise zur Didaktik: Im Kontext statistischer Tests kommt es häufig zu Fehlvorstellungen, die sich teils darum drehen, dass man aus den Beobachtungen schließen könne, dass Hypothese oder Alternative mit gewissen Wahrscheinlichkeiten gelten oder Hypothese bzw. Alternative gar bewiesen seien. Oder etwa, dass man die Wahrscheinlichkeit berechnet hätte, die Hypothese fälschlicherweise zu verwerfen. All dies macht hier keinen Sinn. Es können im Kontext unserer Hypothesentests grundsätzlich keine Wahrscheinlichkeiten über Hypothese oder Alternative angegeben werden.
Zu einer gegebenen Beobachtung $x \in S$ kann man unter der Hypothese (also einem Wahrscheinlichkeitsmaß $\mathbb{P}_p$) angeben, wie wahrscheinlich so eine Beobachtung oder eine mindestens so extreme Beobachtung ist. Je kleiner diese Wahrscheinlichkeit, desto weniger plausibel erscheint die Hypothese. Fällt diese Wahrscheinlichkeit unter das Signifikanzniveau, lehnen wir die Hypothese als zu wenig plausibel ab.
Um dies konkreter zu machen, betrachten wir nochmals das Beispiel des Münzwurfs. Die Daten liefern den Wert $\sum_{1 \leq i \leq n} x_i$, also wie oft Kopf geworfen wird. Die Hypothese ist, dass die Münze fair ist. Unter dieser theoretischen Annahme können wir die Wahrscheinlichkeit berechnen, dass wir $\sum_{1 \leq i \leq n} x_i$ oder eine noch größere Abweichung von $n/2$ als es $\sum_{1 \leq i \leq n} x_i$ schon ist, beobachten. Der kritische Bereich ist so

konstruiert, also c so gewählt, dass der Test ablehnt, wenn diese Wahrscheinlichkeit das Signifikanzniveau unterschreitet.
Die Hypothese ist also eine theoretische Annahme, die keine Wahrscheinlichkeit hat. Es ist zu beachten, dass die Begriffe „unwahrscheinlich“ und „wenig plausibel“ im Kontext der Stochastik, speziell der Statistik, nicht annähernd synonym sind. Das Adjektiv „(un-)wahrscheinlich“ nimmt stets Bezug auf ein Wahrscheinlichkeitsmaß. Es kann in diesem Kontext deshalb nur darüber gesprochen werden, wie (un)wahrscheinlich *Ereignisse* sind. Was kein Ereignis ist, hat auch keine Wahrscheinlichkeit und kann deshalb auch nicht mehr oder weniger „wahrscheinlich“ sein.
Grundsätzlich kann man aus den Daten gegebenenfalls zwar schließen, dass die Hypothese wenig plausibel ist, nämlich eben, falls unter der Hypothese die Beobachtung oder eine mindestens so extreme Beobachtung entsprechend unwahrscheinlich ist. Aus den Daten jedoch auf die Hypothese zu schließen, ist nicht möglich. Die Hypothese kann höchstens nicht verworfen werden. (Prinzip der Falsifizierbarkeit)
Fehlvorstellungen sowie falsche Ausdrucksweisen rund um statistische Tests werden auch in einem fiktiven Dialog zwischen zwei Studierenden und einem Dozenten in [7, Seiten 107-110] diskutiert. Siehe auch [5, Abschnitt 30.15].

Beispiel: Fishers exakter Test

Das folgende Beispiel ist [3] entnommen: Der Vergleich des Erfolgs zweier Therapiemethoden $\mathbf{T}_1$ und $\mathbf{T}_2$ wurde in zwei unterschiedlichen Formen präsentiert. In Form $\mathbf{F}_\mathsf{A}$ wurde herausgestellt, wie groß jeweils der Anteil der Patient:innen ist, bei denen Behandlung $\mathbf{T}_1$ erfolglos bzw. Behandlung $\mathbf{T}_2$ erfolgreich war. In Form $\mathbf{F}_\mathsf{B}$ wurde dies gerade umgekehrt dargestellt. (Dieser Anteil spielt in der folgenden Diskussion keine Rolle, es geht lediglich um die unterschiedlichen Formen der Präsentation ein und desselben Sachverhalts.)

Von insgesamt 167 Ärzt:innen, die an einer Sommerschule teilnahmen, wurden rein zufällig 80 ausgewählt, denen die Information in Form $\mathbf{F}_\mathsf{A}$ präsentiert wurde, die restlichen 87 bekamen die Information in Form $\mathbf{F}_\mathsf{B}$. Jede:r der Ärzt:innen hatte sich daraufhin für die Bevorzugung einer der beiden Therapiemethoden $\mathbf{T}_1$ oder $\mathbf{T}_2$ zu entscheiden. Das Ergebnis zeigt die folgende Vierfeldertafel:

	Methode $\mathbf{T}_1$	Methode $\mathbf{T}_2$	Total
Form $\mathbf{F}_\mathsf{A}$	40	40	80
Form $\mathbf{F}_\mathsf{B}$	73	14	87
Total	113	54	167

Die Daten zeigen, dass in der Gruppe, die die Information in From $\mathbf{F}_\mathsf{A}$ präsentiert bekam, im Verhältnis weniger Befürworter:innen der Therapiemethode $\mathbf{T}_1$ waren als in der Gruppe, die die Information in Form $\mathbf{F}_\mathsf{B}$ präsentiert bekam (nämlich 40:40 gegen 73:14).

Wir wollen untersuchen, ob sich die Ärzt:innen in ihrer Entscheidung durch die Form der Darstellung beeinflussen ließen und formulieren dazu ein Testproblem:

Unsere Hypothese ist, dass die Form, in der die Information präsentiert wird, keinen Einfluss auf die Entscheidung der Ärzt:innen habe.
Unter der Hypothese ergeben sich die 113 Personen, die sich für Therapiemethode $\mathbf{T}_1$ entschieden haben, rein zufällig aus den 167 Ärzt:innen, denn die Form der Präsentation hat keinen Einfluss auf deren Entscheidung. Wir modellieren dies mit einer Urne mit $n = 167$ Kugeln, davon $s = 80$ schwarze und $w = 87$ weiße Kugeln, aus der wir $k = 113$ Kugeln ohne Zurücklegen ziehen. Unter der Hypothese werden diese Kugeln also gleichwahrscheinlich aus der Urne gezogen wie in unserem Beispiel 3.12. Bezeichnet X die Anzahl dabei gezogener schwarzer Kugeln, so hat X die hypergeometrische Verteilung, siehe Beispiel 7.6. Insbesondere ist also

$$\mathbb{P}(X = \ell) = h(\ell; 113, 167, 80) = \frac{\binom{80}{\ell}\binom{87}{113-\ell}}{\binom{167}{113}}, \tag{29}$$

siehe auch (5). Aus Beispiel 8.8 wissen wir, dass $\mathbb{E}[X] = 113 \cdot \frac{80}{167} \approx 54{,}1$. Unsere 40 beobachteten schwarzen Kugeln liegen also um $14{,}1$ vom Erwartungwert entfernt. Wir erhalten damit einen p-Wert von

$$\begin{aligned} p(40) &= \mathbb{P}(|X - 54{,}1| \geq 14{,}1) \\ &= \mathbb{P}(X \leq 40) + \mathbb{P}(X \geq 68{,}2) \\ &= \mathbb{P}(X \leq 40) + 1 - \mathbb{P}(X \leq 68). \end{aligned}$$

Da wir die Verteilung von X aus (29) kennen, können wir den p-Wert explizit ermitteln und erhalten $p(40) \approx 2{,}98 \cdot 10^{-6}$. Auch R liefert den gesuchten p-Wert über die Verteilungsfunktion der hypergeometrischen Verteilung, vgl. (14) und die R Anweisung direkt über (14), mit

```
phyper(40,113,54,80) + 1 - phyper(68,113,54,80))
```

Unter der Hypothese tritt also ein Ergebnis, das so extrem ist wie das Beobachtete, etwa 3 Mal in einer Million auf. Ein Test lehnt die Hypothese damit ab bis zu einem Signifikanzniveau von $\alpha = p(40) < 3 \cdot 10^{-6}$. Damit wird die Hypothese mehr als fragwürdig.

Die eben beschriebene Vorgehensweise, bekannt als *Fishers exakter Test auf Unabhängigkeit*, ist ein typisches Beispiel eines Permutationstests: Man schreibt die beobachteten Daten in einem Gedankenexperiment dem reinen Zufall zu, indem man jede andere Aufteilung der Ärzt:innen mit Informationen in Form von $\mathbf{F}_A$ bzw. $\mathbf{F}_B$ auf die Befürworter:innen der Therapiemethoden $\mathbf{T}_1$ bzw. $\mathbf{T}_2$ (jede andere Permutation der 167 Personen) als ebenso wahrscheinlich ansieht wie die beobachtete.

7 Probeklausur

Die folgende Probleklausur besteht aus 6 Aufgaben, auf die jeweils maximal 8 Punkte vergeben werden. Bestanden ist die Klausur ab 24 von 48 möglichen Klausurpunkten. Es sind keine Hilfsmittel zugelassen, also insbesondere kein Skript und kein Taschenrechner.

15 Aufgaben

Aufgabe 1. Ein Glücksrad habe 5 gleichwahrscheinliche Ausgänge, nämlich 2, 3, 6, 7, 8. Es werde einmal gedreht.

(a) Modellieren Sie das obige Experiment als geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$, indem Sie Ω und die Wahrscheinlichkeiten der Elementarereignisse angeben.

(b) $A \subset \Omega$ sei das Ereignis, nach Drehen eine durch 3 teilbare Zahl zu erhalten. Geben Sie die Menge A und die Wahrscheinlichkeit $\mathbb{P}(A)$ an.

Eine Person geht folgende Wette ein: Tritt A ein, bekommt sie 4€, andernfalls verliert sie 3€.

(c) Modellieren Sie den Gewinn bzw. Verlust der Person mit einer Zufallsvariablen $X \colon \Omega \to \mathbb{R}$.

(d) Bestimmen Sie den Erwartungswert $\mathbb{E}[X]$.

Aufgabe 2. Kreuzen Sie jeweils das richtige Ergebnis an. Jede Frage hat genau eine richtige Antwort.

(a) In einem See befinden sich 19 Karpfen und 26 Barsche (Fischarten). Ein Fischer angelt 10 Fische. Was ist die Wahrscheinlichkeit, dass er 7 Barsche und 3 Karpfen geangelt hat.

$\square \ \frac{\binom{19}{3}\binom{45}{7}}{\binom{26}{10}}$ $\quad \square \ \frac{\binom{26}{3}\binom{26}{7}}{\binom{45}{10}}$ $\quad \square \ \frac{\binom{26}{3}\binom{19}{7}}{\binom{45}{10}}$ $\quad \square \ \frac{\binom{19}{3}\binom{26}{7}}{\binom{45}{10}}$ $\quad \square \ \frac{\binom{19}{10}\binom{26}{7}}{\binom{45}{3}}$

(b) Auf wie viele Arten können 2 blaue, 3 rote und 2 grüne Kugeln in eine Reihe gelegt werden

☐ 200 ☐ 220 ☐ 190 ☐ 230 ☐ 210

(c) In einem Beutel befinden sich 4 blaue und 5 rote Kugeln. Es werden zufällig 2 ohne Zurücklegen gezogen. Mit welcher Wahrscheinlichkeit werden dabei 2 Kugeln derselben Farbe gezogen?

$\square \ \frac{7}{8}$ $\quad \square \ \frac{6}{11}$ $\quad \square \ \frac{5}{8}$ $\quad \square \ \frac{4}{9}$ $\quad \square \ \frac{5}{11}$

(d) Firma A produziert 10% defekte Produkte, Firma B produziert 15% defekte Produkte und Firma C produziert 8% defekte Produkte. Wie wahrscheinlich ist es, ein defektes Produkt zu kaufen, wenn man eine Firma zufällig gleichverteilt wählt?

☐ 10% ☐ 11% ☐ 13% ☐ 20% ☐ 8%

Aufgabe 3.
Geben Sie zu den folgenden Szenarien jeweils an, um welchen Typ von Abzählproblem es sich handelt und wieviele Möglichkeiten es gibt. Geben Sie das Ergebnis als natürliche Zahl an.
Für die volle Punktzahl ist kein Rechenweg oder explizites Angeben von Wahrscheinlichkeitsräumen notwendig.

(a) Eine Bäckerei hat 7 verschiedene Donut-Sorten. Wie viele Möglichkeiten gibt es für einen Kauf von 3 Donuts?

(b) An einem Basketballturnier nehmen 8 Teams teil. Wie viele Möglichkeiten gibt es für die Top 3?

(c) Ein Kleinunternehmen aus 21 Mitarbeiterinnen soll 3 Repräsentantinnen zu einer Wirtschaftstagung schicken. Wie viele Möglichkeiten gibt es?

(d) Ein Geschäftsmann hat je 4 verschiedene Krawatten, Sakkos und Anzugshosen. Wie viele mögliche Outfits hat er?

Aufgabe 4.
Eine Kochin schläft jede Nacht mit 10%-iger Wahrscheinlichkeit schlecht. In diesem Falle versalzt sie mit 15%-iger Wahrscheinlichkeit am nächsten Tag die Suppe. Hat sie gut geschlafen, so versalzt sie sie nur mit 1% Wahrscheinlichkeit.

(a) Modellieren Sie das Zufallsexperiment als zweistufigen Baum und geben Sie die zugehörige Vierfeldertafel an.

(b) Welches sind bedingte, welches unbedingte Wahrscheinlichkeiten?

(c) Mit welcher Wahrscheinlichkeit versalzt die Köchin heute die Suppe?

(d) Die Köchin hat heute die Suppe tatsächlich versalzen. Mit welcher Wahrscheinlichkeit hat sie schlecht geschlafen?

Bei den Teilen (c) und (d) ist der Rechenweg erforderlich.

Aufgabe 5. Ein Torwart hält im Durchschnitt $\frac{1}{3}$ aller Elfmeter. Es werden 4 Elfmeter geschossen. Bestimmen Sie die Wahrscheinlichkeiten der folgenden Ereignisse.

(a) Er hält mindestens 3 Elfmeter.

(b) Er hält genau 2 Elfmeter.

Es werden nun soviele Elfmeter geschossen, bis er erstmals hält. Bestimmen Sie die Wahrscheinlichkeiten der folgenden Ereignisse:

(c) Es werden mindestens 4 Elfmeter geschossen.

(d) Es werden höchstens 2 Elfmeter geschossen.

Stellen Sie jeweils den Zusammenhang zur Binomialverteilung bzw. zur geometrischen Verteilung her und geben Sie die Ergebnisse als gekürzte Brüche an.

Aufgabe 6.
Der Bombardier CRJ900 (Flugzeug) hat 90 Sitzplätze. Eine Fluggesellschaft erlaubt bis zu 100 Buchungen, um möglichen Stornierungen entgegenzuwirken. Man geht von einer Ausfallrate von 10% pro Person aus.
Schätzen Sie mit Normalapproximation (mit und ohne Korrekturtermen $\pm 1/2$) die Wahrscheinlichkeit dafür ab, dass es zu einer Überbuchung (mindestens 91 Personen erscheinen) kommt. Gehen Sie davon aus, dass ursprünglich 100 Buchungen getätigt wurden. Geben Sie das Ergebnis als Differenz der Verteilungsfunktion Φ der Standardnormalverteilung an.

16 Musterlösungen

Lösung zu Aufgabe 1.

(a) Der Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ mit $\Omega = \{2, 3, 6, 7, 8\}$ und $\mathbb{P}(\{\omega\}) = 1/5$ für jedes $\omega \in \Omega$ modelliert das gegebene Glücksrad.

(b) Es ist $A = \{3, 6\}$, und es gilt $\mathbb{P}(\{3, 6\}) = 2/5$.

(c) Die Zufallsvariable $X \colon \Omega \to \mathbb{R}$ hat die Form

$$X(\omega) = \begin{cases} 4, & \text{falls } \omega \in \{3, 6\}, \\ -3, & \text{falls } \omega \in \{2, 7, 8\}. \end{cases}$$

(d) Es gilt $\mathbb{E}[X] = 4 \cdot \frac{2}{5} - 3 \cdot \frac{3}{5} = -\frac{1}{5}$.

Lösung zu Aufgabe 2.
Eine Begründung ist **nicht** notwendig für die volle Punktzahl.

(a) Es handelt sich um eine hypergeometrische Wahrscheinlichkeit, die als „Günstige“ durch „Mögliche“ berechnet werden kann. Von 45 Fischen werden 10 gezogen, also $\binom{45}{10}$ „Mögliche“ (im Nenner). Von den 19 Karpfen und 26 Barschen werden je 3 bzw. 7 gezogen, also $\binom{19}{3}$ mal $\binom{26}{7}$ „Günstige“ (im Zähler).

(b) Insgesamt gäbe es 7! Permutationen der Kugeln, wären diese unterscheidbar. Da jeweils die Kugeln derselben Farbe nicht unterscheidbar sind, muss durch 2! 3! 2! geteilt werden. Also ergibt sich $\frac{7!}{2!3!2!} = 210$.

(c) Es ist $\mathbb{P}(\{2 \text{ mal blau}\}) + \mathbb{P}(\{2 \text{ mal rot}\}) = \frac{4}{9} \cdot \frac{3}{8} + \frac{5}{9} \cdot \frac{4}{8} = \frac{4}{9}$.

(d) Bezeichne A, B, C die Ereignisse, dass man das Produkt in den entsprechenden Firmen kauft sowie D das Ereignis, dass es defekt ist. Dann ergibt sich mit dem Satz von der totalen Wahrscheinlichkeit, dass

$$\begin{aligned}\mathbb{P}(D) &= \mathbb{P}(D|A)\mathbb{P}(A) + \mathbb{P}(D|B)\mathbb{P}(B) + \mathbb{P}(D|C)\mathbb{P}(C) \\ &= 10\%\frac{1}{3} + 15\%\frac{1}{3} + 8\%\frac{1}{3} = 11\%.\end{aligned}$$

Lösung zu Aufgabe 3.

(a) Stichprobe ohne Beachtung der Reihenfolge, mit Zurücklegen. Die Größe der Grundgesamtheit ist durch die $n = 7$ Donut-Sorten gegeben, $k = 3$ Ziehungen (Modell IV). Es ergeben sich $\binom{7+3-1}{3} = \binom{9}{3} = 84$ mögliche Käufe.

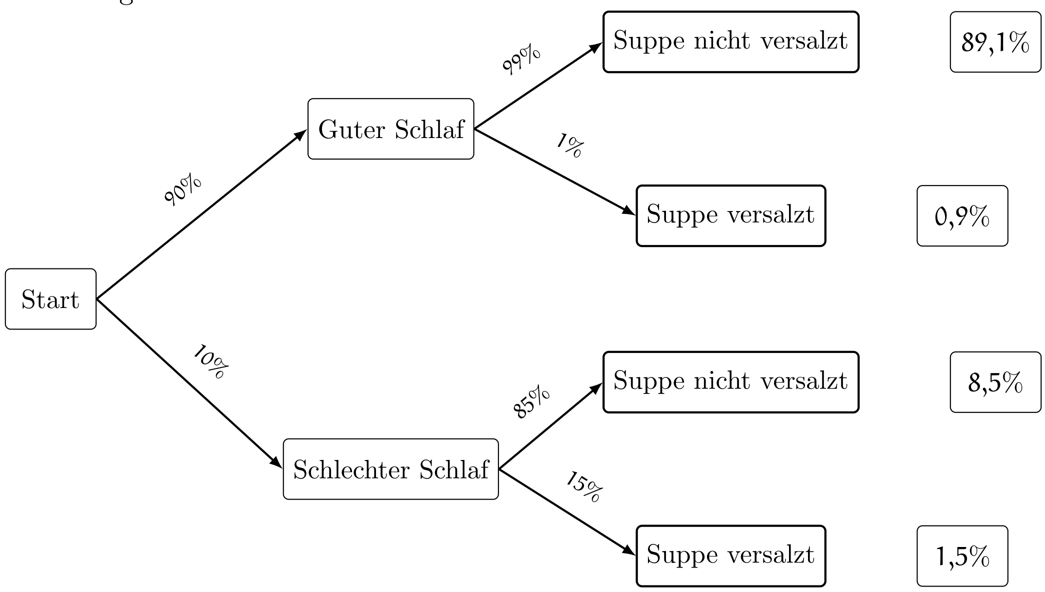
(b) Stichprobe mit Beachtung der Reihenfolge, ohne Zurücklegen. Die Größe der Grundgesamtheit ist durch $n = 8$ Teams gegeben, $k = 3$ Ziehungen (Modell II). Es ergeben sich also $8 \cdot 7 \cdot 6 = 336$ Möglichkeiten für die Top 3.

(c) Stichprobe ohne Beachtung der Reihenfolge, ohne Zurücklegen. Die Größe der Grundgesamtheit ist durch $n = 21$ Mitarbeiterinnen gegeben, $k = 3$ Ziehungen (Modell III). Also ergeben sich $\binom{21}{3} = 1330$ Möglichkeiten.

(d) Stichprobe mit Beachtung der Reihenfolge, mit Zurücklegen. Die Größe der Grundgesamtheit ist durch $n = 4$ Arten der jeweiligen Kleidungsstücke gegeben, $k = 3$ Ziehungen (Modell I). Es ergeben sich $4^3 = 64$ Möglichkeiten.

Lösung zu Aufgabe 4.

(a) Baumdiagramm



Vierfeldertafel

	Guter Schlaf	Schlechter Schlaf	Total
Suppe nicht versalzt	89,1%	8,5%	97,6%
Suppe versalzt	0,9%	1,5%	2,4%
Total	90%	10%	100%

(b) Der Wert 10% ist eine unbedingte Wahrscheinlichkeit, die Werte 15% und 1% sind bedingte Wahrscheinlichkeiten.

(c) Es bezeichne S das Ereignis, dass die Köchin die Suppe versalzt, und G das Ereignis, dass die Köchin zuvor gut geschlafen hat. Der Satz von der totalen Wahrscheinlichkeit ergibt

$$\mathbb{P}(\mathsf{S}) = \mathbb{P}(\mathsf{G})\mathbb{P}(\mathsf{S} \mid \mathsf{G}) + \mathbb{P}(\mathsf{G}^\mathsf{C})\mathbb{P}(\mathsf{S} \mid \mathsf{G}^\mathsf{C}) = 90\% \cdot 1\% + 10\% \cdot 15\% = 2{,}4\%.$$

(d) Die Ereignisse S und G seien wie in (c). Der Satz von Bayes liefert

$$\mathbb{P}(\mathsf{G}^\mathsf{C} \mid \mathsf{S}) = \frac{\mathbb{P}(\mathsf{G}^\mathsf{C})\mathbb{P}(\mathsf{S} \mid \mathsf{G}^\mathsf{C})}{\mathbb{P}(\mathsf{G})\mathbb{P}(\mathsf{S} \mid \mathsf{G}) + \mathbb{P}(\mathsf{G}^\mathsf{C})\mathbb{P}(\mathsf{S} \mid \mathsf{G}^\mathsf{C})} = \frac{10\%\, 15\%}{2{,}4\%} = \frac{5}{8}.$$

Lösung zu Aufgabe 5.

Es handelt sich um die unabhängige Hintereinanderausführung eines Bernoulli-Experiments. Erfolg bedeutet, dass der Torwart den Elfmeter hält. Die Erfolgswahrscheinlichkeit beträgt also $\mathsf{p} = \frac{1}{3}$.

Die Zufallsvariable X zähle die Anzahl gehaltener Elfmeter. Also ist X binomial $\mathsf{B}(4, \frac{1}{3})$ verteilt. Die Anzahl geschossener Elfmeter in (c) und (d) werde mit Y bezeichnet. Damit ist Y ist geometrisch $g_{\frac{1}{3}}$ verteilt Für die Wahrscheinlichkeiten ergibt sich:

(a) $\mathbb{P}(X \geq 3) = \binom{4}{3}(\frac{1}{3})^3\frac{2}{3} + \binom{4}{4}(\frac{1}{3})^4 = \frac{1}{9}$.

(b) $\mathbb{P}(X = 2) = \binom{4}{2}(\frac{1}{3})^2(\frac{2}{3})^2 = \frac{24}{81} = \frac{8}{27}$.

(c) $\mathbb{P}(Y \geq 4) = 1 - \mathbb{P}(Y \leq 3) = 1 - (\frac{1}{3} + \frac{2}{3}\frac{1}{3} + (\frac{2}{3})^2\frac{1}{3}) = \frac{8}{27}$.

Das Ereignis kann auch dadurch beschrieben werden, dass die ersten drei Elfmeter nicht gehalten werden, so dass man auch mit $(2/3)^3$ zum Ergebnis kommt.

(d) $\mathbb{P}(Y \leq 2) = \frac{1}{3} + \frac{2}{3}\frac{1}{3} = \frac{5}{9}$.

Lösung zu Aufgabe 6.
Bezeichne X die Anzahl nicht stornierter Buchungen. Dann ist X binomial $\mathsf{B}(100, \frac{9}{10})$ verteilt. Also gilt

$$\mu := \mathbb{E}[X] = 100 \cdot \frac{9}{10} = 90,$$
$$\sigma := \sqrt{\mathrm{Var}(X)} = \sqrt{100 \cdot \frac{1}{10} \cdot \frac{9}{10}} = 3.$$

Die Abschätzung der Wahrscheinlichkeit einer Überbuchung mittels Normalapproximation ohne Korrekturterme ergibt sich wie folgt:

$$\begin{aligned}\mathbb{P}(\text{Überbuchung}) &= \mathbb{P}(91 \leq X \leq 100) = \mathbb{P}\left(\frac{91-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{100-\mu}{\sigma}\right)\\ &= \mathbb{P}\left(\frac{1}{3} \leq X^* \leq \frac{10}{3}\right) \approx \Phi\left(\frac{10}{3}\right) - \Phi\left(\frac{1}{3}\right),\end{aligned}$$

wobei X^* die standardisierte Form von X bezeichnet.
Die Abschätzung der Wahrscheinlichkeit einer Überbuchung mittels Normalapproximation mit Korrekturtermen ergibt sich wie folgt:

$$\begin{aligned}\mathbb{P}(\text{Überbuchung}) &= \mathbb{P}(91 \leq X \leq 100) = \mathbb{P}(90{,}5 \leq X \leq 100{,}5)\\ &= \mathbb{P}\left(\frac{90{,}5-90}{3} \leq \frac{X-\mu}{\sigma} \leq \frac{100{,}5-90}{3}\right)\\ &= \mathbb{P}\left(\frac{1}{6} \leq X^* \leq \frac{21}{6}\right) \approx \Phi\left(\frac{21}{6}\right) - \Phi\left(\frac{1}{6}\right).\end{aligned}$$

Literatur

[1] Andreas Büchter and Hans-Wolfgang Henn. *Elementare Stochastik.* Mathematik für das Lehramt. Springer-Verlag Berlin Heidelberg, 2nd. edition, 2007. Eine Einführung in die Mathematik der Daten und des Zufalls.

[2] Hermann Dinges and Hermann Rost. *Prinzipien der Stochastik.* Teuber Studienbücher Mathematik. Vieweg+Teubner Verlag, 1982.

[3] David Freedman, Robert Pisani, and Roger Purves. *Statistics.* New York, NY: W. W. Norton & Co., 3rd edition edition, 1998.

[4] Hans-Otto Georgii. *Stochastik.* de Gruyter Lehrbuch. Walter de Gruyter & Co., Berlin, expanded edition, 2009. Einführung in die Wahrscheinlichkeitstheorie und Statistik.

[5] Norbert Henze. *Stochastik für Einsteiger.* Vieweg + Teubner Verlag, Wiesbaden, expanded edition, 2012. Eine Einführung in die faszinierende Welt des Zufalls.

[6] Götz Kersting and Anton Wakolbinger. *Elementare Stochastik.* Mathematik Kompakt. Birkhäuser/Springer, Basel, second edition, 2010.

[7] Michael Messer and Gaby Schneider. *Statistik.* Springer Spektrum, 2019. Theorie und Praxis im Dialog.