Wahrscheinlichkeits-Rechnung und Statistik

— Eine Einführung —

Unterlagen für das Sommersemester 2018

Prof. Dr. Karl Stroetmann

2. Mai 2018

Dieses Skript ist einschließlich der LaTeX-Quellen sowie der in diesem Skript diskutierten Programme unter

`https://github.com/karlstroetmann/Statistik`

im Netz verfügbar. Das Skript selbst finden Sie in dem Unterverzeichnis Skript. Dort ist das Skript in der Datei `statistik.pdf` abgespeichert. Wenn Sie auf Ihrem Rechner `git` installieren und mein Repository mit Hilfe des Befehls

```
git clone https://github.com/karlstroetmann/Statistik.git
```

klonen, dann können Sie durch den Befehl

```
git pull
```

die aktuelle Version meines Skriptes aus dem Netz laden. Falls Sie Fehler in dem Skript finden, so bin ich für Hinweise an die Email-Adresse

`karl.stroetmann@dhbw-mannheim.de`

dankbar. Wer sich mit git und LaTeX auskennt, darf mir auch gerne einen Pull-Request schicken.

Inhaltsverzeichnis

Kapitel 1

Einführung

Die Vorlesung "Wahrscheinlichkeits-Theorie und Statistik" beschäftigt sich sowohl mit der Wahrscheinlichkeits-Theorie, als auch mit der Anwendung dieser Theorie in der induktiven Statistik. Die induktive Statistik hat im wesentlichen zwei Aufgaben:

1. Das Schätzen von Parameter.

   Nehmen Sie an, Sie möchten die Lebensdauer einer LED-Lampe herausfinden. Eine Möglichkeit, diese Lebensdauer zu bestimmen, besteht darin, eine Menge verschiedene Lampen solange brennen zu lassen, bis sie defekt sind. Aus der Lebensdauer der einzelnen Lampen versuchen wir dann die durchschnittliche Lebensdauer von LED-Lampen abzuleiten. Hier stellt sich die Frage, wieviele Lampen zu testen sind, damit wir eine ausreichende Sicherheit bei der Schätzung der Lebensdauer haben.

2. Das Testen von Hypothesen.

   Wird ein neues Medikament entwickelt, so muss unter anderem untersucht werden, ob das Medikament einen Beitrag zur Heilung liefert oder ob der Patient ohne das Medikament genauso schnell gesund geworden wäre.

Statistik und Wahrscheinlichkeits-Theorie sind eine der Grundlagen des maschinellen Lernens, das in den letzten Jahren das Gebiet der künstlichen Intelligenz revolutioniert hat. Neben der induktiven Statistik gibt es noch die deskriptive Statistik, deren Aufgabe es ist, umfangreiche Daten in Kennzahlen zusammenzufassen und diese Daten in Grafiken übersichtlich anzuordnen. Aus Zeitgründen werden wir uns mit diesem Teilgebiet der Statistik nur am Rande beschäftigen können.

Wir verdeutlichen das grundsätzliche Vorgehen der induktiven Statistik an einem einfachen Beispiel. Ein Biologe hat im Wald einen Ameisenhaufen entdeckt und möchte wissen, wieviele Ameisen in dieser Kolonie wohnen. Da es zu aufwendig ist, alle Ameisen zu zählen, wählt er ein statistisches Verfahren: Er fängt $1\,000$ Ameisen und markiert diese Ameisen mit einem Farbtupfer. Anschließend entlässt er die markierten Ameisen in die Freiheit. Am nächsten Tag kommt er wieder und fängt 200 Ameisen. Er stellt fest, dass sich unter diesen Ameisen 5 Tiere befinden, die einen Farbtupfer haben. Unter der Annahme, dass sich die Ameisen in der Nacht gut durchmischt haben und dass folglich die Wahrscheinlichkeit für eine Ameise, bei der zweiten Zählung gefangen zu werden, unabhängig davon ist, ob die Ameise farblich markiert ist, kann der Biologe nun die Anzahl aller Ameisen schätzen:

1. Bezeichnet n die Gesamtzahl der Ameisen, so ist der Prozentsatz p der markierten Ameisen durch die Formel

   $$p = \frac{1000}{n} \tag{1.1}$$

   gegeben.

2. Aus der Voraussetzung, dass die Wahrscheinlichkeit dafür, dass eine Ameise bei der zweiten Zählung gefangen wird, unabhängig von der Markierung ist, folgt, dass für die Zahl der gefangenen Tiere, die farblich markiert sind, in etwa

$$5 = p \cdot 200 \tag{1.2}$$

gilt.

Setzen wir den Wert von p aus Gleichung (1.1) in Gleichung (1.2) ein, so finden wir

$$5 = \frac{1000}{n} \cdot 200 \quad \Leftrightarrow \quad n = \frac{1000 \cdot 200}{5} = 40\,000. \tag{1.3}$$

Der Biologe wird also vermuten, dass die Kolonie von etwa $40\,000$ Ameisen bewohnt wird. Die hier abgeleitete Formel für die Anzahl der Ameisen wird in der Literatur als Lincoln-Petersen Schätzer bezeichnet.

Neben dem Schätzwert für die Anzahl der Ameisen beantwortet die Statistik zusätzlich die Frage, wie genau der angegebene Schätzwert ist. Wir werden im Laufe der Vorlesung ein Verfahren entwickeln, mit dessen Hilfe wir ein Intervall $[n_1, n_2]$ berechnen können, so dass für die Anzahl n der Ameisen die Aussage

$$n \in [n_1, n_2]$$

mit einer Konfidenz von $95,0\%$ gilt. Um den Begriff eines solchen Kondidenz-Intervalls exakt fassen zu können, müssen wir uns zunächst mit der Wahrscheinlichkeits-Theorie auseinander setzen.

1.1 Anwendungen der Theorie der Statistik

Eine der wichtigsten Anwendungen von Wahrscheinlichkeits-Theorie und Statistik ist das maschinelle Lernen, das im Laufe der letzten Jahre die Grundlage verschiedener spektakulärer Durchbrüche im Bereich der künstlichen Intelligenz war. Stellvertretend für andere Erfolge möchte ich hier zuerst das Programm AlphaGo nennen, das im Jahre 2017 die Nummer 1 der Weltrangliste in Go, Ke Jie besiegt hat. Ein anderer Bereich, in dem maschinelles Lernen einen Durchbruch gebracht hat, ist die Entwicklung selbstfahrender Kraftfahrzeuge, die nach den Plänen führender Autohersteller im Laufe der nächsten fünf Jahre Marktreife erlangen werden. Neben diesen spektakulären Anwendungen der Statistik der letzten Jahre gibt es eine große Zahl von Bereichen, in denen Statistik schon seit langem eingesetzt wird. An Stelle einer vollständigen Liste solcher Anwendungen, die leicht mehrere Seiten füllen würde, möchte einige wenige Anwendungen herausgreifen, bei denen Statistik entscheidend ist.

1. Bei der Messung der Wirksamkeit von Medikamenten und medizinischen Therapien werden seit langem statistische Tests durchgeführt. Wird ein neues Medikament eingeführt, so reicht es bei der Zulassung nicht aus, jemanden zu kennen, der behauptet, dass das Medikament der Tante seines Nachbarn das Leben gerettet hat. Stattdessen werden umfangreiche statistische Tests durchgeführt, mit denen die Wirksamkeit eines Medikaments oder einer Behandlungsmethode untersucht werden. Beispielsweise wurde im Jahre 2012 eine umfangreiche statistische Studie in Norwegen durchgeführt, bei der festgestellt wurde, dass das Brustkrebs-Screening in der bisher durchgeführten Form sehr fragwürdig ist und häufig zu sinnlosen Behandlungen geführt hat.

2. Im Versicherungswesen müssen Risiken bewertet werden. Wahrscheinlichkeits-Theorie und Statistik bilden die Grundlage einer solchen Bewertung.

3. Die Vorhersage eines Wahlausgangs beruht auf statistischen Erhebungen und wahrscheinlichkeitstheoretischen Berechnungen. Bei der US-amerikanischen Präsidentschaftswahl 2016 war es entscheidend, das Trump sein Geld genau in den Staaten für Werbung eingesetzt hat, die bei der Wahl das Zünglein an der Waage waren. Bei einer demokratischen Wahl in einem Land wie den USA kommt es entscheidend darauf an, wer die besseren Statistiker hat. Einen interessanten Artikel dazu finden Sie unter

https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win

Das Interessante ist, das Trump bei der Wahl 2016 nur 957,6 Millionen Dollar ausgegeben hat, während Hillary Clinton über 1,4 Milliarden Dollar verfügen konnte, siehe hier. Er konnte die Wahl trotzdem gewinnen, weil sein Team genau ausgerechnet hatte, welche Wahlbezirke mit wieviel Geld zu gewinnen waren. Das erklärt auch, warum insgesamt deutlich weniger Wähler für Trump gestimmt haben als für Clinton. Im US-amerikanischen Wahlsystem kommt es nicht auf die Gesamtzahl aller Stimmen an, sondern auf die Gesamtzahl der gewonnenen Wahlbezirke und da lag Donald Trump am Ende der Auszählung dann vorne.

4. In vielen Mannschaft-Sportarten wird heute Statistik eingesetzt, um die Teams optimal zusammen zu stellen. In dem Bestseller Moneyball [Lew03] beschreibt Michael Lewis, wie es dem Manger Billy Beane der Oakland Athletics Baseball-Mannschaft gelang, das Team aus Oakland konkurrenzfähig zu den New York Yankees zu machen. Das Buch wurde 2011 verfilmt.

5. Die Anfänge der Statistik gehen auf Untersuchungen von Glücksspielen zurück. Da Glücksspiele relative leicht zu verstehen sind, werden wir uns bei der Einführung der Wahrscheinlichkeits-Theorie bei den Anwendungen auf die Untersuchung von Glücksspielen beschränken.

1.2 Literatur

Es folgt eine subjektive Zusammenstellung von Literatur zur Statistik. Zusätzlich zu dem vorliegenden Skript kann ich Ihnen die folgende Literatur empfehlen.

1. Das meiner Ansicht nach beste Buch zur Wahrscheinlichkeits-Theorie ist das Buch Introduction to Probability von Joseph K. Blitzstein und Jessica Hwang [BH14]. Professor Blitzstein hat die Vorlesung, die diesem Buch zu Grunde liegt, auf Video aufgenommen und auf stat110.net zur Verfügung gestellt.

2. Ein weiteres Buch mit dem Titel Introduction to Probability stammt von den Autoren Charles M. Grinstead und Laurie J. Snell [GS97]. Dieses Buch ist ebenfalls sehr empfehlenswert und kann unter der Adresse

   https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/amsbook.mac.pdf

   kostenlos aus dem Netz geladen werden.

3. Im Gegensatz zu den beiden vorgenannten Büchern behandelt das Buch Principles of Statistics von Michael Bulmer [Bul79] sowohl die Wahrscheinlichkeits-Theorie als auch die Anwendung dieser Theorie in der Statistik. Amazon bietet eine preiswerte Kindle-Edition für € 9.99 an.

4. Das Buch Statistik — Der Weg zur Datenanalyse von Ludwig Fahrmeir und anderen [FHK+16] ist ein gutes deutsches Buch zur Statistik. Es enthält allerdings wesentlich mehr Stoff als wir im Rahmen der Vorlesung behandeln können. Dieses Buch ist in unserer Bibliothek verfügbar.

5. Als Nachschlagewerk zur Statistik eignet sich das Buch Engineering Statistics [NIS12], das vom US-amerikanischen National Institute of Standards herausgegeben wird.

Kapitel 2

Einführung in die Wahrscheinlichkeits-Rechnung

Das Leben ist voller Unwägbarkeiten und Risiken. Schon die Kelten wussten, dass ihnen jederzeit der Himmel auf den Kopf fallen konnte. Glücklicherweise passiert so etwas nicht sehr häufig. Trotzdem stellt sich die Frage, wie wahrscheinlich ein solches Ereignis ist. Zur Beantwortung dieser Frage ist zunächst der Begriff der Wahrscheinlichkeit festzulegen. Intuitiv sagen wir, dass ein Ereignis A mit einer Wahrscheinlichkeit p eintritt, wenn bei n-maliger Wiederholung des Experiments das Ereignis A in etwa mit der Häufikeit $p \cdot N$ auftritt. Um der Wahrscheinlichkeits-Begriff genauer festzulegen, definieren wir den Begriff des Zufalls-Experiments. Darunter verstehen wir ein Experiment, dessen Ausgang nicht eindeutig vorbestimmt ist. Ein Beispiel für ein solches Experiment wäre der Wurf einer Münze, bei dem als Ergebnis entweder *Wappen* oder *Zahl* auftritt. Ein anderes Zufalls-Experiment wäre der Wurf eines Würfels, dessen Seiten mit den Zahlen 1 bis 6 beschriftet sind. Hier können als Ergebnisse die natürlichen Zahlen 1 bis 6 auftreten. Mathematisch werden solche Zufalls-Experimente durch den Begriff des Wahrscheinlichkeits-Raums erfasst. Wir betrachten zunächst den Fall, dass die Menge der Ergebnisse endlich oder abzählbar unendlich ist. In diesem Fall wollen wir von einen diskreten Wahrscheinlichkeits-Raum sprechen. Später werden wir noch den allgemeinen Fall untersuchen, bei dem die Menge der Ergebnisse überabzählbar ist. Dieser Fall tritt meistens dann auf, wenn es sich bei den Ergebnissen des Zufalls-Experiments um reelle Zahlen handelt.

2.1 Diskrete Wahrscheinlichkeits-Räume

Definition 1 (Wahrscheinlichkeits-Raum)
Ein diskreter Wahrscheinlichkeits-Raum ist ein Tripel $\langle \Omega, 2^\Omega, P \rangle$, für das gilt:

1. *Ω ist eine Menge, die entweder endlich ist, dann gilt also*

   $$\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\},$$

   oder Ω ist abzählbar unendlich, dann gilt

   $$\Omega = \{\omega_n \mid n \in \mathbb{N}\}.$$

   Die Elemente von Ω bezeichnen wir als die Ergebnisse eines Zufalls-Experiments und die Menge Ω nennen wir die Ergebnis-Menge.

2. *2^Ω ist die Potenzmenge von Ω, also die Menge aller Teilmengen der Menge Ω. Diese Teilmengen bezeichnen wir auch als Ereignisse und die Menge 2^Ω nennen wir die Ereignis-Algebra. Mengen der Form $\{\omega_i\}$, die genau ein Element aus Ω enthalten, nennen wir Elementar-Ereignisse.*

3. *$P : 2^\Omega \to \mathbb{R}$ ist eine Abbildung, die jedem Ereignis $A \subseteq \Omega$ eine reelle Zahl $P(A)$ zuordnet. Die Zahl $P(A)$ bezeichnen wir als die Wahrscheinlichkeit, mit der das Ereignis A eintritt. Die Wahr-*

scheinlichkeit P muss den folgenden Kolmogorow-Axiomen (Andrei Nikolajewitsch Kolmogorow; 1903 – 1987) genügen:

(a) $0 \leq P(A) \leq 1$ *für alle* $A \subseteq \Omega$.
Da wir die Wahrscheinlichkeit $P(A)$ als den Bruchteil aller Fälle, in denen das Ereignis A im Durchschnitt eintritt, interpretieren wollen, muss $P(A)$ eine nicht-negative Zahl sein, die nicht größer als 1 sein kann.

(b) $P(\emptyset) = 0$.
Die leere Menge $\emptyset$ bezeichnen wir als das unmögliche Ereignis.

(c) $P(\Omega) = 1$.
Die Menge Ω bezeichnen wir als das sichere Ereignis.

(d) $A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B)$ *für alle* $A, B \subseteq \Omega$.
Schließen sich zwei Ereignisse A und B gegenseitig aus, gilt also $A \cap B = \emptyset$, so nennen wir diese Ereignisse unvereinbar. Sind A und B unvereinbare Ereignisse, so ergibt sich die Wahrscheinlichkeit des Ereignisses $A \cup B$ als die Summe der Wahrscheinlichkeiten der einzelnen Ereignisse.

Die Funktion P bezeichnen wir als das Wahrscheinlichkeits-Maß.

Schreibweise: Das vierte Kolmogorow-Axiom bezeichnen wir als die Additivität des Wahrscheinlichkeits-Maßes. Um dieses Axiom einfacher schreiben zu können, vereinbaren wir folgende Schreibweise: Sind A und B zwei disjunkte Mengen, so schreiben wir die Vereinigung von A und B als $A \uplus B$. Der Term $A \uplus B$ steht also für zweierlei:

1. Für die Vereinigungs-Menge $A \cup B$.
2. Für die Aussage $A \cap B = \emptyset$.

Mit dieser Schreibweise lautet das Axiom der Additivität

$$P(A \uplus B) = P(A) + P(B).$$

Diese Gleichung gilt genau dann, wenn $A \uplus B$ definiert ist und das ist genau dann der Fall, wenn $A \cap B = \emptyset$ gilt.

Beispiel: Ein möglicher Wahrscheinlichkeits-Raum für das Zufalls-Experiment "Würfeln mit einem kubischen Würfel" ist $\langle \Omega, 2^{\Omega}, P \rangle$ wobei Ω und P wie folgt definiert sind:

- $\Omega = \{1, 2, 3, 4, 5, 6\}$,
- $P(A) = \dfrac{|A|}{|\Omega|} = \dfrac{1}{6} \cdot |A|$.

  Hier bezeichnet $|A|$ die Anzahl der Elemente der Menge A. Beachten Sie, dass wir im ersten Semester für die Anzahl der Elemente die Schreibweise *card*(A) verwendet haben. Diese Schreibweise ist mir jetzt zu aufwendig.

Das Ereignis "es wurde eine gerade Zahl gewürfelt" wird dann durch die Menge $G = \{2, 4, 6\}$ beschrieben. Für die Wahrscheinlichkeit von G gilt

$$P(G) = \frac{1}{6} \cdot |G| = \frac{1}{6} \cdot 3 = \frac{1}{2}.$$

Bei dem oben angegebenen Wahrscheinlichkeits-Raum sind wir davon ausgegangen, dass alle Elementar-Ereignisse dieselbe Wahrscheinlichkeit haben. Diese Annahme ist aus Symmetrie-Gründen naheliegend. In diesem Fall nennen wir das Wahrscheinlichkeits-Maß P auch gleichmäßig, das zugehörige Zufalls-Experiment heißt dann ein Laplace-Experiment (Pierre Simon Laplace; 1749 – 1827). Falls die Wahrscheinlichkeit für alle Seiten eines Würfels den selben Wert hat, so sprechen wir von einem Laplace-Würfel. ◇

Definition 2 (Laplace-Experiment) *Ein Zufalls-Experiment ist ein Laplace-Experiment falls der Wahrscheinlichkeits-Raum die Form $\langle \Omega, 2^\Omega, P \rangle$ hat, wobei Ω eine endliche Menge ist und und das Wahrscheinlichkeits-Maß P für eine Ereignis $A \subseteq \Omega$ nach der Formel*

$$P(A) = \frac{|A|}{|\Omega|}$$

berechnet wird. Die Gültigkeit der Kolmogorow-Axiome kann in diesem Fall sofort nachgerechnet werden. ⋄

Aufgabe 1: Es sei Ω eine endliche Menge. Für alle $A \subseteq \Omega$ definieren wir

$$P(A) := \frac{|A|}{|\Omega|}.$$

Beweisen Sie, dass das Tripel $\langle \Omega, 2^\Omega, P \rangle$ dann ein diskreter Wahrscheinlichkeits-Raum ist. ⋄

Die Annahme der Gleichmäßigkeit des Wahrscheinlichkeits-Maßes ist logisch nicht zwingend. Wenn der Würfel beispielsweise auf einer Seite mit Blei beschwert ist, so könnte das Wahrscheinlichkeits-Maß auch wie folgt gegeben sein:

$$P\big(\{1\}\big) = 0.5, \quad P\big(\{2\}\big) = P\big(\{3\}\big) = P\big(\{4\}\big) = P\big(\{5\}\big) = P\big(\{6\}\big) = 0.1.$$

Dann würde für das Ereignis "es wurde eine gerade Zahl gewürfelt" gelten:

$$P(G) = P\big(\{2\}\big) + P\big(\{4\}\big) + P\big(\{6\}\big) = 0.1 + 0.1 + 0.1 = 0.3.$$

2.2 Additionssätze

Die Kolmogorow-Axiome geben an, wie sich die Wahrscheinlichkeit eines Ereignisses $A \cup B$ dann aus den Wahrscheinlichkeiten der Ereignisse A und B berechnen lässt, wenn die Ereignisse A und B unvereinbar sind. Wir wollen jetzt den Fall untersuchen, in dem $A \cap B \neq \emptyset$ ist, in dem also die Ereignisse A und B gleichzeitig auftreten können. In diesem Fall zerlegen wir die Menge $A \cup B$ in drei Teilmengen:

$$A \cup B = (A \backslash B) \uplus (B \backslash A) \uplus (A \cap B)$$

Da die drei Mengen $A \backslash B$, $B \backslash A$ und $A \cap B$ paarweise disjunkt sind, gilt

$$P(A \cup B) = P(A \backslash B) + P(B \backslash A) + P(A \cap B) \tag{2.1}$$

Außerdem gilt

$$A = (A \backslash B) \uplus (A \cap B) \quad \text{und} \quad B = (B \backslash A) \uplus (A \cap B).$$

Daraus folgt sofort

$$P(A) = P(A \backslash B) + P(A \cap B) \qquad \text{und} \qquad P(B) = P(B \backslash A) + P(A \cap B).$$

Subtrahieren wir auf beiden Seiten dieser Gleichungen den Term $P(A \cap B)$, so erhalten wir

$$P(A) - P(A \cap B) = P(A \backslash B) \qquad \text{und} \qquad P(B) - P(A \cap B) = P(B \backslash A).$$

Aus Gleichung (2.1) folgt nun

$$\begin{aligned} P(A \cup B) &= P(A \backslash B) + P(B \backslash A) + P(A \cap B) \\ &= \big(P(A) - P(A \cap B)\big) + \big(P(B) - P(A \cap B)\big) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned} \tag{2.2}$$

Es gibt eine ganz analoge Formel zur Berechnung der Anzahl der Elemente einer Menge. Bezeichnen wir für eine Menge M die Anzahl ihrer Elemente mit $|M|$, so gilt

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Diese Formel lässt sich wie folgt interpretieren: Wenn wir zuerst die Elemente von A zählen und anschließend die Elemente von B zählen, so zählen wir die Elemente der Schnittmenge $A \cap B$ doppelt und müssen daher die Anzahl dieser Elemente abziehen.

Die Gleichung (2.2) lässt sich verallgemeinern. Betrachten wir die Vereinigung dreier Mengen A, B und C so finden wir

$$\begin{aligned}
& P(A \cup B \cup C) \\
= \;& P\big((A \cup B) \cup C\big) \\
= \;& P(A \cup B) + P(C) - P\big((A \cup B) \cap C\big) \\
& \text{Gleichung (2.2) auf } A \cup B \text{ anwenden} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - P\big((A \cup B) \cap C\big) \\
& \text{Distributiv-Gesetz } (A \cup B) \cap C = (A \cap C) \cup (B \cap C) \text{ berücksichtigen} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - P\big((A \cap C) \cup (B \cap C)\big) \\
& \text{Gleichung (2.2) auf } (A \cap C) \cup (B \cap C) \text{ anwenden} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - \big(P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C))\big) \\
& (A \cap C) \cap (B \cap C) = A \cap B \cap C \text{ berücksichtigen} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - \big(P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)\big) \\
& \text{Klammer auflösen und umsortieren} \\
= \;& P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)
\end{aligned}$$

In ähnlicher Weise lässt sich eine Formel herleiten, die die Wahrscheinlichkeit einer Vereinigung von n Ereignissen angeben.

Satz 3 (Siebformel von Poincaré und Sylvester) *Ist* $[A_1, A_2, \cdots, A_n]$ *eine Liste von Ereignissen, so gilt*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \cdot \sum_{\substack{I \subseteq \{1,\cdots,n\} \\ |I|=k}} P\left(\bigcap_{i \in I} A_i\right).$$

Aufgabe 2: Beweisen Sie die Siebformel von Poincaré und Sylvester durch vollständige Induktion. ◇

Ist ein Wahrscheinlichkeits-Raum $\langle \Omega, 2^{\Omega}, P \rangle$ gegeben, so definieren wir für ein Ereignis $A \in 2^{\Omega}$ das Komplement von A als

$$A^{\mathrm{c}} = \Omega \backslash A.$$

Wegen $A \uplus A^{\mathrm{c}} = \Omega$ und $P(\Omega) = 1$ gilt dann

$$P(A) + P(A^{\mathrm{c}}) = 1 \quad \text{und daraus folgt} \quad P(A^{\mathrm{c}}) = 1 - P(A).$$

Für das Komplement von Mengen gelten die beiden folgenden De Morganschen Gesetze:

$$(A \cup B)^{\mathrm{c}} = A^{\mathrm{c}} \cap B^{\mathrm{c}} \quad \text{und} \quad (A \cap B)^{\mathrm{c}} = A^{\mathrm{c}} \cup B^{\mathrm{c}}.$$

Diese Gesetze lassen sich wie folgt verallgemeinern:

$$\left(\bigcup_{i=1}^{n} A_i\right)^{\mathrm{c}} = \bigcap_{i=1}^{n} A_i^{\mathrm{c}} \quad \text{und} \quad \left(\bigcap_{i=1}^{n} A_i\right)^{\mathrm{c}} = \bigcup_{i=1}^{n} A_i^{\mathrm{c}}$$

Gelegentlich ist es so, dass es schwer fällt, die Wahrscheinlichkeit für ein gegebenes Ereignis A unmittelbar zu berechnen. Manchmal ist es in solchen Fällen möglich, stattdessen die Wahrscheinlichkeit für das Ereignis A^{c} zu berechnen. Die gesuchte Wahrscheinlichkeit $P(A)$ ergibt sich dann nach der Formel

$P(A) = 1 - P(A^c)$. Die nächste Aufgabe liefert ein Beispiel für eine solche Situation.

Aufgabe 3: An einer Schule können die Schüler Spanisch, Englisch und Französisch lernen.

(a) 40% der Schüler lernen Spanisch.

(b) 60% der Schüler lernen Englisch.

(c) 55% der Schüler lernen Französisch.

(d) 30% der Schüler lernen Spanisch und Englisch.

(e) 20% der Schüler lernen Spanisch und Französisch.

(f) 35% der Schüler lernen Französisch und Englisch.

(g) 10% der Schüler lernen Spanisch, Französisch und Englisch.

Wie groß ist der Prozentsatz der Schüler, die überhaupt keine Fremdsprache lernen?

Lösung: Wir führen folgende Bezeichnungen ein.

1. S: Menge der Schüler, die Spanisch lernt.
2. F: Menge der Schüler, die Französisch lernt.
3. E: Menge der Schüler, die Englisch lernt.
4. K: Menge der Schüler, die keine Fremdsprache lernt.

Dann gilt

$$\begin{aligned} P(K) &= P\big((S \cup F \cup E)^c\big) \\ &= 1 - P(S \cup F \cup E) \\ &= 1 - P(S) - P(F) - P(E) + P(S \cap F) + P(S \cap E) + P(F \cap E) - P(S \cap E \cap F) \\ &= 1 - 0.4 - 0.55 - 0.6 + 0.2 + 0.3 + 0.35 - 0.1 \\ &= 0.2 \end{aligned}$$

Also lernen 20% der Schüler überhaupt keine Fremdsprache. □

Aufgabe 4: Zur Ermittelung der Noten würfelt ein Lehrer mit zwei Laplace-Würfeln. Die Note ergibt sich dann als das Minimum der gewürfelten Augenzahlen.

(a) Geben Sie den Wahrscheinlichkeits-Raum für dieses Zufalls-Experiment an.

(b) Geben Sie die für die Notenvergabe relevanten Ereignisse an.

Lösung: Wir definieren

$$\Omega = \big\{\langle i, j\rangle \mid i \in \{1, 2, 3, 4, 5, 6\},\ j \in \{1, 2, 3, 4, 5, 6\}\big\}$$

Da es sich um zwei Laplace-Würfel handelt, sind alle Ereignisse gleich wahrscheinlich. Da $|\Omega| = 36$ ist, hat das Wahrscheinlichkeits-Maß für eine beliebige Menge $A \in 2^\Omega$ den Wert

$$P(A) = \frac{1}{36} \cdot |A|.$$

Die für die Notenvergabe interessanten Ereignisse sind:

1. Vergabe einer 1:
$$A_1 = \big\{\langle 1, n\rangle \mid n \in \{1, \cdots, 6\}\big\} \cup \big\{\langle n, 1\rangle \mid n \in \{1, \cdots, 6\}\big\}$$
Die Menge A_1 besteht aus $6 + 6 - 1 = 11$ Elementen, denn
$$\big\{\langle 1, n\rangle \mid n \in \{1, \cdots, 6\}\big\} \cap \big\{\langle n, 1\rangle \mid n \in \{1, \cdots, 6\}\big\} = \big\{\langle 1, 1\rangle\big\}.$$
Also hat die Wahrscheinlichkeit für die Note 1 den Wert $\frac{11}{36} = 0.30\overline{5}$.

2. Vergabe einer 2:
$$A_2 = \big\{\langle 2, n\rangle \mid n \in \{2, \cdots, 6\}\big\} \cup \big\{\langle n, 2\rangle \mid n \in \{2, \cdots, 6\}\big\}$$
Die Menge A_2 besteht aus $5 + 5 - 1 = 9$ Elementen. Also hat die Wahrscheinlichkeit für die Note 2 den Wert $\frac{9}{36} = 0.25$.

3. Vergabe einer 3:
$$A_3 = \big\{\langle 3, n\rangle \mid n \in \{3, 4, 5, 6\}\big\} \cup \big\{\langle n, 3\rangle \mid n \in \{3, 4, 5, 6\}\big\}$$
Die Menge A_3 besteht aus $4 + 4 - 1 = 7$ Elementen. Also hat die Wahrscheinlichkeit für die Note 3 den Wert $\frac{7}{36} = 0.19\overline{4}$.

4. Vergabe einer 4:
$$A_4 = \big\{\langle 4, n\rangle \mid n \in \{4, 5, 6\}\big\} \cup \big\{\langle n, 4\rangle \mid n \in \{4, 5, 6\}\big\}$$
Die Menge A_4 besteht aus $3 + 3 - 1 = 5$ Elementen. Also hat die Wahrscheinlichkeit für die Note 4 den Wert $\frac{5}{36} = 0.13\overline{8}$.

5. Vergabe einer 5:
$$A_5 = \big\{\langle 5, n\rangle \mid n \in \{5, 6\}\big\} \cup \big\{\langle n, 5\rangle \mid n \in \{5, 6\}\big\}$$
Die Menge A_5 besteht aus $2 + 2 - 1 = 3$ Elementen. Also hat die Wahrscheinlichkeit für die Note 5 den Wert $\frac{3}{36} = 0.08\overline{3}$.

6. Vergabe einer 6:
$$A_6 = \big\{\langle 6, 6\rangle\big\}$$
Die Menge A_6 enthält offenbar nur ein Element. Also hat die Wahrscheinlichkeit für die Note 6 den Wert $\frac{1}{36} = 0.02\overline{7}$. □

Aufgabe 5: Der Lehrer aus der letzten Aufgabe stellt fest, dass sich bei seinem Verfahren ein zu guter Notendurchschnitt ergibt. Daher ändert er das Verfahren ab: Zur Ermittelung der Note wird jetzt die Summe der Augenzahlen durch zwei geteilt. Falls das Ergebnis keine ganze Zahl ist, wird aufgerundet. Geben Sie die für die Notenvergabe relevanten Ereignisse an und berechnen Sie die Wahrscheinlichkeiten für das Auftreten dieser Ereignisse. □

2.3 Kombinatorik

Die letzten beiden Aufgaben zeigen, dass die Berechnung von Wahrscheinlichkeiten oft auf die Berechnung der Anzahl der Elemente einer Menge zurück geführt werden kann. Dies liegt daran, dass es sich bei den zu Grunde liegenden Zufalls-Experimenten jeweils um ein Laplace-Experiment handelt. Bei allen Laplace-Experimenten läuft die Berechnung der Wahrscheinlichkeiten auf die Bestimmung der Anzahl der Elemente einer Menge heraus. Diese Anzahl nennen wir auch die Kardinalitäten der Menge. Wir werden jetzt die Kardinalitäten für diejenigen Mengen-Typen, die in der Praxis häufig vorkommen, berechnen.

Wir nehmen als erstes an, dass wir n Mengen A_1, A_2, $\cdots$, A_n gegeben haben. Wir untersuchen die Menge M aller Listen der Länge n, für die das i-te Element ein Element der Menge M_i ist. Die

Menge M ist nichts anderes als das kartesische Produkt der Mengen A_1, A_2, $\cdots$, A_n, es gilt also

$$M = A_1 \times A_2 \times \cdots \times A_n = \big\{ [x_1, \cdots, x_n] \bigm| \forall i \in \{1, \cdots, n\} : x_i \in A_i \big\}.$$

Für die Mächtigkeit dieser Menge gilt

$$|M| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_n| = \prod_{i=1}^{n} |A_i|, \tag{2.3}$$

denn für das erste Element einer Liste gibt es $|A_1|$ Möglichkeiten, für das zweite Element gibt es $|A_2|$ Möglichkeiten und für das letzte Element gibt es $|A_n|$ Möglichkeiten. Da die einzelnen Möglichkeiten beliebig kombiniert werden können, ist die Gesamtzahl aller Möglichkeiten durch das Produkt gegeben. Die Gleichung (2.3) bezeichnen wir daher als Produkt-Regel.

Anzahl der k-Tupel mit Wiederholung: Ist M eine Menge der Mächtigkeit n, so gibt es nach der Produkt-Regel insgesamt n^k Möglichkeiten, ein Tupel der Länge k mit Elementen aus M zu bilden, denn es gilt

$$\big|M^k\big| = |M|^k = n^k. \tag{2.4}$$

Anzahl der k-Tupel ohne Wiederholung: Oft sind nur solche k-Tupel interessant, die kein Element mehrfach enthalten. Solche k-Tupel bezeichnen wir als Permutationen. Ist M eine Menge und $k \in \mathbb{N}$, so definieren wir die Menge der k-Permutationen aus M als

$$P(M, k) = \big\{ [x_1, \cdots, x_k] \in M^k \bigm| i \neq j \rightarrow x_i \neq x_j \big\}.$$

Gilt $|M| = n$, so haben wir für das Element x_1 insgesamt n Möglichkeiten. Für das zweite Element x_2 einer Permutation haben wir eine Möglichkeit weniger, also nur noch $n-1$ Möglichkeiten der Auswahl. Allgemein haben wir für das i-te Element x_i nur noch $n - (i - 1)$ Möglichkeiten, denn wir haben ja schon $i - 1$ Elemente vorher ausgewählt. Damit ergibt sich

$$\big|P(M, k)\big| = n \cdot (n-1) \cdot \cdots \cdot \big(n - (k-1)\big) = \frac{n!}{(n-k)!}. \tag{2.5}$$

Wenn wir für k den Wert n einsetzen, ergibt sich als Spezialfall die Formel

$$\big|P(M, n)\big| = n!.$$

Damit gibt es insgesamt $n!$ Möglichkeiten um die Elemente einer Menge der Mächtigkeit n so in einer Liste anzuordnen, dass jedes Element der Menge genau einmal in der Liste auftritt.

Anzahl der k-Kombinationen ohne Wiederholung: Wir bestimmen jetzt die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Dazu definieren wir $C(M, k)$ als die Menge aller k-elementigen Teilmengen von M:

$$C(M, k) = \big\{ N \in 2^M \bigm| |N| = k \big\}.$$

Diese Teilmengen bezeichnen wir auch als die k-elementigen Kombinationen der Menge M. Um die Mächtigkeit von $C(M, k)$ zu bestimmen, überlegen wir uns, wie die Mengen $P(M, k)$ und $C(M, k)$ zusammenhängen. Ist eine Kombination

$$\{x_1, x_2, \cdots, x_k\} \in C(M, k)$$

gegeben, so gibt es $k!$ Möglichkeiten, um diese k Elemente in einem Tupel anzuordnen. Daher erhalten wir die Anzahl der k-Permutationen aus der Anzahl der k-Kombinationen durch Multiplikation mit der Zahl $k!$ und sehen, dass

$$|P(M, k)| = |C(M, k)| \cdot k!$$

gilt. Setzen wir hier den Wert ein, den wir in Gleichung (2.5) für die Anzahl der k-Permutationen gefunden haben, so erhalten wir

$$\frac{n!}{(n-k)!} = |C(M,k)| \cdot k!.$$

Division dieser Gleichung durch $k!$ liefert

$$|C(M,k)| = \frac{n!}{k! \cdot (n-k)!}. \tag{2.6}$$

Der Ausdruck, den wir hier auf der rechten Seite der Gleichung erhalten haben, tritt in der Mathematik an vielen Stellen auf. Wir definieren daher für $n \in \mathbb{N}$ und $k \in \{0, \cdots, n\}$

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

Der Ausdruck $\displaystyle\binom{n}{k}$ wird als n über k gelesen und als Binomial-Koeffizient bezeichnet. Falls $k > n$ ist, wollen wir zusätzlich definieren

$$\binom{n}{k} := 0 \quad \text{falls } k > n,$$

denn es ist nicht möglich, aus einer n-elementigen Menge eine Teilmenge mit k Elementen auszuwählen, falls $k > n$ ist. Der Ausdruck $\binom{n}{k}$ genügt der folgenden Rekursions-Gleichung:

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

Beweis: Wir wollen uns anschaulich überlegen, warum diese Gleichung wahr ist. Die Zahl auf der rechten Seite der obigen Gleichung gibt an, auf wieviel verschiedene Möglichkeiten wir aus einer Menge von $n+1$ Personen ein Team von $k+1$ Personen auswählen können. Wir nehmen nun an, dass sich unter den $n+1$ Personen genau eine Frau befindet. Dann gibt es zwei Möglichkeiten:

1. Die Frau wird für das Team ausgewählt. Dann müssen wir dann aus den verbleibenden n Personen noch k Personen auswählen. Dafür gibt es $\displaystyle\binom{n}{k}$ Möglichkeiten.
2. Die Frau wird nicht für das Team ausgewählt. In diesem Fall müssen wir aus den n Männern genau $k+1$ Männer auswählen. Dafür gibt es $\displaystyle\binom{n}{k+1}$ Möglichkeiten.

Damit gibt es insgesamt

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

Möglichkeiten, um aus $n+1$ Personen ein Team von $k+1$ Personen auszuwählen. Diese Zahl muss folglich gleich $\displaystyle\binom{n+1}{k+1}$ sein.

Aufgabe 6: Beweisen Sie die Gleichung

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

auf algebraischem Wege indem Sie für $\displaystyle\binom{n}{k}$ den Wert $\displaystyle\frac{n!}{k! \cdot (n-k)!}$ einsetzen. $\diamond$

Der Name "Binomial-Koeffizient" für den Ausdruck $\binom{n}{k}$ rührt daher, dass dieser Ausdruck in dem Binomischen Lehrsatz auftritt.

Satz 4 (Binomischer Lehrsatz, Alessandro Binomi) *Ist $n \in \mathbb{N}$ und sind $x, y \in \mathbb{R}$, so gilt*

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

Beweis: Falls wir die Rekursions-Gleichung $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ verwenden, kann der binomische Lehrsatz ohne große Mühe durch vollständige Induktion nach n bewiesen werden. Ein solcher Beweis liefert aber keine Erklärung dafür, warum in dem binomischen Lehrsatz der Ausdruck $\binom{n}{k}$ auftritt. Wir geben daher einen anderen Beweis. Dazu betrachten wir das Produkt

$$(x_1 + y_1) \cdot (x_2 + y_2) \cdot \dots \cdot (x_n + y_n) = \prod_{k=1}^{n} (x_k + y_k).$$

Beachten Sie, dass dieses Produkt mit dem Ausdruck $(x + y)^n$ übereinstimmt, wenn

$$x_1 = x_2 = \dots = x_n = x \quad \text{und} \quad y_1 = y_2 = \dots = y_n = y$$

gilt. Um das obige Produkt auszumultiplizieren, müssen wir alle Produkte aus der Menge

$$\bigl\{ z_1 \cdot z_2 \cdot \dots \cdot z_n \bigm| \forall i \in \{1, \cdots, n\} : z_i \in \{x_i, y_i\} \bigr\}$$

bilden, wobei wir für z_i jeweils einen der Werte x_i oder y_i einsetzen. Betrachten wir nun alle Möglichkeiten ein Produkt auszuwählen, bei denen k der z_i den Wert x_i haben. Dies sind $\binom{n}{k}$ Möglichkeiten, denn wir müssen aus der Menge der n Indizes $\{1, \cdots, n\}$ insgesamt k Indizes auswählen. Nehmen wir jetzt an, dass $x_1 = x_2 = \dots = x_n = x$ und $y_1 = y_2 = \dots = y_n = y$ ist, so hat das entsprechende Produkt die Form $x^k \cdot y^{n-k}$, denn alle $n - k$ der z_i, die nicht gleich x_i sind, müssen gleich y_i sein. Zusammen liefern diese Produkte den Term

$$\binom{n}{k} \cdot x^k \cdot y^{n-k}.$$

Wenn wir diese Terme aufsummieren, erhalten wir den binomischen Lehrsatz. □

Satz 5 (Vandermonde-Gleichung, Alexandre-Théophile Vandermont, 1735-1796)
Sind m, n und k natürliche Zahlen, so gilt

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i} \cdot \binom{n}{k-i}.$$

Beweis: Zwar kann die Vandermonde-Gleichung algebraisch bewiesen werden, aber es ist instruktiver, wenn wir versuchen zu verstehen, was die beiden Seiten der Gleichung eigentlich aussagen. Die linke Seite der Gleichung gibt an, wie viele Möglichkeiten es gibt, aus einer Menge mit $m + n$ Personen ein Team von k Personen auszuwählen. Wir nehmen nun an, dass die ersten m Personen männlich sind, während es sich bei den restlichen n Personen um Frauen handelt. Zur Vereinfachung betrachten wir zunächst nur den Fall, dass sowohl m als auch n mindestens so groß wie k ist. Das Team aus k Personen, das wir auswählen, enthält i Männer, wobei i eine Zahl von 0 bis maximal k ist. Nehmen wir an das Team besteht aus i Männern und folglich $k - i$ Frauen. Dann gibt es $\binom{m}{i}$ Möglichkeiten, die i Männer aus der Menge von insgesamt m Männern auswählen und außerdem $\binom{n}{k-i}$ Möglichkeiten, um die $k - i$ Frauen aus den insgesamt n Frauen auszuwählen. Das Produkt

$$\binom{m}{i} \cdot \binom{n}{k-i}$$

gibt also an, wie viele Möglichkeiten es gibt, aus einer Gruppe, die aus m Männern und n Frauen besteht, ein Team zusammen zu stellen, dass aus i Männern und $k - i$ Frauen besteht. Summieren wir diese Möglichkeiten für alle in Betracht kommenden Werte von i auf, so haben wir insgesamt die Anzahl der Möglichkeiten, aus $m + n$ Personen ein Team von k Personen zu bilden.

Beachten Sie, dass die obige Argumentation auch dann noch richtig ist, wenn m oder n kleiner als k ist. Das liegt daran, dass beispielsweise $\binom{m}{k} = 0$ ist wenn $m < k$ gilt. □

Aufgabe 7: Zeigen Sie die folgenden Gleichungen indem Sie darlegen, wie die beiden Seiten der jeweiligen Gleichungen als Auswahl von Mengen interpretiert werden können.

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

   Bemerkung: Diese Gleichung wird als Symmetrie-Eigenschaft der Binomial-Koeffizienten bezeichnet.

2. $(n+1) \cdot \binom{n}{k} = (k+1) \cdot \binom{n+1}{k+1}$

   Hinweis: Nehmen Sie an, dass Sie aus einer Gruppe von $n+1$ Personen ein Projektteam von $k+1$ Personen zusammen stellen sollen. Zusätzlich müssen Sie eines der Teammitglieder zum Projekleiter ernennen.

3. $\sum\limits_{i=1}^{n} i = \binom{n+1}{2}$

   Hinweis: Stellen Sie sich vor, Sie veranstalten eine Party mit $n+1$ Gästen, bei der jeder Gast jedem anderen Gast die Hand gibt und stellen Sie sich die Frage, wie viele Hände insgesamt geschüttelt werden. $\diamond$

Aufgabe 8: In einem Swinger-Club besuchen n Paare eine Veranstaltung. Die n Herren werden nummeriert und für jeden Herren wird ein Chip mit der Nummer dieses Herren in einen Hut gelegt. Anschließend ziehen die Damen eine Nummer aus dem Hut. Diese Nummer legt fest, mit welchem der Herren die Damen den Rest des Abends verbringen. Die Damen wären sehr enttäuscht, wenn Sie die Nummer Ihres Gemahls ziehen würden, denn dann hätten Sie ja auch gleich zu Hause bleiben können. Wie groß ist die Wahrscheinlichkeit, dass keine der Damen die Nummer Ihres Gemahls zieht? Untersuchen Sie außerdem, ob diese Zahl für große n gegen einen Grenzwert strebt.

Hinweis: Benutzen Sie die Siebformel von Poincaré und Sylvester.

Lösung: Wir überlegen uns zunächst, wie die Ergebnis-Menge Ω für unser Zufalls-Experiment aussieht. Wir wollen die Ziehung aller Chips durch die Damen als ein Ergebnis des Zufalls-Experiments auffassen. Nummerieren wir die Damen analog zu den Herren, so können wir ein solches Ergebnis als eine Liste der Zahlen von 1 bis n darstellen. Beispielsweise würde die Liste

$$[2, 3, 1]$$

festlegen, dass die erste Dame den Abend mit dem Herren Nummer 2 verbringt, die zweite Dame verbringt den Abend mit dem Herren Nummer 3 und die dritte Dame verbringt den Abend mit dem Herren Nummer 1. Wir definieren also

$$\Omega = \big\{[x_1, \cdots, x_n] \;\big|\; \forall i \in \{1, \cdots, n\} : x_i \in \{1, \cdots, n\}\big\}.$$

Wir haben oben gesehen, dass es insgesamt $n!$ Listen gibt, in denen die Zahlen von 1 bis n jeweils einmal auftreten, es gilt also

$$|\Omega| = n!.$$

Weiter definieren wir für $k = 1, \cdots, n$ das Ereignis A_k als das Ereignis, bei dem die k-te Dame Glück hat und nicht die Nummer des k-ten Herren aus dem Hut zieht. Es gilt dann

$$A_k := \{L \in \Omega \,|\, L[k] \neq k\}.$$

Das Ereignis E, für das wir uns interessieren, ist das Ereignis, bei dem alle Damen Glück haben, es gilt also

$$E = \big\{L \in \Omega \;\big|\; \forall k \in \{1, \cdots, n\} : L[k] \neq k\big\} = \bigcap_{k=1}^{n} A_k.$$

Da nicht zu sehen ist, wie $P(E)$ unmittelbar berechnet werden kann, betrachten wir zunächst das Komplementär-Ereignis

$$E^{\mathrm{c}} = \Omega \backslash E$$

und berechnen $P(E)$ dann nach der Formel

$$P(E) = 1 - P(E^c).$$

Nach De Morgan gilt

$$E^c = \Omega \backslash \bigcap_{k=1}^{n} A_k = \bigcup_{k=1}^{n} A_k^c.$$

Hier ist A_k^c das Ereignis, bei dem die k-te Dame Pech hat, es gilt also

$$A_k^c = \{L \in \Omega \mid L[k] = k\}.$$

Da wir E^c als Vereinigung von Mengen dargestellt haben, erfolgt die Berechnung von $P\big(E^c\big)$ über die Formel von Poincaré und Sylvester. Also gilt

$$P\big(E^c\big) = \sum_{j=1}^{n} (-1)^{j+1} \cdot \sum_{\substack{I \subseteq \{1,\cdots,n\} \\ |I|=j}} P\left(\bigcap_{k\in I} A_k^c\right).$$

Ist I eine Teilmenge der Menge $\{1, \cdots, n\}$, so ist die Menge

$$\bigcap_{k\in I} A_k^c$$

gerade das Ereignis, bei dem alle Damen aus der Menge I die Nummer Ihres Gemahls aus dem Hut ziehen, es gilt

$$\bigcap_{k\in I} A_k^c = \big\{L \in \Omega \;\big|\; \forall k \in I : L[k] = k\big\} =: B_I.$$

Wenn wir die Kardinalität der Menge B_I bestimmen wollen, dann müssen wir uns überlegen, wie viele Möglichkeiten wir haben, um die Elemente der Listen L festzulegen. Falls die Menge I aus $j := |I|$ Elementen besteht, haben wir also noch $n-j$ Elemente, die wir auf die restlichen $n-j$ Listen-Positionen verteilen können. Da jedes Element genau einmal vorkommen darf, gibt es dafür $(n-j)!$ verschiedene Möglichkeiten. Also haben wir

$$|B_I| = (n-j)! \quad \text{mit } j = |I|.$$

Als nächstes müssen wir uns überlegen, wie viele Möglichkeiten es gibt, um eine Menge $I \subseteq \{1, \cdots, n\}$ mit $|I| = j$ zu bilden: Dies ist die Anzahl der Teilmengen der Menge $\{1, \cdots, n\}$, die aus j verschiedenen Elementen besteht. Nach dem, was wir in dieser Vorlesung bisher gezeigt haben, gibt es dafür $\binom{n}{j}$ Möglichkeiten. Da $|\Omega| = n!$ ist, haben wir damit insgesamt:

$$\begin{aligned}
P(E^c) &= \sum_{j=1}^{n} (-1)^{j+1} \cdot \sum_{\substack{I \subseteq \{1,\cdots,n\} \\ |I|=j}} P\left(\bigcap_{k\in I} A_k^c\right) \\
&= \sum_{j=1}^{n} (-1)^{j+1} \cdot \binom{n}{j} \cdot \frac{(n-j)!}{n!} \\
&= \sum_{j=1}^{n} (-1)^{j+1} \cdot \frac{n!}{j! \cdot (n-j)!} \cdot \frac{(n-j)!}{n!} \\
&= \sum_{j=1}^{n} (-1)^{j+1} \cdot \frac{1}{j!}
\end{aligned}$$

Damit finden wir

$$P(E) = 1 - P(E^c) = 1 - \sum_{j=1}^{n} (-1)^{j+1} \cdot \frac{1}{j!} = \frac{(-1)^0}{0!} + \sum_{j=1}^{n} \frac{(-1)^j}{j!} = \sum_{j=0}^{n} \frac{(-1)^j}{j!}.$$

Interessant ist, wie sich $P(E)$ verhält, wenn n groß wird. Aus der Analysis kennen wir die Exponential-

Funktion

$$\exp(x) = \sum_{j=0}^{\infty} \frac{x^n}{n!}$$

und wissen, dass $\exp(x) = \mathrm{e}^x$ ist. Damit sehen wir

$$\lim_{n\to\infty} P(E) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} = \exp(-1) = \mathrm{e}^{-1} = \frac{1}{\mathrm{e}}.$$

Wir stellen fest, dass die Eulersche Zahl e in der Natur allgegenwärtig ist. ◇

Anzahl der k-Kombinationen mit Wiederholung: Als letztes stellen wir uns die Frage, wie viele Möglichkeiten es gibt, aus einer Menge M, die aus n Elementen besteht, eine Multimenge auszuwählen, die aus insgesamt k Elementen besteht. Den Begriff der Multimenge müssen wir zunächst formal definieren.

Definition 6 (Multimenge) *Eine Multimenge ist eine Zusammenfassung von Objekten, für die folgendes gilt:*

1. *Die Reihenfolge der Elemente in einer Multimenge spielt keine Rolle.*

   In dieser Hinsicht unterscheiden sich Multimengen also nicht von Mengen.

2. *In einer Multimenge können Objekte auch mehrfach auftreten.*

   Dieser Aspekt ist der wesentliche Unterschied zwischen Mengen und Multimengen, den eine Menge enthält jedes Objekt höchstens einmal. ◇

Multimengen werden ähnlich wie Mengen mit Hilfe von geschweiften Klammern notiert. Zur Unterscheidung zwischen Mengen und Multimengen setzen wir den Index $_\mathrm{m}$ an die schließende Klammer, wir schreiben also beispielsweise

$$A = \{1, 2, 2, 2, 2, 3\}_\mathrm{m}$$

um auszudrücken, dass A eine Multimenge ist, die die Elemente 1, 2 und 3 enthält und bei der außerdem das Element 2 mit der Vielfachheit 4 auftritt. Die Mächtigkeit von A ist dann 6. Analog benutzen wir für die Vereinigung von Multimengen den Operator $\cup_\mathrm{m}$, es gilt also beispielsweise

$$\{1, 2, 2, 3\}_\mathrm{m} \cup_\mathrm{m} \{1, 2, 3, 4, 4\}_\mathrm{m} = \{1, 1, 2, 2, 2, 3, 3, 4, 4\}_\mathrm{m}$$

Mit dieser Definition der Vereinigung von Multimengen gilt offenbar

$$|A \cup_\mathrm{m} B| = |A| + |B| \quad \text{für beliebige Multimengen } A \text{ und } B,$$

denn wenn ein Element sowohl in A als auch in B vorkommt, dann kommt es in der Multimenge $A \cup_\mathrm{m} B$ mehrfach vor. Die Menge aller Multimengen der Mächtigkeit k mit Elementen aus der Menge M bezeichnen wir mit

$$C_\mathrm{m}(M, k).$$

Da wir nur an der Anzahl der Elemente der Menge $C_\mathrm{m}(M, k)$ interessiert sind, können wir O.B.d.A. annehmen, dass die Menge M aus den ersten n natürlichen Zahlen besteht, es gilt also

$$M = \{1, 2, 3, \cdots, n\}.$$

Wir müssen uns überlegen, wie viele Multimengen mit k Elementen wir aus den Elementen der Menge M bilden können. Am einfachsten wird dieses Problem, wenn wir diese Multimengen als geordnete Listen aufschreiben, wobei wir die verschiedenen Zahlen durch senkrechte Striche trennen. Mit dieser Vereinbarung schreibt sich die Multimenge

$$\{1, 2, 2, 4, 4\}_\mathrm{m}$$

als die Liste

$$[1\,|\,2\,2\,|\,|\,4\,4\,4].$$

Beachten Sie, dass wir auch die (nicht vorhandenen) Dreien durch senkrechte Striche von den Zweien und den Vieren getrennt haben und deswegen zwischen den Zweien und Vieren zwei senkrechte Striche sind. Aufgrund der Tatsache, dass die obige Liste geordnet ist, liefern die Zahlen in der Liste keine Information mehr. Wir können daher die Darstellung weiter vereinfachen, wenn wir die Zahlen durch das Symbol $\star$ ersetzen. Zusätzlich lassen wir noch die eckigen Klammern weg. Damit wird die Multimenge $\{1,2,2,4,4\}_{\mathrm{m}}$ also in den String

$$\star|\star\star||\star\star\star$$

überführt. Wir sehen, dass die senkrechten Striche die Sternchen so in Gruppen einteilen, dass wir daraus die ursprüngliche Liste wiederherstellen können, denn die Sternchen der ersten Gruppe entsprechen den Einsen, die Sternchen der zweiten Gruppe entsprechen den Zweien und so weiter. Insgesamt besteht die Liste nun aus $k+n-1$ Zeichen: Zunächst haben wir jedes der k Elemente der Multimenge durch einen Stern ersetzt und dann haben wir zwischen den n Gruppen von Elementen noch jeweils einen senkrechten Strich eingefügt. Wir überlegen uns jetzt, wie viele Strings es gibt, die insgesamt die Länge $k+n-1$ haben und die nur aus den beiden Zeichen $\star$ und $|$ bestehen. Um einen String dieser Art zu spezifizieren reicht es aus, wenn wir wissen, an welchen Positionen sich die $n-1$ senkrechten Striche befinden, welche die Gruppen aus Sternchen trennen, denn an den anderen k Positionen müssen dann ja Sternchen stehen. Die Positionen der senkrechten Striche sind Zahlen aus der Menge $\{1,\cdots,k+n-1\}$, denn der String hat ja insgesamt die Länge $k+n-1$. Der String ist also dann eindeutig festgelegt, wenn wir eine Teilmenge der Menge $\{1,\cdots,k+n-1\}$ mit genau $n-1$ Elementen festlegen, welche die Positionen der senkrechten Striche spezifiziert. Eine Menge mit $k+n-1$ Elementen hat nach dem, was wir früher bewwiesen haben,

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

Teilmengen, die aus $n-1$ Elementen bestehen. Aufgrund der Symmetrie der Binomial-Koeffizienten gilt

$$\binom{k+n-1}{n-1} = \binom{k+n-1}{(k+n-1)-(n-1)} = \binom{n+k-1}{k}.$$

Damit gibt es für eine Menge M mit n Elementen genau

$$|C_{\mathrm{m}}(M,k)| = \binom{n+k-1}{k} \tag{2.7}$$

Multimengen, die Teilmengen von M sind und aus insgesamt k Elementen bestehen. In der Wahrscheinlichkeits-Theorie werden solche Multimengen auch als Kombinationen mit Wiederholung bezeichnet.

Aufgabe 9:

(a) Wie viele Möglichkeiten gibt es, ein Tripel natürliche Zahlen $\langle x,y,z\rangle$ zu wählen, so dass

$$x+y+z=5$$

gilt? Schreiben Sie ein Programm, dass die Menge dieser Tripel explizit berechnet.

(b) Es sei $s\in\mathbb{N}$. Wie viele Möglichkeiten gibt es, n natürliche Zahlen $x_1,x_2,\cdots,x_n$ so zu wählen, dass

$$\sum_{k=1}^{n} x_k = s$$

gilt?

Hinweis: Stellen Sie die Summe durch einen String dar, der aus s Sternchen und $n-1$ senkrechten Strichen besteht. Beispielsweise können Sie die Summe $3+2+0+1$ durch den String $\star\star\star|\star\star||\star$ darstellen.

(c) Wieder sei $t \in \mathbb{N}$. Wie viele Möglichkeiten gibt es, n positive natürliche Zahlen $y_1, y_2, \cdots, y_n$ so zu wählen, dass

$$\sum_{k=1}^{n} y_k = t$$

gilt?

Hinweis: Führen Sie diese Teilaufgabe durch eine geeignete Transformation auf Teil (b) dieser Aufgabe zurück. ◇

Aufgabe 10: Wie viele Möglichkeiten gibt es, mit drei nummerierten Würfeln in der Summe die Zahl 8 zu würfeln? Die Würfel sind nummeriert, damit wir beispielsweise die beiden folgenden Ergebnisse unterscheiden können:

1. Der erste Würfel zeigt eine 3, der zweite Würfel zeigt eine 4 und der dritte Würfel zeigt eine 1.
2. Der erste Würfel zeigt eine 1, der zweite Würfel zeigt eine 3 und der dritte Würfel zeigt eine 4.

Hinweis: Sie können ein Programm schreiben, dass die Menge aller Tripel der Form $\langle x, y, z\rangle$ berechnet, für die einerseits $x, y, z \in \{1, \cdots, 6\}$ und andererseits $x + y + z = 8$ gilt. ◇

Wenn die Anzahl der Würfel wesentlich größer wird als in der letzten Aufgabe, dann kann es schnell sehr aufwendig werden, die Anzahl der Fälle zu berechnen, mit denen eine vorgegebene Zahl als Summe gewürfelt werden kann. Der folgende Satz liefert daher eine explizite Formel, mit der diese Anzahl berechnet werden kann.

Satz 7 (Würfel-Summen-Satz) *Es seien n und s positive natürliche Zahlen. Wir definieren*

$$W(s) := \bigl\{\langle x_1, \cdots, x_n\rangle \in \{1, \cdots, 6\}^n \bigm| x_1 + \cdots + x_n = s\bigr\}.$$

Die Menge $W(s)$ beschreibt gerade die Menge der Möglichkeiten, die es gibt, um mit n Würfeln in der Summe die Zahl s zu würfeln. Dann gilt

$$|W(s)| = \sum_{k=0}^{n} (-1)^k \cdot \binom{n}{k} \cdot \binom{s - 6 \cdot k - 1}{s - 6 \cdot k - n}.$$

Beweis: Wir betrachten zunächst ein etwas einfacheres Problem: Wir nehmen an, dass die Würfel anstatt mit den Zahlen von Eins bis Sechs mit den Zahlen von Null bis Fünf beschriftet sind und in der Summe die Zahl t ergeben. Wir definieren dazu die Menge

$$V(t) := \bigl\{\langle x_1, \cdots, x_n\rangle \in \{0, \cdots, 5\}^n \bigm| x_1 + \cdots + x_n = t\bigr\}$$

und berechnen zunächst $|V(t)|$. Das ursprüngliche Problem kann auf dieses Problem zurück geführt werden, denn zwischen den Mengen $V(s-n)$ und $W(s)$ besteht der Zusammenhang

$$\langle x_1, \cdots, x_n\rangle \in W(s) \;\Leftrightarrow\; \langle x_1 - 1, \cdots, x_n - 1\rangle \in V(s-n),$$

denn wenn wir in dem ursprünglichen Problem von jedem der n Würfel die Zahl 1 abziehen, sind einerseits die Würfel mit den Zahlen von 0 bis 5 beschriftet und andererseits zeigen die Würfel dann in der Summe die Zahl $s - n$. Diese Gleichung zeigt, dass

$$|W(s)| = |V(s-n)|$$

gilt. Wir müssen also in der Lösung des modifizierten Problems später nur für t die Zahl $s-n$ einsetzen und haben dann das ursprüngliche Problem gelöst. Wir definieren

$$B(t) := \bigl\{\langle x_1, \cdots, x_n\rangle \in \mathbb{N}^n \bigm| x_1 + \cdots + x_n = t\bigr\}.$$

Aus der letzten Aufgabe wissen wir, dass es $\binom{t+n-1}{t}$ Möglichkeiten gibt, n Variablen x_1, x_2, $\cdots$, x_n so mit natürlichen Zahlen zu belegen, dass die Summe $x_1 + x_2 + \cdots + x_n = t$ ist, es gilt also

$$|B(t)| = \binom{t+n-1}{t}.$$

Setzen wir hier für t den Wert $s-n$ ein, so sehen wir, dass die Anzahl $|B(t)|$ gerade dem Summand mit dem Index $k=0$ in der zu beweisenden Summen-Formel entspricht. Die Menge $B(t)$ kann im Allgemeinen auch solche Tupel $\langle x_1, \cdots, x_n \rangle$ enthalten, bei denen ein x_i größer als 5 ist. Wir müssen nun die Anzahl der Möglichkeiten berechnen, bei denen die Summe $x_1 + \cdots + x_n = t$ ist, aber wenigstens eine der Variablen x_i echt größer als 5 ist, denn diese Anzahl müssen wir von $|B(t)|$ wieder abziehen. Wir definieren daher für $i = 1, \cdots, n$ die Mengen $A_i(t)$ wie folgt:

$$A_i(t) := \bigl\{ \langle x_1, \cdots, x_n \rangle \in B(t) \bigm| x_i \geq 6 \bigr\}.$$

Die Menge A_i fasst also die Fälle zusammen, bei denen die i-te Variable mindestens 6 ist. Zwischen der gesuchten Menge $V(t)$ und den Mengen $B(t)$ und $A_i(t)$ besteht der folgende Zusammenhang

$$V(t) = B(t) \backslash \bigcup_{i=1}^{n} A_i(t).$$

Da die Mengen $A_i(t)$ Teilmengen der Menge $B(t)$ sind, folgt daraus für die Kardinalität von $V(t)$ die Formel

$$|V(t)| = |B(t)| - \left| \bigcup_{i=1}^{n} A_i(t) \right|$$

Da die verschiedenen Mengen $A_i(t)$ nicht paarweise disjunkt sind, müssen wir die Siebformel von Poincaré und Sylvester verwenden, um die Kardinalität der Vereinigungs-Menge zur berechnen. Es gilt

$$\left| \bigcup_{i=1}^{n} A_i(t) \right| = \sum_{k=1}^{n} (-1)^{k+1} \cdot \sum_{\substack{I \subseteq \{1, \cdots, n\} \\ |I| = k}} \left| \bigcap_{i \in I} A_i(t) \right|.$$

Um für eine gegebene Teilmenge $I \subseteq \{1, \cdots, n\}$ die Kardinalität der Schnitt-Menge

$$A_I(t) := \bigcap_{i \in I} A_i(t)$$

berechnen zu können, überlegen wir uns, dass $A_I(t)$ gerade die Tupel $\langle x_1, \cdots, x_n \rangle \in B(t)$ enthält, für die für alle $i \in I$ die Ungleichung $x_i \geq 6$ gilt, wir haben also

$$A_I(t) := \bigl\{ \langle x_1, \cdots, x_n \rangle \in B(t) \bigm| \forall i \in I : x_i \geq 6 \bigr\}.$$

Diese Menge transformieren wir in die Menge $A'_I(t)$ bei der wir von den Tupel aus $A_I(t)$ in den Komponenten x_i mit $i \in I$ jeweils 6 abziehen. Diese Menge können wir am einfachsten formal definieren, wenn wir vorher die Hilfsfunktion

$$\texttt{ite} : \mathbb{B} \times \mathbb{Z} \times \mathbb{Z}$$

wie folgt definieren:

$$\texttt{ite}(c, x, y) := \begin{cases} x & \text{falls } c = \texttt{true}; \\ y & \text{falls } c = \texttt{false}. \end{cases}$$

Der Name `ite` steht hier als Abkürzung für *if-then-else*.

$$A'_I(t) := \bigl\{ \langle \texttt{ite}(1 \in I, x_1 - 6, x_1), \cdots, \texttt{ite}(n \in I, x_n - 6, x_n \rangle \bigm| \langle x_1, \cdots, x_n \rangle \in A_I(t) \bigr\}.$$

Ist $k := |I|$, so gilt offenbar

$$A'_I(t) = B(t - k \cdot 6)$$

und damit ist klar, dass

$$|A_I(t)| = |B(t - k \cdot 6)| = \binom{t - k \cdot 6 + n - 1}{t - k \cdot 6}$$

gilt. Wir sehen, dass die Kardinalität der Menge $|A_I(t)|$ nicht wirklich von der Index-Menge I sondern nur von der Mächtigkeit $k = |I|$ der Menge I abhängt. Da es insgesamt $\binom{n}{k}$ Möglichkeiten gibt, eine Menge I der Mächtigkeit k als Teilmenge der Menge $\{1, \cdots, n\}$ auszuwählen, haben wir jetzt insgesamt folgende Formel gefunden:

$$\begin{aligned}
|V(t)| &= |B(t)| - \left|\bigcup_{i=1}^{n} A_i(t)\right| \\
&= \binom{t+n-1}{t} - \sum_{k=1}^{n}(-1)^{k+1} \cdot \sum_{\substack{I\subseteq\{1,\cdots,n\}\\ |I|\leq k}} |A_I(t)| \\
&= \binom{t+n-1}{t} - \sum_{k=1}^{n}(-1)^{k+1} \cdot \binom{n}{k} \cdot \binom{t-k\cdot 6+n-1}{t-k\cdot 6} \\
&= \sum_{k=0}^{n}(-1)^{k} \cdot \binom{n}{k} \cdot \binom{t-k\cdot 6+n-1}{t-k\cdot 6}
\end{aligned}$$

Beachten Sie, dass die Zahl $t - k \cdot 6$ negativ werden kann. Für diesen Fall legen wir fest, dass der Binomial-Koeffizient

$$\binom{t-k\cdot 6+n-1}{t-k\cdot 6}$$

den Wert 0 hat. Setzen wir in der Formel für $|V(t)|$ für t den Wert $s - n$ ein, so ergibt sich wegen $|W(s)| = |V(s-n)|$ die Behauptung. □

Aufgabe 11: Es sei eine Gruppe von n Personen gegeben. Berechnen Sie die Wahrscheinlichkeit dafür, dass von diesen n Personen wenigstens zwei Personen am selben Tag Geburtstag haben. Zur Vereinfachung dürfen Sie folgendes annehmen:

1. Keine der Personen hat an einem Schalttag Geburtstag.
2. Für alle anderen Tage ist das Wahrscheinlichkeits-Maß gleichmäßig.
3. Die Geburtstage der verschiedenen Personen sind voneinander unabhängig. Insbesondere gibt es unter den Personen kein Zwillings-Paar.

Wie groß muss die Zahl n sein, damit es sich lohnt darauf zu wetten, dass wenigstens zwei Personen am selben Tag Geburtstag haben.

Lösung: Wir definieren $T := \{1, \cdots, 365\}$. Unser Zufalls-Experiment besteht darin, dass wir eine Liste der Länge n mit Elementen aus T auswählen. Die Menge Ω ist also durch

$$\Omega = \big\{[x_1, \cdots, x_n] \,\big|\, \forall i \in \{1, \cdots, n\} : x_i \in T\big\} = T^n$$

gegeben. Nach dem oben gezeigten gilt also

$$|\Omega| = |T|^n = 365^n.$$

Das uns interessierende Ereignis A_n, dass wenigstens zwei Personen am selben Tag Geburtstag haben, lässt sich nicht unmittelbar angeben. Wir betrachten daher zunächst das komplementäre Ereignis A_n^{c}, das ausdrückt, dass alle n Personen an verschiedenen Tagen Geburtstag haben. Es gilt

$$A_n^{\mathrm{c}} = \big\{[x_1, \cdots, x_n] \,\big|\, \forall i \in \{1, \cdots, n\} : x_i \in T\big) \ \wedge\ \forall i, j \in \{1, \cdots, n\} : \big(i \neq j \rightarrow x_i \neq x_j\big)\big\}$$

Bei der Menge A_n^{c} handelt es sich gerade um die Menge aller n-Permutationen der Menge $T = \{1, \cdots, 365\}$. Daher gilt

$$\big|A_n^{\mathrm{c}}\big| = \frac{365!}{(365-n)!}.$$

Da wir von einem gleichmäßigen Wahrscheinlichkeits-Maß ausgehen, gilt für die Wahrscheinlichkeit des gesuchten Ereignisses A_n

$$P(A_n) = 1 - P(A_n^{\mathrm{c}}) = 1 - \frac{|A_n^{\mathrm{c}}|}{|\Omega|} = 1 - \frac{365!}{(365-n)! \cdot 365^n}.$$

Tragen wir für $n = 1, \cdots, 60$ die Wahrscheinlichkeiten $P(A_n)$ in eine Tabelle ein, so erhalten wir die in Tabelle 2.1 gezeigten Werte. Abbildung 2.1 zeigt diese Werte graphisch. Wir erkennen, dass bereits bei einer Gruppe von 23 Personen die Wahrscheinlichkeit dafür, dass zwei Personen am selben Tag Geburtstag haben, größer als 0.5 ist. Die Wette lohnt sich also ab einer Gruppengröße von 23 Personen. □

Bemerkung: In der Realität gibt es Jahreszeiten, an denen die Wahrscheinlichkeit einer Geburt höher ist als in anderen Jahreszeiten. Das führt dazu, dass die Wahrscheinlichkeit dafür, dass zwei Personen am selben Tag Geburtstag haben, in der Praxis etwas höher als in Tabelle 2.1 angegeben ist.

n	$P(A_n)$	n	$P(A_n)$	n	$P(A_n)$
1	0.0	2	0.0027397260	3	0.0082041659
4	0.0163559124	5	0.0271355737	6	0.0404624837
7	0.0562357031	8	0.0743352924	9	0.0946238339
10	0.1169481777	11	0.1411413783	12	0.1670247888
13	0.1944102752	14	0.2231025120	15	0.2529013198
16	0.2836040053	17	0.3150076653	18	0.3469114179
19	0.3791185260	20	0.4114383836	21	0.4436883352
22	0.4756953077	23	0.5072972343	24	0.5383442579
25	0.5686997040	26	0.5982408201	27	0.6268592823
28	0.6544614723	29	0.6809685375	30	0.7063162427
31	0.7304546337	32	0.7533475278	33	0.7749718542
34	0.7953168646	35	0.8143832389	36	0.8321821064
37	0.8487340082	38	0.8640678211	39	0.8782196644
40	0.8912318098	41	0.9031516115	42	0.9140304716
43	0.9239228557	44	0.9328853686	45	0.9409758995
46	0.9482528434	47	0.9547744028	48	0.9605979729
49	0.9657796093	50	0.9703735796	51	0.9744319933
52	0.9780045093	53	0.9811381135	54	0.9838769628
55	0.9862622888	56	0.9883323549	57	0.9901224593
58	0.9916649794	59	0.9929894484	60	0.9941226609

Tabelle 2.1: Die ersten 60 Werte für das Geburtstags-Problem.

2.4 Die hypergeometrische Verteilung

Aufgabe 12: In einer Ameisen-Kolonie, in der $40\,000$ Ameisen leben, sind $1\,000$ Ameisen mit Farbe markiert worden. Ein Forscher fängt nun zufällig 200 dieser Ameisen. Berechnen Sie die Wahrscheinlichkeit, dass er genau k markierte Tiere fängt. Gehen Sie dabei von einem gleichmäßigen Wahrscheinlichkeits-Maß aus, nehmen Sie also an, dass die Wahrscheinlichkeit dafür, dass eine Ameise gefangen wird, unabhängig von der farblichen Markierung für alle Ameisen gleich groß ist. Tabellieren Sie die Werte für $k = 0, \cdots, 20$ und tragen Sie die gefundenen Werte in einem Diagramm auf.

Lösung: Wir modellieren die Ameisen-Kolonie durch die Menge $M := \{1, \cdots, 40\,000\}$. Die Menge der farblich markierten Ameisen bezeichnen wir mit F. Wir legen willkürlich fest, dass die ersten Tausend Ameisen diejenigen Ameisen sind, die markiert sind. Also gilt

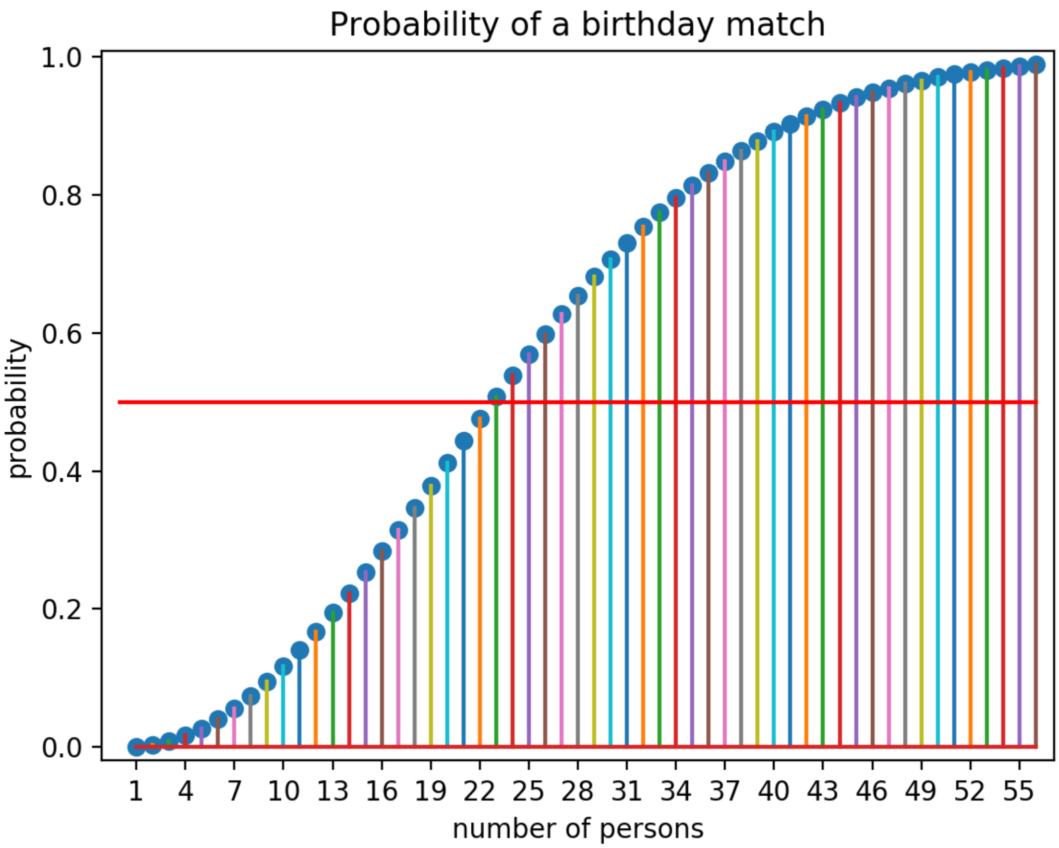

Abbildung 2.1: Das Geburtstags-Problem.

$$F := \{1, \cdots, 1\,000\}$$

Das der Aufgabe zugrunde liegende Zufalls-Experiment besteht darin, dass wir aus der Menge M eine Teilmenge der Größe 200 auswählen. Damit lässt sich die Ergebnis-Menge als

$$\Omega := \left\{A \in 2^M \;\middle|\; |A| = 200\right\}$$

definieren. Ω ist also die Menge aller der Teilmengen von M, die genau 200 Elemente enthalten. Nach Gleichung (2.6) gilt für die Mächtigkeit dieser Menge

$$|\Omega| = |C(M, 200)| = \binom{|M|}{200} = \binom{40\,000}{200}$$

Das uns interessierende Ereignis Λ_k ist das Ereignis, bei dem genau k farblich markierte Ameisen gefangen werden. Wir können daher Λ_k als

$$\Lambda_k := \left\{B \in 2^M \;\middle|\; |B| = 200 \wedge |B \cap F| = k\right\}$$

definieren: Λ_k besteht aus genau den Teilmengen von M, die einerseits 200 Elemente enthalten und die andererseits genau k Elemente aus der Menge $\{1, \cdots, 1\,000\}$ enthalten. Die Berechnung der Wahrscheinlichkeit des Ereignisses Λ_k läuft auf die Berechnung von $|\Lambda_k|$ hinaus. Dazu formen wir die Definition von Λ_k etwas um. Eine Menge B liegt genau dann in Λ_k, wenn B insgesamt k Elemente aus der Menge F und $200 - k$ Elemente aus der Menge $M \backslash F$ enthält. Folglich gilt

$$\begin{aligned}
|\Lambda_k| &= \bigl|\bigl\{C \cup D \bigm| C \subseteq F \wedge D \subseteq M\backslash F \wedge |C| = k \wedge |D| = 200-k\bigr\}\bigr| \\
&= \Bigl|\Bigl\{C \cup D \Bigm| C \in \bigl\{C' \in 2^F \bigm| |C'| = k\bigr\} \;\wedge\; D \in \bigl\{D' \in 2^{M\backslash F} \bigm| |D'| = 200-k\bigr\}\Bigr\}\Bigr| \\
&= \bigl|\{C' \in 2^F \mid |C'| = k\}\bigr| \cdot \bigl|\{D' \in 2^{M\backslash F} \mid |D'| = 200-k\}\bigr| \\
&= \binom{|F|}{k} \cdot \binom{|M\backslash F|}{200-k} \qquad \text{nach Gleichung (2.6)} \\
&= \binom{1000}{k} \cdot \binom{39\,000}{200-k}.
\end{aligned}$$

Damit haben wir für die Wahrscheinlichkeit des Ereignisses Λ_k die Formel

$$P(\Lambda_k) = \frac{|\Lambda_k|}{|\Omega|} = \frac{\binom{1000}{k} \cdot \binom{39\,000}{200-k}}{\binom{40\,000}{200}}$$

gefunden. In Tabelle 2.2 sind die Wahrscheinlichkeiten aufgelistet. Für $k \geq 17$ sinkt die Wahrscheinlichkeit unter 10^{-5} und ist damit vernachlässigbar.

k	$P(\Lambda_k)$	n	$P(\Lambda_k)$	k	$P(\Lambda_k)$
0	0.0062425837	1	0.0321774371	2	0.0824301154
3	0.1399249245	4	0.1770598037	5	0.1781466648
6	0.1484517284	7	0.1053817150	8	0.0650519877
9	0.0354730483	10	0.0173006288	11	0.0076226006
12	0.0030592431	13	0.0011261811	14	0.0003825168
15	0.0001204896	16	0.0000353530	17	0.0000096999

Tabelle 2.2: Wie viele markierte Ameisen werden gefangen?

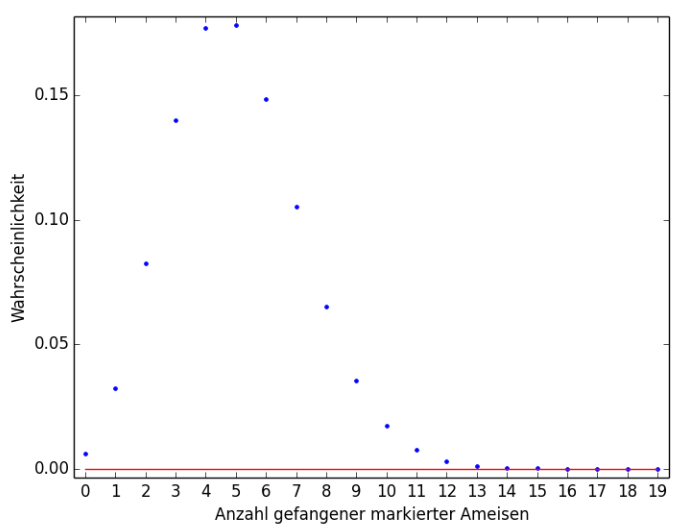

Abbildung 2.2: Die Wahrscheinlichkeit, k markierte Ameisen zu fangen.

Die in der letzten Aufgabe behandelte Situation kommt in der Praxis häufig vor. Ein Beispiel liefert die Überwachung der Qualität der Produktion von elektronischen Bauteilen. Eine Fertigungsanlage produziert am Tag eine bestimmte Zahl N solcher Bauteile. Davon sind K Bauteile defekt, während die restlichen $N-K$ Bauteile funktionieren. Zur Überprüfung der Produktions-Qualität wird eine Stichprobe vom Umfang n genommen. In dieser Stichprobe findet man dann k defekte Bauteile, während die restlichen $n-k$ Bauteile einwandfrei sind. Falls die Zahlen N, K und n gegeben sind, können wir uns fragen, wie wahrscheinlich es ist, dass unter den n Bauteilen der Stichprobe insgesamt k Bauteile defekt sind. Die Situation ist dann dieselbe wie in der letzten Aufgabe:

1. Die Gesamtzahl N entspricht der Anzahl aller Ameisen.
2. Die Anzahl K der defekten Bauteile entspricht der Zahl der markierten Ameisen.
3. Die Umfang n der Stichprobe entspricht der Zahl der gefangenen Ameisen.
4. Die Anzahl k der defekten Bauteile in der Stichprobe entspricht der Anzahl der gefangenen Ameisen, die markiert sind.

Zur Berechnung des Wahrscheinlichkeits-Maßes gehen wir daher wie bei der Lösung der letzten Aufgabe vor und definieren die Menge $\mathcal{M} := \{1, \cdots, N\}$, wobei wir die Bauteile durchnummerieren und also jedes Bauteil durch eine Zahl der Menge $\mathcal{M}$ darstellen. Die Menge der defekten Bauteile bezeichnen wir mit $\mathcal{F}$. Wir gehen ohne Beschränkung der Allgemeinheit davon aus, dass die defekten Bauteile in der Aufzählung aller Bauteile am Anfang stehen, die Bauteile mit den Nummern $1, \cdots, K$ sind also defekt. Folglich gilt

$$\mathcal{F} = \{1, \cdots, K\}.$$

Das der Aufgabe zugrunde liegende Zufalls-Experiment besteht darin, dass wir aus der Menge $\mathcal{M}$ eine Teilmenge der Größe n auswählen. Damit lässt sich die Ergebnis-Menge als

$$\Omega = \big\{A \in 2^{\mathcal{M}} \bigm| |A| = n\big\}$$

Nach Gleichung (2.6) gilt für die Mächtigkeit dieser Menge

$$|\Omega| = |C(\mathcal{M}, n)| = \binom{|\mathcal{M}|}{n} = \binom{N}{n}$$

Das uns interessierende Ereignis Λ_k ist dann durch die Menge

$$\Lambda_k := \big\{B \in 2^{\mathcal{M}} \bigm| |B| = n \wedge |B \cap \mathcal{F}| = k\big\}$$

definiert: Λ_k besteht aus genau den Teilmengen von $\mathcal{M}$, die n Elemente enthalten von denen k defekt sind. Die Berechnung der Wahrscheinlichkeit des Ereignisses Λ_k läuft auf die Berechnung von $|\Lambda_k|$ hinaus. Dazu formen wir die Definition von Λ_k etwas um. Eine Menge B liegt genau dann in Λ_k, wenn B insgesamt k Elemente aus der Menge $\mathcal{F}$ und $n-k$ Elemente aus der Menge $\mathcal{M}\backslash\mathcal{F}$ enthält. Folglich gilt

$$\begin{aligned}
|\Lambda_k| &= \big|\big\{C \cup D \bigm| C \subseteq \mathcal{F} \wedge D \subseteq \mathcal{M}\backslash\mathcal{F} \wedge |C| = k \wedge |D| = n-k\big\}\big| \\
&= \big|\big\{C \cup D \bigm| C \in \{C' \in 2^{\mathcal{F}} \bigm| |C'| = k\} \ \wedge \ D \in \{D' \in 2^{\mathcal{M}\backslash\mathcal{F}} \bigm| |D'| = n-k\}\big\}\big| \\
&= \big|\{C' \in 2^{\mathcal{F}} \bigm| |C'| = k\}\big| \cdot \big|\{D' \in 2^{\mathcal{M}\backslash\mathcal{F}} \bigm| |D'| = n-k\}\big| \\
&= \binom{|\mathcal{F}|}{k} \cdot \binom{|\mathcal{M}\backslash\mathcal{F}|}{n-k} \qquad \text{nach Gleichung (2.6)} \\
&= \binom{n}{k} \cdot \binom{N-K}{n-k}.
\end{aligned}$$

Damit haben wir für das Wahrscheinlichkeits-Maß die Formel

$$P(\Lambda_k) = \frac{|\Lambda_k|}{|\Omega|} = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}} \tag{2.8}$$

gefunden. In der Statistik wird das obige Beispiel abstrahiert. Statt von Bauteilen sprechen wir hier von Kugeln in einer Urne. Von diesen Kugeln sind dann K Kugeln schwarz, was den defekten Bauteilen entspricht, die restlichen Kugeln sind weiß. Dann gibt die obige Formel die Wahrscheinlichkeit dafür an, dass bei einer Entnahme von n Kugeln k Kugeln schwarz sind. Das in Gleichung 2.8 angegebene Wahrscheinlichkeits-Maß wird in der Literatur als hypergeometrische Verteilung bezeichnet.

Aufgabe 13: Eine Firma erhält eine Lieferung von 100 Geräten. Der zuständige Prüfer wählt zufällig 10 Geräte aus. Die Lieferung wird genau dann akzeptiert, wenn dabei kein defektes Gerät gefunden wird. Nehmen Sie an, dass von den gelieferten 100 Geräten 10 Geräte defekt sind. Wie hoch ist die Wahrscheinlichkeit dafür, dass die Lieferung trotzdem akzeptiert wird?

2.5 Die Binomial-Verteilung

Wir greifen das Beispiel aus der Einleitung zur Ameisen-Zählung wieder auf und modifizieren es mit dem Ziel, die Durchführung zu vereinfachen:

1. Am ersten Tag markieren wir wie vorher insgesamt $1\,000$ Ameisen farblich.
2. Der zweite Tag ist schwieriger, denn hier müssten wir tatsächlich erst 200 Ameisen einsammeln bevor wir mit dem Zählen der markierten Ameisen beginnen. Also ändern wir das Experiment so ab, dass wir nacheinander 200 Ameisen untersuchen und jedes Mal überprüfen, ob die Ameise markiert ist.

Am zweiten Tag kann es jetzt durchaus passieren, dass wir dieselbe Ameise mehrmal zählen. Dadurch ändert sich natürlich auch das Wahrscheinlichkeits-Maß. Wenn wir berechnen wollen, mit welcher Wahrscheinlichkeit wir nun k markierte Ameisen finden, werden die Dinge erfreulicherweise einfacher. Wir behandeln gleich den allgemeinen Fall und nehmen folgendes an:

1. Für jede einzelne Ameise hat die Wahrscheinlichkeit, dass die Ameise markiert ist, den Wert
$$p = \frac{K}{N}.$$
Hier bezeichnet N die Gesamtzahl der Ameisen und K ist die Anzahl der insgesamt markierten Ameisen. Die Menge $\mathcal{M}$ der Ameisen hat also die Form
$$\{a_1, a_2, \cdots, a_N\}$$
Ohne Beschränkung der Allgemeinheit gehen wir davon aus, dass die Ameisen a_1 bis a_K markiert sind.
2. Das Zufalls-Experiment besteht darin, dass wir n Ameisen auf ihre Färbung untersuchen. Dabei erhalten wir als Ergebnis eine Liste $[x_1, x_2, \cdots, x_n]$ von Ameisen. Stellen wir die Ameisen durch natürliche Zahlen von 1 bis N dar, so gilt also $x_i \in \{a_1, \cdots, a_N\}$.

Da Listen der Länge n nichts anderes als n-Tupel sind, ist die Ergebnis-Menge unseres Zufalls-Experiments also das n-fache kartesische Produkt der Menge $\mathcal{M}$:
$$\Omega = \mathcal{M}^n.$$
Nach Gleichung (2.4) folgt daraus
$$|\Omega| = |\mathcal{M}|^n = N^n.$$
Wir definieren nun Λ_k als die Menge aller n-Tupel, die aus k markierten Ameisen bestehen. Wir denken uns ein solches n-Tupel aus zwei Teilen bestehend: einem k-Tupel von markierten Ameisen und einem $(n-k)$-Tupel von unmarkierten Ameisen. Es gibt insgesamt K^k solcher k-Tupel und $(N-K)^{n-k}$ solcher $(n-k)$-Tupel. Als nächstes überlegen wir uns, wieviele Möglichkeiten es gibt, aus einem k-Tupel und einem $(n-k)$-Tupel ein n-Tupel zu erzeugen. Betrachten wir zunächst ein konkretes Beispiel: Um aus dem 3-Tupel $[x_1, x_2, x_3]$ und dem 2-Tupel $[y_1, y_2]$ ein 5-Tupel zu erstellen müssen wir die Menge I der Indizes festlegen, an denen wir die Elemente x_i einfügen. Setzen wir beispielsweise $I := \{1, 2, 3\}$, so würden die Elemente x_i am Anfang stehen und wir hätten

$$I = \{1, 2, 3\}: \qquad [x_1, x_2, x_3, y_1, y_2].$$

Für $I = \{1, 3, 4\}$ würde sich

$$I = \{1, 3, 4\}: \qquad [x_1, y_1, x_2, x_3, y_2].$$

ergeben. Jede solche Index-Menge $I \subseteq \{1, \cdots, 5\}$ legt also eindeutig fest, wie wir aus den beiden Tupeln ein 5-Tupel bilden können.

Im allgemeinen Fall gilt einerseits $I \subseteq \{1, \cdots, n\}$ und andererseits $|I| = k$. Damit ist I dann eine k-elementige Teilmenge einer n-elementigen Menge. Nach Gleichung (2.6) gilt also

$$|I| = \binom{n}{k}.$$

Damit ergibt sich für die Mächtigkeit des Ereignisses Λ_k der folgende Ausdruck

$$|\Lambda_k| = |I| \cdot K^k \cdot (N-K)^{n-k} = \binom{n}{k} \cdot K^k \cdot (N-K)^{n-k}.$$

Die Wahrscheinlichkeit, dass wir k markierte Ameisen zählen, ergibt sich jetzt zu

$$\begin{aligned}
P(k) := P(\Lambda_k) &= \frac{|\Lambda_k|}{|\Omega|} \\
&= \frac{\binom{n}{k} \cdot K^k \cdot (N-K)^{n-k}}{N^n} \\
&= \binom{n}{k} \cdot \frac{K^k}{N^k} \cdot \frac{(N-K)^{n-k}}{N^{n-k}} \\
&= \binom{n}{k} \cdot \left(\frac{K}{N}\right)^k \cdot \left(1 - \frac{K}{N}\right)^{n-k} \\
&= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad \text{mit} \quad p := \frac{K}{N}
\end{aligned}$$

Wir abstrahieren nun von den Ameisen und fassen unsere Ergebnisse wie folgt zusammen. Ist ein Zufalls-Experiment dadurch gekennzeichnet, dass n mal ein Experiment durchgeführt wird, bei dem es nur zwei mögliche Ergebnisse gibt, die wir jetzt mit a und b bezeichnen und hat die Wahrscheinlichkeit für das Auftreten von a bei jedem solchen Experiment den selben Wert p, dann ist die Wahrscheinlichkeit dafür, dass bei einer n-fachen Durchführung dieses Experiments k mal das Ergebnis a auftritt, durch

$$P(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \tag{2.9}$$

gegeben. Diese Wahrscheinlichkeits-Funktion bezeichnen wir als Binomial-Verteilung und definieren

$$\mathrm{Bin}(n, k; p) := \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

2.6 Zufalls-Variablen

Es wird Zeit, dass wir einen Begriff formal definieren, der uns in verschiedenen Beispielen schon mehrfach begegnet ist. Dies ist der Begriff der Zufalls-Variable.

Definition 8 (Zufalls-Variable) *Es sei ein Wahrscheinlichkeits-Raum $\langle \Omega, 2^\Omega, P \rangle$ gegeben. Eine Funktion*

$$X : \Omega \to \mathbb{R}$$

bezeichnen wir als Zufalls-Variable. □

Beispiel: Ein einfaches Beispiel für eine Zufalls-Variable wäre die Summe S der Augenzahlen, wenn mit zwei Würfeln gewürfelt wird. Der Ergebnis-Raum Ω ist in diesem Fall

$$\Omega = \bigl\{\langle i,j\rangle \bigm| i,j \in \{1,\cdots,6\}\bigr\}.$$

Wenn wir davon ausgehen, dass es sich bei den Würfeln um Laplace-Würfel handelt, dann ist die Wahrscheinlichkeits-Verteilung $P : 2^\Omega \to \mathbb{R}$ durch die Formel

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{36}\cdot |A|$$

gegeben. Die Zufalls-Variable $S : \Omega \to \mathbb{R}$ definieren wir durch

$$S\bigl(\langle i,j\rangle\bigr) := i + j.$$

Wenn uns die Wahrscheinlichkeit für das Auftreten einer bestimmten Augensumme interessiert, dann müssen wir zunächst die diesbezüglichen Ereignisse definieren. Bei diesen Ereignissen handelt es sich um die Mengen

$$\bigl\{\langle i,j\rangle \in \Omega \bigm| i+j = s\bigr\} \quad \text{für } s = 2,\cdots,12.$$

Ist X eine Zufalls-Variable, die auf einem Ergebnis-Raum Ω definiert ist, so vereinbaren wir zur Abkürzung die folgende Schreibweise:

$$P(X = x) := P\bigl(\{\omega \in \Omega \mid X(\omega) = x\}\bigr).$$

In unserem konkreten Beispiel schreiben wir also

$$P(S = s) = P\bigl(\{\langle i,j\rangle \in \Omega \mid i+j = s\}\bigr).$$

Für $s \leq 7$ gilt

$$\bigl\{\langle i,j\rangle \in \Omega \bigm| i+j = s\bigr\} = \bigl\{\langle k, s-k\rangle \mid k \in \{1,\cdots,s-1\}\bigr\}.$$

Die Bedingung $k \leq s-1$ folgt aus der Ungleichung $1 \leq s-k$. Für $s > 7$ finden wir

$$\bigl\{\langle i,j\rangle \in \Omega \bigm| i+j = s\bigr\} = \bigl\{\langle 7-k, s-7+k\rangle \mid k \in \{1,\cdots,13-s\}\bigr\}.$$

Die Bedingung $k \leq 13-s$ folgt dabei aus der Forderung $s-7+k \leq 6$. Daraus ergibt sich

$$\Bigl|\bigl\{\langle i,j\rangle \in \Omega \bigm| i+j = s\bigr\}\Bigr| = \begin{cases} s-1 & \text{falls } s \leq 7; \\ 13-s & \text{sonst.} \end{cases}$$

Also haben wir

$$P(S = s) = \begin{cases} \dfrac{s-1}{36} & \text{falls } s \leq 7; \\[2ex] \dfrac{13-s}{36} & \text{sonst.} \end{cases}$$

Die Funktion $s \mapsto P(S = s)$ bezeichnen wir als Wahrscheinlichkeits-Funktion der Zufalls-Variablen S. Wir schreiben diese Funktion als f_s, es gilt also

$$f_S(s) = P(S = s).$$

Wir sehen hier, dass die Wahrscheinlichkeit für die verschiedenen möglichen Werte der Summe S nicht mehr gleichmäßig verteilt sind, obwohl die Wahrscheinlichkeits-Verteilung P auf dem zugrunde liegenden Wahrscheinlichkeits-Raum sehr wohl gleichmäßig ist. Die Ursache hierfür ist einfach einzusehen: Es gibt beispielsweise 6 Möglichkeiten, in der Summe eine 6 zu würfeln, aber es gibt nur eine einzige Möglichkeit um in der Summe eine 12 zu würfeln.

Definition 9 (Wahrscheinlichkeits-Funktion) *Ist $\langle \Omega, 2^\Omega, P\rangle$ ein diskreter Wahrscheinlichkeits-Raum und ist $X : \Omega \to \mathbb{R}$ eine Zufalls-Variablen, so bezeichnen wir die Funktion*

$$f_X : \mathrm{range}(X) \to [0,1],$$

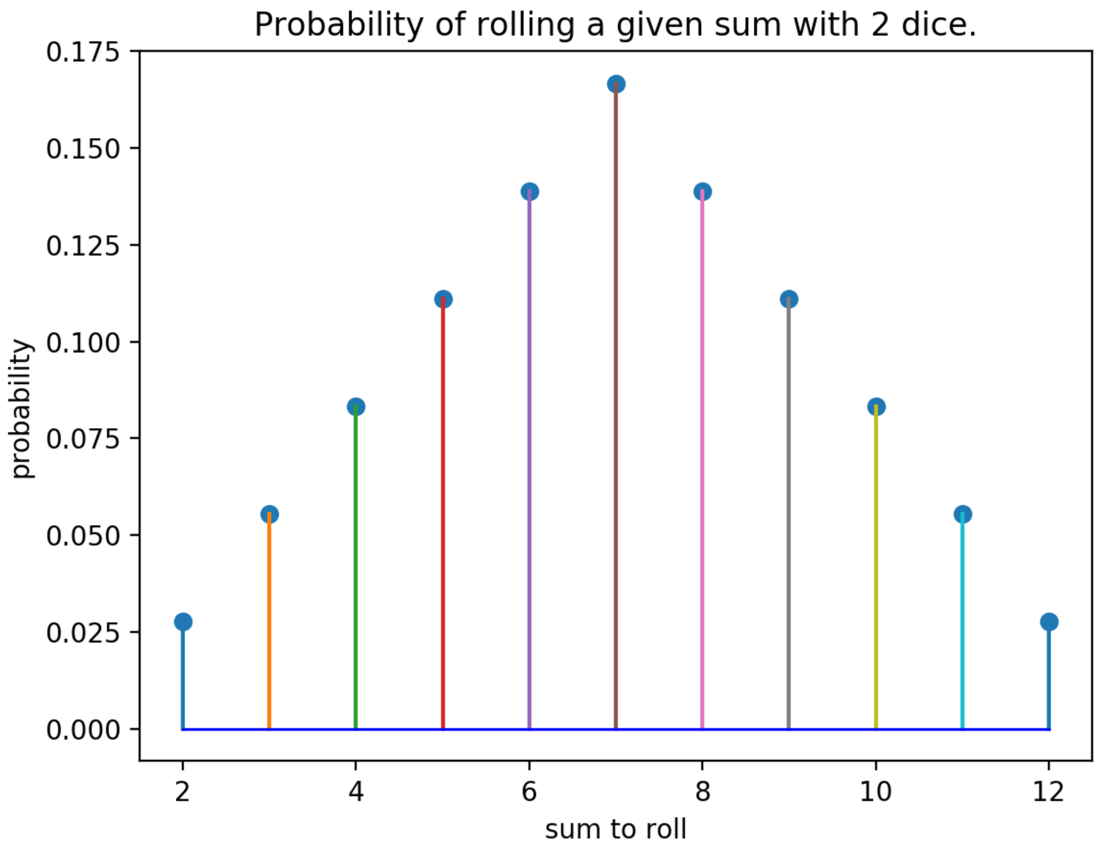

Abbildung 2.3: Wahrscheinlichkeits-Funktion der Zufalls-Variablen S

die durch die Gleichung

$$f_X(x) := P(X = x)$$

definiert ist, als die der Zufalls-Variablen X zugeordnete Wahrscheinlichkeits-Funktion. ◇

Beispiel: Bei dem letzten Beispiel haben wir die Summe S der Augenzahlen beim Würfeln mit zwei Würfeln als Zufalls-Variable betrachtet. Für die Zufalls-Variable S gilt in diesem Fall

$$\text{range}(S) = \{2, \cdots, 12\},$$

denn die Summe der Augenzahl beim Würfeln mit zwei Würfeln ist mindestens 2 und höchstens 12. Für die Wahrscheinlichkeits-Funktion f_S gilt nach dem, was wir oben ausgerechnet haben:

$$f_S(k) = \begin{cases} \dfrac{k-1}{36} & \text{falls } k \leq 7; \\[2ex] \dfrac{13-k}{36} & \text{sonst.} \end{cases}$$

Die Wahrscheinlichkeits-Funktion f_s ist in Abbildung 2.3 gezeigt. ◇

Bemerkung: Ist $\langle \Omega, 2^\Omega, P \rangle$ ein diskreter Wahrscheinlichkeits-Raum, ist $X : \Omega \to \mathbb{R}$ eine Zufalls-Variable mit $\text{range}(X) = \{x_1, \cdots, x_n\}$ und hat X die Wahrscheinlichkeits-Funktion f_X, so muss

$$\sum_{i=1}^{n} f_X(x_i) = 1$$

gelten, denn aus der Gleichung $\mathrm{range}(X) = \{x_1, \cdots, x_n\}$ folgt, dass das Ereignis

$$\bigcup_{i=1}^{n} \{\omega \in \Omega \mid X(\omega) = x_i\}$$

mit dem Ergebnis-Raum Ω identisch ist und folglich die Wahrscheinlichkeit 1 hat. Da wir andererseits

$$f_X(x_i) = P\big(\{\omega \in \Omega \mid X(\omega) = x_i\}\big)$$

haben, muss die Summe dieser Wahrscheinlichkeiten den Wert 1 ergeben. $\diamond$

Aufgabe 14: Betrachten Sie allgemein den Fall, dass mit n Würfeln gewürfelt wird. Der Ergebnis-Raum Ω ist dann das n-fache kartesische Produkt der Menge $\{1, \cdots, 6\}$:

$$\Omega = \{1, \cdots, 6\}^n$$

Berechnen Sie mit Hilfe eines geeigneten Programms (in einer Programmiersprache ihrer Wahl) die Wahrscheinlichkeit dafür, dass die Augensumme aller Würfel den Wert s hat und erstellen Sie den Graphen der Funktion

$$P_n : \{n, \cdots, 6 \cdot n\} \to \mathbb{R}$$

die durch

$$P_n(s) = P\Big(\big\{\langle x_1, \cdots, x_n\rangle \in \Omega \mid s = \sum_{i=1}^{n} x_i\big\}\Big).$$

gegeben ist. Zeichnen Sie diesen Graphen für die Werte $n = 2$, $n = 3$, $n = 5$, $n = 10$, sowie $n = 100$.

Hinweis: Sie können den Wert $P_n(s)$ natürlich mit dem Würfel-Summen-Satz berechnen. Allerdings muss dann die von Ihnen verwendete Programmiersprache in der Lage sein, Ausdrücke der Form $\binom{n}{k}$ auch für sehr große Werte von n exakt berechnen zu können. Wenn diese Voraussetzung nicht erfüllt ist, dann können Sie $P_n(s)$ am besten rekursiv berechnen, indem Sie P_n auf P_{n-1} zurück führen. Allerdings sollten Sie sich darüber im Klaren sein, dass eine naive rekursive Implementierung nur für kleine Werte von n in endlicher Zeit ein Ergebnis berechnet. $\diamond$

Definition 10 (Bernoulli-Verteilung) *Ist $\langle\Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und ist X eine Zufalls-Variable, die nur die beiden Werte 0 und 1 annimmt, wobei der Wert 1 mit der Wahrscheinlichkeit p angenommen wird, gilt also*

$$X : \Omega \to \{0, 1\} \quad \textit{und} \quad P(X = 1) = p,$$

dann sagen wir, dass X eine Bernoulli-Verteilung (Jacob I. Bernoulli, 1655 — 1705) mit Parameter p hat und schreiben

$$X \sim \mathrm{Bern}(p) \;\textit{(lese: } X \textit{ ist Bernoulli-verteilt mit Parameter } p\textit{).}$$ $\diamond$

Hat ein Zufalls-Experiment nur zwei mögliche Ausgänge, von denen wir einen als Erfolg und den anderen als Misserfolg bezeichnen, so können wir die Menge Ω als

$$\Omega := \{\mathrm{E}, \mathrm{M}\}$$

darstellen, wobei das Ergebnis E für "Erfolg" und das Ergebnis M für "Misserfolg" steht. Definieren wir die Zufalls-Variable $X : \Omega \to \{0, 1\}$ durch

$$X(\mathrm{E}) = 1 \quad \text{und} \quad X(\mathrm{M}) = 0$$

und definieren wir weiter

$$p := P(\{\mathrm{E}\}),$$

so gilt $X \sim \mathrm{Bern}(p)$. Ein solches Zufalls-Experiment heißt Bernoulli-Experiment mit Parameter p. Ein einfaches Beispiel für ein solches Bernoulli-Experiment wäre der Wurf einer Münze, wenn wir beispielsweise "Wappen" als Erfolg und "Zahl" als Misserfolg interpretieren. Wiederholen wir ein solches Bernoulli-Experiment mehrmals, wobei die einzelnen Wiederholungen voneinander unabhängig sind, und summieren wir dann die Anzahl der Erfolge, so ist diese Summe eine Zufalls-Variable, die einer Binomial-Verteilung genügt. Diesen Begriff definieren wir nun formal.

Definition 11 (Binomial-Verteilung) *Wird ein Bernoulli-Experiment mit Parameter p insgesamt n mal so wiederholt, dass die einzelnen Ausführungen des Experiments voneinander unabhängig sind, und beschreibt die Zufalls-Variable S die Anzahl der Erfolge, die sich bei der n-maligen Durchführung des Bernoulli-Experiments ergeben, so sagen wir, dass die Zufalls-Variable S binomial mit den Parameter n und p verteilt ist und schreiben*

$$S \sim \mathrm{Bin}(n, p).$$ □

Wir wollen nun die Wahrscheinlichkeits-Funktion einer binomial-verteilten Zufalls-Variable berechnen. Dazu definieren wir zunächst den Ereignis-Raum Ω als die Menge aller n-Tupel, die nur die beiden Elemente E (für "Erfolg") und M (für "Misserfolg") enthalten, es gilt also

$$\Omega = \{\mathrm{E}, \mathrm{M}\}^n.$$

Ist p die Wahrscheinlichkeit dafür, dass bei einem einzelnen der n Bernoulli-Experimente ein Erfolg eintritt, so hängt die Wahrscheinlichkeit $P(\omega)$ für $\omega \in \Omega$ davon ab, wie oft in der Liste ein Erfolg bzw. ein Misserfolg vorliegt. Ist beispielsweise $n = 5$ und betrachten wir das Elementar-Ereignis

$$A := \big\{[\mathrm{E}, \mathrm{M}, \mathrm{E}, \mathrm{E}, \mathrm{M}]\big\},$$

so gilt $P(A) = p^3 \cdot (1-p)^2$, denn bei dem Ereignis A sind 3 der Bernoulli-Experimente erfolgreich, während zwei der Bernoulli-Experimente nicht erfolgreich waren. Ist allgemein B ein Elementar-Ereignis, bei dem k der zugehörigen Bernoulli-Experimente erfolgreich und die restlichen $n-k$ Bernoulli-Experimente nicht erfolgreich sind, so gilt

$$P(B) = p^k \cdot (1-p)^{n-k}.$$

Wir müssen uns nun fragen, wie viele Elementar-Ereignisse es in der Menge Ω gibt, bei denen genau k der insgesamt n Bernoulli-Experimente erfolgreich sind. Diese Zahl ist aber gleich der Anzahl der Möglichkeiten, die wir haben, um aus einer n-elementigen Menge k Elemente auszuwählen und diese Zahl ist gerade $\binom{n}{k}$. Damit haben wir die Wahrscheinlichkeits-Funktion einer binomial verteilten Zufalls-Variable gefunden: Falls $S \sim \mathrm{Bin}(n, p)$ ist, so gilt

$$f_S(k) = P(S = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

Die Abbildungen 2.4 und 2.5 auf Seite 31 zeigen verschiedene Binomial-Verteilungen.

2.7 Erwartungswert und Varianz

Definition 12 (Erwartungswert) *Ist $\langle \Omega, 2^\Omega, P \rangle$ ein diskreter Wahrscheinlichkeits-Raum und ist*

$$X : \Omega \to \mathbb{R}$$

eine Zufalls-Variable, so definieren wir den Erwartungswert $E[X]$ als

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

Hat der Wertebereich von X die Form

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\} = \{x_n \mid n \in \mathbb{N}\},$$

so können wir den Erwartungswert auch durch die Formel

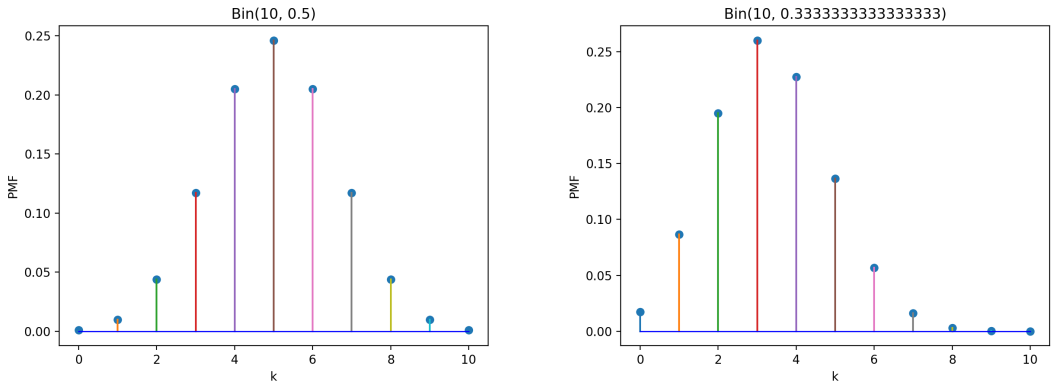

Abbildung 2.4: Die Binomial-Verteilungen $\mathrm{Bin}(10, 1/2)$ und $\mathrm{Bin}(10, 1/3)$.

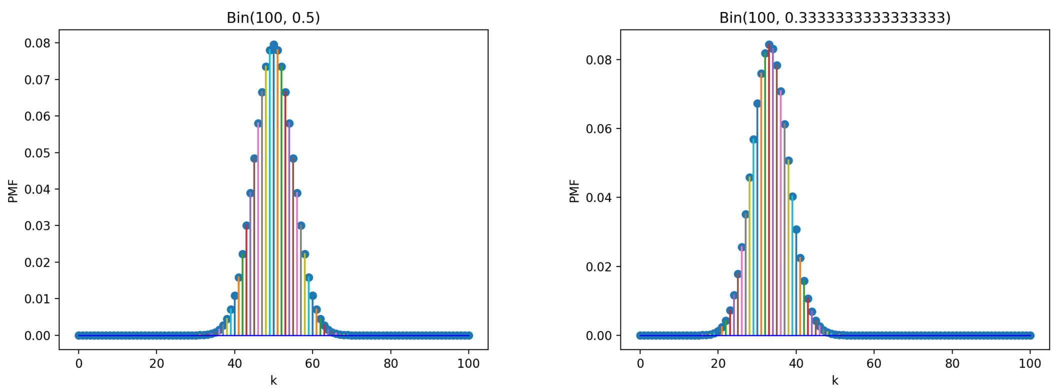

Abbildung 2.5: Die Binomial-Verteilungen $\mathrm{Bin}(100, 1/2)$ und $\mathrm{Bin}(100, 1/3)$.

$$E[X] = \sum_{n=0}^{\infty} P(X = x_n) \cdot x_n$$

berechnen. Eine analoge Formel gilt, wenn die Menge $X(\Omega)$ endlich ist. $\diamond$

Der Erwartungswert einer Zufalls-Variable gibt den durchschnittlichen Wert an, der sich bei einer großen Zahl von Versuchen ergeben würde. Diese Aussage werden wir später noch näher präzisieren und dann auch beweisen können.

Beispiel: Wir berechnen den Erwartungswert der Augenzahl beim Würfeln mit einem Würfel. Es gilt

$$\begin{aligned} E[S] &= \sum_{s=1}^{6} P(S = s) \cdot s \\ &= \sum_{s=1}^{6} \frac{1}{6} \cdot s \;=\; \frac{1}{6} \cdot \sum_{s=1}^{6} s \\ &= \frac{1}{6} \cdot \frac{1}{2} \cdot 6 \cdot (6+1) \;=\; \frac{7}{2} \end{aligned}$$

Als nächstes berechnen wir den Erwartungswert der Augensumme für das Würfeln mit zwei Laplace-Würfeln. Hier gilt

$$\begin{aligned} E[S] &= \sum_{s=2}^{12} P(S=s)\cdot s \\ &= \frac{1}{36}\cdot\sum_{s=2}^{7}(s-1)\cdot s + \frac{1}{36}\cdot\sum_{s=8}^{12}(13-s)\cdot s \\ &= \frac{252}{36} = 7 \end{aligned}$$

Dieses Ergebnis war auch zu erwarten, denn beim Würfeln mit einem Würfel ist der Erwartungswert $\frac{7}{2}$ und der Erwartungswert der Augensumme beim Würfeln mit zwei Würfeln sollte doppelt so groß sein.

◇

Aufgabe 15: Beim Mensch-ärger-dich-nicht müssen Sie am Anfang eine 6 würfeln um das Spiel beginnen zu können. Berechnen Sie den Erwartungswert für die Anzahl der Würfe, die benötigt werden um eine 6 zu würfeln.

Lösung: Die Wahrscheinlichkeit, dass die 6 sofort beim ersten Wurf kommt, beträgt $\frac{1}{6}$, während die Wahrscheinlichkeit, dass beim ersten Wurf keine 6 gewürfelt wird, offenbar den Wert $\frac{5}{6}$ hat. Die Wahrscheinlichkeit dafür, dass im zweiten Wurf die 6 fällt, ist das Produkt aus der Wahrscheinlichkeit, dass im ersten Wurf keine 6 fällt und der Wahrscheinlichkeit, dass im zweiten Wurf eine sechs fällt und hat daher den Wert $\frac{5}{6}\cdot\frac{1}{6}$. Allgemein hat die Wahrscheinlichkeit dafür, dass erst im $n+1$-ten Wurf eine 6 fällt, den Wert

$$\left(\frac{5}{6}\right)^n\cdot\frac{1}{6}.$$

Bezeichnen wir die Zufalls-Variable, welche die Anzahl der benötigten Würfe angibt, mit N, so erhalten wir für den Erwartungswert von N die Formel

$$E[N] = \sum_{n=0}^{\infty}\left(\frac{5}{6}\right)^n\cdot\frac{1}{6}\cdot(n+1) = \frac{1}{6}\cdot\sum_{n=0}^{\infty}\left(\frac{5}{6}\right)^n\cdot(n+1).$$

Diese Summe können wir mit einem Trick auf die geometrische Reihe zurück führen. Wir haben im letzten Semester gesehen, dass für alle $q\in]-1,1[$

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$$

gilt. Differenzieren wir diese Formel nach q, so erhalten wir die Gleichung

$$\sum_{n=1}^{\infty} n\cdot q^{n-1} = \frac{1}{(1-q)^2}.$$

Die Summe auf der linken Seite geht erst bei $n=1$ los, denn das konstante Glied fällt beim Differenzieren weg. Ersetzen wir in dieser Summe n durch $n+1$, so haben wir

$$\sum_{n=0}^{\infty}(n+1)\cdot q^n = \frac{1}{(1-q)^2}.$$

Diese Summe hat aber genau die Form, die oben bei der Berechnung des Erwartungswerts auftritt. Damit gilt

$$E[N] = \frac{1}{6}\cdot\sum_{n=0}^{\infty}\left(\frac{5}{6}\right)^n\cdot(n+1) = \frac{1}{6}\cdot\frac{1}{\left(1-\frac{5}{6}\right)^2} = \frac{1}{6}\cdot\frac{1}{\left(\frac{1}{6}\right)^2} = 6.$$

Wir müssen im Schnitt also 6 mal würfeln bis eine 6 auftritt. Dieses Ergebnis ist intuitiv einleuchtend.

◇

Bemerkung: Neben dem Erwartungswert gibt es noch den Begriff des Medians einer Zufalls-Variable. Ist $X : \Omega \to \{x_1, \cdots, x_n\}$ eine Zufalls-Variable so gilt

$$x \text{ ist Median von } X \text{ g.d.w. } P(X \leq x) \geq \frac{1}{2} \wedge P(X \geq x) \geq \frac{1}{2}.$$

Für diskrete Zufalls-Variablen ist der Median oft nicht eindeutig. Betrachten wir beispielsweise beim Würfeln mit einem Würfel die Zufalls-Variable X, die trivial als

$$X(k) := k \quad \text{für alle } k \in \{1, \cdots, 6\}$$

definiert ist, so gilt beispielsweise

$$P(X \leq 3.1) = \frac{1}{2} \quad \text{und} \quad P(X \geq 3.1) = \frac{1}{2},$$

aber genauso gilt auch

$$P(X \leq 3.9) = \frac{1}{2} \quad \text{und} \quad P(X \geq 3.9) = \frac{1}{2}.$$

Damit ist sowohl die Zahl 3.1 als auch die Zahl 3.9 ein Median der Zufalls-Variable X. Offenbar ist jede Zahl in dem offenen Intervall $(3, 4)$ ein Median der Zufalls-Variable X. In der deskriptiven Statistik spielt der Median eine wichtige Rolle, wenn die Verteilung der Werte sehr unsymmetrisch ist, aber in der Wahrscheinlichkeits-Rechnung ist der Median nicht so wichtig. $\diamond$

Der Erwartungswert gibt den mittleren Wert einer Zufalls-Variable wieder. Damit wissen wir aber noch nichts darüber, wie weit die einzelnen Werte der Zufalls-Variable um diesen Mittelwert streuen. Darüber gibt die Varianz Aufschluss.

Definition 13 (Varianz, Standard-Abweichung)
Ist $\langle \Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und ist

$$X : \Omega \to \mathbb{R}$$

eine Zufalls-Variable, so definieren wir die Varianz $\mathrm{Var}[X]$ als den Erwartungswert der Zufalls-Variable $\omega \mapsto \bigl(X(\omega) - E[X]\bigr)^2$, also gilt

$$\mathrm{Var}[X] := E\bigl[(X - E[X])^2\bigr].$$

Hat der Wertebereich von X die Form

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\} = \{x_n \mid n \in \mathbb{N}\}$$

und setzen wir zur Abkürzung $\mu = E[X]$, so können wir die Varianz auch durch die Formel

$$\mathrm{Var}[X] = \sum_{n=0}^{\infty} P(X = x_n) \cdot (x_n - \mu)^2$$

berechnen. Eine analoge Formel gilt, wenn die Menge $X(\Omega)$ endlich ist.

Die Standard-Abweichung ist als die Quadrat-Wurzel aus der Varianz definiert

$$\sigma(X) := \sqrt{\mathrm{Var}[X]}.$$

Im Gegensatz zur Varianz hat die Standard-Abweichung dieselbe Einheit wie die Zufalls-Variable X. $\diamond$

Beispiel: Wir berechnen die Varianz der Zufalls-Variable S, welche die Augenzahl beim Würfeln mit einem Würfel wiedergibt. Wegen $E[S] = \frac{7}{2}$ gilt

$$\begin{aligned}\mathrm{Var}[S] &= \sum_{s=1}^{6} P(S=s)\cdot\left(s-\frac{7}{2}\right)^2\\ &= \sum_{s=1}^{6}\frac{1}{6}\cdot\left(s-\frac{7}{2}\right)^2 = \frac{1}{6}\cdot\sum_{s=1}^{6}\left(s-\frac{7}{2}\right)^2 = \frac{35}{12}.\end{aligned}$$

Damit gilt für die Standard-Abweichung

$$\sigma(X) = \sqrt{\frac{35}{12}} \approx 1.707825129.$$

Führen wir dieselbe Rechnung für das Experiment "Würfeln mit zwei Würfeln" durch, so erhalten wir für die Varianz

$$\begin{aligned}\mathit{Var}[S] &= \sum_{s=2}^{12} P(S=s)\cdot(s-7)^2\\ &= \frac{1}{36}\cdot\sum_{s=2}^{7}(s-1)\cdot(s-7)^2 + \frac{1}{36}\cdot\sum_{s=8}^{12}(13-s)\cdot(s-7)^2 = \frac{35}{6}\end{aligned}$$

Wir sehen, dass die Varianz jetzt genau doppelt so groß ist wie beim Würfeln mit einem Würfel. Diese Beobachtung werden wir später verallgemeinern und beweisen. □

Kapitel 3

Berechnung von Fakultäten und Binomial-Koeffizienten

In der Wahrscheinlichkeits-Rechnung verfolgen uns die Fakultät und die Binomial-Koeffizienten auf Schritt und Tritt. Für kleine Werte von n können wir $n!$ und $\binom{n}{k}$ problemlos mit dem Taschenrechner ausrechnen. Aber schon bei der Lösung von Aufgabe 5 stößt der Taschenrechner an seine Grenzen, denn der Ausdruck

$$\binom{40\,000}{200},$$

der bei der Lösung dieser Aufgabe im Nenner auftritt, liefert eine ganze Zahl mit 545 Stellen. Zwar kann SetlX solche Zahlen noch mühelos berechnen, aber in vielen andern Programmier-Sprachen können Sie mit solchen Zahlen nicht mehr arbeiten. Wir stellen daher in diesem Abschnitt eine Näherungs-Formel für die Fakultät und für den Binomial-Koeffizienten vor. Wir beginnen mit einer Approximation der Fakultät. Die klassische Näherung von Stirling (James Stirling; 1692 - 1770) lautet

$$n! \approx \sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{\mathrm{e}}\right)^n.$$

Hier bezeichnet e die Eulersche Zahl. Genauer ist die Formel des ungarischen Mathematikers Cornelius Lanczos (1893 - 1974). Die Stirling-Formel auf der rechten Seite approximiert die Fakultät in dem folgenden Sinne: Es gilt

$$\lim_{n\to\infty} \frac{\sqrt{2\cdot\pi\cdot n}\cdot\left(\frac{n}{\mathrm{e}}\right)^n}{n!} = 1.$$

Für eine Herleitung dieser Formeln bleibt uns leider nicht die Zeit. Oft sind die Fakultäten so groß, dass sie nicht mehr auf dem Taschenrechner dargestellt werden können. Sie treten dann meist in Brüchen auf und der zu berechnende Bruch ist durchaus noch darstellbar. Dann kann es hilfreich sein, zum Logarithmus überzugehen. Um beispielsweise $\binom{n}{k}$ zu berechnen, gehen wir wie folgt vor:

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!\cdot(n-k)!} \\
&= \exp\left(\ln\left(\frac{n!}{k!\cdot(n-k)!}\right)\right) \\
&= \exp\Big(\ln\big(n!\big) - \ln\big(k!\big) - \ln\big((n-k)!\big)\Big) \qquad (3.1)
\end{aligned}$$

Zu Berechnung der verschiedenen Fakultäten in dieser Formel wenden wir auf die Stirling-Formel den Logarithmus an. Das liefert

$$\ln\big(n!\big) \approx \frac{1}{2}\cdot\ln(2\cdot\pi\cdot n) + n\cdot\ln(n) - n.$$

Setzen wir $\ln(k!)$, $\ln((n-k)!)$ und $\ln(n!)$ die entsprechenden Wert in Gleichung (3.1) ein, so können wir den Binomial-Koeffizienten approximieren. Berechnen wir $\binom{100}{50}$ auf diese Weise, so erhalten wir

$$\binom{100}{50} \approx 101\,143\,884\,241\,463\,730\,000\,000\,000\,000.$$

Der exakte Wert ist

$$\binom{100}{50} = 100\,891\,344\,545\,564\,193\,334\,812\,497\,256$$

und der relative Fehler liegt bei $2.5‰$. Bei der Berechnung von Wahrscheinlichkeiten sind solche Genauigkeiten meistens ausreichend.

Noch einfacher können Binomial-Koeffizienten über die Formel

$$\binom{n}{k} \approx \sqrt{\frac{2}{\pi \cdot n}} \cdot 2^n \cdot \exp\left(-\frac{\left(k - \frac{1}{2} \cdot n\right)^2}{\frac{1}{2} \cdot n}\right).$$

berechnet werden. Diese Formel geht auf Abraham de Moivre (1667 - 1754) zurück. Diese Näherung liefert brauchbare Werte sobald die Bedingung $n > 36$ erfüllt ist. Die Werte sind am genauesten für $k \approx \frac{n}{2}$. Berechnen wir mit dieser Formel eine Näherung für $\binom{100}{50}$, so erhalten wir

$$\binom{100}{50} \approx 1.0114388424145895 \cdot 10^{29}.$$

Auch diesmal liegt der relative Fehler bei $2.5‰$.

Der Binomial-Koeffizient tritt oft in Termen der Form

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

auf. Für diesen Fall gibt es eine Approximations-Formel, die noch einfacher zu handhaben ist. Die Formel lautet

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \approx \frac{1}{\sqrt{2 \cdot \pi \cdot n \cdot p \cdot (1-p)}} \cdot \exp\left(-\frac{(k - n \cdot p)^2}{2 \cdot n \cdot p \cdot (1-p)}\right)$$

und geht auf Pierre Simon Laplace (1749 - 1827) zurück. Diese Näherung liefert brauchbare Werte sobald die Bedingung

$$n \cdot p \cdot (1-p) > 9$$

erfüllt ist. Für $p = \frac{1}{2}$ geht diese Formel in die von de Moivre angegebene Formel über. Die Werte, die mit dieser Formel berechnet werden, sind am brauchbarsten für die Werte von k, die in der Nähe von $n \cdot p$ liegen.

Normalerweise interessiert uns nicht eine einzelne Wahrscheinlichkeit der Form

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k},$$

sondern wir wollen vielmehr wissen, welchen Wert ein Ausdruck der Form

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

annimmt, denn dieser Ausdruck gibt die Wahrscheinlichkeit dafür an, dass bei einer Binomial-Verteilung bei einer n-fachen Durchführung eines Experiments ein Ergeignis höchstens k-mal auftritt. Wir wollen daher jetzt eine Näherung für diese Summe herleiten. Dazu definieren wir zunächst

$$F_p^n(k) := \sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}.$$

Die Funktion $F_p^n(k)$ bezeichnen wir auch als die Verteilungsfunktion. Um die nachfolgende Rechnung zu vereinfachen, definieren wir zur Abkürzung

$$\mu := n \cdot p, \quad q := 1 - p \quad \text{und} \quad \sigma := \sqrt{n \cdot p \cdot q}.$$

Allgemein lässt sich eine Summe der Form $\sum_{i=0}^{k} f(i)$ wie folgt durch ein Integral approximieren:

$$\sum_{i=a}^{b} f(i) \approx \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} f(t)\,\mathrm{d}t$$

Wir approximieren nun in der Definition der Verteilungsfunktion $F_p^n(k)$ einerseits den Binomial-Koeffizienten durch die Formel von Laplace und andererseits nähern wir die Summe durch ein Integral an. Dann erhalten wir

$$\begin{aligned}
F_p^n(k) &= \sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} \\
&\approx \sum_{i=0}^{k} \frac{1}{\sqrt{2\,\pi\,n\,p\,(1-p)}} \cdot \exp\left(-\frac{(i-n\cdot p)^2}{2\,n\,p\,(1-p)}\right) \\
&= \sum_{i=0}^{k} \frac{1}{\sqrt{2\,\pi}\,\cdot\sigma} \cdot \exp\left(-\frac{(i-\mu)^2}{2\,\sigma^2}\right) \\
&\approx \int_{-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{\sqrt{2\,\pi}\,\cdot\sigma} \cdot \exp\left(-\frac{(u-\mu)^2}{2\,\sigma^2}\right)\,\mathrm{d}u \\
&= \frac{1}{\sqrt{2\,\pi}\,\cdot\sigma} \cdot \int_{-\frac{1}{2}}^{k+\frac{1}{2}} \exp\left(-\frac{(u-\mu)^2}{2\,\sigma^2}\right)\,\mathrm{d}u
\end{aligned}$$

Das hier auftretende Integral vereinfachen wir, indem wir die Variablen-Substitution

$$t(u) = \frac{u-\mu}{\sigma}, \quad \text{also } \mathrm{d}t = \tfrac{1}{\sigma}\,\mathrm{d}u \quad \sigma\,\mathrm{d}t = \mathrm{d}u$$

durchführen. Setzen wir hier für u die Grenzen $-\frac{1}{2}$ bzw. $k+\frac{1}{2}$ ein, so erhalten wir

$$\begin{aligned}
F_p^n(k) &= \frac{1}{\sqrt{2\,\pi}\,\cdot\sigma} \cdot \int\limits_{(-\frac{1}{2}-\mu)/\sigma}^{(k+\frac{1}{2}-\mu)/\sigma} \exp\left(-\frac{t^2}{2}\right)\,\sigma\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2\,\pi}} \cdot \int\limits_{(-\frac{1}{2}-\mu)/\sigma}^{(k+\frac{1}{2}-\mu)/\sigma} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t
\end{aligned}$$

Um an dieser Stelle weiter zu kommen, benötigen wir die Gauß'sche Integralfunktion $\Phi(x)$. Diese Funktion wird wie folgt definiert:

$$\Phi(x) = \frac{1}{\sqrt{2\,\pi}} \cdot \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t$$

Damit haben für die Verteilungsfunktion $F_p^n(k)$ die Näherung

$$F_p^n(k) = \Phi\left(\frac{k+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{-\frac{1}{2}-\mu}{\sigma}\right)$$

gefunden. Berücksichtigen wir noch, dass $\mu = n\cdot p$ und $\sigma = \sqrt{n\,p\,q}$ gilt und bedenken, dass wir eine Näherung für große Werte von n suchen, so gilt

$$\Phi\left(\frac{-\frac{1}{2}-\mu}{\sigma}\right) = \Phi\left(\frac{-\frac{1}{2}-n\,p}{\sqrt{n\,p\,q}}\right) \underset{n\to\infty}{\longrightarrow} \Phi\left(-\infty\right) = 0$$

Damit lautet unsere Näherung für die Verteilungsfunktion $F_p^n(k)$

$$F_p^n(k) \approx \Phi\left(\frac{k+\frac{1}{2}-n\,p}{\sqrt{n\,p\,q}}\right).$$

In dieser Formel wird häufig die sogenannte *Stetigkeits-Korrektur* $\frac{1}{2}$ weggelassen, denn gegenüber dem Term $n\,p$ fällt diese kaum ins Gewicht. Wir werden also im folgenden die Näherung

$$\boxed{F_p^n(k) \approx \Phi\left(\frac{k - n\,p}{\sqrt{n\,p\,q}}\right)}$$

verwenden. Die Φ-Funktion ist in $\textsc{SetlX}$ vordefiniert und hat dort den Namen `stat_normalCDF`. Diese Funktion wird mit drei Parametern aufgerufen. Es gilt

$$\Phi(x) = \texttt{stat\_normalCDF}(x, 0, 1).$$

Damit sind wir jetzt in der Lage, auch interessantere Aufgaben aus der Wahrscheinlichkeits-Rechnung numerisch zu lösen.

Aufgabe 16: Angenommen, wir würfeln n mal mit einem Laplace-Würfel. Wie gross muss n gewählt werden, damit wir mit einer Wahrscheinlichkeit von 99% davon ausgehen können, dass mindestens $\frac{n}{7}$ mal eine Sechs gewürfelt wird?

Kapitel 4

Bedingte Wahrscheinlichkeiten

In diesem Abschnitt beantworten wir die folgende Frage: Es sei ein Wahrscheinlichkeits-Raum $\langle\Omega, 2^\Omega, P\rangle$ gegeben. Wir betrachten ein Ereignis $A \in 2^\Omega$. Angenommen wir erfahren nun, dass ein Ereignis B eingetreten ist. Wie verändert sich durch diese Information die Wahrscheinlichkeit für das Eintreten des Ereignisses A? Die Wahrscheinlichkeit für das Eintreten von A unter der Voraussetzung, dass B bereits eingetreten ist, bezeichnen wir mit

$P(A|B)$ (lese: Wahrscheinlichkeit von A gegeben B).

Beispiel: Wir betrachten das Zufalls-Experiment "Wurf eines Laplace-Würfels" mit dem Ergebnis-Raum $\Omega = \{1, \cdots, 6\}$. Es sei A das Ereignis, dass eine 6 gewürfelt wird, also $A = \{6\}$. Da wir vorausgesetzt haben, dass es sich um einen Laplace-Würfel handelt, gilt

$$P(A) = \frac{1}{6}.$$

Es sei weiterhin B das Ereignis, dass die Augenzahl größer als 4 ist, also $B = \{5, 6\}$. Wir nehmen nun an, dass der Würfel geworfen wird und wir gesagt bekommen, dass das Ereignis B eingetreten ist. Das genaue Ergebnis des Zufalls-Experiments ist uns allerdings nicht bekannt. Da dann nur noch zwei Möglichkeiten für das Ergebnis bleiben, nämlich die Zahlen 5 und 6, würden wir in dieser neuen Situation dem Ereignis A die Wahrscheinlichkeit $\frac{1}{2}$ zuordnen, also gilt

$$P(A|B) = \frac{1}{2}. \qquad \diamond$$

Beispiel: Aus der bayerischen Sterbetafel 2013/2015, in der das Lebensalter von $100\,000$ verstorbenen Frauen verzeichnet ist, entnehmen wir die Information, dass $95\,132$ aller Frauen das Alter von mindestens 60 Jahren erreichen, während $72\,885$ ein Alter von 80 oder mehr Jahren erreichen. Angenommen, eine Frau wird 60 Jahre alt. Mit welcher Wahrscheinlichkeit erreicht Sie dann ein Alter von 80 Jahren?

Wir bezeichnen das Ereignis, dass eine Frau das Alter von 60 Jahren erreicht, mit A, während das Ereignis, dass eine Frau das Alter von 80 Jahren erreicht, mit B bezeichnet wird. Offenbar ist B eine Teilmenge von A und es ist klar, dass der Anteil der sechzigjährigen Frauen, die auch noch das Alter von 80 Jahren erreichen, durch den Bruch $\frac{|B|}{|A|}$ gegeben wird. Also gilt

$$P(B|A) = \frac{|B|}{|A|} = \frac{72\,885}{95\,132} \approx 76.6\%.$$

Damit beträgt also die Wahrscheinlichkeit dafür, dass eine sechzigjährige Frau das Alter von 80 Jahren erreicht, 76.6%. $\diamond$

Wir verallgemeinern die obigen Beispiele. Wir gehen davon aus, dass ein Wahrscheinlichkeits-Raum

$$\langle\Omega, 2^\Omega, P\rangle$$

mit einer gleichmäßigen Wahrscheinlichkeits-Verteilung P gegeben ist, es gilt also

$$P(E) = \frac{|E|}{|\Omega|} \quad \text{für alle } E \in 2^{\Omega}.$$

Wir betrachten zwei Ereignisse A und B und berechnen die bedingte Wahrscheinlichkeit für das Eintreten von B, wenn wir bereits wissen, dass A eingetreten ist. Wenn A eingetreten ist, kommen für das Eintreten von B nur noch die Ergebnisse in Frage, die in $B \cap A$ liegen. Nehmen wir an, dass diese Ergebnisse nach wie vor untereinander dieselbe Wahrscheinlichkeit haben, dann haben wir einen neuen Wahrscheinlichkeits-Raum, dessen Ereignis-Raum die Menge A ist. Folglich gilt

$$P(B|A) = \frac{|B \cap A|}{|A|} = \frac{\frac{|B\cap A|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{P(B \cap A)}{P(A)}. \tag{4.1}$$

Diese Gleichung für die bedingte Wahrscheinlichkeit gilt auch im allgemeinen Fall. Sind zwei Ereignisse A und B gegeben und führen wir das zugrunde liegende Zufalls-Experiment n-mal aus, so erwarten wir, dass für große n ein Ereignis E etwa $n \cdot P(E)$ mal eintritt. Ist das Ereignis A bereits eingetreten, so tritt das Ereignis B genau dann ein, wenn das Ereignis $B \cap A$ eintritt. Also ist die relative Häufigkeit für das Eintreten des Ereignisses B unter der Annahme, dass A bereits eingetreten ist, durch den Quotienten

$$\frac{n \cdot P(B \cap A)}{n \cdot P(A)} = \frac{P(B \cap A)}{P(A)}$$

gegeben und daher definieren wir die bedingte Wahrscheinlichkeit immer als diesen Quotienten.

Aufgabe 17: Eine Lieferung von Glühbirnen enthalte erfahrungsgemäß drei Arten von Glühbirnen:

1. Glühbirnen, die bereits defekt sind. Der Anteil dieser Glühbirnen betrage 10%.
2. Glühbirnen, die zwar funktionieren, aber nur eine Lebensdauer von weniger als 10 Tagen haben. Hier beträgt der Anteil 20%.
3. Glühbirnen, die voll funktionsfähig sind.

Angenommen, Sie testen eine Glühbirne und stellen fest, dass diese Birne noch nicht defekt ist. Wie hoch ist dann die Wahrscheinlichkeit dafür, dass diese Glühbirne voll funktionsfähig ist?

Lösung: Wir bezeichnen das Ereignis "Glühbirne defekt" mit D, das Ereignis "Glühbirne hat kurze Lebensdauer" mit K und das Ereignis "Glühbirne voll funktionsfähig" mit F. Gefragt ist dann nach der Wahrscheinlichkeit, dass die Glühbirne voll funktionsfähig ist unter der Bedingung, dass die Glühbirne nicht defekt ist. Das ist genau die bedingte Wahrscheinlichkeit $P(F|D^c)$:

$$\begin{aligned}
P(F|D^c) &= \frac{P(F \cap D^c)}{P(D^c)} \\
&= \frac{P\big(F \cap (K \cup F)\big)}{P(K \cup F)} && \text{wegen } D^c = K \cup F \\
&= \frac{P\big((F \cap K) \cup (F \cap F)\big)}{P(K \cup F)} \\
&= \frac{P\big(\emptyset \cup F\big)}{P(K \cup F)} && \text{wegen } F \cap K = \emptyset \\
&= \frac{P(F)}{P(K) + P(F)} && \text{wegen } K \cap F = \emptyset \\
&= \frac{0.7}{0.2 + 0.7} && \text{wegen } P(F) = 1 - P(D) - P(K) \\
&= \frac{7}{9}
\end{aligned}$$

Damit hat die gesuchte Wahrscheinlichkeit den Wert $0.\overline{7}$. $\diamond$

Die Gleichung (4.1) zur Berechnung der bedingten Wahrscheinlichkeit kann wie folgt umgestellt werden:

$$P(B \cap A) = P(B|A) \cdot P(A) \tag{4.2}$$

Aufgabe 18: Wir nehmen an, dass bei der Produktion von Glühbirnen die Wahrscheinlichkeit dafür, dass eine Glühbirne defekt ist, den Wert 0.1 hat. Diese Glühbirnen werden anschließend in Kisten verpackt und ausgeliefert. Die Wahrscheinlichkeit, dass eine solche Kiste beim Transport hinfällt, betrage 5%. Außerdem gehen wir davon aus, dass beim Hinfallen einer Kiste 30% der intakten Glühbirnen zerstört werden. Berechnen Sie die Wahrscheinlichkeit dafür, dass eine gelieferte Glühbirne funktionsfähig ist.

4.1 Die totale Wahrscheinlichkeit und die Formel von Bayes

Es sei ein Wahrscheinlichkeits-Raum $\langle \Omega, 2^\Omega, P \rangle$ gegeben. Wir betrachten zwei Ereignisse A und B. Nach Definition des komplementären Ereignisses gilt

$$B \cup B^c = \Omega.$$

Wegen $A \cap \Omega = A$ folgt daraus

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Da die beiden Mengen $A \cap B$ und $A \cap B^c$ disjunkt sind, können wir damit die Wahrscheinlichkeit des Ereignisses A durch die folgende Formel berechnen:

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \quad \text{nach Gleichung (4.2)} \end{aligned}$$

Dieses Ergebnis lässt sich verallgemeinern. Ist eine Familie $B_1, \cdots, B_n$ von Ereignissen gegeben, so dass

1. $B_1 \cup \cdots \cup B_n = \Omega$ und
2. $B_i \cap B_j = \emptyset$ für alle $i \neq j$

gilt, so haben wir

$$\begin{aligned} P(A) &= P(A \cap B_1) + \cdots + P(A \cap B_n) \\ &= P(A|B_1) \cdot P(B_1) + \cdots + P(A|B_n) \cdot P(B_n) \\ &= \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i). \end{aligned} \tag{4.3}$$

Dies ist die Formel von der totalen Wahrscheinlichkeit. Eine Familie $B_1, \cdots, B_n$ von Ereignissen mit den oben angegebenen Eigenschaften bezeichnen wir als Zerlegung oder auch Partition von Ω.

Aufgabe 19: Erfahrungsgemäß sind etwa 8% aller Männer farbenblind, während nur 0.6% aller Frauen farbenblind sind. Nehmen Sie an, dass der Anteil der Männer in der Gesamtbevölkerung 47% beträgt und berechnen Sie die Wahrscheinlichkeit dafür, dass eine beliebige Person farbenblind ist.

Lösung: Es sei B das Ereignis, das eine Person farbenblind ist. Das Ereignis, dass eine Person männlich ist bezeichnen wir mit M und das dazu komplementäre Ereignis bezeichnen wir mit F. Dann gilt

$$P(B) = P(B|M) \cdot P(M) + P(B|F) \cdot P(F) = 0.08 \cdot 0.47 + 0.006 \cdot 0.53 \approx 0.04078.$$

Also sind etwa 4.1% aller Personen farbenblind. $\diamond$

Aus der Gleichung $P(B \cap A) = P(B|A) \cdot P(A)$ können wir eine weitere Formel ableiten:

$$\begin{aligned} P(B|A) \cdot P(A) &= P(B \cap A) \\ &= P(A \cap B) \\ &= P(A|B) \cdot P(B) \end{aligned}$$

Teilen wir diese Formel durch $P(A)$, so ergibt sich

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

Nehmen wir nun an, dass $B_1, \cdots, B_n$ eine Zerlegung von Ω ist, so gilt nach dem Satz von der totalen Wahrscheinlichkeit

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i).$$

Gilt nun $B = B_k$ für ein $k \in \{1, \cdots, n\}$, so haben wir insgesamt:

$$\boxed{P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum\limits_{i=1}^{n} P(A|B_i) \cdot P(B_i)}} \tag{4.4}$$

Dies ist der Satz von Bayes (Thomas Bayes, 1701 – 1761). Diese Formel führt die Wahrscheinlichkeit dafür, dass ein Ereignis B_k unter der Bedingung A eintritt auf die Wahrscheinlichkeiten, dass das Ereignis A unter der Bedingung B_i eintritt, zurück. Die Formel von Bayes ist sehr wichtig im Bereich der Medizin und der Rechtssprechung. Wir betrachten entsprechende Beispiele.

Beispiel: Ein (hypothetischer) Bluttest erkennt das Vorliegen von Malaria mit einer Wahrscheinlichkeit von 99.9%. Mit einer Wahrscheinlichkeit von $0.5‰$ liefert der Test aber auch dann ein positives Ergebnis, wenn keine Malaria vorliegt. Die Wahrscheinlichkeit, dass eine Person aus Deutschland an Malaria erkrankt ist, liegt bei etwa 10^{-6}. Angenommen, Sie machen diesen Bluttest und erhalten ein positives Ergebnis. Wie hoch ist dann die Wahrscheinlichkeit, dass Sie Malaria haben?

Lösung: Wir bezeichnen das Ereignis, dass eine in Deutschland lebende Person Malaria hat, mit B_1 und das dazu komplementäre Ereignis mit B_2. Weiterhin sei A das Ereignis, dass der Bluttest ein positives Result ergibt. Nach der Aufgabenstellung haben wir

$$P(B_1) = 10^{-6}, \quad P(B_2) = 1 - 10^{-6}, \quad P(A|B_1) = 0.999 \quad \text{und} \quad P(A|B_2) = 0.0005.$$

Gesucht ist die bedingte Wahrscheinlichkeit $P(B_1|A)$. Nach der Formel von Bayes gilt

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1) \cdot P(B_1)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)} \\ &= \frac{0.999 \cdot 10^{-6}}{0.999 \cdot 10^{-6} + 0.0005 \cdot (1 - 10^{-6})} \\ &\approx 0.001994017946 \end{aligned}$$

Damit liegt die Wahrscheinlichkeit dafür, dass Sie tatsächlich an Malaria erkrankt sind, unter $2‰$. Dieses Beispiel zeigt, dass die bedingte Wahrscheinlichkeit für das Eintreten eines Ereignisses A unter einer Bedingung B völlig verschieden ist von der Wahrscheinlichkeit des Eintretens des Ereignisses B unter der Bedingung A! ◇

Das bei der letzten Aufgabe erhaltene Ergebnis erscheint zunächst kontraintuitiv. Wir beleuchten den Sachverhalt daher noch aus einer anderen Perspektive. Wir gehen davon aus, dass in Deutschland etwa $83\,000\,000$ Personen leben. Von diesen Personen haben dann 83 Personen Malaria. Würden wir nun alle Personen aus Deutschland testen, so würden wir zwei Gruppen von Personen positiv testen:

1. Die 83 tatsächlich an Malaria erkrankte Person werden (mit hoher Wahrscheinlichkeit) alle positiv getestet.

2. Bei den $82\,999\,917$ gesunden Personen schlägt der Test mit einer Wahrscheinlichkeit von $0.5‰$ an. Daher erhalten wir also zusätzlich etwa $0.0005 \cdot 82\,999\,917 \approx 41\,500$ weitere positive Testergebnisse.

Da es insgesamt $41\,583$ positive Tests gibt, ist das Verhältnis von positiven Tests, die korrekt das Vorliegen von Malaria erkennen, zu den positiven Tests, die fälschlicherweise das Vorliegen von Malaria behaupten, durch den Bruch $\frac{83}{41\,583}$ geben und das sind etwa 2 Promille.

Bemerkung: Ein immer wiederkehrender Fehler in der Statistik besteht darin, die beiden Wahrscheinlichkeiten $P(A|B)$ und $P(B|A)$ zu verwechslen. Leider kommt dieser Fehler auch in der Praxis häufig vor und hat in der Vergangenheit schon mehrfach zu Justizirrtümern geführt. Daher heißt dieser Fehler im Angelsächsischen auch Prosecutors Fallacy. Am bekanntesten ist der Fall von Sally Clark. Diese hatte zwei Kinder durch plötzlichen Kindestod verloren. Die Wahrscheinlichkeit T, ein Kind durch plötzlichen Kindestod zu verlieren, beträgt etwa $\frac{1}{8543}$. Der Sachverständige in dem Prozess gab an, dass daher die Wahrscheinlichkeit für einen doppelten plötzlichen Kindestod bei etwa $\frac{1}{73\,000\,000}$ läge. Dies war der erste statistische Fehler in dem Prozess, denn hier wurde fälschlich angenommen, dass die beiden Todesfälle voneinander unabhängig sind, was Unsinn ist.

Der zweite Fehler bestand in der Verwechselung zweier Wahrscheinlichkeiten. Bezeichnen wir das Ereignis, dass eine Mutter zwei Kinder durch doppelten Kindestod verliert, mit D und das Ereignis, dass Ereignis, dass eine Mutter ihre Kinder nicht umbringt und unschuldig ist, mit U, so würde bei statistischer Unabhängigkeit der Todesfälle zwar

$$P(D|U) = \frac{1}{73\,000\,000}$$

gelten, aber die Wahrscheinlichkeit $P(U|D)$, also die Wahrscheinlichkeit, dass eine Mutter unschuldig ist, wenn Sie zwei Kinder verliert, ist von der Wahrscheinlichkeit $P(D|U)$ grundverschieden! In dem ersten Prozess von Sally Clark wurde fälschlicherweise angenommen, dass $P(D|U) = P(U|D)$ ist und dass folglich die Wahrscheinlichkeit, dass Sally Clark unschuldig ist, nur $\frac{1}{73\,000\,000}$ beträgt. Sie wurde daher zu einer lebenslänglichen Freiheitsstrafe verurteilt, die erst vier Jahre später aufgehoben wurde. Sally Clark starb vier Jahre nach ihrer Freilassung an den Folgen einer akuten Alkohlvergiftung. $\diamond$

Aufgabe 20: Auf einer Insel ist die Tochter des dort amtierenden Königs einem Verbrechen zum Opfer gefallen. Es gibt zunächst keinen Verdächtigen, aber dafür kann am Tatort eine DNA-Probe des Täters sichergestellt werden. Da es zu teuer ist, alle $100\,000$ auf der Insel lebenden Männer zu testen, werden zufällig 100 Männer für einen DNA-Test ausgewählt. Bei einem dieser Männer ist der Test in der Tat positiv. Bei dem besagten Test wird ein spezielles Gen verglichen. Die Chance, dass dieses Gen bei zwei zufällig ausgewählten Männern identisch ist, liegt bei 1 zu $50\,000$. Wie groß ist die Wahrscheinlichkeit, dass es sich bei dem positiv getesteten Mann um den Täter handelt? Beantworten Sie dieselbe Frage für den Fall, dass auf der Insel $1\,000\,000$ Männer leben. $\diamond$

Bemerkung: Forensiche DNA-Tests sind keineswegs so sicher, wie dies in der ländlichen Bevölkerung gemeinhin angenommen wird. Einen schönen Überblick über die Probleme, die es bei der Auswertung solcher Tests gibt, finden Sie in diesem Artikel. Zusammenfassend lässt sich sagen, dass es sich bei den DNA-Tests ähnlich verhält, wie bei bei den Gottesurteilen im Mittelalter: Während die Gerichte damals blind auf den lieben Gott vetraut haben, vertrauen Sie nun blind einer Wissenschaft, die sie in den meisten Fällen selber nicht verstehen. Ein besonders abschreckendes Beispiel ist der Fall des Obdachlosen Lukis Anderson, dessen DNA unter den Fingernägeln eines Mordopfers gefunden wurde. Obwohl Lukis Anderson auf Grund übermäßigen Alkohlkonsums zum Zeitpunkt des Mordes bewusstlos im Krankenhaus lag, musste er trotzdem fünf Monate im Gefängnis verbringen, bis herauskam, dass dieselben Sanitäter, die das Mordopfer untersucht hatten, zwei Stunden vorher Lukis Anderson wegen einer akuten Alkoholvergiftung ins Krankenhaus gebracht hatten. Diese Sanitäter hatten ein Testgerät an den Finger von Anderson angeschlossen. Dasselbe Gerät wurde später auch an den Fingern des Mordopfers angeschlossen. Mehr dazu können Sie hier lesen. Eine kritische wissenschaftliche Analyse forensischer DNA-Tests finden Sie in dem Buch “Misleading DNA Evidence” von Peter Gill [Gil14].

Aufgabe 21: Eine Familie hat zwei Kinder, über die sonst nichts bekannt ist. Für die restlichen Teilaufgaben sollten Sie zur Vereinfachung annehmen, dass gleich viele Jungen wie Mädchen geboren werden.

(a) Angenommen, Sie wissen, dass eines der beiden Kinder ein Junge ist. Wie groß ist dann die Wahrscheinlichkeit, dass beide Kinder Jungen sind?

(b) Angenommen, Sie wissen, dass das ältere der beiden Kinder ein Junge ist. Wie groß ist nun die Wahrscheinlichkeit, dass beide Kinder Jungen sind?

(c) Angenommen, Sie wissen, dass eines der Kinder im Winter geboren wurde und außerdem ein Junge ist. Wie groß ist in diesem Fall die Wahrscheinlichkeit, dass beide Kinder Jungen sind?

Zur Vereinfachung sollten Sie davon ausgehen, dass die Wahrscheinlichkeit einer Geburt für alle Jahreszeiten dieselbe ist. $\diamond$

4.2 Das Monty-Hall-Problem

Aufgabe 22: Das folgende Problem wird in der Literatur als das Monty-Hall-Problem bezeichnet.

> *Bei einer Fernseh-Show steht der Kandidat vor der Auswahl, eine von drei Türen zu öffnen. Hinter einer der Türen befindet sich ein Auto, das der Kandidat gewinnt, wenn er diese Tür öffnet. Hinter den anderen beiden Türen befindet sich jeweils eine Ziege. Der Kandidat hat keinerlei Informationen, wo sich das Auto befindet und wählt zufällig eine der Türen aus. Nachdem der Kandidat seine Wahl getroffen hat, gibt er diese Wahl bekannt und der Show-Master Monty Hall tritt in Aktion. Dieser weiß, hinter welcher Tür sich das Auto befindet und öffnet eine Tür, die einerseits verschieden ist von der Tür, die der Kandidat gewählt hat und hinter der andererseits eine Ziege steht. Falls es hier zwei Möglichkeiten gibt, trifft der Show-Master die Wahl zufällig, wobei er beide Türen mit derselben Wahrscheinlichkeit wählt. Der Kandidat erhält jetzt die Möglichkeit, seine ursprüngliche Wahl zu revidieren und die andere, noch verbleibende Tür auszuwählen. Angenommen, der Kandidat hat die erste Tür gewählt und der Show-Master hat die zweite Tür geöffnet um dem Kandidaten die dahinter verborgende Ziege zu zeigen. Ist es für den Kandidaten vorteilhaft, seine Wahl revidieren?*

Lösung: Wir überlegen uns zunächst, wie der Ergebnisraum Ω aussieht und definieren

$$M = \{1, 2, 3\} \quad \text{und} \quad \Omega = \bigl\{\langle a, w, o\rangle \mid a \in M \wedge w \in M \wedge o \in M\backslash\{a, w\}\bigr\}$$

Hier gibt a die Tür an, hinter der das Auto steht, w gibt die Tür an, die der Kandidat gewählt hat und o gibt die Tür an, die vom Show-Master geöffnet wurde. Die Wahrscheinlichkeits-Verteilung P, die dem Problem zugrunde liegt, ist **keine** gleichmäßige Verteilung. Der Grund dafür ist die dritte Komponente o der Trippel $\langle a, w, o\rangle$. Zunächst sind alle Werte von a und w tatsächlich gleich wahrscheinlich. Für gegebene Werte $a, w \in M$ definieren wir das Ereignis $E(a, w)$ als

$$E(a, w) = \bigl\{\langle a, w, o\rangle \in M^3 \bigm| o \neq a \wedge o \neq w\bigr\}.$$

Das Ereignis $E(a, w)$ fasst alle die Ergebnisse aus Ω zusammen, bei denen die Werte von a und w fest sind. Wir nehmen an, dass diese Ereignisse alle dieselbe Wahrscheinlichkeit haben. Das es insgesamt 9 Paare $\langle a, w\rangle$ gibt, gilt

$$\forall a, w \in M : P\bigl(E(a, w)\bigr) = \frac{1}{9}.$$

Wenn $a \neq w$ ist, dann enthält die Menge $E(a, w)$ genau ein Element, aber wenn $a = w$ ist, was dem Fall entspricht, dass der Kandidat die Tür gewählt hat, hinter der tatsächlich das Auto steht, dann enthält die Menge $E(a, w)$ zwei Elemente, denn dann hat der Show-Master zwei Türen zur Verfügung, die er öffnen kann. Wir haben festgelegt, dass der Show-Master in diesem Fall beide Türen mit derselben Wahrscheinlichkeit auswählt. Daher gilt insgesamt

$$P\Bigl(\bigl\{\langle a,w,o\rangle\bigr\}\Bigr) = \begin{cases} \dfrac{1}{9} & \text{falls } a \neq w; \\[2ex] \dfrac{1}{18} & \text{sonst.} \end{cases}$$

Um das Problem zu lösen, definieren wir nun eine Reihe von Ereignissen.

1. $A_n := \bigl\{\langle a,w,o\rangle \in \Omega \bigm| a = n\bigr\}$ für $n = 1,2,3$.

   Das Ereignis A_n gibt an, dass das Auto hinter der Tür mit der Nummer n steht.

2. $W_n := \bigl\{\langle a,w,o\rangle \in \Omega \bigm| w = n\bigr\}$ für $n = 1,2,3$.

   Das Ereignis W_n gibt an, dass der Kandidat die Tür mit der Nummer n gewählt hat.

3. $O_n := \bigl\{\langle a,w,o\rangle \in \Omega \bigm| o = n\bigr\}$ für $n = 1,2,3$.

   Das Ereignis O_n gibt an, dass der Show-Master die Tür mit der Nummer n geöffnet hat.

Um die Aufgabe zu lösen müssen wir die bedingten Wahrscheinlichkeiten

$$P(A_1|W_1 \cap O_2) \quad \text{und} \quad P(A_3|W_1 \cap O_2)$$

berechnen, denn dies sind die Wahrscheinlichkeiten dafür, dass das Auto hinter der ersten bzw. hinter der zweiten Tür steht. Die Bedingung $W_1 \cap O_2$ drückt dabei unser bisheriges Wissen aus: Der Kandiat hat die Tür 1 gewählt und der Show-Master hat die Tür 2 geöffnet.

Wir beginnen mit der Berechnung von $P(A_3|W_1 \cap O_2)$. Nach Gleichung (4.1) gilt

$$P(A_3|W_1 \cap O_2) = \frac{P(A_3 \cap W_1 \cap O_2)}{P(W_1 \cap O_2)}. \tag{4.5}$$

Nun haben wir $A_3 \cap W_1 \cap O_2 = \bigl\{\langle 3,1,2\rangle\bigr\}$, also gilt

$$P(A_3 \cap W_1 \cap O_2) = P\Bigl(\bigl\{\langle 3,1,2\rangle\bigr\}\Bigr) = \frac{1}{9}. \tag{4.6}$$

Die Wahrscheinlichkeit $P(W_1 \cap O_2)$ berechnen wir mit Hilfe des Satzes von der totalen Wahrscheinlichkeit, wobei wir als Zerlegung von Ω die Mengen A_1, A_2 und A_3 wählen:

$$P(W_1 \cap O_2) = \sum_{i=1}^{3} P(W_1 \cap O_2|A_i) \cdot P(A_i) \tag{4.7}$$

Die Ausdrücke $P(W_1 \cap O_2|A_i) \cdot P(A_i)$ berechnen wir für $i = 1,2,3$ mit Hilfe von Gleichung (4.2).

1. $P(W_1 \cap O_2|A_1) \cdot P(A_1) = P(W_1 \cap O_2 \cap A_1) = P\Bigl(\bigl\{\langle 1,1,2\rangle\bigr\}\Bigr) = \dfrac{1}{18}.$

2. $P(W_1 \cap O_2|A_2) \cdot P(A_2) = P(W_1 \cap O_2 \cap A_2) = P\Bigl(\bigl\{\bigr\}\Bigr) = 0.$

3. $P(W_1 \cap O_2|A_3) \cdot P(A_3) = P(W_1 \cap O_2 \cap A_3) = P\Bigl(\bigl\{\langle 3,1,2\rangle\bigr\}\Bigr) = \dfrac{1}{9}.$

Damit haben wir nun nach Gleichung (4.7)

$$P(W_1 \cap O_2) = \frac{1}{18} + 0 + \frac{1}{9} = \frac{3}{18} = \frac{1}{6}. \tag{4.8}$$

Setzen wir die Ergebnisse der Gleichungen (4.6) und (4.8) in Gleichung (4.5) ein, so erhalten wir

$$P(A_3|W_1 \cap O_2) = \frac{\frac{1}{9}}{\frac{1}{6}} = \frac{2}{3}. \tag{4.9}$$

Eine analoge Rechnung liefert $P(A_1|W_1 \cap O_2) = \frac{1}{3}$. Wir brauchen die Rechnung nicht durchzuführen, denn es muss

$$P(A_1|W_1 \cap O_2) + P(A_2|W_1 \cap O_2) + P(A_3|W_1 \cap O_2) = 1$$

gelten und wegen $P(A_2|W_1 \cap O_2) = 0$ wissen wir

$$P(A_1|W_1 \cap O_2) = 1 - P(A_3|W_1 \cap O_2) = 1 - \frac{2}{3} = \frac{1}{3}.$$

Folglich ist es für den Kandiaten vorteilhaft, seine Entscheidung zu revidieren. $\diamond$

4.3 Das Simpson Paradoxon

Nun möchte ich Ihnen eine wahre Geschichte erzählen, die sich so oder so ähnlich in einem nur wenige Lichtjahre entfernten Paralleluniversum zugetragen hat. In einem Krankenhaus waren zwei Chirurgen als Chefärzte beschäftigt. Da war zum einen Dr. Frank Frankenstein und zum anderen Dr. Jacky Jackall. Dr. Frankenstein erhielt ein jährliches Salär von einer Millionen Euro, während seine Kollegin sich mit einer kümmerlichen halben Millionen begnügen musste. Da 80% aller von Dr. Jackall durchgeführten Operationen erfolgreich waren, während Ihr Kollege nur in 55% einen Erfolg verzeichnen konnte, fühlte Dr. Jackall sich diskriminiert und reichte eine Klage bei Gericht ein. Vor Gericht wurden die durchgeführten Operationen genauer aufgeschlüsselt. In Tabelle 4.1 finden Sie die detailierten Zahlen.

	Dr. Frankenstein		Dr. Jackall	
	erfolgreich	nicht erfolgreich	erfolgreich	nicht erfolgreich
Hirntransplantationen	45	45	0	10
Abtreibungen	10	0	80	10
Summe	55	45	80	20

Tabelle 4.1: Behandlungserfolge von Dr. Frankenstein und Dr. Jackall

Der Tabelle entnehmen wir, dass Dr. Frankensteins Spezialgebiet die Neurochirugie ist: Er hat 90 Hirntransplantationen durchgeführt, von denen leider nur die Hälfte erfolgreich war. Angesichts der technischen Herausforderungen einer solchen Operation ist dies schon ein beachtliches Ergebnis. Zusätzlich hat Dr. Frankenstein noch 10 Abtreibungen zu verbuchen, die alle erfolgreich waren. Auch Dr. Jackall hat sich an Hirntransplantationen versucht, allerdings waren ihre Bemühungen nicht von Erfolg gekrönt. Der Schwerpunkt der Tätigkeit von Dr. Jackall waren Abtreibungen. Von den 90 durchgeführten Abtreibungen waren 80 erfolgreich. Wir sehen nun, dass Dr. Frankenstein sowohl bei Hirntransplantationen, als auch bei Abtreibungen eine bessere Bilanz aufzuweisen hat als Dr. Jackall. Trotzdem ist seine Gesamtbilanz deutlich schlechter als die seiner Kollegin. Der Grund ist, dass die von Dr. Frankenstein hauptsächlich durchgeführten Operationen wesentlich schwieriger sind als die Operationen, die Dr. Jackall durchgeführt hat. Dieses auf den ersten Blick paradoxe Phänomen wird als das Simpson Paradoxon bezeichnet.

An der Universität von Californien in Berkeley gab es einen ähnlichen Fall: Dort hatte eine Studentin geklagt, die sich auf einen Master-Studienplatz beworben hatte und abgelehnt worden war. Die Studentin hatte festgestellt, dass die Ablehnungsquote bei weiblichen Studenten wesentlich höher lag als bei männlichen Studenten und fühlte sich daher diskriminiert. Bei der gerichtlichen Untersuchung kam heraus, dass Frauen sich bevorzugt auf Studienplätze für Medizin beworben hatten. Dort ist das Auswahlverfahren sehr selektiv und nur wenige Bewerber werden genommen. Die Männer hatten sich dagegen vor allem auf technische Studienplätze und Informatik beworben. Wie Sie selber aus eigener Erfahrung wissen, wird dort praktisch jeder Bewerber genommen. Dadurch waren dann im Endergebnis prozentual mehr Frauen als Männer abgelehnt worden, obwohl in fast allen Studiengängen Frauen bevorzugt immatrikuliert wurden.

Die Ursache für das Auftreten des Simpson Paradoxons ist das Vorhandensein einer verborgenen Variablen. In unserem ersten Beispiel ist die verborgene Variable die Schwierigkeit der Operation, im zweiten Beispiel ist es die Fakultät, in der ein Studienplatz gesucht wurde. Die beiden Beispiele zeigen, dass die Angabe summarischer Wahrscheinlichkeiten irreführend sein kann, wenn dort grundsätzlich verschiedene Ereignisse zusammengefasst werden.

Aufgabe 23: Eine Studie, die den Behandlungserfolg von Chefärzten mit dem Behandlungserfolg von Oberärzten vergleicht, kommt zu dem Schluss, dass Chefärzte bei den gleichen Operationen statistisch einen geringeren Behandlungserfolg haben als die Oberärzte. Überlegen Sie, woran das liegen könnte. ◇

4.4 Unabhängige Ereignisse

Es gibt viele Situationen in denen das Wissen, dass ein Ereignis E eingetreten ist, die Wahrscheinlichkeit für das Auftreten eines anderen Ereignisses F nicht beeinflusst. In diesem Fall gilt

$$P(F|E) = P(F). \tag{4.10}$$

Ein einfaches Beispiel dafür wäre das Würfeln mit zwei Würfeln. Wenn F das Ereignis ist, dass im ersten Wurf eine Sechs gewürfelt wird und E das Ereignis ist, das der im zweiten Wurf eine Drei gewürfelt wird, also

$$F = \bigl\{\langle 6, i\rangle \bigm| i \in \{1, \cdots, 6\}\bigr\} \quad \text{und} \quad E = \bigl\{\langle i, 3\rangle \bigm| i \in \{1, \cdots, 6\}\bigr\},$$

dann ist offensichtlich, dass das Ereignis E das Ereignis F nicht beeinflussen kann, denn die beiden Ereignisse betreffen ja verschiedene Würfel.

Ersetzen wir in Gleichung (4.10) die bedingte Wahrscheinlichkeit durch den in Gleichung (4.1) gegebenen Wert, so haben wir

$$\frac{P(F \cap E)}{P(E)} = P(F).$$

Multiplikation dieser Gleichung mit $P(E)$ liefert

$$P(F \cap E) = P(E) \cdot P(F). \tag{4.11}$$

Wir bezeichnen zwei Ereignisse E und F als unabhängig, wenn Gleichung (4.11) erfüllt ist. Wegen $F \cap E = E \cap F$ ist diese Definition symmetrisch: Die Ereignisse E und F sind genau dann unabhängig, wenn die Ereignisse F und E unabhängig sind.

Unabhängige Ereignisse treten dann auf, wenn ein Zufalls-Experiment durchgeführt wird, dass aus zwei unabhängigen Zufalls-Experimenten besteht. Ein einfaches Beispiel dafür ist das Würfeln mit zwei Würfeln. Um diese Situation formal beschreiben zu können, führen wir das Produkt zweier Wahrscheinlichkeits-Räume ein.

Definition 14 (Produkt-Raum) *Sind $\mathcal{W}_1 = \langle \Omega_1, 2^{\Omega_1}, P_1\rangle$ und $\mathcal{W}_2 = \langle \Omega_2, 2^{\Omega_2}, P_2\rangle$ zwei Wahrscheinlichkeits-Räume, so definieren wir*

$$\mathcal{W}_1 \times \mathcal{W}_2 := \langle \Omega_1 \times \Omega_2, 2^{\Omega_1 \times \Omega_2}, P\rangle,$$

wobei die neue Wahrscheinlichkeits-Verteilung P dadurch definiert wird, dass wir die Wahrscheinlichkeiten der Elementar-Ereignisse angeben:

$$P\bigl(\{\langle x, y\rangle\}\bigr) := P_1(\{x\}) \cdot P_2(\{y\}) \quad \text{für alle } x \in \Omega_1 \text{ und alle } y \in \Omega_2.$$

Für beliebige Ereignisse E ist die Wahrscheinlichkeits-Verteilung dann durch

$$P(E) = \sum_{\omega \in E} P\bigl(\{\omega\}\bigr)$$

gegeben. ◇

Definition 15 *Ist*

$$\langle \Omega, 2^{\Omega}, P\rangle = \mathcal{W}_1 \times \mathcal{W}_2 = \langle \Omega_1, 2^{\Omega_1}, P_1\rangle \times \langle \Omega_2, 2^{\Omega_2}, P_2\rangle$$

ein Produkt-Raum, und ist E ein Ereignis aus diesem Raum, so sagen wir, dass E durch die erste Komponente bestimmt ist, falls es eine Menge $\widehat{E}$ gibt, so dass

$$E = \bigl\{\langle x, y\rangle \bigm| x \in \widehat{E} \wedge y \in \Omega_2\bigr\} = \widehat{E} \times \Omega_2$$

gilt. Analog sagen wir, dass E durch die zweite Komponente bestimmt ist, wenn es eine Menge $\widehat{E}$ gibt, so dass

$$E = \bigl\{\langle x, y\rangle \bigm| x \in \Omega_1 \wedge y \in \widehat{E}\bigr\} = \Omega_1 \times \widehat{E}$$

gilt. ◇

Beispiel: Wir betrachten das Zufalls-Experiment "Würfeln mit zwei Würfeln" und definieren zunächst $\Omega_1 := \{1, \cdots, 6\}$,

$$P_1(A) := \frac{1}{6} \cdot |A| \quad \text{und} \quad \mathcal{W}_1 := \langle \Omega_1, 2^{\Omega_1}, P_1\rangle.$$

Weiter sei $\mathcal{W} := \mathcal{W}_1 \times \mathcal{W}_1$. Dann ist das Ereignis

$$\bigl\{\langle 6, n\rangle \bigm| n \in \{1, \cdots, 6\}\bigr\}$$

durch die erste Komponente bestimmt, während das Ereignis

$$\bigl\{\langle m, n\rangle \bigm| m \in \{1, \cdots, 6\} \wedge n \in \{2, 4, 6\}\bigr\}$$

durch die zweite Komponente bestimmt wird. ◇

Satz 16 *Ist $\mathcal{W} = \langle \Omega_1, 2^{\Omega_1}, P_1\rangle \times \langle \Omega_2, 2^{\Omega_2}, P_2\rangle$ ein Produkt-Raum und sind E und F Ereignisse, so dass E durch die erste und F durch die zweite Komponente bestimmt ist, so sind die Ereignisse E und F unabhängig.*

Beweis: Nach Voraussetzung gibt es Mengen $\widehat{E}$ und $\widehat{F}$, so dass

$$E = \bigl\{\langle x, y\rangle \in \Omega_1 \times \Omega_2 \mid x \in \widehat{E} \wedge y \in \Omega_2\bigr\} \quad \text{und} \quad F = \bigl\{\langle x, y\rangle \in \Omega_1 \times \Omega_2 \mid x \in \Omega_1 \wedge y \in \widehat{F}\bigr\}$$

gilt. Es ist zu zeigen, dass

$$P(E \cap F) = P(E) \cdot P(F)$$

gilt. Nach Definition der Wahrscheinlichkeits-Verteilung P gilt

$$\begin{aligned} P(E \cap F) &= \sum_{\omega \in E \cap F} P\bigl(\{\omega\}\bigr) \\ &= \sum_{\langle x, y\rangle \in E \cap F} P\bigl(\{\langle x, y\rangle\}\bigr) \end{aligned}$$

Die Bedingung $\langle x, y\rangle \in E \cap F$ formen wir um:

$$\begin{aligned} & \langle x, y\rangle \in E \cap F \\ \Leftrightarrow \quad & \langle x, y\rangle \in E \wedge \langle x, y\rangle \in F \\ \Leftrightarrow \quad & \Bigl(x \in \widehat{E} \wedge y \in \Omega_2\Bigr) \wedge \Bigl(x \in \Omega_1 \wedge y \in \widehat{F}\Bigr) \\ \Leftrightarrow \quad & x \in \widehat{E} \wedge y \in \widehat{F} \end{aligned}$$

Damit können wir die Summe in der Gleichung für $P(E \cap F)$ umschreiben:

$$P(E \cap F) = \sum_{x \in \widehat{E},\, y \in \widehat{F}} P\bigl(\{\langle x, y\rangle\}\bigr).$$

Nach Definition des Produkt-Raums gilt $P\bigl(\{\langle x, y\rangle\}\bigr) = P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr)$. Damit haben wir

$$P(E \cap F) = \sum_{x \in \widehat{E},\, y \in \widehat{F}} P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr).$$

Jetzt können wir die Summe, die oben über x und y läuft, in zwei getrennte Summen aufspalten und finden

$$\begin{aligned}
P(E \cap F) &= \sum_{x \in \widehat{E},\, y \in \widehat{F}} P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr) \\
&= \sum_{x \in \widehat{E}} \sum_{y \in \widehat{F}} P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr) \\
&= \sum_{x \in \widehat{E}} P_1\bigl(\{x\}\bigr) \cdot \sum_{y \in \widehat{F}} P_2\bigl(\{y\}\bigr) \\
&= \left(\sum_{x \in \widehat{E}} P_1\bigl(\{x\}\bigr)\right) \cdot \left(\sum_{y \in \widehat{F}} P_2\bigl(\{y\}\bigr)\right) \\
&= P_1\bigl(\widehat{E}\bigr) \cdot P_2\bigl(\widehat{F}\bigr)
\end{aligned}$$

Der Beweis ist abgeschlossen, wenn wir zeigen können, dass $P_1(\widehat{E}) = P(E)$ und $P_2(\widehat{F}) = P(F)$ gilt. Dies folgt aus

$$\begin{aligned}
P(E) &= P\bigl(\bigl\{\langle x, y\rangle \in \Omega_1 \times \Omega_2 \mid x \in \widehat{E} \wedge y \in \Omega_2\bigr\}\bigr) \\
&= \sum_{\langle x,y\rangle \in \{\langle x,y\rangle \in \Omega_1 \times \Omega_2 \,\mid\, x \in \widehat{E} \,\wedge\, y \in \Omega_2\}} P\bigl(\bigl\{\langle x, y\rangle\bigr\}\bigr) \\
&= \sum_{x \in \widehat{E},\, y \in \Omega_2} P\bigl(\bigl\{\langle x, y\rangle\bigr\}\bigr) \\
&= \sum_{x \in \widehat{E},\, y \in \Omega_2} P_1\bigl(\bigl\{x\bigr\}\bigr) \cdot P_2\bigl(\bigl\{y\bigr\}\bigr) \\
&= \sum_{x \in \widehat{E}} \sum_{y \in \Omega_2} P_1\bigl(\bigl\{x\bigr\}\bigr) \cdot P_2\bigl(\bigl\{y\bigr\}\bigr) \\
&= \sum_{x \in \widehat{E}} P_1\bigl(\bigl\{x\bigr\}\bigr) \cdot \sum_{y \in \Omega_2} P_2\bigl(\bigl\{y\bigr\}\bigr) \\
&= \left(\sum_{x \in \widehat{E}} P_1\bigl(\{x\}\bigr)\right) \cdot \left(\sum_{y \in \Omega_2} P_2\bigl(\{y\}\bigr)\right) \\
&= P_1\bigl(\widehat{E}\bigr) \cdot P_2\bigl(\Omega_2\bigr) \\
&= P_1\bigl(\widehat{E}\bigr) \cdot 1 \\
&= P_1\bigl(\widehat{E}\bigr)
\end{aligned}$$

Damit haben wir $P_1(\widehat{E}) = P(E)$ gezeigt. Der Nachweis von $P_2(\widehat{F}) = P(F)$ verläuft völlig analog, so dass wir auf die Details verzichten können. □

Es ist offensichtlich, dass der Begriff des Produkt-Raums auch auf Produkte von mehr als zwei Faktoren erweitert werden kann. Als Anwendung der bisher präsentierten Theorie zeigen wir eine alternative Lösung des Monty-Hall-Problems, bei der wir den zugrunde liegenden Wahrscheinlichkeits-Raum nicht explizit angeben. Stattdessen können wir uns auf die Betrachtung von Ereignissen beschränken.

Beispiel: Diesmal definieren wir die Ereignisse A_i, W_i und O_i unmittelbar ohne Rückgriff auf einen zugrunde liegenden Ergebnisraum Ω wie folgt:

1. A_n: "Das Auto steht hinter der Tür mit der Nummer n".

   Da die Ereignisse A_1, A_2 und A_3 alle dieselbe Wahrscheinlichkeit haben, gilt

$$P(A_n) = \frac{1}{3} \quad \text{für alle } n \in \{1, 2, 3\}.$$

2. W_n: "Der Kandidat hat die Tür mit der Nummer n gewählt".

   Da die Ereignisse W_1, W_2 und W_3 alle dieselbe Wahrscheinlichkeit haben, gilt

   $$P(W_n) = \frac{1}{3} \quad \text{für alle } n \in \{1, 2, 3\}.$$

3. O_n: "Der Showmaster hat die Tür mit der Nummer n geöffnet".

   Die Wahrscheinlichkeit $P(O_n)$ lässt sich nicht unmittelbar angeben, denn das Ereignis O_n wird offenbar von den anderen Ereignissen beeinflusst.

Wir berechnen wieder die bedingte Wahrscheinlichkeit $P(A_3|W_1 \cap O_2)$:

$$P(A_3|W_1 \cap O_2) = \frac{P(A_3 \cap W_1 \cap O_2)}{P(W_1 \cap O_2)}$$

Um an dieser Stelle weitermachen zu können, müssen wir $P(A_3 \cap W_1 \cap O_2)$ berechnen. Wir stellen diese Wahrscheinlichkeit als bedingte Wahrscheinlichkeit dar:

$$P(A_3 \cap W_1 \cap O_2) = P(O_2 \cap A_3 \cap W_1) = P(O_2|A_3 \cap W_1) \cdot P(A_3 \cap W_1)$$

Die Ereignisse A_3 und W_1 sind offenbar unabhängig, daher gilt

$$P(A_3 \cap W_1) = P(A_3) \cdot P(W_1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

Die bedingte Wahrscheinlichkeit $P(O_2|A_3 \cap W_1)$ hat den Wert 1, denn wenn das Auto hinter der dritten Tür steht und der Kandidat die erste Tür wählt, dann hat der Showmaster keine Wahlmöglichkeit und öffnet immer die zweite Tür. Damit haben wir also

$$P(A_3 \cap W_1 \cap O_2) = P(O_2|A_3 \cap W_1) \cdot P(A_3 \cap W_1) = \frac{1}{9}$$

Jetzt müssen wir noch die Wahrscheinlichkeit $P(W_1 \cap O_2)$ bestimmen. Das geht so:

$$\begin{aligned}
P(W_1 \cap O_2) &= P\big(W_1 \cap O_2 \cap (A_1 \cup A_2 \cup A_3)\big) \\
&= P\big((W_1 \cap O_2 \cap A_1) \cup (W_1 \cap O_2 \cap A_2) \cup (W_1 \cap O_2 \cap A_3)\big) \\
&= P(W_1 \cap O_2 \cap A_1) + P(W_1 \cap O_2 \cap A_2) + P(W_1 \cap O_2 \cap A_3)
\end{aligned}$$

Es bleibt die Aufgabe, die Wahrscheinlichkeiten $P(W_1 \cap O_2 \cap A_i)$ für $i = 1, 2, 3$ zu berechnen. Damit das möglich ist, müssen wir diese Wahrscheinlichkeiten als bedingte Wahrscheinlichkeiten schreiben. Wir beginnen mit $P(W_1 \cap O_2 \cap A_1)$:

$$P(W_1 \cap O_2 \cap A_1) = P(O_2 \cap A_1 \cap W_1) = P(O_2|A_1 \cap W_1) \cdot P(A_1 \cap W_1)$$

Die Ereignisse A_1 und W_1 sind offenbar unabhängig, daher gilt

$$P(A_1 \cap W_1) = P(A_1) \cdot P(W_1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

Die bedingte Wahrscheinlichkeit $P(O_2|A_1 \cap W_1)$ hat aus Symmetrie-Gründen den selben Wert wie die Wahrscheinlichkeit $P(O_3|A_1 \cap W_1)$, denn wenn das Auto hinter der ersten Tür steht und der Kandidat diese Tür wählt, dann kann der Showmaster entweder die zweite oder die dritte Tür öffnen. Da diese beiden Wahrscheinlichkeiten zusammen den Wert 1 ergeben müssen, folgt

$$P(O_2|A_1 \cap W_1) = \frac{1}{2}.$$

Damit haben wir also

$$P(W_1 \cap O_2 \cap A_1) = P(O_2|A_1 \cap W_1) \cdot P(A_1 \cap W_1) = \frac{1}{2} \cdot \frac{1}{9} = \frac{1}{18}.$$

Jetzt berechnen wir $P(W_1 \cap O_2 \cap A_2)$. Hier müssen wir nicht lange überlegen, denn wenn das Auto hinter der zweiten Tür steht, dann wird der Showmaster die zweite Tür nicht öffnen, also gilt

$$W_1 \cap O_2 \cap A_2 = \emptyset, \quad \text{folglich ist} \quad P(W_1 \cap O_2 \cap A_2) = 0.$$

Jetzt benötigen wir $P(W_1 \cap O_2 \cap A_3)$. Es gilt aber $W_1 \cap O_2 \cap A_3 = A_3 \cap W_1 \cap O_2$ und für dieses Ereignis haben wir oben bereits die Wahrscheinlichkeit $\frac{1}{9}$ gefunden. Damit haben wir insgesamt

$$\begin{aligned}
P(A_3|W_1 \cap O_2) &= \frac{P(A_3 \cap W_1 \cap O_2)}{P(W_1 \cap O_2)} \\
&= \frac{\frac{1}{9}}{P(W_1 \cap O_2 \cap A_1) + P(W_1 \cap O_2 \cap A_2) + P(W_1 \cap O_2 \cap A_3)} \\
&= \frac{\frac{1}{9}}{\frac{1}{18} + 0 + \frac{1}{9}} \\
&= \frac{2}{1+2} \\
&= \frac{2}{3}
\end{aligned}$$

Das ist dasselbe Ergebnis, was wir auch schon früher gefunden haben. $\diamond$

Aufgabe 24: Anton, Bruno und Charlie haben sich unsterblich in dieselbe Frau verliebt und beschließen, ein Dreier-Duell durchzuführen. Dieses wird folgendermaßen durchgeführt: Die drei Duellanten stellen sich so auf, dass zwischen jedem Paar ein Abstand von 25 Metern besteht. Es wird ausgelost, wer als erstes einen Schuß abgeben darf, danach wird immer in der Reihenfolge Anton, Bruno, Charlie geschossen, mit der kleinen Einschränkung, dass jemand, der tot ist, nicht mehr mitspielen darf. Wenn jemand einen Schuß abgeben darf, so steht es ihm frei, auf welchen Kontrahenten er schießt. Außerdem hat er auch die Möglichkeit, in die Luft zu schießen. Wir wissen weiterhin, dass Anton nichts mehr haßt als Munition zu verschwenden und daher nie in die Luft schießen wird. Die Bewaffnung der drei Duellanten ist unterschiedlich.

1. Anton verfügt über eine Maschinen-Pistole vom Typ Kalaschnikow. Seine Trefferwahrscheinlichkeit liegt daher bei 100%.
2. Bruno setzt eine Pumpgun ein und hat eine Trefferwahrscheinlichkeit von 80%.
3. Charlie besitzt einen klassischen Fünfundvierziger Peacemaker mit dem er eine Trefferwahrscheinlichkeit von 50% erzielt.

Lösen Sie die folgenden Aufgaben:

(a) Überlegen Sie, welche Strategie für die einzelnen Kontrahenten optimal ist.

(b) Berechnen Sie die Überlebenswahrscheinlichkeit für jeden der Duellanten.

Hinweis: Definieren Sie folgende Ereignisse:

1. A_{A}: Anton beginnt,
2. A_{B}: Bruno beginnt,
3. A_{C}: Charlie beginnt,
4. U_{A}: Anton überlebt,
5. U_{B}: Bruno überlebt,
6. U_{C}: Charlie überlebt.

Betrachten Sie die Fälle, dass Anton, Bruno oder Charlie das Duell beginnt, getrennt und benutzen Sie dann den Satz von der totalen Wahrscheinlichkeit in der folgenden Form

$$P(U_\mathrm{A}) = P(U_\mathrm{A}|A_\mathrm{A}) \cdot P(A_\mathrm{A}) + P(U_\mathrm{A}|A_\mathrm{B}) \cdot P(A_\mathrm{B}) + P(U_\mathrm{A}|A_\mathrm{C}) \cdot P(A_\mathrm{C}).$$

Analoge Gleichungen gelten natürlich für $P(U_\mathrm{B})$ und $P(U_\mathrm{C})$. In dem Fall, dass Bruno beginnt, werden Sie feststellen, dass er dann die größten Chancen auf den Endsieg hat, wenn er versucht, Anton zu eliminieren. Bezeichnen Sie das Ereignis, dass ihm dies gelingt, mit E. Dann ist es sinnvoll, im zweiten Fall eine weitere Fallunterscheidung danach durchzuführen, ob Bruno den Anton eliminiert oder nicht.

Die am schwierigsten zu berechnende Situation ist jetzt die, wenn Bruno den Anton erschossen hat und Charlie und Bruno sich abwechselnd beschießen. Für diesen Fall definieren wir die folgenden Ereignisse:

1. B: Bruno gewinnt das Duell,
2. C: Charlie gewinnt das Duell,
3. S: Charlie trifft bei seinem ersten Schuß,
4. T: Bruno trifft bei seinem ersten Schuß.

Beachten Sie, dass wir wieder in der Ausgangssituation angekommen sind, wenn sowohl Bruno als auch Charlie ihren ersten Schuß daneben setzen.

Bei der Aufgabe können Sie die folgende Verallgemeinerung des Satzes von der totalen Wahrscheinlichkeit benutzen:

$$P(A|B) = P(A|B \cap C) \cdot P(C) + P(A|B \cap C^\mathrm{c}) \cdot P(C^\mathrm{c}). \qquad \diamond$$

Der Begriff der Unabhängigkeit von Ereignissen kann auf mehr als zwei Ereignisse ausgedehnt werden. Wir sagen, dass $\{A_1, \cdots, A_n\}$ eine Menge unabhängiger Ereignisse ist, falls für jede nicht-leere Teilmenge $I \subseteq \{1, \cdots, n\}$ gilt:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

4.5 Unabhängige Zufalls-Variablen

Wir verallgemeinern die Begriffsbildung des letzten Abschnitts und betrachten nun die Unabhängigkeit von Zufalls-Variablen.

Definition 17 (Unabhängige Zufalls-Variablen) *Es sei*

1. $\langle \Omega, 2^\Omega, P \rangle$ *ein Wahrscheinlichkeits-Raum,*
2. $X : \Omega \to \mathbb{R}$ *und* $Y : \Omega \to \mathbb{R}$ *seien zwei Zufalls-Variablen.*

Dann sind die Zufalls-Variablen X und Y unabhängig, wenn für alle $x, y \in \mathbb{R}$ die Ereignisse

$$\{\omega \in \Omega \mid X(\omega) = x\} \quad \text{und} \quad \{\omega \in \Omega \mid Y(\omega) = y\}.$$

unabhängig sind. Nach Definition der Unabhängigkeit zweier Ereignisse ist das äquivalent zu der Forderung, dass für alle $x, y \in \mathbb{R}$ die Gleichung

$$\begin{aligned} & P\Big(\{\omega \in \Omega \mid X(\omega) = x\} \cap \{\omega \in \Omega \mid Y(\omega) = y\}\Big) \\ = \; & P\Big(\{\omega \in \Omega \mid X(\omega) = x\}\Big) \cdot P\Big(\{\omega \in \Omega \mid Y(\omega) = y\}\Big) \end{aligned}$$

gilt. Diese Gleichung können wir kürzer in der Form

$$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$$

schreiben. $\diamond$

Beispiel: Angenommen, wir würfeln mit zwei Laplace-Würfel. Der Ergebnis-Raum hat dann die Form

$$\Omega = \bigl\{\langle i, j\rangle \bigm| i, j \in \{1, \cdots, 6\}\bigr\}.$$

Definieren wir die Zufalls-Variablen X und Y durch

$$X\bigl(\langle i, j\rangle\bigr) = i \quad \text{und } Y\bigl(\langle i, j\rangle\bigr) = j,$$

so sind diese Zufalls-Variablen unabhängig, denn wählen wir beispielsweise $x = 2$ und $y = 5$, so müssen wir zeigen, dass

$$P(X = 2 \wedge Y = 5) = P(X = 2) \cdot P(Y = 5)$$

gilt. Dies rechnen wir sofort nach, denn einerseits gilt

$$P(X = 2 \wedge Y = 5) = P\bigl(\{\langle 2, 5\rangle\}\bigr) = \frac{1}{36},$$

andererseits haben wir

$$\begin{aligned}
P(X = 2) &= P\bigl(\{\langle i, j\rangle \in \Omega \mid X(\langle i, j\rangle) = 2\}\bigr) \\
&= P\bigl(\{\langle i, j\rangle \in \Omega \mid i = 2\}\bigr) \\
&= P\bigl(\{\langle 2, j\rangle \mid j \in \{1, \cdots, 6\}\bigr) \\
&= \frac{\bigl|\{\langle 2, j\rangle \mid j \in \{1, \cdots, 6\}\bigr|}{36} \\
&= \frac{6}{36} \\
&= \frac{1}{6}.
\end{aligned}$$

Genau so sehen wir $P(Y = 5) = \frac{1}{6}$ und daher gilt insgesamt

$$P(X = 2 \wedge Y = 5) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(X = 2) \cdot P(Y = 5).$$

Die Tatsache, dass wir hier mit $x = 2$ und $y = 5$ gearbeitet haben, ist völlig unerheblich, denn wir hätten bei jedem anderen Wert dasselbe Ergebnis bekommen. Daher sind die beiden Zufalls-Variablen X und Y unabhängig. $\diamond$

Um das letzte Beispiel verallgemeinern zu können, fehlt noch eine Definition.

Definition 18 *Es seien*

1. $\mathcal{W}_1 = \langle \Omega_1, 2^{\Omega_1}, P_1\rangle$ *und* $\mathcal{W}_2 = \langle \Omega_2, 2^{\Omega_2}, P_2\rangle$ *Wahrscheinlichkeits-Räume.*
2. $\mathcal{W} = \mathcal{W}_1 \times \mathcal{W}_2 = \langle \Omega, 2^{\Omega}, P\rangle$ *der aus* $\mathcal{W}_1$ *und* $\mathcal{W}_2$ *gebildete Produkt-Raum.*
3. $X : \Omega \to \mathbb{R}$ *eine Zufalls-Variable auf* $\mathcal{W}$*.*

Dann wird X durch die erste Komponente bestimmt, falls

$$\forall \omega_1 \in \Omega_1 : \forall \omega_2, \omega_3 \in \Omega_2 : X\bigl(\langle \omega_1, \omega_2\rangle\bigr) = X\bigl(\langle \omega_1, \omega_3\rangle\bigr)$$

gilt. Bei der Auswertung der Zufalls-Variable X spielt also die zweite Komponente keine Rolle. Analog sagen wir, dass X durch die zweite Komponente bestimmt wird, wenn die erste Komponente bei der Auswertung von X keine Rolle spielt, wenn also gilt:

$$\forall \omega_2 \in \Omega_2 : \forall \omega_1, \omega_3 \in \Omega_1 : X\bigl(\langle \omega_1, \omega_2\rangle\bigr) = X\bigl(\langle \omega_3, \omega_2\rangle\bigr). \qquad \diamond$$

Satz 19 *Es seien*

1. *$\mathcal{W}_1 = \langle \Omega_1, 2^{\Omega_1}, P_1 \rangle$ und $\mathcal{W}_2 = \langle \Omega_2, 2^{\Omega_2}, P_2 \rangle$ Wahrscheinlichkeits-Räume.*
2. *$\mathcal{W} = \mathcal{W}_1 \times \mathcal{W}_2 = \langle \Omega, 2^{\Omega}, P \rangle$ der aus $\mathcal{W}_1$ und $\mathcal{W}_2$ gebildete Produkt-Raum.*
3. *$X_1 : \Omega \to \mathbb{R}$ und $X_2 : \Omega \to \mathbb{R}$ Zufalls-Variablen auf $\mathcal{W}$.*

Falls nun X_1 durch die erste Komponente bestimmt wird und X_2 durch die zweite Komponente bestimmt wird, dann sind die Zufalls-Variablen X_1 und X_2 unabhängig.

Beweis: Wir müssen zeigen, dass für beliebige Zahlen $x_1, x_2 \in \mathbb{R}$ die Gleichung

$$P(X_1 = x_1 \wedge X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$$

gilt. Dies ist gleichbedeutend damit, dass die beiden Ereignisse

$$E_1 := \{\omega \in \Omega \mid X_1(\omega) = x_1\} \quad \text{und} \quad E_2 := \{\omega \in \Omega \mid X_2(\omega) = x_2\}$$

unabhängig sind. Dies folgt aus Satz 16, falls wir zeigen können, dass E_1 auf die erste und E_2 auf zweite Komponente beschränkt ist. Wir führen den Nachweis für das Ereignis E_1, der Nachweis für das Ereignis E_2 verläuft völlig analog. Dazu definieren wir eine Menge $\widehat{E}_1$ durch

$$\widehat{E}_1 := \{\omega_1 \in \Omega_1 \mid \exists \omega_2 \in \Omega_2 : X_1(\langle \omega_1, \omega_2 \rangle) = x_1\}$$

und zeigen, dass $E_1 = \widehat{E_1} \times \Omega_2$ gilt. Den Nachweis, dass diese beiden Mengen gleich sind führen wir, indem wir zeigen, dass

$$\langle \omega_1, \omega_2 \rangle \in \{\omega \in \Omega \mid X_1(\omega) = x_1\} \;\Leftrightarrow\; \langle \omega_1, \omega_2 \rangle \in \{\langle \omega_1, \omega_2 \rangle \in \Omega \mid \omega_1 \in \widehat{E}_1 \wedge \omega_2 \in \Omega_2\}$$

gilt. Wir betrachten zunächst die linke Seite:

$$\begin{aligned} & \langle \omega_1, \omega_2 \rangle \in \{\omega \in \Omega \mid X_1(\omega) = x_1\} \\ \Leftrightarrow \quad & X_1(\langle \omega_1, \omega_2 \rangle) = x_1 \end{aligned} \tag{4.12}$$

Jetzt formen wir die rechte Seite um:

$$\begin{aligned} & \langle \omega_1, \omega_2 \rangle \in \{\langle \omega_1, \omega_2 \rangle \in \Omega \mid \omega_1 \in \widehat{E}_1 \wedge \omega_2 \in \Omega_2\} \\ \Leftrightarrow \quad & \omega_1 \in \widehat{E}_1 \wedge \omega_2 \in \Omega_2 \\ \Leftrightarrow \quad & \omega_1 \in \widehat{E}_1 \\ \Leftrightarrow \quad & \omega_1 \in \{\omega_1 \in \Omega_1 \mid \exists \omega_3 \in \Omega_2 : X_1(\langle \omega_1, \omega_3 \rangle) = x_1\} \\ \Leftrightarrow \quad & \exists \omega_3 \in \Omega_2 : X_1(\langle \omega_1, \omega_3 \rangle) = x_1 \end{aligned} \tag{4.13}$$

Um zu zeigen, dass die Bedingungen (4.13) und (4.12) äquivalent sind, bemerken wir zunächst, dass die Richtung

$$X_1(\langle \omega_1, \omega_2 \rangle) = x_1 \;\Rightarrow\; \exists \omega_3 \in \Omega_2 : X_1(\langle \omega_1, \omega_3 \rangle) = x_1$$

offensichtlich ist, denn wir können für das ω_3, dessen Existenz auf der rechten Seite gefordert wird, ja ω_2 einsetzen. Um die umgekehrte Richtung

$$\exists \omega_3 \in \Omega_2 : X_1(\langle \omega_1, \omega_3 \rangle) = x_1 \;\Rightarrow\; X_1(\langle \omega_1, \omega_2 \rangle) = x_1$$

zu zeigen, nehmen wir also an, dass für ein $\omega_3 \in \Omega_2$ die Gleichung

$$X_1(\langle \omega_1, \omega_3 \rangle) = x_1$$

gilt. Nun folgt aus der Voraussetzung, dass X_1 durch die erste Komponente bestimmt wird,

$$X_1(\langle \omega_1, \omega_2 \rangle) = X_1(\langle \omega_1, \omega_3 \rangle) = x_1$$

und damit ist der Beweis abgeschlossen. □

Der nächsten Satz ist eine unmittelbare Konsequenz aus der Definition des Begriffs der unabhängigen Zufalls-Variablen.

Satz 20 *Es sei*

1. $\langle \Omega, 2^{\Omega}, P \rangle$ *ein Wahrscheinlichkeits-Raum,*
2. $X : \Omega \to \mathbb{R}$ *und* $Y : \Omega \to \mathbb{R}$ *seien zwei unabhängige Zufalls-Variablen.*
3. $A, B \subseteq \mathbb{R}$.

Dann sind die Ereignisse

$$\bigl\{\omega \in \Omega \bigm| X(\omega) \in A\bigr\} \quad \textit{und} \quad \bigl\{\omega \in \Omega \bigm| Y(\omega) \in B\bigr\}$$

unabhängig, es gilt also

$$P(X \in A \wedge Y \in B) = P(X \in A) \cdot P(Y \in B).$$

Beweis: Zunächst eine Vorbemerkung. Ist eine Menge

$$C = \{c_n \mid n \in \mathbb{N}\}$$

gegeben, wobei wir stillschweigend voraussetzen, dass die Elemente c_n paarweise voneinander verschieden sind, es gilt also $m \neq n \to c_m \neq n$, und ist weiter Z eine Zufalls-Variable auf Ω, so kann die Wahrscheinlichkeit $P(Z \in C)$ wie folgt berechnet werden:

$$\begin{array}{lcll}
P(Z \in C) & = & P\bigl(\{\omega \in \Omega \mid Z(\omega) \in C\}\bigr) & \\
& = & P\left(\biguplus\limits_{n=0}^{\infty} \{\omega \in \Omega \mid Z(\omega) = c_n\}\right) & \text{denn die } c_n \text{ sind paarweise verschieden} \\
& = & \sum\limits_{n=0}^{\infty} P\bigl(\{\omega \in \Omega \mid Z(\omega) = c_n\}\bigr) & \\
& = & \sum\limits_{n=0}^{\infty} P\bigl(Z = c_n\bigr). &
\end{array}$$

Diese Identität werden wir weiter unten benötigen.

Wir nehmen nun an, dass wir die Mengen A und B in der Form

$$A = \{a_n \mid n \in \mathbb{N}\} \quad \text{und} \quad B = \{b_n \mid n \in \mathbb{N}\},$$

schreiben können, wobei wir stillschweigend $a_m \neq a_n$ für $n \neq m$ und $b_m \neq b_n$ für $n \neq m$ voraussetzen, wir betrachten also nur den Fall, dass die beiden Mengen A und B unendlich viele Elemente enthalten. Dann gilt in Analogie zu der oben gezeigten Identität:

$$\begin{array}{lcll}
P(X \in A \wedge Y \in B) & = & \sum\limits_{m=0}^{\infty}\sum\limits_{n=0}^{\infty} P(X = a_m \wedge Y = b_n) & \\
& = & \sum\limits_{m=0}^{\infty}\sum\limits_{n=0}^{\infty} P(X = a_m) \cdot P(Y = b_n) & X \text{ und } Y \text{ sind unabhängig} \\
& = & \sum\limits_{m=0}^{\infty} P(X = a_m) \cdot \sum\limits_{n=0}^{\infty} P(Y = b_n) & \text{Distributiv-Gesetz} \\
& = & \left(\sum\limits_{m=0}^{\infty} P(X = a_m)\right) \cdot \left(\sum\limits_{n=0}^{\infty} P(Y = b_n)\right) & \text{Distributiv-Gesetz} \\
& = & P(X \in A) \cdot P(Y \in B) &
\end{array}$$

□

Genau wie wir den Begriff der Unabhängigkeit auch für mehr als zwei Ereignisse definieren konnten, so können wir auch den Begriff der Unabhängigkeit von Zufalls-Variablen für mehr als zwei Zufalls-Variablen definieren. Eine Menge $\{X_1, \cdots, X_n\}$ von Zufalls-Variablen ist unabhängig, falls für jede nicht-leere Teilmenge $I \subseteq \{1, \cdots, n\}$ gilt:

$$P\Big(\bigwedge_{i \in I} X_i = x_i\Big) = \prod_{i \in I} P(X_i = x_i).$$

4.6 Eigenschaften von Erwartungswert und Varianz

Der Erwartungswert hat die folgende Linearitäts-Eigenschaft.

Satz 21 (Linearität des Erwartungswerts) *Ist $\langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum und sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei Zufalls-Variablen, so können wir für beliebige $\alpha, \beta \in \mathbb{R}$ eine Zufalls-Variable $Z : \Omega \to \mathbb{R}$ durch*

$$Z(\omega) := \alpha \cdot X(\omega) + \beta \cdot Y(\omega)$$

definieren. Für den Erwartungswert dieser Zufalls-Variable gilt:

$$E\big[Z\big] = \alpha \cdot E\big[X\big] + \beta \cdot E\big[Y\big].$$

Beweis: Dieser Satz wird bewiesen, indem wir die Definition des Erwartungswerts expandieren:

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

□

Aufgabe 25: Beweisen Sie: Ist $\langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum und definieren wir für ein beliebiges $\beta \in \mathbb{R}$ die Zufalls-Variable $X : \Omega \to \mathbb{R}$ als

$$X(\omega) := \beta,$$

so gilt $E[X] = \beta$. ◇

Im Gegensatz zu dem Erwartungswert ist die Varianz kein linearer Operator. Es gilt aber der folgende Satz.

Satz 22 (Verschiebungs-Satz) *Ist $\langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum und ist $X : \Omega \to \mathbb{R}$ ein Zufalls-Variable, so können wir für beliebige $\alpha, \beta \in \mathbb{R}$ eine Zufalls-Variable $Z : \Omega \to \mathbb{R}$ durch*

$$Z(\omega) = \alpha \cdot X(\omega) + \beta$$

definieren. Für die Varianz dieser Zufalls-Variable gilt:

$$\mathrm{Var}\big[Z\big] = \alpha^2 \cdot \mathrm{Var}\big[X\big].$$

Beweis: Zur Abkürzung definieren wir zunächst $\mu := E[X]$. Nach dem gerade bewiesenen Satz gilt

$$E[\alpha \cdot X + \beta] = \alpha \cdot E[X] + \beta \cdot E[1] = \alpha \cdot \mu + \beta.$$

Nun gilt für die Varianz von Z:

$$\begin{aligned}
\mathrm{Var}[Z] &= E\big[(Z - E[Z])^2\big] \\
&= E\Big[\big(\alpha \cdot X + \beta - (\alpha \cdot \mu + \beta)\big)^2\Big] \\
&= E\Big[\big(\alpha \cdot X - \alpha \cdot \mu\big)^2\Big] \\
&= E\Big[\big(\alpha \cdot (X - \mu)\big)^2\Big] \\
&= E\Big[\alpha^2 \cdot (X - \mu)^2\Big] \\
&= \alpha^2 \cdot E\Big[(X - \mu)^2\Big] \\
&= \alpha^2 \cdot \mathrm{Var}[X]
\end{aligned}$$

□

Um den später folgenden Satz beweisen zu können, benötigen wir einen Satz über den Erwartungswerts des Produktes zweier unabhängiger Zufalls-Variablen.

Lemma 23 *Ist $\langle\Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei unabhängige Zufalls-Variablen, so gilt*

$$E\big[X \cdot Y\big] = E\big[X\big] \cdot E\big[Y\big].$$

Beweis: Der Wertebereich der Zufalls-Variablen X und Y sei durch die Mengen

$$X(\Omega) = \{x_n \mid n \in \mathbb{N}\} \quad \text{und} \quad Y(\Omega) = \{y_n \mid n \in \mathbb{N}\}$$

gegeben. Dann ist der Wertebereich der Zufalls-Variable $Z : \Omega \to \mathbb{R}$, die durch

$$Z(\omega) = X(\omega) \cdot Y(\omega)$$

definiert ist, wie folgt gegeben:

$$Z(\Omega) = \{x_m \cdot y_n \mid m, n \in \mathbb{N}\}.$$

Damit finden wir für den Erwartungswert von Z

$$\begin{aligned}
E[Z] &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P(X = x_m \wedge Y = y_n) \cdot x_m \cdot y_n \\
&= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P(X = x_m) \cdot P(Y = y_n) \cdot x_m \cdot y_n && X \text{ und } Y \text{ sind unabhängig} \\
&= \sum_{m=0}^{\infty} P(X = x_m) \cdot x_m \cdot \sum_{n=0}^{\infty} P(Y = y_n) \cdot y_n && \text{Distributiv-Gesetz} \\
&= \left(\sum_{m=0}^{\infty} P(X = x_m) \cdot x_m\right) \cdot \left(\sum_{n=0}^{\infty} P(Y = y_n) \cdot y_n\right) && \text{Distributiv-Gesetz} \\
&= E[X] \cdot E[Y]
\end{aligned}$$

□

Aufgabe 26: Zeigen Sie: Sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei unabhängige Zufalls-Variablen, ist $\alpha \in \mathbb{R}$ und ist die Zufalls-Variable $Z : \Omega \to \mathbb{R}$ durch

$$Z(\omega) := X(\omega) + \alpha$$

definiert, so sind auch die Zufalls-Variablen Z und Y unabhängig. ◇

Der nächste Satz ist für die weitere Theorie von fundamentaler Bedeutung.

Satz 24 *Ist $\langle\Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei unabhängige Zufalls-Variablen, so gilt*

$$\mathrm{Var}\big[X + Y\big] = \mathrm{Var}\big[X\big] + \mathrm{Var}\big[Y\big].$$

Beweis: Zur Abkürzung setzen wir $\mu_X := E[X]$ und $\mu_Y := E[Y]$. Wegen $E[X+Y] = \mu_X + \mu_Y$ gilt dann

$$\begin{aligned}
\mathrm{Var}[X+Y] &= E\big[(X+Y-\mu_X-\mu_Y)^2\big] \\
&= E\Big[\big((X-\mu_X)+(Y-\mu_Y)\big)^2\Big] \\
&= E\big[(X-\mu_X)^2+(Y-\mu_Y)^2+2\cdot(X-\mu_X)\cdot(Y-\mu_Y)\big] \\
&= E\big[(X-\mu_X)^2\big]+E\big[(Y-\mu_Y)^2\big]+2\cdot E\big[(X-\mu_X)\cdot(Y-\mu_Y)\big] \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\cdot E\big[(X-\mu_X)\cdot(Y-\mu_Y)\big] \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\cdot E\big[(X-\mu_X)\big]\cdot E\big[Y-\mu_Y)\big] \\
&\ \text{nach Lemma 23, denn } X-\mu_X \text{ und } Y-\mu_Y \text{ sind unabhängig nach letzter Aufgabe} \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\cdot\big(E[X]-\mu_X\big)\cdot\big(E[Y]-\mu_Y\big) \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\cdot\big(\mu_X-\mu_X\big)\cdot\big(\mu_Y-\mu_Y\big) \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y].
\end{aligned}$$

□

Der letzte Satz lässt sich auf Mengen von unabhängigen Zufalls-Variablen verallgemeinern. Ist $\{X_1, \cdots, X_n\}$ eine Menge unabhängiger Zufalls-Variablen, so gilt

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

In dieser Form hat der Satz eine äußerst wichtige Konsequenz. Wir gehen aus von einem Zufalls-Experiment, das durch einen Wahrscheinlichkeits-Raum der Form

$$\mathcal{W} = \langle \Omega, 2^{\Omega}, P\rangle$$

beschrieben wird. Weiter sei X eine Zufalls-Variable, die bei diesem Experiment eine Rolle spielt. Wiederholen wir das Zufalls-Experiment n-mal so, dass sich die einzelnen Wiederholungen des Experiments nicht beeinflussen, dann haben wir eine Menge $\{X_1, \cdots, X_n\}$ von n Zufalls-Variablen, die unabhängig sind. Wir definieren jetzt die Zufalls-Variable

$$\overline{X} := \frac{1}{n}\cdot\sum_{k=1}^{n} X_k.$$

$\overline{X}$ ist offenbar das arithmetische Mittel der Zufalls-Variablen $X_1, \cdots, X_n$. Für den Erwartungswert von $\overline{X}$ gilt dann

$$\begin{aligned}
E\big[\overline{X}\big] &= E\left[\frac{1}{n}\cdot\sum_{k=1}^{n} X_k\right] \\
&= \frac{1}{n}\cdot\sum_{k=1}^{n} E\left[X_k\right] \\
&= \frac{1}{n}\cdot\sum_{k=1}^{n} E[X] \\
&= \frac{1}{n}\cdot E[X]\cdot\sum_{k=1}^{n} 1 \\
&= \frac{1}{n}\cdot E[X]\cdot n \\
&= E[X]
\end{aligned}$$

Dabei haben wir ausgenutzt, dass die Zufalls-Variablen X_i alle den selben Erwartungswert $E[X]$ haben.

Natürlich ist es nicht weiter verwunderlich, dass das arithmetische Mittel dann ebenfalls den Erwartungswert $E[X]$ hat. Wesentlich interessanter ist die Frage nach der Varianz der Zufalls-Variable $\overline{X}$. Es gilt

$$\begin{aligned}
\mathrm{Var}\left[\overline{X}\right] &= \mathrm{Var}\left[\frac{1}{n} \cdot \sum_{k=1}^{n} X_k\right] \\
&= \left(\frac{1}{n}\right)^2 \cdot \mathrm{Var}\left[\sum_{k=1}^{n} X_k\right] \\
&= \frac{1}{n^2} \cdot \sum_{k=1}^{n} \mathrm{Var}\left[X_k\right] \\
&= \frac{1}{n^2} \cdot \sum_{k=1}^{n} \mathrm{Var}\left[X\right] \\
&= \frac{1}{n^2} \cdot \mathrm{Var}\left[X\right] \cdot \sum_{k=1}^{n} 1 \\
&= \frac{1}{n^2} \cdot \mathrm{Var}\left[X\right] \cdot n \\
&= \frac{1}{n} \cdot \mathrm{Var}\left[X\right]
\end{aligned}$$

Die Standard-Abweichung σ ist die Wurzel der Varianz, es gilt also

$$\boxed{\sigma\left[\overline{X}\right] = \frac{1}{\sqrt{n}} \cdot \sigma[X].}$$

Dieser Zusammenhang wird in der Literatur als das $\sqrt{n}$-Gesetz bezeichnet:

Satz 25 ($\sqrt{n}$-Gesetz) *Es sei*

1. $\mathcal{W} = \langle \Omega, 2^{\Omega}, P \rangle$ *ein Wahrscheinlichkeits-Raum,*
2. $X : \Omega \to \mathbb{R}$ *eine Zufalls-Variable auf* $\mathcal{W}$*,*
3. $\mathcal{W}^n = \underbrace{\mathcal{W} \times \cdots \times \mathcal{W}}_{n}$ *sei das* n*-fache kartesisches Produkt von* $\mathcal{W}$ *mit sich selbst,*
4. $X_i : \Omega^n \to \mathbb{R}$ *sei für* $i = 1, \cdots, n$ *definiert durch*
$$X_i\big(\langle \omega_1, \cdots, \omega_n \rangle\big) : X(\omega_i)$$
5. $\overline{X} : \Omega^n \to \mathbb{R}$ *sei definiert durch*
$$\overline{X}(\omega) := \frac{1}{n} \cdot \sum_{k=1}^{n} X_i(\omega)$$

Dann gilt:

1. $E\left[\overline{X}\right] = E[X]$
2. $\mathrm{Var}\left[\overline{X}\right] = \frac{1}{n} \cdot \mathrm{Var}[X]$ *und* $\sigma\left[\overline{X}\right] = \frac{1}{\sqrt{n}} \cdot \sigma[X]$ □

Das $\sqrt{n}$-Gesetz hat in der Praxis eine sehr wichtige Anwendung bei der Messung. Wird eine physikalische Größe gemessen, so können wir diese Messung als Zufalls-Experiment ansehen. Der Erwartungswert des Zufalls-Experiments ist dann der tatsächliche Wert, während die Standard-Abweichung als die ungefähre Größe des Messfehlers interpretiert werden kann. Das $\sqrt{n}$-Gesetz zeigt uns eine Möglichkeit, um die Genauigkeit unserer Messung zu verbessern: Führen wir dieselbe Messung n-mal durch und bilden den Mittelwert unserer Ergebnisse, so wird die Mess-Genauigkeit um den Faktor $\sqrt{n}$ gesteigert. Ein anderes Beispiel ist das Würfeln: Gibt die Zufalls-Variable X die Augenzahl beim einmaligen Würfeln an, so haben wir für Varianz und Standard-Abweichung die Werte

$$\mathrm{Var}[X] = \frac{35}{12} \quad \text{und} \quad \sigma[X] = \sqrt{\frac{35}{12}} \approx 1.707825128.$$

Würfeln wir hingegen 100-mal und bilden das arithmetische Mittel $\overline{X}$ aller gewürfelten Augenzahlen, so finden wir für die Standard-Abweichung den Wert

$$\sigma\left[\overline{X}\right] = 0.1707825128$$

Im nächsten Abschnitt werden wir sehen, welche konkreten Aussagen über die Wahrscheinlichkeit aus der Standard-Abweichung abgeleitet werden können.

4.7 Satz von Tschebyschow und Gesetz der großen Zahlen

Der folgende Satz, der nach Pafnuti Lwowitsch Tschebyschow (1821 – 1894) benannt ist, zeigt uns, welche konkrete Bedeutung die Standard-Abweichung $\sigma[X]$ hat.

Satz 26 (Ungleichung von Tschebyschow) *Es sei*

1. $\langle \Omega, 2^\Omega, P\rangle$ *ein Wahrscheinlichkeits-Raum,*
2. $X : \Omega \to \mathbb{R}$ *eine Zufalls-Variable mit dem Erwartungswert* $\mu = E[X]$ *und der Varianz* $\sigma^2 = \mathrm{Var}[X]$.

Dann gilt für alle $r \in \mathbb{R}_+$

$$P\big(|X-\mu| \geq r\cdot\sigma\big) \leq \frac{1}{r^2}.$$

Beweis: Zunächst formen wir die Bedingung $|X-\mu| \geq r\cdot\sigma$ um. Es gilt offenbar für beliebige $\omega \in \Omega$:

$$|X(\omega)-\mu| \geq r\cdot\sigma \;\Leftrightarrow\; \big(X(\omega)-\mu\big)^2 \geq r^2\cdot\sigma^2 \;\Leftrightarrow\; \left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2$$

Wir zeigen jetzt zunächst die Ungleichung

$$r^2\cdot P\big(|X(\omega)-\mu| \geq r\cdot\sigma\big) \leq 1.$$

Wenn wir diese Ungleichung durch r^2 dividieren, folgt die Behauptung.

$$\begin{aligned}
r^2\cdot P\big(|X(\omega)-\mu| \geq r\cdot\sigma\big) &= r^2\cdot P\left(\left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2\right) \\
&= r^2\cdot \sum_{\left\{\omega\in\Omega \,\middle|\, \left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2\right\}} P\big(\{\omega\}\big) \\
&= \sum_{\left\{\omega\in\Omega \,\middle|\, \left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2\right\}} P\big(\{\omega\}\big)\cdot r^2 \\
&\leq \sum_{\left\{\omega\in\Omega \,\middle|\, \left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2\right\}} P\big(\{\omega\}\big)\cdot \left(\frac{X(\omega)-\mu}{\sigma}\right)^2
\end{aligned}$$

Im letzten Schritt haben wir ausgenutzt, dass die Summe nur über die ω läuft, für die $\left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2$ gilt, so dass die Summe größer wird, wenn wir r^2 durch $\left(\frac{X(\omega)-\mu}{\sigma}\right)^2$ ersetzen. Im nächsten Schritt vergrößern wir die Summe, indem wir über alle $\omega \in \Omega$ summieren:

$$\begin{aligned}
r^2 \cdot P\big(|X(\omega)-\mu| \geq r\cdot\sigma\big) &\leq \sum_{\left\{\omega\in\Omega \,\middle|\, \left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2\right\}} P\big(\{\omega\}\big)\cdot\left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \\
&\leq \sum_{\omega\in\Omega} P\big(\{\omega\}\big)\cdot\left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \\
&= \frac{1}{\sigma^2}\cdot\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\cdot(X(\omega)-\mu)^2 \\
&= \frac{1}{\sigma^2}\cdot\mathrm{Var}[X] \\
&= 1
\end{aligned}$$

□

Aufgabe 27: Schätzen Sie mit Hilfe der Ungleichung von Tschebyschow ab, wie oft Sie würfeln müssen, damit die Wahrscheinlichkeit, dass der arithmetische Mittelwert der Augenzahlen um mehr als 0.1 vom Erwartungswert abweicht, kleiner als 1% ist?

Lösung: Zunächst formalisieren wir die Aufgabenstellung. Gesucht ist ein n, so dass für den arithmetischen Mittelwert

$$\overline{X}_n := \frac{1}{n}\cdot\sum_{k=1}^{n} X_i$$

die Beziehung

$$P\big(|\overline{X}_n - \mu| \geq 0.1\big) \leq 0.01$$

gilt. Um diesen Ausdruck auf die Ungleichung von Tschebyschow zurück führen zu können, formen wir den Wert 0.1 um:

$$0.1 = 0.1\cdot\frac{1}{\sigma}\cdot\sigma \quad \text{und definieren } r = 0.1\cdot\frac{1}{\sigma}$$

Damit können wir die obige Ungleichung auch in der Form

$$P\big(|\overline{X}_n - \mu| \geq r\cdot\sigma\big) \leq 0.01$$

schreiben. Damit diese Ungleichung aus der Ungleichung von Tschebyschow folgt, muss

$$\frac{1}{r^2} \leq 0.01$$

gelten. Setzen wir hier den Wert für r ein, so haben wir

$$10^2\cdot\sigma^2 \leq 0.01, \quad \text{also} \quad 10\cdot\sigma \leq 0.1, \quad \text{also} \quad \sigma \leq \tfrac{1}{100}.$$

Die Standard-Abweichung $\sigma = \sigma\big(\overline{X}_n\big)$ können wir aus dem $\sqrt{n}$-Gesetz bestimmen, denn da die Augenzahl beim Würfeln mit einem Würfel die Varianz $\frac{35}{12}$ hat, gilt für die Varianz beim n-maligen Würfeln

$$\sigma\big(\overline{X}_n\big) = \sqrt{\frac{1}{n}\cdot\frac{35}{12}}$$

Also haben wir die Ungleichung

$$
\begin{aligned}
& \sqrt{\frac{1}{n}\cdot\frac{35}{12}} \leq \frac{1}{100} \\
\Leftrightarrow\quad & \frac{1}{n}\cdot\frac{35}{12} \leq \frac{1}{10\,000} \\
\Leftrightarrow\quad & 10\,000\cdot\frac{35}{12} \leq n \\
\Leftrightarrow\quad & 29\,166.\overline{6}\cdots \leq n
\end{aligned}
$$

Wenn wir $29\,167$ mal würfeln, dann weicht der arithmetische Mittelwert um weniger als 0.1 vom Erwartungswert ab. ◇

Bemerkung: Die Ungleichung von Tschebyschow ist nicht sehr genau. Wir werden später noch eine wesentlich bessere Abschätzung finden. ◇

Bemerkung: Sie können die letzte Aufgabe auch mit Hilfe der Näherung von Laplace für die kumulative Verteilungsfunktion lösen. Dann ergibt sich ein kleinerer Wert. Das ist nicht weiter verwunderlich, denn die Tschebyschow-Ungleichung gilt für beliebige Verteilungen, während die Näherung von Laplace nur gilt, wenn die Verteilungsfunktion die Form

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

hat. ◇

Wenn wir ein Zufalls-Experiment nur oft genug durchführen und dabei eine Zufalls-Variable X ermitteln, so wird die Wahrscheinlichkeit, dass der arithmetische Mittelwert, den wir bei n Messungen finden, beliebig klein, wenn wir nur n groß genug wählen. Genauer gilt:

Satz 27 (Schwaches Gesetz der großen Zahlen) *Es sei*

1. $\mathcal{W} = \langle\Omega, 2^\Omega, P\rangle$ *ein Wahrscheinlichkeits-Raum,*
2. $X : \Omega \to \mathbb{R}$ *eine Zufalls-Variable auf* $\mathcal{W}$ *mit dem Erwartungswert* $\mu = E[X]$ *und der Standard-Abweichung* $\sigma := \sqrt{\mathrm{Var}[X]}$,
3. $\mathcal{W}^n = \underbrace{\mathcal{W}\times\cdots\times\mathcal{W}}_{n}$ *sei das* n*-fache kartesisches Produkt von* $\mathcal{W}$ *mit sich selbst,*
4. $X_i : \Omega^n \to \mathbb{R}$ *sei für* $i = 1,\cdots,n$ *definiert durch*
$$X_i\big(\langle\omega_1,\cdots,\omega_n\rangle\big) := X(\omega_i)$$
5. $\overline{X}_n : \Omega^n \to \mathbb{R}$ *sei definiert durch*
$$\overline{X}_n(\omega) := \frac{1}{n}\cdot\sum_{i=1}^{n} X_i(\omega)$$

Dann gilt für alle $\varepsilon > 0$*:*

$$\lim_{n\to\infty} P\big(|\overline{X}_n - \mu| \geq \varepsilon\big) = 0.$$

Beweis: Nach dem $\sqrt{n}$-Gesetz haben wir für die Zufalls-Variable $\overline{X}_n$:

$$E\Big[\overline{X}_n\Big] = \mu \quad \text{und} \quad \mathrm{Var}\Big[\overline{X}_n\Big] = \frac{\sigma^2}{n}, \quad \text{also } \sigma\Big[\overline{X}_n\Big] = \frac{\sigma}{\sqrt{n}}.$$

Definieren wir nun

$$r := \frac{\varepsilon}{\sigma}\cdot\sqrt{n},$$

so gilt

$$r \cdot \sigma\Big[\overline{X}_n\Big] = r \cdot \frac{\sigma}{\sqrt{n}} = \frac{\varepsilon}{\sigma} \cdot \sqrt{n} \cdot \frac{\sigma}{\sqrt{n}} = \varepsilon$$

und mit nach der Ungleichung von Tschebyschow folgt

$$P\big(|\overline{X}_n - \mu| \geq \varepsilon\big) \leq \frac{1}{\frac{\varepsilon^2}{\sigma^2} \cdot n}, \quad \text{also } P\big(|\overline{X}_n - \mu| \geq \varepsilon\big) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}.$$

Wegen $\lim\limits_{n\to\infty} \dfrac{\sigma^2}{n \cdot \varepsilon^2} = 0$ folgt die Behauptung. □

Das schwache Gesetz der großen Zahlen zeigt uns, dass der Erwartungswert und das arithmetische Mittel einer Zufalls-Variable eng miteinander verknüpft sind.

4.8 Erwartungswert und Varianz der Binomial-Verteilung

In diesem Abschnitt wollen wir Erwartungswert und Varianz einer binomial-verteilten Zufalls-Variable berechnen. Damit dies einfach möglich ist, präsentieren wir zunächst einen alternativen Zugang zur Binomial-Verteilung. Dazu betrachten wir das folgende Beispiel.

Beispiel: Ein Betrunkener befindet sich in einer Stadt, in der die Straßen ein quadratisches Muster bilden. Dementsprechend lassen sich alle Kreuzungen durch Angabe zweier natürlicher Zahlen k und l spezifizieren, die die Koordinaten dieser Kreuzung angeben. Wir nehmen nun an, dass der Betrunkene zunächst an der Kreuzung steht, die durch das Paar $\langle 0, 0\rangle$ spezifiziert wird. Der Betrunkene geht jetzt zufällig los, wobei er mit der Wahrscheinlichkeit p nach Osten und mit der Wahrscheinlichkeit $1 - p$ nach Norden geht. Jedes Mal, wenn der Betrunkene wieder an eine Kreuzung kommt, geht er wieder mit der Wahrscheinlichkeit p nach Osten und $(1-p)$ nach Norden. Beträgt der Abstand zwischen zwei Kreuzungen eine Längeneinheit, so legt der Betrunkene insgesamt einen Weg von n Längeneinheiten zurück. Wir stellen uns die Frage, mit welcher Wahrscheinlichkeit der Betrunkene dann an der Kreuzung mit den Koordinaten $\langle k, n - k\rangle$ ankommt.

Um die Kreuzung $\langle k, n-k\rangle$ zu erreichen muss der Betrunkene k-mal nach Osten und $n-k$-mal nach Norden gegangen sein. Da wir davon ausgehen, dass die einzelnen Entscheidungen, die der Betrunkene an den Kreuzungen trifft, voneinander unabhängig sind, hat ein bestimmter Weg zur Kreuzung $\langle k, n-k\rangle$ die Wahrscheinlichkeit

$$p^k \cdot (1-p)^{n-k}.$$

Das Problem ist, dass es im Normalfall mehrere Wege gibt, die von der Kreuzung $\langle 0, 0\rangle$ starten, bei der Kreuzung $\langle k, n-k\rangle$ enden und insgesamt eine Länge von n haben. Um die Gesamt-Wahrscheinlichkeit zu berechnen, müssen wir die Wahrscheinlichkeit aller Wege aufsummieren. Da die Wahrscheinlichkeit für jeden Weg dieselbe ist, reicht es aus, wenn wir die Anzahl der Wege der Länge n bestimmten, die von $\langle 0, 0\rangle$ nach $\langle k, n-k\rangle$ führen. Wir definieren

$$s(n, k) := \text{Anzahl der Wege der Länge } n \text{ von } \langle 0, 0\rangle \text{ nach } \langle k, n-k\rangle.$$

Wir berechnen die Funktion $s(n, k)$ durch Induktion über n.

I.A.: $n = 0$. Es gibt genau einen Weg (der Länge 0) von $\langle 0, 0\rangle$ nach $\langle 0, 0\rangle$. Also gilt

$$s(0, 0) := 1.$$

I.S.: $n \mapsto n+1$. Die Kreuzung $\langle k, n+1-k\rangle$ kann der Betrunkene auf zwei Arten erreichen. Entweder er kommt von der Kreuzung $\langle k, n-k\rangle$ und geht von da aus nach Norden, oder er kommt von der Kreuzung $\langle k-1, n+1-k\rangle$ und geht von da aus nach Osten. Die Gesamtzahl aller Wege ergibt sich, wenn wir zu der Anzahl der Wege, die nach $\langle k, n\rangle$ führen, die Anzahl der Wege, die nach $\langle k-1, n+1-k\rangle$ führen, hinzu addieren. Wegen $\langle k-1, n+1-k\rangle = \langle k-1, n-(k-1)\rangle$ gilt also

$$s(n+1, k) = s(n, k) + s(n, k-1).$$

Wir werden zeigen, dass für alle $k, n \in \mathbb{N}$

$$s(n,k) = \binom{n}{k}$$

gilt. Dazu benötigen wir den folgenden Hilfssatz.

Lemma 28 (Pascal'sche Regel) *Die durch*

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

definierten Binomial-Koeffizienten genügen der Rekurrenz-Gleichung

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

Beweis: Der Beweis ergibt sich durch eine einfache Expansion der Definition.

$$\begin{aligned}
\binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(k-1)! \cdot (n-(k-1))!} + \frac{n!}{k! \cdot (n-k)!} \\
&= \frac{n!}{(k-1)! \cdot (n+1-k))!} + \frac{n!}{k! \cdot (n-k)!} \\
&= n! \cdot \left(\frac{k}{k! \cdot (n+1-k)!} + \frac{n+1-k}{k! \cdot (n+1-k)!} \right) \\
&= n! \cdot \frac{k+n+1-k}{k! \cdot (n+1-k)!} \\
&= n! \cdot \frac{n+1}{k! \cdot (n+1-k)!} \\
&= \frac{(n+1)!}{k! \cdot (n+1-k)!} \\
&= \binom{n+1}{k}
\end{aligned}$$

□

Damit können wir jetzt das obige Beispiel zu Ende führen und zeigen, dass für die Anzahl der Wege

$$s(n,k) = \frac{n!}{k! \cdot (n-k)!}$$

gilt. Wir führen den Nachweis durch Induktion nach n.

I.A.: $n = 0$. Der einzige mögliche Wert für k ist $k = 0$, folglich gilt

$$s(0,0) = 1.$$

Andererseits haben wir

$$\binom{0}{0} = \frac{0!}{0! \cdot 0!} = \frac{1}{1 \cdot 1} = 1.$$

I.S.: $n \mapsto n + 1$. Es gilt

$$\begin{aligned} s(n+1,k) &= s(n,k) + s(n,k-1) \\ &\stackrel{IV}{=} \binom{n}{k} + \binom{n}{k-1} \\ &= \binom{n+1}{k} \qquad \text{nach der Pascal'schen Regel} \end{aligned}$$ □

Damit gilt für die Wahrscheinlichkeit $P\big(\{\langle k, n-k\rangle\}\big)$, dass der Betrunkene nach einem Weg der Länge n an der Kreuzung $\langle k, n-k\rangle$ landet, die Formel

$$P\big(\{\langle k, n-k\rangle\}\big) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = B(n,p;k).$$

Wir abstrahieren jetzt von dem Alkohol und destillieren die mathematische Essenz des Beispiels. Dazu definieren wir zunächst den Begriff des Bernoulli-Experiments (Jakob Bernoulli; 1654 - 1705).

Definition 29 (Bernoulli-Experiment)
Wir nennen ein Zufalls-Experiment ein Bernoulli-Experiment (Jacob Bernoulli, 1655–1705) wenn es nur zwei mögliche Ergebnisse des Experiments gibt. Der Wahrscheinlichkeits-Raum kann dann in der Form

$$\mathcal{W} = \big\langle \{a,b\}, 2^{\{a,b\}}, P \big\rangle$$

geschrieben werden. Dann setzen wir

$$p := P\big(\{1\}\big)$$

und nennen p den Parameter des Bernoulli-Experiments. Ist die Zufalls-Variable $X : \{a,b\} \to \{0,1\}$ durch

$$X(a) := 0 \quad \textit{und} \quad X(b) := 1$$

definiert, so sagen wir, dass X eine Bernoulli-verteilte Zufalls-Variable mit Parameter p ist und schreiben

$$X \sim \texttt{Bern}(p).$$ ◇

Wird ein Bernoulli-Experiment n-mal durchgeführt, so sprechen wir von einer Bernoulli-Kette. Ist $\mathcal{W}_B$ der Wahrscheinlichkeits-Raum des einzelnen Bernoulli-Experiments, so ist das n-fache Produkt

$$\mathcal{W} = \mathcal{W}_B^n$$

der Wahrscheinlichkeits-Raum der Bernoulli-Kette. Definieren wir für das i-te Bernoulli-Experiment eine Zufalls-Variable

$$X_i : \{a,b\} \to \mathbb{R} \quad \text{durch} \quad X_i(a) := 0 \quad \text{und} \quad X_i(b) := 1,$$

so gilt $X_i \sim \texttt{Bern}(p)$ und wir können den Erwartungswert und die Varianz von X_i sehr einfach berechnen, denn es gilt:

$$\begin{aligned} E[X_i] &= \sum_{\omega \in \{a,b\}} X_i(\omega) \cdot P\big(\{\omega\}\big) \\ &= X_i(a) \cdot P\big(\{a\}\big) + X_i(b) \cdot P\big(\{b\}\big) \\ &= 0 \cdot (1-p) + 1 \cdot p \\ &= p \end{aligned}$$

Für die Varianz finden wir

$$\begin{aligned}
\mathrm{Var}[X_i] &= E\Big[\big(X_i - E[X_i]\big)^2\Big] \\
&= E\Big[\big(X_i - p\big)^2\Big] \\
&= \sum_{\omega\in\{a,b\}} (X_i(\omega) - p)^2 \cdot P\big(\{\omega\}\big) \\
&= \big(X_i(a) - p\big)^2 \cdot P\big(\{a\}\big) + \big(X_i(b) - p\big)^2 \cdot P\big(\{b\}\big) \\
&= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p \\
&= \big(p + (1-p)\big) \cdot p \cdot (1-p) \\
&= p \cdot (1-p)
\end{aligned}$$

Als nächstes definieren wir für die Bernoulli-Kette die Zufalls-Variable

$$X : \{a,b\}^n \to \mathbb{R} \quad \text{durch} \quad X\big([\omega_1, \cdots, \omega_n]\big) = \sum_{i=1}^{n} X_i(\omega_i).$$

Es gilt genau dann

$$X\big([\omega_1, \cdots, \omega_n]\big) = k,$$

wenn die Menge

$$I := \big\{ i \in \{1, \cdots, n\} \bigm| \omega_i = b \big\}$$

aus genau k Elementen besteht. Da es genau $\binom{n}{k}$ Möglichkeiten gibt, eine Teilmenge I mit k Elementen aus der Menge $\{1, \cdots, n\}$ auszuwählen und jede einzelne Kombination, in denen k mal das Ergebnis b und $n-k$ mal das Ergebnis a auftritt, die Wahrscheinlichkeit

$$p^k \cdot (1-p)^{n-k}$$

hat, gilt insgesamt

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

Also ist die Zufalls-Variable X binomial-verteilt, wir haben

$$X \sim \mathtt{Bin}(n, p).$$

Da X die Summe von n Zufalls-Variablen ist, gilt für den Erwartungswert

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = n \cdot p.$$

Da die Menge der Zufalls-Variablen $\{X_1, \cdots, X_n\}$ unabhängig ist, haben wir im letzten Abschnitt gesehen, dass wir auch die Varianz als Summe der Varianzen der einzelnen Zufalls-Variablen berechnen können. Also gilt

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = \sum_{i=1}^{n} p \cdot (1-p) = n \cdot p \cdot (1-p).$$

Aufgabe 28: Die Augenfarbe wird durch ein einzelnes Paar von Genen vererbt. Jeder Mensch hat zwei dieser Gene. Die Augenfarbe braun ist dominant, d.h. wenn Sie auch nur ein Gen für braune Augen haben, dann sind ihre Augen braun. Nur wenn Sie zwei Gene für blaue Augen haben, haben Sie blaue Augen. Nehmen Sie nun an, dass eine Familie 6 Kinder hat und dass der älteste Sohn blaue Augen hat, während die Eltern beide braune Augen haben. Wir groß ist die Wahrscheinlichkeit, dass insgesamt genau drei Kinder blaue Augen haben?

4.9 Die Poisson-Verteilung

In diesem Abschnitt betrachten wir einen Spezialfall der Binomial-Verteilung, der in der Praxis häufig auftritt. Dieser Spezialfall ist dadurch gekennzeichnet, dass in der Formel

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

einerseits der Wert von n sehr groß wird, andererseits aber der Wert von p sehr klein ist. Wir illustrieren diesen Spezialfall durch ein Beispiel.

Beispiel: Die (hypothetische) Firma Diamond-Connections habe 10 Millionen Kunden. In einem vorgegebenen Zeitintervall von, sagen wir mal, drei Minuten ruft jeder dieser Kunden mit einer Wahrscheinlichkeit p im Call-Center der Firma an. Da die meisten Kunden etwas anderes zu tun haben als im Call-Center anzurufen, kommen im Schnitt pro Zeitintervall fünf Anrufe an, die Wahrscheinlichkeit p hat also den Wert

$$p = \frac{5}{10\,000\,000} = \frac{1}{2\,000\,000}$$

Unser Ziel ist es zu berechnen, wieviele Mitarbeiter die Firma Diamond-Connections einstellen muss, damit sicher gestellt ist, dass die Wahrscheinlichkeit, dass bei einem eingehenden Anruf kein Platz des Call-Centers mehr frei ist, kleiner als 1% ist. Wir wollen zur Vereinfachung weiter voraussetzen, dass alle Gespräche genau ein Zeitintervall andauern und dass zusätzlich alle Gespräche am Anfang eines Zeitintervalls beginnen.

Wir berechnen zunächst die Wahrscheinlichkeit, dass in einem gegebenen Zeitintervalls k Teilnehmer anrufen. Setzen wir voraus, dass verschiedene Teilnehmer unabhängig sind, so ist die Zufalls-Variable K, die die Anzahl der Teilnehmer angibt, binomial-verteilt, es gilt also

$$\begin{aligned} P(K = k) &= B(n, p; k) \\ &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &= \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} \end{aligned}$$

Für die Werte von n, die in der Praxis in Frage kommen, können wir die Formel in der oben angegebenen Form nicht auswerten. Auch die Approximations-Formel von Laplace bringt uns an dieser Stelle nicht weiter, denn diese Formel liefert nur dann brauchbare Ergebnisse, wenn die Bedingung

$$n \cdot p \cdot (1-p) > 9$$

erfüllt ist. Um hier weiterzukommen, führen wir die Abkürzung $\lambda = n \cdot p$ ein, λ ist also der Erwartungswert der Zufalls-Variable K. Dann gilt $p = \frac{\lambda}{n}$ und wenn wir diesen Wert in die obere Formel einsetzen, erhalten wir

$$\begin{aligned} P(K = k) &= \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} \\ &= \frac{n!}{k! \cdot (n-k)!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{1}{k!} \cdot \frac{n \cdot (n-1) \cdot \cdots \cdot \big(n - (k-1)\big)}{n^k} \cdot \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{1}{k!} \cdot 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \cdots \cdot \left(1 - \frac{k-1}{n}\right) \cdot \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

Wir überlegen uns nun, wie sich dieser Ausdruck für einen festen Wert von k verhält, wenn n sehr große Werte annimmt. Zunächst haben wir

$$\lim_{n\to\infty} 1 - \frac{i}{n} = 1 \quad \text{für alle } i = 1, \cdots, k-1, \text{ also auch}$$

$$\lim_{n\to\infty} 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \cdots \cdot \left(1 - \frac{k-1}{n}\right) = 1.$$

Weiter gilt

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

Diese Gleichung haben wir im zweiten Semester gezeigt, indem wir die Umformungen

$$\left(1 - \frac{\lambda}{n}\right)^n = \exp\left(\ln\left(1 - \frac{\lambda}{n}\right)^n\right) = \exp\left(n \cdot \ln\left(1 - \frac{\lambda}{n}\right)\right) = \exp\left(\frac{\ln\left(1 - \frac{\lambda}{n}\right)}{\frac{1}{n}}\right)$$

benutzt haben. Anschließend konnten wir den Grenzwert mit Hilfe der Regel von L'Hospital berechnen. Schließlich gilt auch

$$\lim_{n\to\infty} 1 - \frac{\lambda}{n} = 1 \quad \text{und also für festes } k \text{ auch} \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1.$$

Damit haben wir für große n und $p = \frac{\lambda}{n}$ jetzt die Näherung

$$P(K = k) = B\Big(n, \frac{\lambda}{n}; k\Big) \approx \frac{\lambda^k}{k!} \cdot \mathrm{e}^{-\lambda} \tag{4.14}$$

gefunden. Die durch Gleichung (4.14) definierte Wahrscheinlichkeits-Verteilung heißt Poisson-Verteilung nach Simón Denis Poisson (1781 - 1840). Wir schreiben

$$X \sim \texttt{Pois}(\lambda),$$

wenn für die Zufalls-Variable X die Gleichung

$$P(X = k) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

gilt.

Jetzt können wir die eingangs gestellte Frage nach der Anzahl der Mitarbeiter des Call-Centers beantworten. Hat das Call-Center m Mitarbeiter, so können auch nur m Anrufer bedient werden. Wir müssen daher m so groß wählen, dass die Wahrscheinlichkeit, dass mehr als m Kunden in einem Zeitintervall anrufen, kleiner als 1% ist. Es gilt

$$P(K > m) = 1 - P(K \le m) = 1 - \sum_{k=0}^{m} \frac{\lambda^k}{k!} \cdot \mathrm{e}^{-\lambda} = 1 - \mathrm{e}^{-\lambda} \cdot \sum_{k=0}^{m} \frac{\lambda^k}{k!}$$

Für $\lambda = n \cdot p = 10\,000\,000 \cdot \frac{5}{10\,000\,000} = 5$ finden wir folgende Zahlen:

$$P(K > 10) \approx 0.0136952688, \quad \text{aber } P(K > 11) \approx 0.0054530920.$$

Die Wahrscheinlichkeit, dass mehr als 11 Kunden in einem Zeitintervall anrufen, liegt also bei 0.5% und damit reichen 11 Mitarbeiter aus. □

Die obigen Betrachtungen des Call-Centers sind aus mehreren Gründen unrealistisch. Das größte Problem in unserer Modellierung des Call-Centers ist die Annahme, dass die einzelnen Kunden ihre Anrufe unabhängig von einander tätigen. Das ist zum Beispiel dann sicher nicht mehr der Fall, wenn beispielsweise durch einen Fehler in der Software zur Erstellung von Rechnungen ein Fehler auftritt, denn in einer solchen Situation werden schlagartig sehr viele Kunden im Call-Center anrufen. Phänomene wie die oben beschriebenen werden in der Theorie der Warteschlangen genauer untersucht, siehe z.B. [GH85].

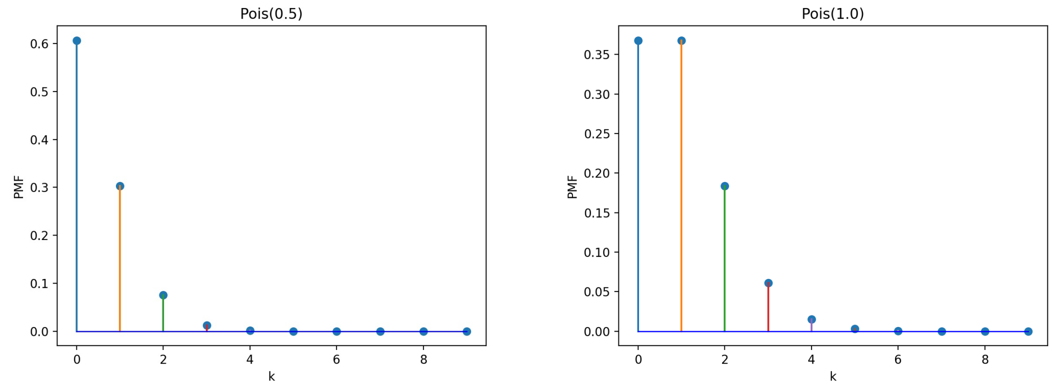

Abbildung 4.1: Die Poisson-Verteilungen $\mathrm{Pois}(1/2)$ und $\mathrm{Pois}(1)$.

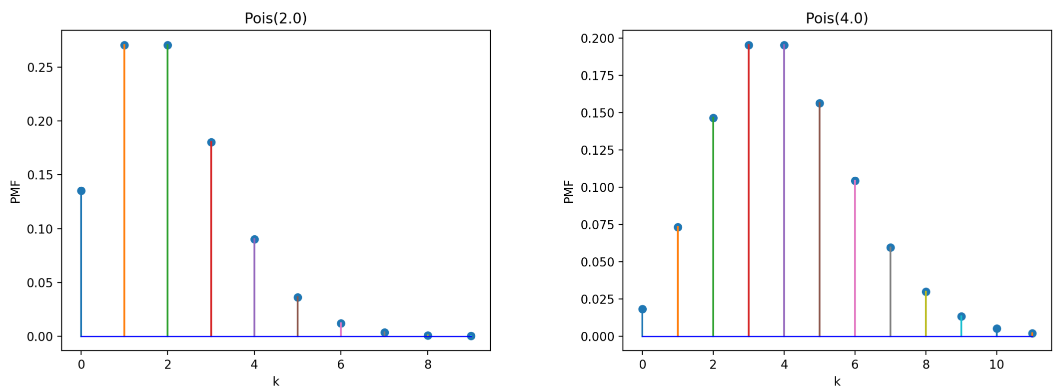

Abbildung 4.2: Die Poisson-Verteilungen $\mathrm{Pois}(2)$ und $\mathrm{Pois}(4)$.

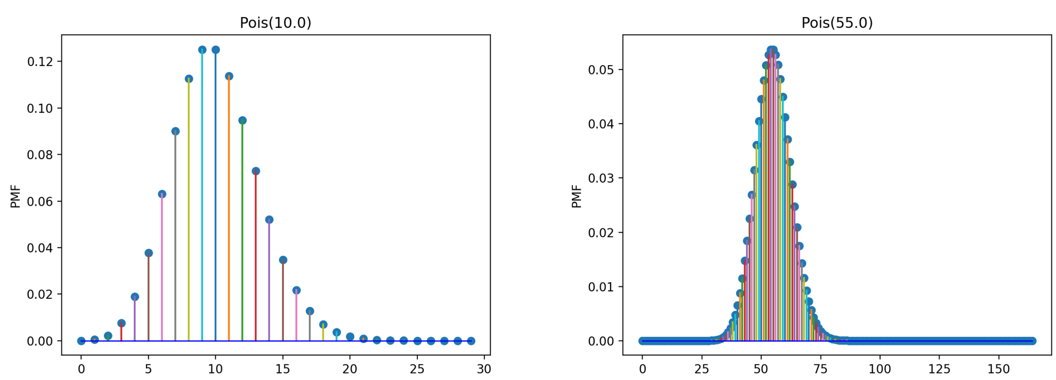

Abbildung 4.3: Die Poisson-Verteilungen $\mathrm{Pois}(10)$ und $\mathrm{Pois}(55)$.

Aufgabe 29: In Singapur gab es im Jahre 1996 insgesamt 76 Exekutionen. Aus tiefer Reue über ihre Missetaten haben alle Delinquenten ihre Organe zur Verfügung gestellt. Bestimmen Sie die Wahrscheinlichkeit, dass der örtliche Organhändler seinem amerikanischen Geschäftspartner in einem zufällig ausgewählten Monat des Jahres 1996 mindestens drei Sätze Innereien feilbieten konnte. Nehmen Sie dazu an, dass die Zufalls-Variable X, welche die Anzahl der Exekutionen in einem Monat angibt, Poisson-verteilt ist.

4.9.1 Erwartungswert und Varianz einer Poisson-verteilten Zufalls-Variable

Es sei $\langle \Omega, 2^\Omega, P\rangle$ ein diskreter Wahrscheinlichkeits-Raum $K : \Omega \to \mathbb{N}$ eine Zufalls-Variable, die Poisson-verteilt ist, für die also

$$P(K = k) = \frac{\lambda^k}{k!} \cdot \mathrm{e}^{-\lambda}$$

gilt. Wir wollen nun den Erwartungswert und die Varianz von K berechnen. Am einfachsten geht das, wenn wir uns erinnern, wie wir die Poisson-Verteilung hergeleitet haben: Die Poisson-Verteilung ist aus der Binomial-Verteilung hergeleitet worden, indem wir dort $\lambda = n \cdot p$ gesetzt und dann n gegen Unendlich laufen gelassen haben. Dann gilt $p = \frac{\lambda}{n}$ und wir haben

$$\lim_{n\to\infty} B\Big(n, \frac{\lambda}{n}; k\Big) = \frac{\lambda^k}{k!} \cdot \mathrm{e}^{-\lambda}.$$

Definieren wir also binomial-verteilte Zufalls-Variablen X_n so, dass

$$P(X_n = k) = B\Big(n, \frac{\lambda}{n}; k\Big),$$

dann sollte gelten

$$E[K] = \lim_{n\to\infty} E[X_n] \quad \text{und} \quad \mathrm{Var}[K] = \lim_{n\to\infty} \mathrm{Var}[X_n].$$

Erwartungswert und Varianz einer binomial-verteilten Zufalls-Variable X haben wir im letzen Abschnitt berechnet und $E[X] = n \cdot p$ und $\mathrm{Var}[X] = n \cdot p \cdot (1 - q)$ gefunden. Also haben wir

$$E[K] = \lim_{n\to\infty} E[X_n] = \lim_{n\to\infty} n \cdot \frac{\lambda}{n} = \lambda.$$

Für die Varianz finden wir

$$\mathrm{Var}[K] = \lim_{n\to\infty} \mathrm{Var}[X_n] = \lim_{n\to\infty} n \cdot \frac{\lambda}{n} \cdot \left(1 - \frac{\lambda}{n}\right) = \lambda \cdot \left(1 - \lim_{n\to\infty} \frac{\lambda}{n}\right) = \lambda$$

Also haben sowohl der Erwartungswert als auch die Varianz einer Poisson-verteilten Zufalls-Variable den Wert λ.

Aufgabe 30: Berechnen Sie den Erwartungswert und die Varianz einer Poisson-verteilten Zufalls-Variable durch direkte Anwendung der Definition von Erwartungswert und Varianz ohne auf die Binomial-Verteilung zurückzugreifen.

4.9.2 Die Summe Poisson-verteilter Zufalls-Variablen

Wir betrachten zwei unabhängige Zufalls-Variablen X_1 und X_2, die beide Poisson-verteilt sind mit den Parametern λ_1 und λ_2, es gilt also

$$P(X_1 = k) = \frac{\lambda_1^k}{k!} \cdot \mathrm{e}^{-\lambda_1} \quad \text{und} \quad P(X_2 = k) = \frac{\lambda_2^k}{k!} \cdot \mathrm{e}^{-\lambda_2}.$$

Wir berechnen die Verteilung der Zufalls-Variable $Z := X_1 + X_2$. Es gilt

$$\begin{aligned}
P(Z=k) &= P(X_1+X_2=k) \\
&= \sum_{i=0}^{k} P(X_1=i \wedge X_2=k-i) \\
&= \sum_{i=0}^{k} P(X_1=i)\cdot P(X_2=k-i) \\
&\quad \text{wegen der Unabhängigkeit von } X_1 \text{ und } X_2 \\
&= \sum_{i=0}^{k} \frac{\lambda_1^i}{i!}\cdot \mathrm{e}^{-\lambda_1}\cdot \frac{\lambda_2^{k-i}}{(k-i)!}\cdot \mathrm{e}^{-\lambda_2} \\
&= \mathrm{e}^{-(\lambda_1+\lambda_2)}\cdot \sum_{i=0}^{k} \frac{\lambda_1^i}{i!}\cdot \frac{\lambda_2^{k-i}}{(k-i)!} \\
&= \mathrm{e}^{-(\lambda_1+\lambda_2)}\cdot \frac{1}{k!}\cdot \sum_{i=0}^{k} \frac{k!}{i!\cdot(k-i)!}\cdot \lambda_1^i\cdot\lambda_2^{k-i} \\
&= \mathrm{e}^{-(\lambda_1+\lambda_2)}\cdot \frac{1}{k!}\cdot \sum_{i=0}^{k} \binom{k}{i}\cdot \lambda_1^i\cdot\lambda_2^{k-i} \\
&= \mathrm{e}^{-(\lambda_1+\lambda_2)}\cdot \frac{1}{k!}\cdot (\lambda_1+\lambda_2)^k
\end{aligned}$$

Dabei haben wir im letzten Schritt den Binomischen Lehrsatz

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

verwendet. Unsere Herleitung zeigt, dass für unabhängige Poisson-verteilte Zufalls-Variablen mit den Parametern λ_1 und λ_2 auch die Summe X_1+X_2 Poisson-verteilt ist und zwar mit dem Parameter $\lambda_1+\lambda_2$.

Aufgabe 31: Zwei Call-Center erhalten im Schnitt jeweils fünf Anrufe pro Zeitintervall. Wir hatten in einer früheren Aufgabe gesehen, dass dann 11 Mitarbeiter pro Call-Center ausreichen, um mit einer Wahrscheinlichkeit von 99% garantieren zu können, dass alle eingehenden Anrufe von einem freien Mitarbeiter angenommen werden können. Nun werden die beiden Call-Center zusammengelegt. In dem neuen Call-Center kommen im Schnitt also 10 Anrufe pro Minute an. Berechnen Sie die Anzahl der Mitarbeiter, die in dem neuen Call-Center benötigt werden, um garantieren zu können, dass mit einer Wahrscheinlichkeit von 99% alle eingehenden Anrufe von einem freien Mitarbeiter angenommen werden können. ◇

4.10 Kovarianz

Die folgende Definition verallgemeinert den Begriff der Varianz.

Definition 30 (Kovarianz)
Ist $\langle\Omega, 2^\Omega, P\rangle$ ein diskreter Wahrscheinlichkeits-Raum und sind

$$X:\Omega\to\mathbb{R} \quad \textit{und} \quad Y:\Omega\to\mathbb{R}$$

zwei Zufalls-Variablen, so definieren wir die Kovarianz $\mathrm{Cov}[X,Y]$ der Zufalls-Variablen X und Y als den Erwartungswert der Zufalls-Variable

$$\omega \mapsto \big(X(\omega)-E[X]\big)\cdot\big(Y(\omega)-E[Y]\big),$$

es gilt also

$$\mathrm{Cov}[X,Y] := E\big[(X - E[X]) \cdot (Y - E[Y])\big].$$

Haben die Wertebereiche von X und Y die Form

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\} = \{x_n \mid n \in \mathbb{N}\} \quad \textit{und} \quad Y(\Omega) = \{Y(\omega) \mid \omega \in \Omega\} = \{y_n \mid n \in \mathbb{N}\}$$

und setzen wir zur Abkürzung $\mu_X = E[X]$ und $\mu_Y = E[Y]$, so können wir die Kovarianz auch durch die Formel

$$\mathrm{Cov}[X,Y] = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P(X = x_m \wedge Y = y_n) \cdot (x_m - \mu_X) \cdot (y_n - \mu_Y)$$

berechnen. □

Satz 31 *Ist $\langle \Omega, 2^\Omega, P \rangle$ ein diskreter Wahrscheinlichkeits-Raum und sind*

$$X : \Omega \to \mathbb{R} \quad \textit{und} \quad Y : \Omega \to \mathbb{R}$$

zwei unabhängige Zufalls-Variablen, so gilt

$$\mathrm{Cov}[X,Y] = 0.$$ ◇

Aufgabe 32: Beweisen Sie den letzten Satz. ◇

Der folgende Satz liefert eine alternative Möglichkeit die Varianz zu berechnen.

Satz 32 (Verschiebungs-Satz) *Ist $\langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum und sind*

$$X : \Omega \to \mathbb{R} \quad \textit{und} \quad Y : \Omega \to \mathbb{R}$$

zwei Zufalls-Variablen, so gilt

$$\mathrm{Cov}[X,Y] = E[X \cdot Y] - E[X] \cdot E[Y].$$

Aufgabe 33: Beweisen Sie den letzten Satz. ◇

Satz 33 *Ist $\langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum und sind*

$$X : \Omega \to \mathbb{R} \quad \textit{und} \quad Y : \Omega \to \mathbb{R}$$

zwei Zufalls-Variablen, sind weiter $a, b \in \mathbb{R}$, so lässt sich die Varianz der Zufalls-Variable

$$Z := a \cdot X + b \cdot Y$$

nach der Formel

$$\mathrm{Var}[a \cdot X + b \cdot Y] = a^2 \cdot \mathrm{Var}[X] + b^2 \cdot \mathrm{Var}[Y] + 2 \cdot a \cdot b \cdot \mathrm{Cov}[X,Y]$$

berechnen.

Beweis: Dieser Beweis lässt sich durch einfaches Nachrechnen führen:

$$
\begin{aligned}
& \operatorname{Var}[a \cdot X + b \cdot Y] \\
= \; & E[(a \cdot X + b \cdot Y)^2] - (E[a \cdot X + b \cdot Y])^2 \\
= \; & a^2 \cdot E[X^2] + 2 \cdot a \cdot b \cdot E[X \cdot Y] + b^2 \cdot E[Y^2] - (a \cdot E[X] + b \cdot E[Y])^2 \\
= \; & a^2 \cdot E[X^2] + 2 \cdot a \cdot b \cdot E[X \cdot Y] + b^2 \cdot E[Y^2] - a^2 \cdot E[X]^2 - 2 \cdot a \cdot b \cdot E[X] \cdot E[Y] - b^2 \cdot E[Y]^2 \\
= \; & a^2 \cdot (E[X^2] - E[X]^2) + b^2 \cdot (E[Y^2] - E[Y]^2) + 2 \cdot a \cdot b \cdot (E[X \cdot Y] - E[X] \cdot E[Y]) \\
= \; & a^2 \cdot \operatorname{Var}[X] + b^2 \cdot \operatorname{Var}[Y] + 2 \cdot a \cdot b \cdot \operatorname{Cov}[X, Y].
\end{aligned}
$$

□

Kapitel 5

Stetige Zufalls-Variablen

Bisher haben wir nur mit diskreten Wahrscheinlichkeits-Räumen $\langle \Omega, 2^\Omega, P \rangle$ gearbeitet, Ω war dabei eine höchstens abzählbare Menge diskreter Ergebnisse. Bei vielen Zufalls-Experimenten ist das Ergebnis aber nicht eine natürliche Zahl, sondern eine beliebige reelle Zahl oder sogar ein Tupel reeller Zahlen. In diesen Fällen hat der Wahrscheinlichkeits-Raum die Form

$$\langle \Omega, \mathfrak{A}, P \rangle.$$

Die Komponenten dieses Tripels haben dabei die folgenden Eigenschaften:

1. Die erste Komponente Ω bezeichnen wir als den Ergebnis-Raum. Es gilt

   $\Omega \subseteq \mathbb{R}$ oder $\Omega \subseteq \mathbb{R}^n$ für ein $n \in \mathbb{N}$.

2. Die zweite Komponente $\mathfrak{A}$ bezeichnen wir als den Ereignis-Raum. Es gilt

   $$\mathfrak{A} \subseteq 2^\Omega.$$

   Der Ereignisraum $\mathfrak{A}$ ist jetzt nicht mehr mit der Potenzmenge von Ω identisch, sondern ist nur noch eine Teilmenge dieser Potenzmenge. Der Grund dafür ist, vereinfacht gesagt, dass die Potenzmenge von Ω bestimmte Mengen enthält, die so kompliziert sind, dass wir diesen Mengen keine Wahrscheinlichkeit zuordnen können. Daher können diese Mengen auch nicht Elemente unseres Ereignisraums sein. Diese Mengen sind die so genannten nicht messbaren Mengen. Die Existenz solcher pathologischen Mengen ist Gegenstand des Satzes von Vitali.

   Um mit dem Ereignisraum arbeiten zu können müssen wir fordern, dass der Ereignis-Raum bestimmten Abschuss-Eigenschaften genügt. Genauer muss gelten:

   (a) $\Omega \in \mathfrak{A}$.
   (b) Aus $A \in \mathfrak{A}$ folgt $A^c \in \mathfrak{A}$.
   (c) Gilt für alle $n \in \mathbb{N}$ dass $A_n \in \mathfrak{A}$ ist, dann gilt auch

   $$\bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{A}.$$

   Ein Mengensystem, dass diesen Abschuss-Eigenschaften genügt, bezeichnen wir auch als eine σ-Algebra.

   Wenn $\Omega \subseteq \mathbb{R}$ ist, dann reicht es für unsere Zwecke aus, wenn $\mathfrak{A}$ einerseits alle Intervalle $[a, b]$ enthält, für die $[a, b] \subseteq \Omega$ ist und andererseits die oben angegebenen Abschuss-Eigenschaften erfüllt sind.

3. Die dritte Komponente P ist eine Abbildung, die den selben Axiomen wie im diskreten Fall genügt:

   (a) $0 \leq P(A) \leq 1 \quad$ für alle $A \subseteq \mathfrak{A}$.

(b) $P(\emptyset) = 0$.

(c) $P(\Omega) = 1$.

(d) Ist $(A_n)_{n\in\mathbb{N}}$ eine Folge paarweise disjunkter Mengen aus $\mathfrak{A}$, gilt also

$$\forall n \in \mathbb{N} : A_n \in \mathfrak{A} \quad \text{und} \quad \forall m, n \in \mathbb{N} : m \neq n \to A_m \cap A_n = \emptyset$$

so folgt $\displaystyle P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} P(A_n)$.

Die Funktion P bezeichnen wir als die Wahrscheinlichkeits-Maß.

Um Zufalls-Experimente beschreiben zu können, führen wir den Begriff der Zufalls-Variable ein, wobei wir uns auf den eindimensionalen Fall beschränken. Eine Zufalls-Variable X ist eine Abbildung

$$X : \Omega \to \mathbb{R}.$$

Hat Ω die Form $\Omega = [a, b]$ mit

1. $a \in \mathbb{R} \cup \{-\infty\}$,
2. $b \in \mathbb{R} \cup \{+\infty\}$,
3. $a < b$,

so ist die (kumulative) Verteilungs-Funktion[1] F_X der Zufalls-Variable X definiert als

$$F_X : [a, b] \to [0, 1] \quad \text{mit } F_X(x) := P(X \leq x).$$

Wenn die Verteilungs-Funktion F_X differenzierbar ist, dann nennen wir X eine stetige Zufalls-Variable. Aus der Verteilungs-Funktion lässt sich sofort die Wahrscheinlichkeit eines Ereignisses der Form

$$\big\{\omega \in \Omega \;\big|\; u < X(\omega) \leq v\big\}$$

berechnen, denn es gilt

$$P\big(\big\{\omega \in \Omega \;\big|\; u < X(\omega) \leq v\big\}\big) = F_X(v) - F_X(u).$$

Anstelle der Verteilungs-Funktion werden wir oft mit der Wahrscheinlichkeits-Dichte p_X arbeiten. Die Wahrscheinlichkeits-Dichte

$$p_X : \Omega \to \mathbb{R}$$

wird als Grenzwert definiert:

$$p_X(x) := \lim_{h\to 0} \frac{F_X(x+h) - F_X(x)}{h} = \frac{\mathrm{d}F_X}{\mathrm{d}x}.$$

Nach dem Hauptsatz der Differenzial- und Integral-Rechnung kann die Verteilungs-Funktion auf die Wahrscheinlichkeits-Dichte zurückgeführt werden. Hat Ω die Form $[a, b]$, so gilt:

$$F_X(x) = \int_a^x p_X(t)\,\mathrm{d}t.$$

Beispiel: Der einfachste Fall einer stetigen Zufalls-Variable ist eine gleichverteilte Zufalls-Variable. Es seien $a, b \in \mathbb{R}$ und der Wahrscheinlichkeits-Raum habe die Form

$$\langle [a, b], \mathfrak{A}, P\rangle.$$

Die Wahrscheinlichkeits-Verteilung $P(A)$ sei für ein Intervall $[u, v] \subseteq [a, b]$ durch das Verhältnis der Längen definiert:

$$P\big([u, v]\big) = \frac{v-u}{b-a}$$

[1] Das Attribut kumulativ lassen wir im folgenden weg.

Weiter sei die Zufalls-Variable $X : [a, b] \to \mathbb{R}$ trivial definiert durch $X(x) := x$. Die Verteilungs-Funktion F_X der Zufalls-Variable X hat dann die Form

$$F_X(x) = P(X \leq x) = P\bigl([a, x]\bigr) = \frac{x-a}{b-a}.$$

Damit ergibt sich die Wahrscheinlichkeits-Dichte p_X der Zufalls-Variable X als

$$p_X(x) = \frac{\mathrm{d}F_X}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x-a}{b-a}\right) = \frac{1}{b-a}.$$

Sind $a, b \in \mathbb{R}$ mit $a < b$ und ist $X : \Omega \to \{a, b\}$ eine Zufalls-Variable, für die

$$F_X(x) = \frac{x-a}{b-a}$$

gilt, so schreiben wir

$$X \sim \mathtt{Unif}(a, b)$$

und sagen, dass die Zufalls-Variable X auf dem Intervall $[a, b]$ gleichverteilt ist.

Aufgabe 34: In einer fremden Stadt kommen Sie zu einem zufälligen Zeitpunkt an eine Bushaltestelle, an der kein Fahrplan hängt. Sie wissen allerdings, dass Busse im Intervall von 30 Minuten ankommen. Berechnen Sie die Wahrscheinlichkeit, dass Sie länger als 20 Minuten auf den Bus warten müssen.

Lösung: Der Wahrscheinlichkeits-Raum hat die Form

$$\langle [0, 30], \mathfrak{A}, P\rangle$$

und die Zufalls-Variable X, welche die Zahl in Minuten angibt, bis der nächste Bus kommt, ist gleichverteilt. Also ist die Verteilungs-Funktion F_X durch

$$F_X(x) = \frac{x-0}{30-0} = \frac{x}{30}$$

gegeben. Damit haben wir

$$\begin{aligned} P(X \geq 20) &= P(20 \leq X \leq 30) \\ &= F_X(30) - F_X(20) \\ &= \frac{30}{30} - \frac{20}{30} \\ &= 1 - \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

Die Wahrscheinlichkeit, dass Sie länger als 20 Minuten warten müssen, beträgt also etwa 33%. ⋄

Wir betrachten nun ein Beispiel, bei dem zwar die Wahrscheinlichkeits-Verteilung P gleichmäßig ist, in dem aber die Zufalls-Variable selber nicht mehr gleichverteilt ist.

Beispiel: Der Feuerleitrechner eines Artillerie-Geschützes ist durch Feindeinwirkung beschädigt und produziert für den Abschusswinkel φ nur noch gleichverteilte Zufallszahlen aus dem Intervall $\bigl[0, \frac{\pi}{4}\bigr]$. Unser Wahrscheinlichkeits-Raum hat also die Form

$$\Bigl\langle \bigl[0, \frac{\pi}{4}\bigr], \mathfrak{A}, P\Bigr\rangle \quad \text{mit } P\bigl([u, v]\bigr) = \frac{v-u}{\frac{\pi}{4}} = \frac{4}{\pi} \cdot (v-u)$$

Die Zufalls-Variable, an der wir interessiert sind, ist in diesem Fall die Schussweite $R(\varphi)$, die (unter Vernachlässigung des Luftwiderstands) nach der Formel

$$R(\varphi) = \frac{v_0^2}{g} \cdot \sin(2 \cdot \varphi) \tag{5.1}$$

berechnet werden kann. Dabei bezeichnet v_0 die Geschwindigkeit, die das Geschoss beim Austritt aus der Mündung des Geschützes besitzt. Für die weitere Rechnung nehmen wir $v_0 = 300\,\mathrm{m/s}$ an. Die Konstante g steht für die Erdbeschleunigung, die in unseren Breiten etwa den Wert $9.8\,\mathrm{m/s^2}$ besitzt. Wir wollen die Verteilungs-Funktion der Zufalls-Variable "Schussweite" R berechnen. Nach Gleichung (5.1) nimmt die Schussweite ihren Maximalwert

$$R_{\max} = \frac{v_0^2}{g}$$

bei $\varphi = \frac{\pi}{4}$ an. Für die zugehörige Verteilungs-Funktion F_R finden wir

$$\begin{aligned}
F_R(x) &= P(R \leq x) \\
&= P\left(\frac{v_0^2}{g} \cdot \sin(2 \cdot \varphi) \leq x\right) \\
&= P\left(\sin(2 \cdot \varphi) \leq x \cdot \frac{g}{v_0^2}\right) \\
&= P\left(\varphi \leq \frac{1}{2} \cdot \arcsin\Big(x \cdot \frac{g}{v_0^2}\Big)\right) \\
&= \frac{4}{\pi} \cdot \frac{1}{2} \cdot \arcsin\Big(x \cdot \frac{g}{v_0^2}\Big) \\
&= \frac{2}{\pi} \cdot \arcsin\Big(\frac{x}{R_{\max}}\Big)
\end{aligned}$$

Die zugehörige Wahrscheinlichkeits-Dichte finden wir, indem wir die Verteilungs-Funktion nach x differenzieren. Es ergibt sich

$$p_R(x) = \frac{\mathrm{d}F_f}{\mathrm{d}x} = \frac{2}{\pi \cdot R_{\max}} \cdot \frac{1}{\sqrt{1 - \left(\frac{x}{R_{\max}}\right)^2}}.$$

Diese Wahrscheinlichkeits-Dichte ist keinesweg gleichverteilt. Abbildung 5.1 zeigt den Graphen der Wahrscheinlichkeits-Dichte unter der Annahme $v_0 = 300\,\mathrm{m/s}$. An der rechten Intervall-Grenze divergiert die Wahrscheinlichkeits-Dichte, es gilt

$$\lim_{x \to R_{\max}} p_R(x) = \infty.$$

Abbildung 5.1 zeigt die Verteilungs-Funktion und die Wahrscheinlichkeits-Dichte der Zufalls-Variable R.

Der nächste Satz verallgemeinert das obige Beispiel.

Satz 34 (Variablen-Transformation) *Es sei*

1. $\mathcal{W} = \langle \Omega, \mathfrak{A}, P \rangle$ *ein Wahrscheinlichkeits-Raum,*
2. $X : \Omega \to [a, b]$ *eine stetige Zufalls-Variable,*
3. $\varphi : [a, b] \to [c, d]$ *eine streng monoton wachsende Funktion*
4. $\psi : [c, d] \to [a, b]$ *die Umkehrfunktion von* φ*, also*
   $$\psi\big(\varphi(t)\big) = t \quad \text{für alle } t \in [a, b],$$
5. $Y : \Omega \to [c, d]$ *eine Zufalls-Variable, die definiert ist durch*
   $$Y(\omega) := \varphi\big(X(\omega)\big).$$

Dann gilt für die Wahrscheinlichkeits-Dichten p_X *und* p_Y *die Beziehung*

$$p_Y(t) = p_X\big(\psi(t)\big) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t}.$$

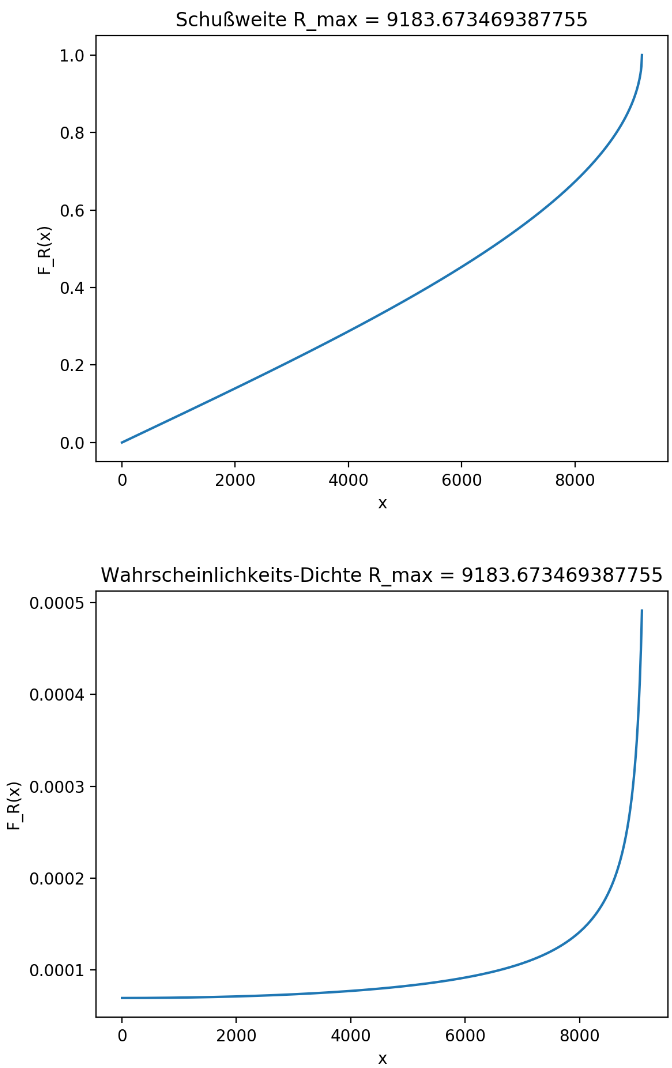

Abbildung 5.1: Verteilungs-Funktion und Wahrscheinlichkeits-Dichte der SchussWeite.

Beweis: Wir betrachten die Verteilungs-Funktion der Zufalls-Variable Y. Es gilt

$$\begin{array}{lcll} F_Y(t) & = & P(Y \leq t) & \\ & = & P(\varphi \circ X \leq t) & \text{nach Definition von } Y \\ & = & P\big(X \leq \psi(t)\big) & \text{denn } \psi \text{ ist Umkehrfunktion von } \varphi \\ & = & F_X\big(\psi(t)\big) & \text{nach Definition der Verteilungs-Funktion von } X \end{array}$$

Wir haben also $F_Y(t) = F_X\big(\psi(t)\big)$. Wenn wir diese Gleichung nach t differenzieren, erhalten wir die Behauptung mit Hilfe der Ketten-Regel:

$$p_Y(t) = \frac{\mathrm{d}}{\mathrm{d}t} F_Y(t) = \frac{\mathrm{d}}{\mathrm{d}t} F_X\big(\psi(t)\big) = \frac{\mathrm{d}F_X}{\mathrm{d}\psi}\big(\psi(t)\big) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t} = p_X\big(\psi(t)\big) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t}. \qquad \square$$

5.1 Erwartungswert und Varianz stetiger Zufalls-Variablen

Es sei $\langle [a,b], \mathfrak{A}, P\rangle$ ein Wahrscheinlichkeits-Raum und

$$X : [a,b] \to [c,d]$$

sei eine stetige Zufalls-Variable mit Wahrscheinlichkeits-Dichte p_X. Wir überlegen uns, wie wir die Definition des Erwartungswerts auf den kontinuierlichen Fall übertragen können. Für eine diskrete Zufalls-Variable Y, deren Wertebereich sich als Menge der Form

$$Y(\Omega) = \bigl\{y_k \bigm| k \in \{1, \cdots, n\}\bigr\}$$

schreiben lässt, hatten wir seinerzeit die Formel

$$E[Y] = \sum_{k=1}^{n} y_k \cdot P(Y = y_k)$$

gefunden. Da der Wertebereich der Zufalls-Variable X kontinuierlich ist, können wir versuchen, die möglichen Werte von X in viele kleine Intervalle der Form

$$[x_k, x_{k+1}] \quad \text{mit } x_k = c + \frac{k}{n}\cdot(d-c) \text{ für } k = 0, \cdots, n-1$$

aufzuteilen. Jedes dieser Intervalle hat die Länge

$$h = x_{k+1} - x_k = \frac{k+1}{n}\cdot(d-c) - \frac{k}{n}\cdot(d-c) = \frac{d-c}{n}.$$

Wenn n groß ist, wird die Länge h diese Intervalle sehr klein, so dass die Zahlen, die in dem selben Intervall liegen, nahezu gleich sind. Dann können wir den Erwartungswert der Zufalls-Variable X näherungsweise durch

$$E[X] \approx \sum_{k=1}^{n} x_k \cdot P(x_{k-1} \le X \le x_k)$$

berechnen. Nun gilt

$$P(x_{k-1} \le X \le x_k) = \int_{x_{k-1}}^{x_k} p_X(x)\,\mathrm{d}x.$$

Also haben wir für den Erwartungswert

$$E[X] \approx \sum_{k=1}^{n} x_k \cdot \int_{x_{k-1}}^{x_k} p_X(x)\,\mathrm{d}x$$

Hier können wir die Konstante x_k in das Intergral hineinziehen und durch die Variable x ersetzen, denn innerhalb eines Intervalls $[x_{k-1}, x_k]$ ändert sich x kaum, wenn n groß und damit die Intervall-Länge $x_k - x_{k-1} = \frac{d-c}{n}$ klein ist. Dann haben wir

$$E[X] \approx \sum_{k=1}^{n} \int_{x_{k-1}}^{x_k} x \cdot p_X(x)\,\mathrm{d}x = \int_c^d x \cdot p_X(x)\,\mathrm{d}x$$

Diese Überlegungen motivieren im Fall stetiger Zufalls-Variablen die Definition des Erwartungswerts einer stetigen Zufalls-Variable X durch die Formel

$$E[X] := \int_c^d x \cdot p_X(x)\,\mathrm{d}x.$$

Hier bezeichnen c und d den minimalen bzw. maximalen Wert, den die Zufalls-Variable X annehmen kann. Die Varianz einer stetigen Zufalls-Variable wird nun wie früher durch

$$\mathrm{Var}[X] := E\Bigl[(X - E[X])^2\Bigr]$$

definiert. Setzen wir $\mu := E[X]$, so heißt dies

$$\mathrm{Var}[X] = \int_c^d (x-\mu)^2 \cdot p_X(x)\,\mathrm{d}x.$$

Aufgabe 35: Berechnen Sie Erwartungswert und Varianz der Zufalls-Variable X, falls für die Verteilungs-Funktion F_X gilt

$$F_X(t) = \begin{cases} 1 - \mathrm{e}^{-\lambda \cdot t} & \text{falls } t \geq 0; \\ 0 & \text{sonst.} \end{cases}$$ ◇

Lösung: Zunächst müssen wir die Wahrscheinlichkeits-Dichte p_X berechnen. Es gilt

$$p_X(t) = \frac{\mathrm{d}}{\mathrm{d}t}(1 - \mathrm{e}^{-\lambda\cdot t}) = \lambda \cdot \mathrm{e}^{-\lambda \cdot t} \quad \text{für } x \geq 0.$$

Damit erhalten wir für den Erwartungswert

$$\begin{aligned}
E[X] &= \int_0^\infty x \cdot \lambda \cdot \mathrm{e}^{-\lambda\cdot x}\,\mathrm{d}x \\
&\quad \text{partielle Integration: } u(x) = x \text{ und } v'(x) = \lambda \cdot \mathrm{e}^{-\lambda\cdot x} \\
&\quad \text{also: } u'(x) = 1 \text{ und } v(x) = -\mathrm{e}^{-\lambda\cdot x} \\
&= \underbrace{-x \cdot \mathrm{e}^{-\lambda\cdot x}\Big|_0^\infty}_{0} + \int_0^\infty \mathrm{e}^{-\lambda\cdot x}\,\mathrm{d}x \\
&\quad \text{denn } \lim_{x\to\infty} x \cdot \mathrm{e}^{-\lambda\cdot x} = 0 \\
&= -\frac{1}{\lambda} \cdot \mathrm{e}^{-\lambda\cdot x}\Big|_0^\infty \\
&= \frac{1}{\lambda}
\end{aligned}$$

Für die Varianz finden wir entsprechend

$$\begin{aligned}
\mathrm{Var}[X] &= \int_0^\infty \Big(x - \frac{1}{\lambda}\Big)^2 \cdot \lambda \cdot \mathrm{e}^{-\lambda\cdot x}\,\mathrm{d}x \\
&\quad \text{partielle Integration: } u(x) = (x - \tfrac{1}{\lambda})^2 \text{ und } v'(x) = \lambda \cdot \mathrm{e}^{-\lambda\cdot x} \\
&\quad \text{also } u'(x) = 2 \cdot (x - \tfrac{1}{\lambda}) \text{ und } v(x) = -\mathrm{e}^{-\lambda\cdot x} \\
&= -\Big(x - \frac{1}{\lambda}\Big)^2 \cdot \mathrm{e}^{-\lambda\cdot x}\Big|_0^\infty + 2 \cdot \int_0^\infty \Big(x - \frac{1}{\lambda}\Big) \cdot \mathrm{e}^{-\lambda\cdot x}\,\mathrm{d}x \\
&= \frac{1}{\lambda^2} + 2 \cdot \int_0^\infty \Big(x - \frac{1}{\lambda}\Big) \cdot \mathrm{e}^{-\lambda\cdot x}\,\mathrm{d}x \\
&\quad \text{partielle Integration: } u(x) = x - \tfrac{1}{\lambda} \text{ und } v'(x) = \mathrm{e}^{-\lambda\cdot x} \\
&\quad \text{also } u'(x) = 1 \text{ und } v(x) = -\tfrac{1}{\lambda} \cdot \mathrm{e}^{-\lambda\cdot x} \\
&= \frac{1}{\lambda^2} - \left(2 \cdot \Big(x - \frac{1}{\lambda}\Big) \cdot \frac{1}{\lambda} \cdot \mathrm{e}^{-\lambda\cdot x}\Big|_0^\infty\right) + \frac{2}{\lambda} \cdot \int_0^\infty \mathrm{e}^{-\lambda\cdot x}\,\mathrm{d}x \\
&= \frac{1}{\lambda^2} - \frac{2}{\lambda^2} - \left(\frac{2}{\lambda^2} \cdot \mathrm{e}^{-\lambda\cdot x}\Big|_0^\infty\right) \\
&= \frac{1}{\lambda^2} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2} \\
&= \frac{1}{\lambda^2}.
\end{aligned}$$ □

Bemerkung: Ein Zufalls-Variable $X : \Omega \to \mathbb{R}$, deren Verteilungs-Funktion die Form

$$F_X(t) = \begin{cases} 1 - \mathrm{e}^{-\lambda \cdot t} & \text{falls } t \geq 0, \\ 0 & \text{sonst} \end{cases}$$

hat, nennen wir eine exponential-verteilte Zufalls-Variable mit Parameter λ und schreiben

$$X \sim \mathrm{Exp}(\lambda).$$ ◇

Definition 35 (Logistische Verteilung) *Eine Zufalls-Variable X ist logistisch mit den Parametern μ und σ verteilt, falls die Verteilungs-Funktion F_X die Form*

$$F_X(x) = \frac{1}{1 + \exp\Bigl(-\dfrac{x-\mu}{s}\Bigr)}$$

hat. In diesem Fall schreiben wir $X \sim \mathrm{Log}(\mu, s)$. ◇

Bemerkung: Die logistische Verteilung spielt später in der logistischen Regression und als Aktivierungsfunktion bei den Neuronen eines neuronalen Netzes eine wichtige Rolle. Abbildung 5.3 auf Seite 83 zeigt die Verteilungs-Funktion F_X einer Zufalls-Variable X mit $X \sim \mathrm{Log}(1, 0$, die zugehörige Wahrscheinlichkeits-Dichte p_X ist Abbildung 5.3 auf Seite 83 geplottet. ◇

Aufgabe 36: Berechnen Sie den Erwartungswert und die Varianz einer logistisch verteilten Zufalls-Variable. Beachten Sie dabei die folgenden **Hinweise**.

(a) Es ist am einfachsten, wenn Sie zunächst den Erwartungs-Wert und die Varianz einer Zufalls-Variable X berechnen, für die $X \sim \mathrm{Log}(0, 1)$ gilt. Ist Y dann eine Zufalls-Variable, für die $Y \sim \mathrm{Log}(\mu, s)$ gilt, so zeigt der Satz 34, dass $Y = \varphi \circ X$ ist, falls Sie

$$\varphi(x) := s \cdot x + \mu$$

definieren. Mit Hilfe des Verschiebungs-Satzes können Sie anschließend $E[Y]$ und $\mathrm{Var}[Y]$ bestimmen.

(b) Bei der Berechnung von Erwartungs-Wert und Varianz der Zufalls-Variable X sollten Sie ausnutzen, dass die Funktion p_X eine gerade Funktion ist, es gilt also

$$p_X(-x) = p_X(x).$$

In der Analysis hatten wir gesehen, dass

$$\int\limits_{-a}^{a} f(x)\mathrm{d}x = 2 \cdot \int\limits_{0}^{a} f(x)\mathrm{d}x$$

ist, falls die Funktion f gerade ist. Ist die Funktion f hingegen ungerade, gilt also $f(-x) = -f(x)$, so haben wir

$$\int\limits_{-a}^{a} f(x)\mathrm{d}x = 0.$$

(c) Die Berechnung des Erwartungswerts $E[X]$ verläuft ohne Komplikationen. Bei der Berechnung der Varianz sollten Sie sich an die Formel für die geometrische Reihe erinnern:

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q} \qquad \text{falls } |q| < 1 \text{ ist.}$$

Leiten wir diese Formel auf beiden Seiten nach q ab, so erhalten wir

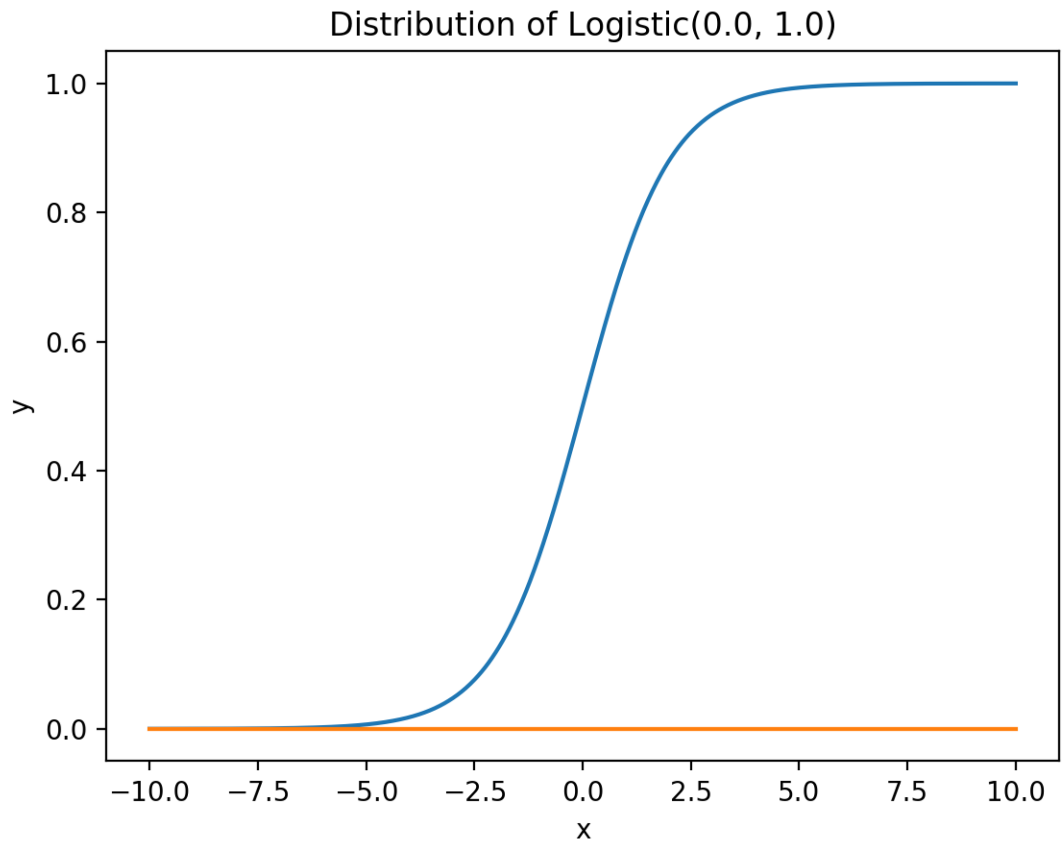

Abbildung 5.2: Verteilungs-Funktion und Wahrscheinlichkeits-Dichte einer logistisch verteilten Zufalls-Variable.

$$\sum_{n=1}^{\infty} n \cdot q^{n-1} = \frac{1}{(1-q)^2} \qquad \text{falls } |q| < 1 \text{ ist.}$$

Diese Formel müssen Sie an geeigneter Stelle rückwärts, also von rechts nach links, anwenden.

(d) Schließlich sollten Sie sich daran erinnern, dass wir am Ende des zweiten Semesters die Formel

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

bewiesen haben. Aus dieser Formel lässt sich die Summenformel

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12}$$

herleiten und $\mathrm{Var}[X]$ kann auf diese Summe zurück geführt werden. $\diamond$

Satz 36 *Es sei $X : \Omega \to [a, b]$ eine Zufalls-Variable mit Wahrscheinlichkeits-Dichte $p_X(t)$. Dann kann die Varianz $\mathrm{Var}[X]$ wie folgt berechnet werden:*

$$\mathrm{Var}[X] = E\big[X^2\big] - \big(E[X]\big)^2.$$

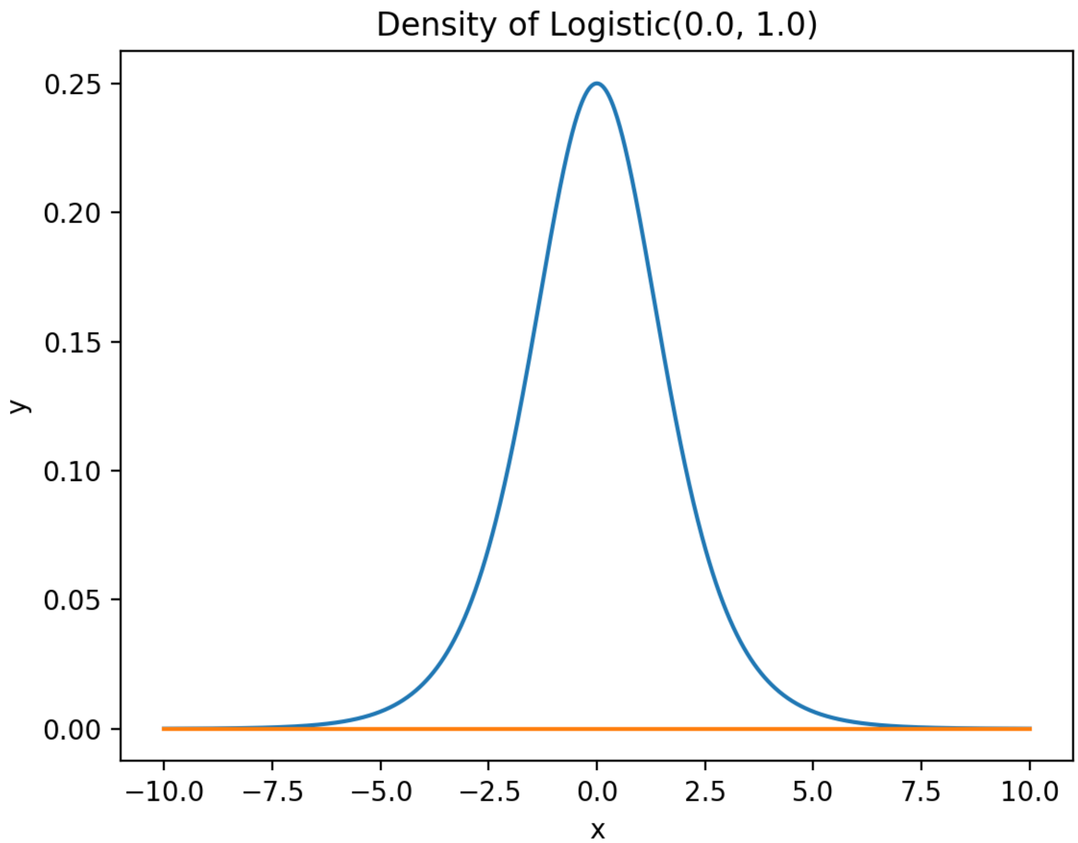

Abbildung 5.3: Verteilungs-Funktion und Wahrscheinlichkeits-Dichte einer logistisch verteilten Zufalls-Variable.

Beweis: Es sei $\mu := E[X] = \displaystyle\int_a^b x \cdot p_X(x)\, \mathrm{d}x$. Dann gilt

$$\begin{aligned}
\mathrm{Var}[X] &= E\Big[\big(X - E[X]\big)^2\Big] \\
&= \int_a^b (x-\mu)^2 \cdot p_X(x)\, \mathrm{d}x \\
&= \int_a^b \big(x^2 - 2\cdot\mu\cdot x + \mu^2\big) \cdot p_X(x)\, \mathrm{d}x \\
&= \int_a^b x^2 \cdot p_X(x)\, \mathrm{d}x \; - \; 2\cdot\mu\cdot \underbrace{\int_a^b x\cdot p_X(x)\, \mathrm{d}x}_{E[X]} + \mu^2 \cdot \underbrace{\int_a^b p_X(x)\, \mathrm{d}x}_{1} \\
&= E\big[X^2\big] - 2\cdot\mu^2 + \mu^2 \\
&= E\big[X^2\big] - \big(E[X]\big)^2
\end{aligned}$$

□

Satz 37 (Verhalten des Erwartungswerts bei Variablen-Transformation) *Es sei*

1. $\mathcal{W} = \langle \Omega, \mathfrak{A}, P \rangle$ *ein Wahrscheinlichkeits-Raum,*
2. $X : \Omega \to [a, b]$ *eine stetige Zufalls-Variable,*
3. $\varphi : [a, b] \to [c, d]$ *eine stetige Funktion*
4. $Y : \Omega \to [c, d]$ *eine Zufalls-Variable, die definiert ist durch*

$$Y := \varphi \circ X, \quad \textit{also} \quad Y(\omega) := \varphi\bigl(X(\omega)\bigr) \quad \textit{für alle } \omega \in \Omega.$$

Dann gilt für den Erwartungswert von Y

$$E[Y] = E[\varphi \circ X] = \int_a^b \varphi(x) \cdot p_X(x)\ dx$$

Beweis: Wir setzen zusätzlich voraus, dass die Funktion φ streng monoton steigend ist und dass $\psi : [c, d] \to [a, b]$ die Umkehrfunktion von φ ist. Nach dem Satz über Variablen-Transformation gilt dann

$$p_Y(t) = p_X\bigl(\psi(t)\bigr) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t}.$$

Nach Definition des Erwartungswerts gilt

$$\begin{aligned}
E[Y] &= \int_c^d y \cdot p_Y(y)\ \mathrm{d}y && \text{nach Definition von } E[Y] \\
&= \int_c^d y \cdot p_X\bigl(\psi(y)\bigr) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}y}\ \mathrm{d}y && \text{nach Satz 34 über Variablen-Transformationen} \\
&= \int_c^d \varphi(\psi(y)) \cdot p_X\bigl(\psi(y)\bigr) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}y}\ \mathrm{d}y && \text{denn } \psi \text{ ist die Umkehrfunktion von } \varphi \\
&= \int_a^b \varphi(\psi) \cdot p_X(\psi)\, \mathrm{d}\psi && \text{Substitutionsregel der Analysis} \\
&= \int_a^b \varphi(x) \cdot p_X(x)\, \mathrm{d}x && \text{ob die Integrations-Variable } x \text{ oder } \psi \text{ heißt, ist egal}
\end{aligned}$$

Im allgemeinen Fall kann der Beweis dadurch geführt werden, dass wir das Intervall $[a, b]$ in Intervalle zerlegen, in denen die Funktion φ streng monoton ist. □

Bemerkung: Im Englischen hat dieser Satz den Namen *law of the unconscious statistician*, abgekürzt LOTUS. ◇

5.2 Moment-erzeugende Funktion

In diesem Abschnitt führen wir ein Hilfsmittel ein, mit dem wir später die Wahrscheinlichkeits-Dichte für eine Reihe wichtiger Zufalls-Variablen berechnen können. Dieses Hilfsmittel ist die moment-erzeugende Funktion einer Zufalls-Variable X. Der Begriff der moment-erzeugenden Funktion wird außerdem zum Beweis des zentralen Grenzwert-Satzes benötigt.

Definition 38 (Moment-erzeugende Funktion) *Es sei* $X : [a, b] \to [c, d]$ *eine stetige Zufalls-Variable. Dann definieren wir die moment-erzeugende Funktion*

$$M_X : \mathbb{R} \to \mathbb{R}$$

für alle $t \in \mathbb{R}$ *als den folgenden Erwartungswert:*

$$M_X(t) := E\bigl[\exp(t \cdot X)\bigr] = \int_c^d \exp\bigl(t \cdot x\bigr) \cdot p_X(x)\ dx.$$

Dabei haben wir bei der letzten Gleichung Gebrauch von Satz 37 (LOTUS) gemacht. ◇

Satz 39 (Verschiebungs-Satz für moment-erzeugende Funktionen) *Es sei $X : \Omega \to \mathbb{R}$ eine Zufalls-Variable, $a \in \mathbb{R}$ und $b \in \mathbb{R}\backslash\{0\}$. Definieren wir die Zufalls-Variable $Y : \Omega \to \mathbb{R}$ als*

$$Y(\omega) = \frac{X(\omega) + a}{b},$$

so gilt

$$M_Y(t) = M_X\Big(\frac{t}{b}\Big) \cdot \exp\Big(\frac{a \cdot t}{b}\Big).$$

Aufgabe 37: Beweisen Sie den Verschiebungs-Satz für moment-erzeugende Funktionen.
Hinweis: Verwenden Sie Satz 34.

Beispiel: Es sei $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$. Die Zufalls-Variable $X : \Omega \to \mathbb{R}$ habe die Wahrscheinlichkeits-Dichte

$$p_X(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right) \quad \text{mit } \sigma > 0.$$

Eine solche Zufalls-Variable heißt normal-verteilt mit Parametern μ und σ, wir schreiben

$$X \sim \mathcal{N}(\mu, \sigma).$$

Wir werden später sehen, dass μ der Erwartungswert von X ist. Den Parameter σ werden wir als die Standard-Abweichung von X erkennen.

Abbildung 5.4 auf Seite 86 zeigt die Wahrscheinlichkeits-Dichte einer Standard-Normal-Verteilung. Abbildung 5.5 auf Seite 86 zeigt eine Banknote mit dem Portrait von Carl Friederich Gauss, auf dem diese Wahrscheinlichkeits-Dichte ebenfalls abgebildet ist.

Falls Y eine Zufalls-Variable mit $Y \sim \mathcal{N}(0, 1)$ ist, sprechen wir von einer standard-normal-verteilten Zufalls-Variable. Um X und Y in Beziehung zu setzen, definieren wir eine lineare Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ durch

$$\varphi(x) = \frac{x - \mu}{\sigma}.$$

Die Umkehrfunktion ψ von φ ist dann

$$\psi(y) = y \cdot \sigma + \mu.$$

Dann gilt $Y := \varphi \circ X$, denn nach Satz 34 haben wir für die Wahrscheinlichkeits-Dichte p_Y:

$$p_Y(y) = p_X(y \cdot \sigma + \mu) \cdot \sigma = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{y^2}{2}\right).$$

Die moment-erzeugende Funktion von Y erhalten wir wie folgt:

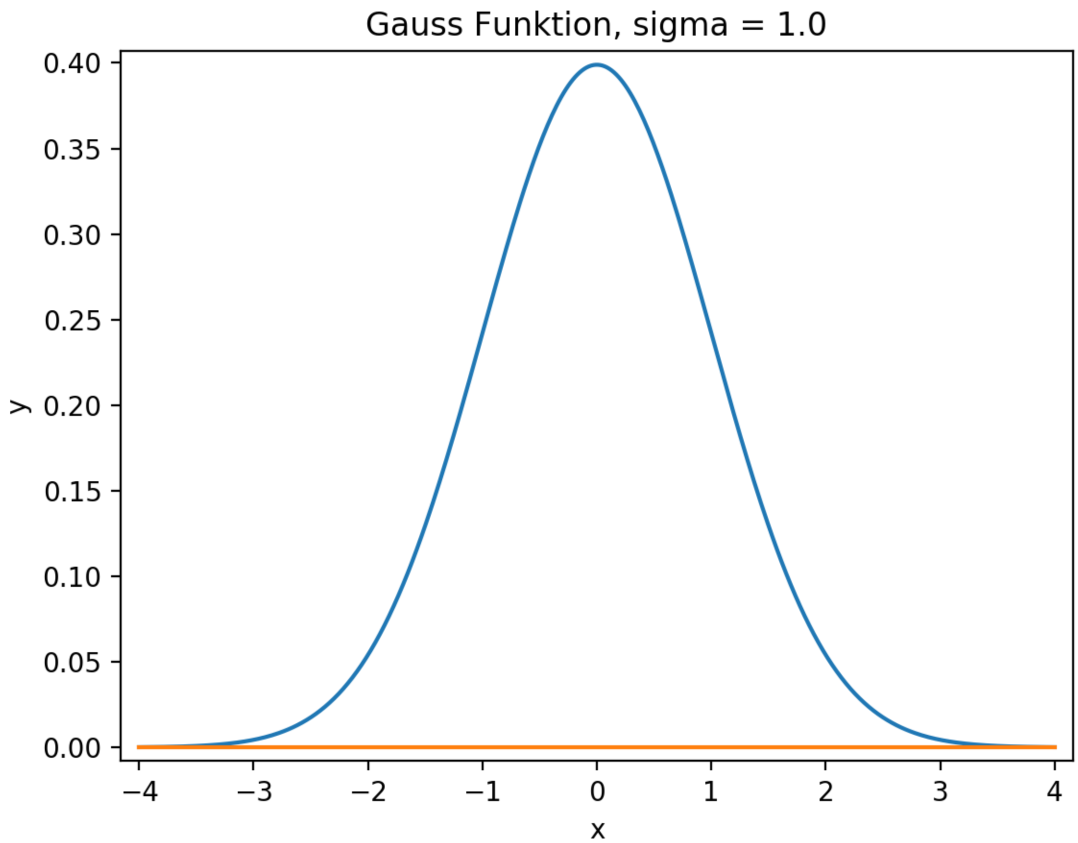

Abbildung 5.4: Wahrscheinlichkeits-Dichte einer $\mathcal{N}(0, 1)$ verteilten Zufalls-Variable.

Abbildung 5.5: Vorderseite eines Zehn-Mark-Scheins.

$$\begin{aligned}
M_Y(t) &= \int_{-\infty}^{\infty} \exp(t\cdot x)\cdot \frac{1}{\sqrt{2\cdot\pi}}\cdot \exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\cdot\pi}}\cdot \int_{-\infty}^{\infty} \exp(t\cdot x)\cdot \exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\cdot\pi}}\cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\cdot(x^2-2\cdot t\cdot x)\right)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\cdot\pi}}\cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\cdot(x^2-2\cdot t\cdot x+t^2)+\frac{1}{2}\cdot t^2\right)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\cdot\pi}}\cdot \exp\left(\frac{1}{2}\cdot t^2\right)\cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\cdot(x-t)^2\right)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\cdot\pi}}\cdot \exp\left(\frac{1}{2}\cdot t^2\right)\cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\cdot y^2\right)\,\mathrm{d}y \\
&\ \text{mit der Variablen-Transformation } x \mapsto y+t \\
&= \exp\left(\frac{1}{2}\cdot t^2\right) \\
&\ \text{denn } \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\cdot y^2\right)\,\mathrm{d}y = \sqrt{2\cdot\pi}.
\end{aligned}$$

Zwischen der Zufalls-Variable X und der Zufalls-Variable Y besteht der Zusammenhang

$$X = \psi\circ Y = Y\cdot\sigma+\mu = \left(Y+\frac{\mu}{\sigma}\right)\cdot\sigma = \left(Y+\frac{\mu}{\sigma}\right)\cdot\frac{1}{\frac{1}{\sigma}}$$

Nach dem Verschiebungs-Satz für moment-erzeugende Funktionen gilt daher

$$\begin{aligned}
M_X(t) &= \exp(\mu\cdot t)\cdot M_Y\big(t\cdot\sigma\big) \\
&= \exp(\mu\cdot t)\cdot\exp\left(\frac{1}{2}\cdot t^2\cdot\sigma^2\right) \\
&= \exp\left(\frac{1}{2}\cdot t^2\cdot\sigma^2+\mu\cdot t\right)
\end{aligned}$$ ◇

Aufgabe 38: Beweisen Sie die beiden folgenden Eigenschaften einer normal-verteilten Zufalls-Variablen $X \sim \mathcal{N}(\mu,\sigma)$:

$$E\big[X\big] = \mu \quad \text{und} \quad \mathrm{Var}\big[X\big] = \sigma^2.$$ ◇

Satz 40 (Eindeutigkeits-Satz) *Es seien $X:\Omega\to\mathbb{R}$ und $Y:\Omega\to\mathbb{R}$ zwei Zufalls-Variablen und für die moment-erzeugenden Funktionen M_X und M_Y gelte*

$$M_X(t) = M_Y(t) \quad \text{für alle } t\in\mathbb{R}.$$

Dann sind die Wahrscheinlichkeits-Dichten p_X und p_Y gleich:

$$p_X(t) = p_Y(t) \quad \text{für alle } t\in\mathbb{R}.$$ ◇

Der Beweis dieses Satzes gelingt, indem die moment-erzeugende Funktion einer Zufalls-Variable auf die Fourier-Transformierte zurückgeführt wird. Da wir die Fourier-Transformation in der Analysis nicht mehr behandelt haben, können wir hier die Details dieses Beweises nicht diskutieren. Trotzdem ist der Satz für unsere weitere Arbeit ein wichtiges Hilfsmittel, denn er gestattet uns, die Wahrscheinlichkeits-Dichten bestimmter Zufalls-Variablen durch Berechnung der moment-erzeugenden Funktion zu berechnen.

Satz 41 *Es seien $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei unabhängige Zufalls-Variablen. Dann lässt sich die moment-erzeugende Funktion der Zufalls-Variable $X + Y$ als Produkt der moment-erzeugenden Funktionen von X und Y schreiben:*

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) \quad \text{für alle } t \in \mathbb{R}.$$

Beweis: Es gilt

$$\begin{aligned}
M_{X+Y}(t) &= E\big[\exp\big(t \cdot (X+Y)\big)\big] \\
&= E\left[\exp(t \cdot X) \cdot \exp(t \cdot Y)\right] \\
&= E\left[\exp(t \cdot X)\right] \cdot E\left[\exp(t \cdot Y)\right] \\
& \text{denn wenn } X \text{ und } Y \text{ unabhängig sind,} \\
& \text{dann sind auch } \exp(t \cdot X) \text{ und } \exp(t \cdot Y) \text{ unabhängig} \\
&= M_X(t) \cdot M_Y(t)
\end{aligned}$$

Hier haben wir im vorletzten Schritt ausgenutzt, dass mit X und Y auch die Zufalls-Variablen $\exp(t \cdot X)$ und $\exp(t \cdot Y)$ unabhängig sind. Genau wir im diskreten Fall auch gilt für den Erwartungswert des Produkts unabhängiger Zufalls-Variablen X und Y die Gleichung

$$E[X \cdot Y] = E[X] \cdot E[Y].$$

□

Definition 42 (Moment) *Es sei X eine Zufalls-Variable. Das n-te Moment von X ist als*

$$\mathrm{Mt}(X, n) := E[X^n]$$

definiert. Gilt $\mu := E[X]$, so ist das n-te zentrale Moment von X als

$$\mathrm{ZM}(X, n) := E\big[(X - \mu)^n\big]$$

definiert. □

Definition 43 (Schiefe) *Es sei X eine Zufalls-Variable mit der Standard-Abweichung σ. Die Schiefe von X ist dann als*

$$v(X) := \frac{\mathrm{ZM}(X, 3)}{\sigma^3} = \frac{E\big[(X - \mu)^3\big]}{\sigma^3}$$

definiert. Die Schiefe misst, wie unsymmetrisch die Verteilung der Zufalls-Variable X ist.

Satz 44 *Ist $f(t) = M_X(t)$ die moment-erzeugende Funktion der Zufalls-Variable X und die Funktion sei f in einer Taylor-Reihe um 0 entwickelbar. Dann gilt*

$$\mathrm{Mt}(X, n) = \frac{\mathrm{d}^n f}{\mathrm{d}t^n}(0).$$

Beweis: Es sei $X : \Omega \to [a, b]$. Dann gilt

$$\begin{aligned}
M_X(t) &= \int_a^b \exp(t \cdot x) \cdot p_X(x)\, \mathrm{d}x \\
&= \int_a^b \sum_{n=0}^{\infty} \frac{(t \cdot x)^n}{n!} \cdot p_X(x)\, \mathrm{d}x \qquad \text{denn } \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{t^n}{n!} \cdot \int_a^b x^n \cdot p_X(x)\, \mathrm{d}x \\
&= \sum_{n=0}^{\infty} \frac{t^n}{n!} \cdot E\big[X^n\big]
\end{aligned}$$

Auf der anderen Seite hat die Taylor-Reihe von $f(t)$ die Form

$$f(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \cdot f^{(n)}(0)$$

Durch einen Koeffizienten-Vergleich der beiden Reihen erkennen wir, dass

$$E[X^n] = f^{(n)}(0) \quad \text{für alle } n \in \mathbb{N} \text{ gilt.}$$ □

Bemerkung: An dieser Stelle sehen wir, wie nützlich die Kenntnis der Moment-erzeugenden Funktion einer Zufalls-Variable X ist, denn der Erwartungswert, die Varianz und die Schiefe von X lassen sich aus den Momenten berechnen, genauer gilt:

(a) $E[X] = \mathrm{Mt}(X, 1)$,

(b) $\mathrm{Var}[X] = \mathrm{Mt}(X, 2) - \mathrm{Mt}(X, 1)^2$,

(c) $v(X) = \dfrac{\mathrm{ZM}(X, 3)}{\mathrm{Var}[X]^{3/2}}$.

Dabei kann das dritte zentrale Moment $\mathtt{ZM}(X, 3)$ wie folgt aus dem ersten, zweiten und dritten Moment berechnet werden:

$$\mathrm{ZM}(X, 3) = \mathrm{Mt}(X, 3) - 3 \cdot \mathrm{Mt}(X, 2) \cdot \mathrm{Mt}(X, 1) + 2 \cdot \mathrm{Mt}(X, 1)^3.$$ ⋄

Aufgabe 39: Beweisen Sie diese Formel.

Satz 45 (Momente der Standard-Normal-Verteilung)
Es sei $X : \Omega \to \mathbb{R}$ eine standard-normal-verteilte Zufalls-Variable, es gelte also $X \sim \mathcal{N}(0, 1)$. Dann können die Momente von X wie folgt berechnet werden:

1. $\mathrm{Mt}(X, 2 \cdot n + 1) = 0$ *für alle* $n \in \mathbb{N}$.
2. $\mathrm{Mt}(X, 2 \cdot n) = (2 \cdot n - 1)!!$ *für alle* $n \in \mathbb{N}$ *mit* $n > 0$.
   Dabei ist für eine ungerade Zahl $2 \cdot m + 1$ der Ausdruck $(2 \cdot m + 1)!!$ als
   $$(2 \cdot m + 1)!! := 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2 \cdot m - 1) \cdot (2 \cdot m + 1) = \prod_{k=0}^{m} (2 \cdot k + 1)$$
   definiert. Dieser Ausdruck wird auch als Doppel-Fakultät bezeichnet.

Beweis: Wir haben früher für die Moment-erzeugende Funktion $M_X(t)$ einer standard-normal-verteilten Zufalls-Variable X den Ausdruck

$$M_X(t) = \exp\left(\frac{t^2}{2}\right)$$

gefunden. Wir wissen, dass sich die Exponential-Funktion als

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

schreiben lässt. Setzen wir hier für x den Ausdruck $t^2/2$ ein, so erhalten wir

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^{2 \cdot n}}{n! \cdot 2^n} = \sum_{n=0}^{\infty} \frac{t^{2 \cdot n}}{(2 \cdot n)!} \cdot \frac{(2 \cdot n)!}{n! \cdot 2^n} = \sum_{n=0}^{\infty} \frac{t^{2 \cdot n}}{(2 \cdot n)!} \cdot (2 \cdot n - 1)!!,$$

denn Sie können durch eine elementare vollständige Induktion zeigen, dass

$$\frac{(2 \cdot n)!}{n! \cdot 2^n} = (2 \cdot n - 1)!!$$

gilt. Der Beweis des letzten Satzes zeigt, dass das n-te Moment einer Zufalls-Variable X gerade der

Koeffizient von $t^n/n!$ in der Taylor-Reihe von $M_X(t)$ ist. Daraus folgt die Behauptung, denn in der Reihe

$$\sum_{n=0}^{\infty} \frac{t^{2\cdot n}}{(2\cdot n)!} \cdot (2\cdot n - 1)!!$$

gibt es keine ungeraden Koeffizienten und der Koeffizient von $t^{2\cdot n}/(2\cdot n)!$ ist offenbar $(2\cdot n-1)!!$. □

5.3 Der zentrale Grenzwert-Satz

Wir kommen nun zu dem wichtigsten Ergebnis der Wahrscheinlichkeits-Rechnung.

Satz 46 (Zentraler Grenzwertsatz) *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger stetiger Zufalls-Variablen, die alle dieselbe Wahrscheinlichkeits-Dichte haben, es gilt also*

$$\forall m, n \in \mathbb{N} : p_{X_m}(t) = p_{X_n}(t).$$

Weiter gelte

$$\mu = E[X_n] \quad \textit{und} \quad \sigma^2 = \mathrm{Var}[X_n] \quad \textit{für alle } n \in \mathbb{N}.$$

Definieren wir für $n \in \mathbb{N}$ die Zufalls-Variablen S_n durch

$$S_n = \frac{1}{\sigma\cdot\sqrt{n}} \cdot \sum_{k=1}^{n} (X_k - \mu),$$

so konvergieren die Wahrscheinlichkeits-Dichten p_{S_n} der Zufalls-Variablen S_n gegen die Normal-Verteilung, es gilt also

$$\lim_{n\to\infty} p_{S_n}(t) = \frac{1}{\sqrt{2\cdot\pi}} \cdot \exp\left(-\frac{1}{2}\cdot t^2\right).$$

Beweis: Wir führen den Beweis, indem wir zeigen, dass die moment-erzeugenden Funktionen M_{S_n} der Zufalls-Variablen S_n gegen die moment-erzeugende Funktion einer normal-verteilten Zufalls-Variable mit Mittelwert 0 und Varianz 1 konvergieren. Die Behauptung folgt dann aus dem Eindeutigkeits-Satz 40. Es sei f_k die moment-erzeugende Funktion der Zufalls-Variable

$$Y_k := \frac{X_k - \mu}{\sigma\cdot\sqrt{n}}, \quad \text{also}$$

$$f_k(t) := M_{Y_k}(t) = E\left[\exp\left(t\cdot\frac{1}{\sigma\cdot\sqrt{n}}\cdot(X_k-\mu)\right)\right].$$

Da die Zufalls-Variablen X_n alle dieselbe Wahrscheinlichkeits-Dichte haben, sind auch die moment-erzeugenden Funktionen M_{Y_k} für alle k gleich:

$$\forall k, l \in \mathbb{N} : f_k = f_l.$$

Daher können wir den Index k bei der Funktion f_k weglassen und

$$f := f_k$$

definieren. Für die moment-erzeugende Funktion M_{S_n} finden wir jetzt das Folgende:

$$\begin{aligned}
M_{S_n}(t) &= E\left[\exp\left(t \cdot \frac{1}{\sigma \cdot \sqrt{n}} \cdot \sum_{k=1}^{n} (X_k - \mu)\right)\right] \\
&= E\left[\prod_{k=1}^{n} \exp\left(t \cdot \frac{1}{\sigma \cdot \sqrt{n}} \cdot (X_k - \mu)\right)\right] \\
&= \prod_{k=1}^{n} E\left[\exp\left(t \cdot \frac{1}{\sigma \cdot \sqrt{n}} \cdot (X_k - \mu)\right)\right] \\
&\quad \text{denn die Zufalls-Variablen } X_k \text{ sind unabhängig} \\
&= \prod_{k=1}^{n} f(t) \\
&= f^n(t)
\end{aligned}$$

In der Definition der moment-erzeugenden Funktion f ersetzen wir die Exponential-Funktion durch die Taylor-Reihe

$$\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

und erhalten

$$\begin{aligned}
f(t) &= E\left[\sum_{i=0}^{\infty} \frac{1}{i!} \cdot \left(\frac{t}{\sigma \cdot \sqrt{n}}\right)^i \cdot (X_k - \mu)^i\right] \\
&= \sum_{i=0}^{\infty} \frac{1}{i!} \cdot \left(\frac{t}{\sigma \cdot \sqrt{n}}\right)^i \cdot E\big[(X_k - \mu)^i\big] \\
&= 1 + \frac{t}{\sigma \cdot \sqrt{n}} \cdot \underbrace{E[X_k - \mu]}_{0} + \frac{1}{2} \cdot \frac{t^2}{\sigma^2 \cdot n} \cdot E\big[(X_k - \mu)^2\big] \\
&\quad + \frac{1}{6} \cdot \frac{t^3}{\sigma^3 \cdot n \cdot \sqrt{n}} \cdot E\big[(X_k - \mu)^3\big] + \cdots \\
&\xrightarrow{n \to \infty} 1 + \frac{1}{2} \cdot \frac{t^2}{\sigma^2 \cdot n} \cdot E\big[(X_k - \mu)^2\big] \\
&= 1 + \frac{1}{2} \cdot \frac{t^2}{\sigma^2 \cdot n} \cdot \sigma^2 \\
&= 1 + \frac{1}{2} \cdot \frac{t^2}{n}
\end{aligned}$$

Dabei haben wir im letzten Schritt ausgenutzt, dass für $n \to \infty$ die Terme, bei denen $n^{i/2}$ mit $i > 2$ im Nenner steht, gegenüber dem Term mit $\frac{1}{n}$ vernachlässigt werden können. Daher haben wir

$$\begin{aligned}
\lim_{n\to\infty} M_{S_n}(t) &= \lim_{n\to\infty} \left(1 + \frac{1}{2} \cdot \frac{t^2}{n}\right)^n \\
&= \exp\left(\frac{1}{2} \cdot t^2\right),
\end{aligned}$$

denn wir erinnern uns, dass wir in der Analysis gezeigt haben, dass

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = \exp(x)$$

gilt. Die Funktion

$$x \mapsto \exp\Big(\frac{1}{2} \cdot t^2\Big)$$

ist aber gerade die moment-erzeugende Funktion einer Zufalls-Variable X, für die $X \sim \mathcal{N}(0,1)$ gilt. Aus dem Eindeutigkeits-Satz 40 für die moment-erzeugenden Funktionen folgt nun

$$\lim_{n\to\infty} p_{S_n}(x) = \frac{1}{\sqrt{2\cdot\pi}} \cdot \exp\left(-\frac{1}{2} \cdot x^2\right). \qquad \square$$

Die Voraussetzungen des zentralen Grenzwertsatzes können noch abgeschwächt werden. Es ist nicht erforderlich, dass die einzelnen Zufalls-Variablen X_k alle dieselbe Wahrscheinlichkeits-Dichte haben. Im wesentlichen reicht es aus, wenn die Erwartungswerte und die Varianzen dieser Zufalls-Variablen beschränkt sind. Darüber hinaus muss dann noch eine weitere recht technische Bedingung erfüllt sein. Der zentrale Grenzwertsatz ist der Grund dafür, dass in der Praxis viele Zufalls-Variablen normal-verteilt sind. Diese liegt einfach daran, dass viele Größen, die wir beobachten, durch eine große Zahl von Faktoren bestimmt werden. Auch wenn die einzelnen Faktoren beliebig verteilt sind, so ist die Summe dieser Faktoren doch annähernd normal-verteilt.

Aufgabe 40: Die Zufalls-Variablen X_1 und X_2 seien normal-verteilt mit den Erwartungswerten μ_1 und μ_2 und den Varianzen σ_1^2 und σ_2^2, es gilt also

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1) \quad \text{und} \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2).$$

Außerdem seien X_1 und X_2 unabhängig. Berechnen Sie die Wahrscheinlichkeits-Dichte der Zufalls-Variable $X_1 + X_2$.

Hinweis: Orientieren Sie sich am Beweis des zentralen Grenzwertsatzes. ◇

5.4 Die Chi-Quadrat-Verteilung

In diesem Abschnitt gehen wir davon aus, dass $X_1, X_2, \cdots, X_n$ unabhängige Zufalls-Variablen sind, die alle einer standardisierten Normalverteilung genügen, für die einzelnen Zufalls-Variablen X_k gilt also $X_k \sim \mathcal{N}(0,1)$ für alle $k = 1, \cdots, n$. Diesen Sachverhalt können wir auch als

$$P(X_k \leq x) = \frac{1}{\sqrt{2\,\pi}} \cdot \int_{-\infty}^{x} \exp\Big(-\frac{t^2}{2}\Big) \mathrm{d}t \quad \text{für alle } k = 1, \cdots, n$$

schreiben. Definieren wir die Zufalls-Variable Z als

$$Z(\omega) = \sum_{i=1}^{n} X_i^2(\omega),$$

so hat Z eine χ^2-Verteilung mit dem Freiheitsgrad n, wir schreiben

$$Z \sim \chi^2(n).$$

Unser Ziel ist es, die Wahrscheinlichkeits-Dichte p_Z für die Zufalls-Variable Z zu berechnen. Wir beginnen mit dem Spezialfall $n = 1$.

5.4.1 Die Chi-Quadrat-Verteilung mit dem Freiheitsgrad Eins

Es sei X eine Zufalls-Variable, die der Standard-Normal-Verteilung genügt, für die Wahrscheinlichkeits-Dichte $p_X(x)$ gilt also

$$p_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\Big(-\frac{x^2}{2}\Big).$$

Wir stellen uns die Frage, wie die Zufalls-Variable X^2 verteilt ist. Wir berechnen zunächst die kumulative Verteilungs-Funktion von X^2, die wir mit F_{X^2} bezeichnen. Ist Ω der zu Grunde liegende Ergebnis-Raum, so haben wir

$$\begin{aligned}
F_{X^2}(x) &= P(X^2 \leq x) \\
&= P\Bigl(\{\omega \in \Omega \mid X^2(\omega) \leq x\}\Bigr) \\
&= P\Bigl(\{\omega \in \Omega \mid -\sqrt{x} \leq X(\omega) \leq \sqrt{x}\}\Bigr) \\
&= P\Bigl(\{\omega \in \Omega \mid X(\omega) \leq \sqrt{x}\} \setminus \{\omega \in \Omega \mid X(\omega) < -\sqrt{x}\}\Bigr) \\
&= \Phi\bigl(\sqrt{x}\bigr) - \Phi\bigl(-\sqrt{x}\bigr),
\end{aligned}$$

denn die Verteilungs-Funktion einer Zufalls-Variable mit Normalverteilung ist das gaußsche Fehlerintegral (benannt nach Carl Friederich Gauss)

$$\Phi(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int\limits_{-\infty}^{x} \exp\bigl(-t^2/2\bigr)\,\mathrm{d}t.$$

Auf Grund der Symmetrie des Integranden gilt die Gleichung

$$\Phi(x) + \Phi(-x) = 1,$$

denn wir haben

$$\begin{aligned}
1 &= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int\limits_{-\infty}^{\infty} \exp\bigl(-t^2/2\bigr)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int\limits_{-\infty}^{x} \exp\bigl(-t^2/2\bigr)\,\mathrm{d}t + \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int\limits_{x}^{\infty} \exp\bigl(-t^2/2\bigr)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int\limits_{-\infty}^{x} \exp\bigl(-t^2/2\bigr)\,\mathrm{d}t + \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int\limits_{-\infty}^{-x} \exp\bigl(-t^2/2\bigr)\,\mathrm{d}t \\
&= \Phi(x) + \Phi(-x).
\end{aligned}$$

Aus $\Phi(x) + \Phi(-x)$ folgt $\Phi(-x) = 1 - \Phi(x)$ und damit finden wir für die Verteilungs-Funktion von X^2 die Gleichung

$$F_{X^2}(x) = \Phi\bigl(\sqrt{x}\bigr) - \Phi\bigl(-\sqrt{x}\bigr) = \Phi\bigl(\sqrt{x}\bigr) - \bigl(1 - \Phi\bigl(\sqrt{x}\bigr)\bigr) = 2 \cdot \Phi\bigl(\sqrt{x}\bigr) - 1.$$

Die Wahrscheinlichkeits-Dichte p_{X^2} der Zufalls-Variable X^2 erhalten wir, indem wir die Verteilungs-Funktion nach x differenzieren. Wegen

$$\frac{\mathrm{d}\Phi}{\mathrm{d}x} = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\bigl(-x^2/2\bigr)$$

finden wir:

$$\begin{aligned}
p_{X^2}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\Bigl(2 \cdot \Phi\bigl(\sqrt{x}\bigr) - 1\Bigr) \\
&= 2 \cdot \frac{1}{2 \cdot \sqrt{x}} \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp(-x/2) \\
&= \frac{1}{\sqrt{2 \cdot \pi \cdot x}} \cdot \exp(-x/2)
\end{aligned}$$

Abbildung 5.6 auf Seite 94 zeigt diese Funktion.

Aufgabe 41: Berechnen Sie die moment-erzeugende Funktion der Zufalls-Variable X^2.

Lösung: Die Wahrscheinlichkeits-Dichte von X^2 haben wir eben berechnet. Setzen wir das Ergebnis in die Definition der Moment-Erzeugenden Funktion ein, so erhalten wir:

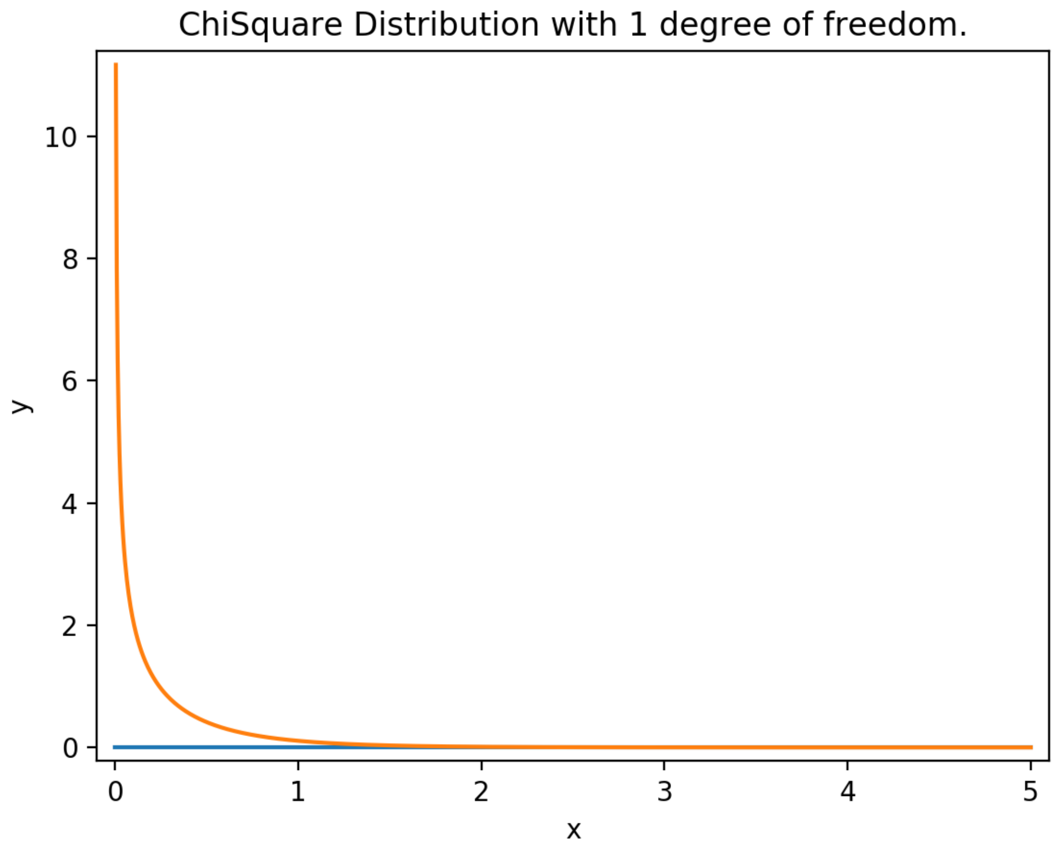

Abbildung 5.6: Die χ^2-Verteilung mit einem Freiheitsgrad.

$$\begin{aligned}
M_{X^2}(t) &= \int_0^\infty \exp(t\cdot x)\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{x}}\cdot\exp\left(-\frac{x}{2}\right)\;\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\cdot\int_0^\infty \frac{1}{\sqrt{x}}\cdot\exp\left(t\cdot x-\frac{x}{2}\right)\;\mathrm{d}x \\
&\quad\text{mit der Variablen-Transformation } x=y^2\text{, also } \mathrm{d}x = 2\cdot y\;\mathrm{d}y \text{ erhalten wir} \\
&= \frac{1}{\sqrt{2\pi}}\cdot\int_0^\infty \frac{1}{y}\cdot\exp\left(\left(t-\frac{1}{2}\right)\cdot y^2\right)\cdot 2\cdot y\;\mathrm{d}y \\
&= \frac{2}{\sqrt{2\pi}}\cdot\int_0^\infty \exp\left(-(1-2\cdot t)\cdot\frac{y^2}{2}\right)\;\mathrm{d}y \\
&\quad\text{mit der Variablen-Transformation } y=\frac{1}{\sqrt{1-2\cdot t}}\cdot z \text{ wird daraus} \\
&= \frac{2}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{1-2\cdot t}}\cdot\int_0^\infty \exp\left(-\frac{z^2}{2}\right)\;\mathrm{d}z \\
&= \frac{2}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{1-2\cdot t}}\cdot\frac{1}{2}\cdot\sqrt{2\pi} \\
&\quad\text{denn } \int_0^\infty \exp\left(-\frac{z^2}{2}\right)\;\mathrm{d}z = \frac{1}{2}\cdot\int_{-\infty}^\infty \exp\left(-\frac{z^2}{2}\right)\;\mathrm{d}z = \frac{1}{2}\cdot\sqrt{2\pi} \\
&= \frac{1}{\sqrt{1-2\cdot t}}
\end{aligned}$$

Wir halten das Ergebnis fest:

$$M_{X^2}(t) = \frac{1}{\sqrt{1-2\cdot t}} \tag{5.2}$$

5.4.2 Der allgemeine Fall

Um unser eigentliches Problem lösen zu können, benötigen wir zwei Definitionen.

Definition 47 (Gamma-Funktion) *Die Gamma-Funktion ist für positive reelle Zahlen x als das folgende Integral definiert:*

$$\Gamma(x) := \int_0^\infty t^{x-1} \cdot \mathrm{e}^{-t}\; dt. \qquad \diamond$$

Aufgabe 42: Weisen Sie die folgenden Eigenschaften der Gamma-Funktion nach:

(a) $\Gamma(1) = 1$.

(b) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Hinweis: Führen Sie das Integral durch eine geeignete Substitution auf das Integral

$$\int_0^\infty \exp\left(-\frac{1}{2}\cdot t^2\right)\; \mathrm{d}t$$

zurück.

(c) $\Gamma(x+1) = x\cdot\Gamma(x)$ falls $x > 0$ ist.

Hinweis: Partielle Integration.

(d) $\Gamma(n+1) = n!$ für alle $n \in \mathbb{N}$.

Hinweis: Vollständige Induktion.

Die letzte Eigenschaft zeigt, dass die Gamma-Funktion als eine Erweiterung der Fakultäts-Funktion auf die positiven reellen Zahlen aufgefasst werden kann. $\diamond$

Definition 48 (Gamma-Verteilung) *Eine Zufalls-Variable Y genügt einer Gamma-Verteilung mit Parametern α und β, falls für die Wahrscheinlichkeits-Dichte p_Y gilt:*

$$p_Y(x) = \begin{cases} \dfrac{1}{\beta^\alpha\cdot\Gamma(\alpha)}\cdot x^{\alpha-1}\cdot\exp\Bigl(-\dfrac{x}{\beta}\Bigr) & \textit{für } x > 0; \\ 0 & \textit{für } x \leq 0. \end{cases}$$

Die Parameter α und β sind dabei positive reelle Zahlen. In diesem Fall schreiben wir $Y \sim \Gamma(\alpha,\beta)$. $\diamond$

Aufgabe 43: Nehmen Sie an, dass Y eine gamma-verteilte Zufalls-Variable mit den Parametern α und β ist, es gilt also

$$Y \sim \Gamma(\alpha,\beta)$$

Berechnen Sie die moment-erzeugende Funktion von Y.

Lösung: Nach Definition der moment-erzeugenden Funktion p_Y gilt:

$$\begin{aligned}
M_Y(t) &= \int_0^\infty \exp(t\cdot x)\cdot p_Y(x)\,\mathrm{d}x \\
&= \int_0^\infty \exp(t\cdot x)\cdot \frac{x^{\alpha-1}\cdot\exp\left(-x/\beta\right)}{\beta^\alpha\cdot\Gamma(\alpha)}\,\mathrm{d}x \\
&= \frac{1}{\beta^\alpha\cdot\Gamma(\alpha)}\cdot\int_0^\infty x^{\alpha-1}\cdot\exp\left(-\frac{x}{\beta}+t\cdot x\right)\,\mathrm{d}x \\
&= \frac{1}{\beta^\alpha\cdot\Gamma(\alpha)}\cdot\int_0^\infty x^{\alpha-1}\cdot\exp\left(-\Big(\frac{1}{\beta}-t\Big)\cdot x\right)\,\mathrm{d}x
\end{aligned}$$

Mit der Variablen-Transformation $x = \dfrac{1}{\left(\frac{1}{\beta}-t\right)}\cdot y$ wird daraus:

$$\begin{aligned}
M_Y(t) &= \frac{1}{\beta^\alpha\cdot\Gamma(\alpha)}\cdot\frac{1}{\left(\frac{1}{\beta}-t\right)}\cdot\frac{1}{\left(\frac{1}{\beta}-t\right)^{\alpha-1}}\cdot\int_0^\infty y^{\alpha-1}\cdot\exp(-y)\,\mathrm{d}y \\
&= \frac{1}{\beta^\alpha\cdot\Gamma(\alpha)}\cdot\frac{1}{\left(\frac{1}{\beta}-t\right)^{\alpha}}\cdot\Gamma(\alpha) \\
&= \frac{1}{\beta^\alpha}\cdot\frac{1}{\left(\frac{1}{\beta}-t\right)^{\alpha}} \\
&= \frac{1}{\beta^\alpha}\cdot\left(\frac{\beta}{1-\beta\cdot t}\right)^{\alpha} \\
&= \big(1-\beta\cdot t\big)^{-\alpha}. \qquad \Box
\end{aligned}$$

Also hat die moment-erzeugende Funktion für eine Zufalls-Variable Y, die einer Gamma-Verteilung mit den Parametern α und β genügt, die Form

$$M_Y(t) = (1-\beta\cdot t)^{-\alpha}. \tag{5.3}$$

Wir lösen nun unser eigentliches Problem, nämlich die Berechnung der Wahrscheinlichkeits-Dichte einer Zufalls-Variable Z, die aus n unabhängigen, standard-normal-verteilten Zufalls-Variablen X_i nach der Formel

$$Z = \sum_{i=1}^{n} X_i^2$$

berechnet wird. In diesem Fall schreiben wir

$$Z \sim \chi^2(n)$$

und sagen, dass Z eine χ^2-Verteilung mit n Freiheitsgraden hat. Wir berechnen zunächst die moment-erzeugende Funktion M_Z von Z. Nach Satz 41 und Gleichung (5.2) gilt

$$M_Z(t) = \Big(M_{X^2}(t)\Big)^n = (1-2\cdot t)^{-\frac{n}{2}}$$

Nach Gleichung (5.3) hat eine Zufalls-Variable, die einer Gamma-Verteilung mit den Parametern $\alpha = \frac{n}{2}$ und $\beta = 2$ genügt, dieselbe moment-erzeugende Funktion. Aus dem Eindeutigkeits-Satz 40 für die moment-erzeugende Funktion folgt daher, dass die Wahrscheinlichkeits-Dichte der Zufalls-Variable $Z = \sum\limits_{i=1}^{n} X_i^2$ die Form

$$p_Z(x) = \frac{1}{A_n}\cdot x^{n/2-1}\cdot\exp\left(-\frac{x}{2}\right) \quad \text{mit } A_n = 2^{n/2}\cdot\Gamma\left(\frac{n}{2}\right) \quad \text{für } x>0 \text{ hat.}$$

Wir formulieren dieses Ergebnis als Satz. Vorher benötigen wir noch eine Definition.

Satz 49 *Es seien X_1, X_2, $\cdots$, X_n unabhängige Zufalls-Variablen, die standard-normal-verteilt sind. Die Zufalls-Variable Z sei definiert als*

$$Z := \sum_{i=1}^{n} X_i^2$$

Dann hat die Zufalls-Variable Z die Wahrscheinlichkeits-Dichte

$$p_Z(x) = \frac{1}{A_n} \cdot x^{n/2-1} \cdot \exp\left(-\frac{x}{2}\right) \quad \textit{mit } A_n = 2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right).$$

Zusammenfassend können wir auch sagen: Ist Z eine Zufalls-Variable, so gilt

$$Z \sim \chi^2(n) \;\Rightarrow\; Z \sim \Gamma\Big(\frac{n}{2}, 2\Big). \qquad \diamond$$

Die Abbildungen 5.7, 5.8, 5.9 und 5.10 auf den folgenden Seiten zeigen die Wahrscheinlichkeits-Dichten der χ^2-Verteilungen für die Freiheitsgrade 1, 2, 3, 4, 5, 10 und 30. Wir sehen, dass sich diese Wahrscheinlichkeits-Dichten mit wachsender Zahl von Freiheitsgraden an die Wahrscheinlichkeits-Dichte einer Normalverteilung annähern. Bei der Diskussion des χ^2-Tests im nächsten Kapitel benötigen wir die beiden folgenden Sätze.

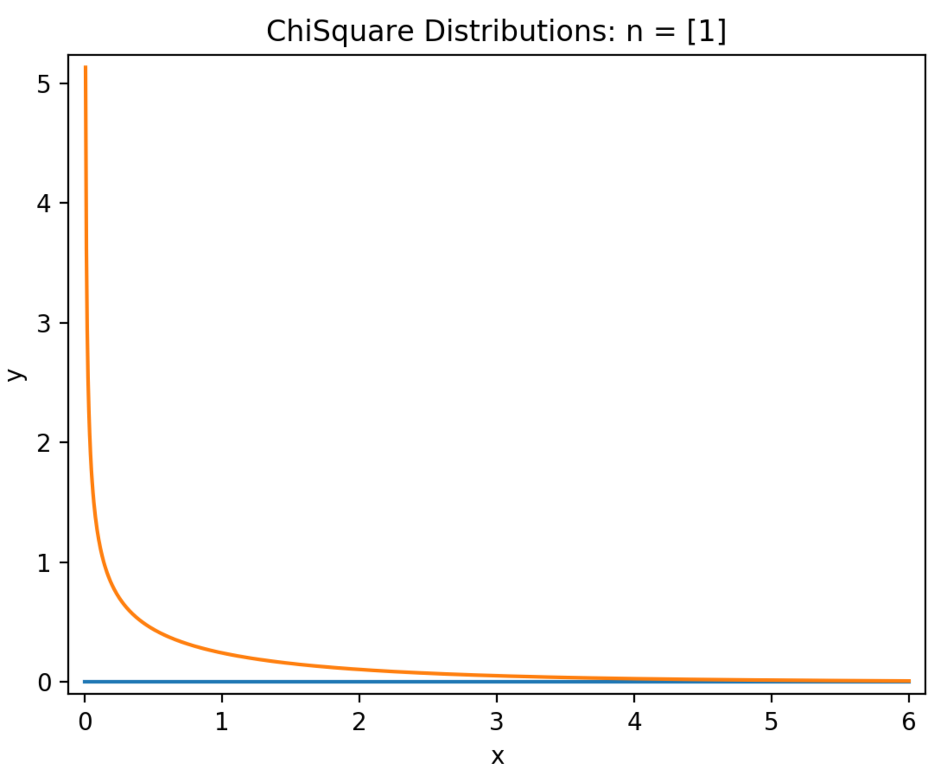

Abbildung 5.7: Die χ^2-Verteilungen mit dem Freiheitsgrad Eins.

Satz 50 *Es seien U, V und W Zufalls-Variablen mit folgenden Eigenschaften:*

1. *U ist χ^2-verteilt mit m Freiheitsgraden,*
2. *V ist χ^2-verteilt mit n Freiheitsgraden,*
3. *U und V sind unabhängig,*

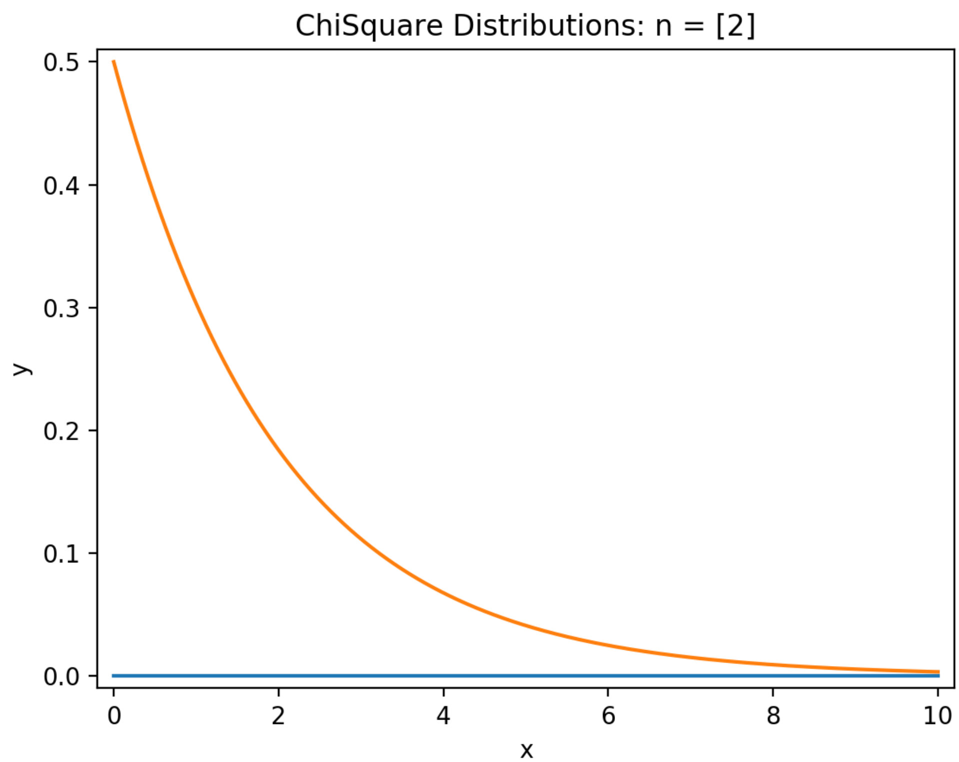

Abbildung 5.8: Die χ^2-Verteilungen mit dem Freiheitsgrad Zwei.

4. $W = U + V$.

Dann genügt die Zufalls-Variable W einer χ^2-Verteilung mit $m + n$ Freiheitsgraden.

Beweis: Aus dem Beweis des letzten Satzes können wir ablesen, wie die moment-erzeugenden Funktionen von U und V aussehen müssen. Es gilt

$$M_U(t) = (1 - 2 \cdot t)^{-\frac{m}{2}} \quad \text{und} \quad M_V(t) = (1 - 2 \cdot t)^{-\frac{n}{2}}.$$

Nach Satz 41 gilt daher für die moment-erzeugende Funktion M_W:

$$M_W(t) = M_U(t) \cdot M_V(t) = (1 - 2 \cdot t)^{-\frac{m}{2}} \cdot (1 - 2 \cdot t)^{-\frac{n}{2}} = (1 - 2 \cdot t)^{-\frac{m+n}{2}}.$$

Dies ist aber genau die moment-erzeugende Funktion einer Zufalls-Variable, die einer χ^2-Verteilung mit $m + n$ Freiheitsgraden genügt. Daher folgt die Behauptung aus dem Eindeutigkeits-Satz 40. □

Bemerkung: Der obige Satz lässt sich leicht auf die Summe von mehr als zwei χ^2-verteilten Zufalls-Variablen verallgemeinern. Sind $U_1, \cdots, U_k$ unabhängige Zufalls-Variablen, so dass für alle $i = 1, \cdots, k$ die Zufalls-Variable U_i einer χ^2-Verteilung mit m_i Freiheitsgraden genügt, dann genügt die Summe

$$W := \sum_{i=1}^{k} U_i$$

einer χ^2-Verteilung mit $m_1 + \cdots + m_k$ Freiheitsgraden. Diese Aussage kann mit Hilfe des letzten Satzes durch eine triviale Induktion nach k gezeigt werden. □

Der nächste Satz zeigt, dass sich der letzte Satz auch umkehren lässt.

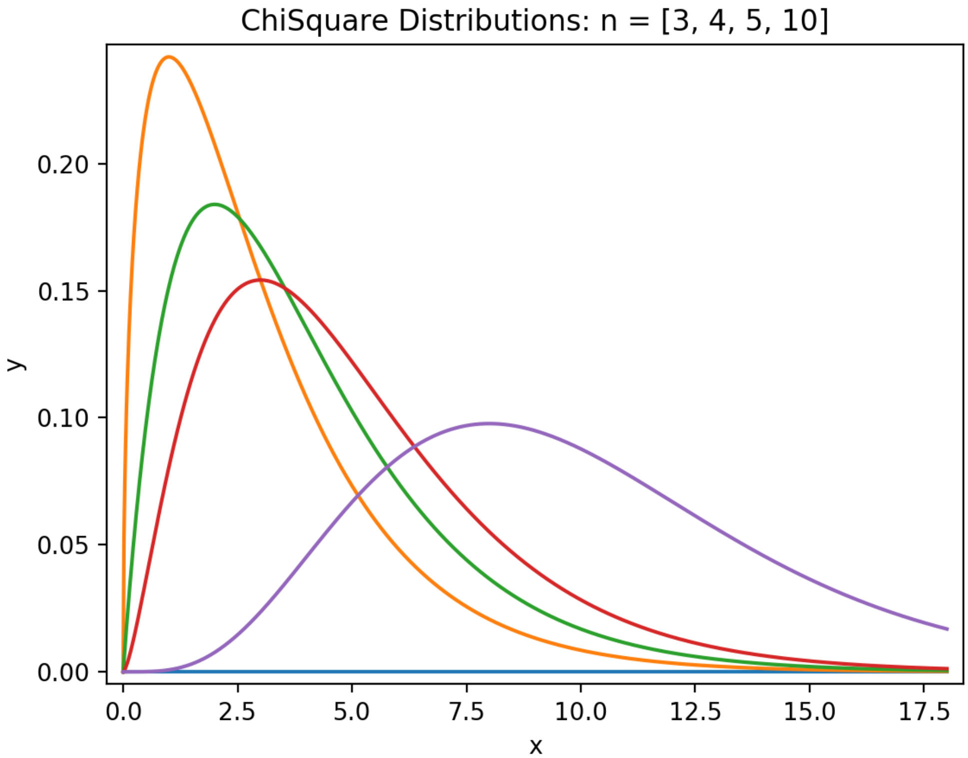

Abbildung 5.9: Die χ^2-Verteilungen mit den Freiheitsgraden 3, 4, 5 und 10.

Satz 51 *Es seien U, V und W Zufalls-Variablen mit folgenden Eigenschaften:*

1. *U ist χ^2-verteilt mit m Freiheitsgraden,*
2. *W ist χ^2-verteilt mit $m+n$ Freiheitsgraden,*
3. *U und V sind unabhängig,*
4. *$W = U + V$.*

Dann genügt die Zufalls-Variable V einer χ^2-Verteilung mit n Freiheitsgraden.

Beweis: Wir führen den Beweis indem wir zeigen, dass die moment-erzeugende Funktion M_V mit der moment-erzeugenden Funktion einer χ^2-verteilten Zufalls-Variable mit n Freiheitsgraden übereinstimmt.

Aus den ersten beiden Voraussetzungen folgt, dass die moment-erzeugenden Funktionen der Zufalls-Variablen U und W die folgende Gestalt haben:

$$M_U(t) = (1 - 2 \cdot t)^{-\frac{m}{2}} \quad \text{und} \quad M_W(t) = (1 - 2 \cdot t)^{-\frac{m+n}{2}}.$$

Nach Satz 41 gilt für die moment-erzeugenden Funktionen

$$M_W(t) = M_U(t) \cdot M_V(t)$$

Setzen wir hier die obigen Werte ein, so folgt

$$(1 - 2 \cdot t)^{-\frac{m+n}{2}} = (1 - 2 \cdot t)^{-\frac{m}{2}} \cdot M_V(t)$$

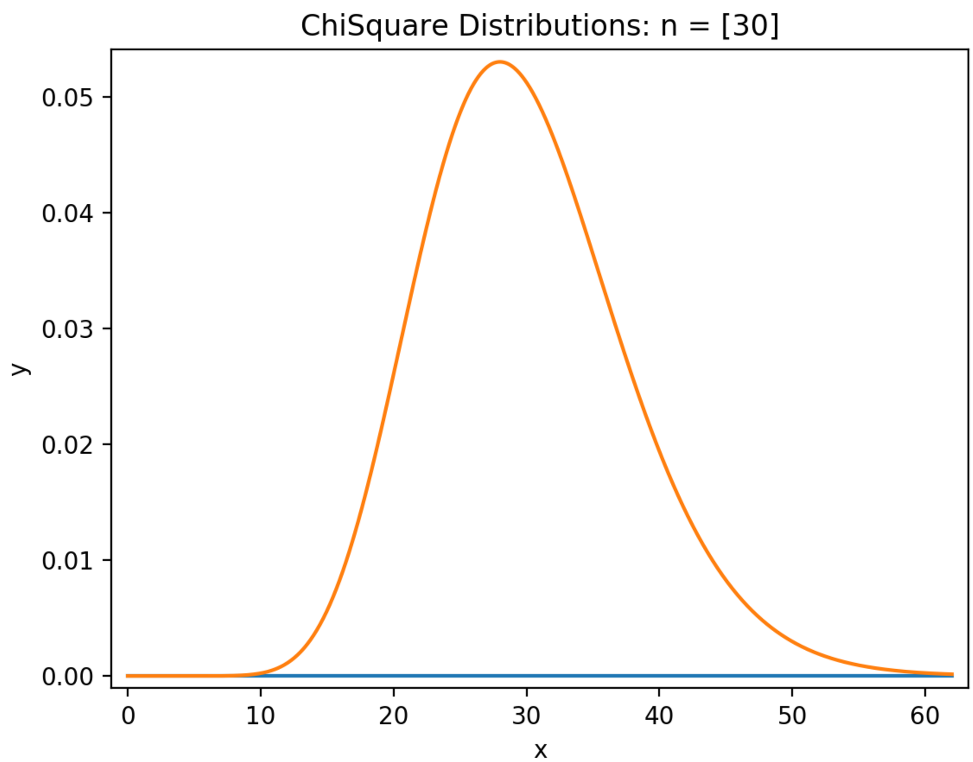

Abbildung 5.10: Die χ^2-Verteilungen mit dem Freiheitsgrad 30.

und das impliziert

$$M_V(t) = (1 - 2 \cdot t)^{-\frac{n}{2}}.$$

Dies ist aber genau die moment-erzeugende Funktion einer χ^2-verteilten Zufalls-Variable vom Freiheitsgrad n. Nach dem Eindeutigkeits-Satz 40 hat die Zufalls-Variable V also eine χ^2-Verteilung mit dem Freiheitsgrad n. □

Aufgabe 44: Nehmen Sie an, dass die Zufalls-Variable Z einer χ^2-Verteilung mit n Freiheitsgraden genügt. Berechnen Sie den Erwartungswert und die Varianz von X.
Hinweis: Berechnen Sie zunächst den Erwartungswert und die Varianz einer Zufalls-Variable Z, für die $Z \sim \chi^2(1)$ gilt. ◇

Aufgabe 45: Es sei X eine Zufalls-Variable, für die $X \sim \chi^2(n)$ gilt. An welcher Stelle nimmt die Wahrscheinlichkeits-Dichte $p_X(x)$ ihren maximalen Wert an?

Aufgabe 46: Es sei X eine Zufalls-Variable, die einer χ^2-Verteilung mit 4 Freiheitsgraden genügt. Berechnen Sie die Verteilungs-Funktion F_X. ◇

Kapitel 6

Induktive Statistik

Zwei wesentliche Aufgaben der Statistik sind das Schätzen von Parametern und das Testen von Hypothesen. Wir besprechen diese beiden Aufgaben in den beiden folgenden Abschnitten.

6.1 Parameter-Schätzung

Bisher haben wir uns im wesentlichen mit der Wahrscheinlichkeits-Rechnung beschäftigt. Bei der Wahrscheinlichkeits-Rechnung haben wir ein Model eines Prozesses in Form eines Wahrscheinlichkeits-Raums. Mit Hilfe dieses Models berechnen wir dann die Wahrscheinlichkeit für das Auftreten bestimmter Ereignisse. Damit das möglich ist, müssen uns die Parameter des Models bekannt sein. Ein solcher Parameter wäre beispielsweise der Erwartungswert λ einer Poisson-Verteilung oder der Erwartungswert μ und die Varianz σ^2 einer normal-verteilten Zufalls-Variable.

Bei der Statistik ist die Situation anders: Wir haben zwar meistens ebenfalls ein Model in Form eines Wahrscheinlichkeits-Raums, allerdings sind diesmal die Parameter, welche die Verteilung der Zufalls-Variablen festlegen, unbekannt. Um diese Parameter bestimmen zu können, beobachten wir Ereignisse und versuchen dann, mit Hilfe der Beobachtungen die Parameter zu erschließen. Wir haben diesen Prozess bereits in der Einführung am Beispiel der Ameisenzählung demonstriert. Wir wollen diesen Prozess jetzt formalisieren. Dazu benötigen wir einige Definitionen.

Definition 52 (Stichprobe) *Gegeben sei eine endliche oder unendliche Menge Ω, die wir im folgenden als Population (oder auch Grundgesamtheit) bezeichnen wollen. Ein Merkmal X dieser Population ist eine Funktion*

$$X : \Omega \to \mathbb{R}.$$

Eine Stichprobe vom Umfang n ist eine Teilmenge

$$\{\omega_1, \cdots, \omega_n\} \subseteq \Omega$$

die genau n Elemente enthält. Setzen wir $x_i := X(\omega_i)$ für $i = 1, \cdots, n$, so werden die x_i auch als Stichproben-Werte bezeichnet. □

6.1.1 Schätzung des Erwartungswertes einer Zufalls-Variable

Ist eine Stichprobe gegeben, so können wir versuchen, von der Stichprobe auf die Grundgesamtheit zu schließen. Ist beispielsweise $\{\omega_1, \cdots, \omega_n\}$ eine Stichprobe und X eine Zufalls-Variable, so wird der arithmetische Mittelwert von X mit $\overline{X}$ bezeichnet und durch

$$\overline{X} := \frac{1}{n} \cdot \sum_{i=1}^{n} X(\omega_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

definiert. Um das Verhalten von $\overline{X}$ mit den Mitteln der Wahrscheinlichkeits-Rechnung untersuchen zu können, gehen wir aus von dem Wahrscheinlichkeits-Raum

$$\mathcal{V} = \langle \Omega, \mathfrak{A}, P \rangle.$$

Hier ist Ω die Population, $\mathfrak{A}$ ist eine geeignete Σ-Algebra und P ein Wahrscheinlichkeits-Maß. Dann definieren wir das n-fache kartesische Produkt

$$\mathcal{W} := \mathcal{V}^n = \langle \Omega^n, \mathfrak{A}^n, \widehat{P} \rangle.$$

Für ein Tupel von Ereignissen $\langle A_1, \cdots, A_n \rangle$ ist das Wahrscheinlichkeits-Maß $\widehat{P}$ als

$$\widehat{P}\big(\langle A_1, \cdots, A_n \rangle\big) := P(A_1) \cdot \ldots \cdot P(A_n)$$

definiert. Dann können wir auf Ω^n die Zufalls-Variablen X_i als

$$X_i\big(\langle \omega_1, \cdots, \omega_i, \cdots, \omega_n \rangle\big) := X(\omega_i)$$

definieren, X_i ist also gerade die Anwendung von X auf die i-te Komponente der Stichprobe. Die Wahrscheinlichkeits-Verteilungen der Zufalls-Variablen X_i sind dann alle gleich der Wahrscheinlichkeits-Verteilung der Zufalls-Variable X auf dem ursprünglichen Wahrscheinlichkeits-Raum $\mathcal{V}$.

Mit den X_i ist auch der arithmetische Mittelwert $\overline{X}$ eine Zufalls-Variable auf dem Stichproben-Raum Ω^n. Wir können den Erwartungswert von $\overline{X}$ berechnen und finden

$$\begin{aligned}
E\Big[\overline{X}\Big] &= E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} E\left[X_i\right] \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} E\left[X\right] \\
&= \frac{1}{n} \cdot E\left[X\right] \cdot \sum_{i=1}^{n} 1 \\
&= \frac{1}{n} \cdot E\left[X\right] \cdot n \\
&= E\left[X\right].
\end{aligned}$$

Diese Rechnung zeigt, dass wir den arithmetischen Mittelwert einer Stichprobe benutzen können um den Erwartungswert der Zufalls-Variable X zu schätzen. Ist

$$\langle x_1, \cdots, x_n \rangle = \langle X(\omega_1), \cdots, X(\omega_n) \rangle$$

ein Stichproben-Ergebnis, so liefert uns dieses Ergebnis eine Schätzung des Mittelwerts $E[X]$:

$$E[X] \approx \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \overline{X}\big(\langle \omega_1, \cdots, \omega_n \rangle\big).$$

Wir untersuchen noch die Varianz der Zufalls-Variable $\overline{X}$, denn diese gibt uns Aufschluss über die Genauigkeit unserer Schätzung. Mit einer Rechnung, die analog zur Herleitung des $\sqrt{n}$-Gesetzes ist, kann gezeigt werden, dass

$$\mathrm{Var}\Big[\overline{X}\Big] = \frac{1}{n} \cdot \mathrm{Var}[X] \quad \text{und damit} \quad \sigma\Big[\overline{X}\Big] = \frac{1}{\sqrt{n}} \cdot \sigma[X]$$

gilt. Dies zeigt, wie die Genauigkeit mit wachsender Größe n der Stichprobe zunimmt. Der zentrale Grenzwertsatz zeigt, dass $\overline{X}$ für große Werte von n annähernd normal verteilt ist. In der Praxis zeigt sich, dass die Wahrscheinlichkeits-Verteilung von $\overline{X}$ für $n \geq 30$ hinreichend gut durch eine Normal-Verteilung approximiert wird.

6.1.2 Schätzung der Varianz einer Zufalls-Variable

Als nächstes suchen wir eine Zufalls-Variable, mit der wir die Varianz einer Zufalls-Variable X abschätzen können. Dazu definieren wir die Stichproben-Varianz S^2 einer Stichprobe vom Umfang n als

$$S^2 := \frac{1}{n} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

Die Zufalls-Variable S^2 ist genau wie die Zufalls-Variable $\overline{X}$ auf dem Stichproben-Raum Ω^n definiert. Wir werden im folgenden die Wahrscheinlichkeits-Verteilung von S^2 berechnen. In der Literatur wird die Stichproben-Varianz gelegentlich mit dem Faktor $\frac{1}{n-1}$ anstelle des Faktors $\frac{1}{n}$ definiert. Wir folgen bei der obigen Definition der Darstellung von Spiegel [SS00].

Wir berechnen jetzt den Erwartungswert der Zufalls-Variable S^2. Dazu definieren wir

$$\mu := E[X] \quad \text{und} \quad \sigma^2 := \mathrm{Var}[X].$$

Zunächst formen wir die Summe $\sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ um:

$$\begin{aligned}
& \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \\
= \; & \sum_{i=1}^{n} \left(X_i - \mu + \mu - \overline{X}\right)^2 \\
= \; & \sum_{i=1}^{n} \Big(\left(X_i - \mu\right)^2 + \left(\mu - \overline{X}\right)^2 + 2 \cdot (\mu - \overline{X}) \cdot (X_i - \mu)\Big) \\
= \; & \sum_{i=1}^{n} \left(X_i - \mu\right)^2 + \sum_{i=1}^{n} \left(\mu - \overline{X}\right)^2 + \sum_{i=1}^{n} 2 \cdot (\mu - \overline{X}) \cdot (X_i - \mu) \\
= \; & \sum_{i=1}^{n} \left(X_i - \mu\right)^2 + \sum_{i=1}^{n} \left(\mu - \overline{X}\right)^2 + 2 \cdot (\mu - \overline{X}) \cdot \sum_{i=1}^{n} (X_i - \mu) \\
= \; & \sum_{i=1}^{n} \left(X_i - \mu\right)^2 + n \cdot \left(\mu - \overline{X}\right)^2 + 2 \cdot (\mu - \overline{X}) \cdot n \cdot (\overline{X} - \mu) \\
& \text{denn } \sum_{i=1}^{n} X_i = n \cdot \overline{X} \text{ und } \sum_{i=1}^{n} \mu = n \cdot \mu \\
= \; & \sum_{i=1}^{n} \left(X_i - \mu\right)^2 + n \cdot \left(\mu - \overline{X}\right)^2 - 2 \cdot n \cdot \left(\mu - \overline{X}\right)^2 \\
= \; & \sum_{i=1}^{n} \left(X_i - \mu\right)^2 - n \cdot \left(\mu - \overline{X}\right)^2
\end{aligned}$$

Wir haben also

$$\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} \left(X_i - \mu\right)^2 - n \cdot \left(\mu - \overline{X}\right)^2 \tag{6.1}$$

gezeigt. Damit können wir $E\left[S^2\right]$ berechnen:

$$
\begin{aligned}
E\big[S^2\big] &= \frac{1}{n}\cdot E\left[\sum_{i=1}^{n}\big(X_i-\overline{X}\big)^2\right] \\
&= \frac{1}{n}\cdot E\left[\left(\sum_{i=1}^{n}\big(X_i-\mu\big)^2\right) - n\cdot\big(\mu-\overline{X}\big)^2\right] \\
&= \frac{1}{n}\cdot \left(\left(\sum_{i=1}^{n} E\left[\big(X_i-\mu\big)^2\right]\right) - n\cdot E\left[\big(\mu-\overline{X}\big)^2\right]\right) \\
&= \frac{1}{n}\cdot \left(\sum_{i=1}^{n} E\left[\big(X_i-\mu\big)^2\right] \; - \; \sum_{i=1}^{n} E\left[\big(\mu-\overline{X}\big)^2\right]\right)
\end{aligned}
$$

Nun gilt einerseits

$$E\left[\big(X_i-\mu\big)^2\right] = E\left[\big(X-\mu\big)^2\right] = \sigma^2,$$

denn die Zufalls-Variablen X_i haben ja alle dieselbe Verteilung wie die Zufalls-Variable X, andererseits haben wir wegen $E\big[\overline{X}\big] = E[X] = \mu$

$$E\left[\big(\mu-\overline{X}\big)^2\right] = \mathrm{Var}\big[\overline{X}\big] = \frac{1}{n}\cdot\sigma^2.$$

Das ergibt insgesamt

$$
\begin{aligned}
E\big[S^2\big] &= \frac{1}{n}\cdot \left(\sum_{i=1}^{n} E\left[\big(X_i-\mu\big)^2\right] \; - \; \sum_{i=1}^{n} E\left[\big(\mu-\overline{X}\big)^2\right]\right) \\
&= \frac{1}{n}\cdot\left(\sum_{i=1}^{n}\sigma^2 \; - \; \sum_{i=1}^{n}\frac{1}{n}\cdot\sigma^2\right) \\
&= \frac{1}{n}\cdot\big(n\cdot\sigma^2-\sigma^2\big) \\
&= \frac{n-1}{n}\cdot\sigma^2
\end{aligned}
$$

Auf den ersten Blick mag dieses Ergebnis verblüffen, denn es zeigt, dass die Stichproben-Varianz S^2 nicht dazu geeignet ist, die Varianz der Zufalls-Variable S zu schätzen. Es gilt aber

$$E\left[\frac{n}{n-1}\cdot S^2\right] = \sigma^2,$$

so dass wir die Zufalls-Variable

$$\widehat{S}^2 = \frac{n}{n-1}\cdot S^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}\big(X_i-\overline{X}\big)^2$$

zur Schätzung der Varianz verwenden können.

Der nächste Satz gibt uns Aufschluss über die Verteilung der Zufalls-Variable S^2 für den Fall, dass die zugrundeliegende Zufalls-Variable X normal verteilt ist und zeigt gleichzeitig die Bedeutung der χ^2-Verteilung.

Satz 53 *Ist die Zufalls-Variable X normal verteilt mit Erwartungswert μ und Varianz σ^2, und wird für eine n-elementige Stichprobe die Zufalls-Variable S^2 durch*

$$S^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}\big(X_i-\overline{X}\big)^2$$

definiert, so hat die Zufalls-Variable

$$\frac{n}{\sigma^2} \cdot S^2 = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

eine χ^2-Verteilung mit $n-1$ Freiheitsgraden.

Beweis: Wir definieren drei Zufalls-Variablen U, V und W durch

$$U := \frac{n}{\sigma^2} \cdot \left(\overline{X} - \mu\right)^2, \; V := \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \quad \text{und} \quad W := \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(X_i - \mu\right)^2.$$

Um zu sehen, wie diese Zufalls-Variablen zusammenhängen, schreiben wir Gleichung (6.1) um:

$$\sum_{i=1}^{n} \left(X_i - \mu\right)^2 = n \cdot \left(\mu - \overline{X}\right)^2 \; + \; \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

Wenn wir diese Gleichung durch σ^2 dividieren und die Gleichung $(\mu - \overline{X})^2 = (\overline{X} - \mu)^2$ berücksichtigen, dann sehen wir, dass

$$W = U + V$$

gilt. Wir untersuchen jetzt, wie die Zufalls-Variablen U, V und W verteilt sind. Wir beginnen mit der Zufalls-Variable W. Zunächst ist mit X_i auch $X_i - \mu$ normal verteilt und die Zufalls-Variable $(X_i - \mu)/\sigma$ ist normal verteilt mit Erwartungswert 0 und Varianz 1, also standard-normal-verteilt. Nach Satz 49 hat die Zufalls-Variable W daher eine χ^2-Verteilung mit n Freiheitsgraden.

Als nächstes untersuchen wir die Verteilung der Zufalls-Variable U. Nach Aufgabe 40 ist die Summe zweier normal verteilter Zufalls-Variablen selber wieder normal verteilt. Damit ist dann auch die Summe von n Zufalls-Variablen normal verteilt. Also ist die Zufalls-Variable $\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ normal verteilt. Also ist auch $\overline{X} - \mu$ normal verteilt. Der Erwartungswert von $\overline{X} - \mu$ ist 0 und die Varianz ist nach dem $\sqrt{n}$-Gesetz σ^2/n. Also hat die Zufalls-Variable

$$\frac{\sqrt{n}}{\sigma} \cdot \left(\overline{X} - \mu\right)$$

den Erwartungswert 0 und die Varianz 1. Wenn wir nun Satz 49 für den Fall $n = 1$ anwenden, sehen wir, dass die Zufalls-Variable U, die ja das Quadrat der Zufalls-Variable $\frac{\sqrt{n}}{\sigma} \cdot (\overline{X} - \mu)$ ist, einer χ^2-Verteilung mit Freiheitsgrad $n = 1$ genügt.

Leider können wir an dieser Stelle den folgenden Sachverhalt nicht beweisen:

Die Zufalls-Variablen U und V sind unabhängig.

Damit können wir jetzt Satz 51 auf die Zufalls-Variablen U, V und W anwenden und folgern, dass die Zufalls-Variable V einer χ^2-Verteilung mit $n-1$ Freiheitsgraden genügt. □

6.2 Testen von Hypothesen

Das Testen von Hypothesen ist eine der beiden Hauptaufgaben der induktiven Statistik, die beispielsweise in der Medizin häufig angewendet wird. Wir erläutern das Testen von Hypothesen an einem einführenden Beispiel. Wir wollen untersuchen, ob sich beim Werfen einer gegeben Münze die Ergebnisse "Wappen" und "Zahl" mit derselben Wahrscheinlichkeit einstellen. Wir kodieren das Ergebnis "Wappen" als 0 und das Ergebnis "Zahl" als 1. Unsere Hypothese ist, dass bei einem Wurf für die Wahrscheinlichkeit

$$P(\{0\}) = P(\{1\}) = \frac{1}{2}$$

gilt. Diese Hypothese bezeichnen wir als Null-Hypothese, denn es gibt Null Unterschied zwischen den beiden Wahrscheinlichkeiten. Um diese Hypothese zu überprüfen, werfen wir die Münze 100 mal. Angenommen, wir erhalten dabei 75 mal "Wappen" und 25 mal "Zahl". Dann können wir uns fragen, wie wahrscheinlich ein solches Ereignis ist, wenn wir die Null-Hypothese annehmen. Wenn diese Wahrscheinlichkeit kleiner ist als ein vorgegebener Wert α, dann würden wir die Null-Hypothese ablehnen und sagen, dass der Test auf dem Niveau α signifikant war. In der Praxis wird für α oft ein Wert von 5% oder 1% gewählt.

Bezeichnen wir die Häufigkeit, mit der das Ereignis "Wappen" eintritt ist, mit X_0 und die Häufigkeit, mit der das Ereignis "Zahl" eintritt, mit X_1, und ist n die Anzahl der Münzwürfe, so misst die Größe

$$C := \frac{(X_0 - n\cdot\frac{1}{2})^2}{n\cdot\frac{1}{2}} + \frac{(X_1 - n\cdot\frac{1}{2})^2}{n\cdot\frac{1}{2}}$$

wie stark X_0 und X_1 von dem Erwartungswert $n\cdot\frac{1}{2}$ abweichen. Wir teilen hier die Quadrate durch $n\cdot\frac{1}{2}$, weil wir später zeigen wollen, dass die Zufalls-Variable

$$2\cdot\frac{(X_0 - n\cdot\frac{1}{2})^2}{n\cdot\frac{1}{2}} = \frac{(X_0 - n\cdot\frac{1}{2})^2}{n\cdot\frac{1}{4}}$$

eine Chi-Quadrat-Verteilung mit dem Freiheitsgrad 1 hat. Die Zufalls-Variable C bezeichnen wir als eine Statistik. Für die oben genannten Werte von $X_0 = 75$ und $X_1 = 25$ ergibt sich für die Statistik C

$$C = \frac{(75-50)^2}{50} + \frac{(25-50)^2}{50} = 2\cdot\frac{25^2}{50} = 25$$

Um diesen Wert interpretieren zu können, berechnen wir die Wahrscheinlichkeit, dass die Zufalls-Variable C einen Wert größer oder gleich 25 annimmt, unter der Voraussetzung, dass die Null-Hypothese wahr ist. Dazu formen wir den Ausdruck für C um:

$$\begin{aligned}
C &= \frac{\left(X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} + \frac{\left(X_1 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} \\
&= \frac{\left(X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} + \frac{\left(n - X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} && \text{denn } X_1 = n - X_0 \\
&= \frac{1}{n\cdot\frac{1}{2}}\cdot\left(\left(X_0 - n\cdot\frac{1}{2}\right)^2 + \left(n - X_0 - n\cdot\frac{1}{2}\right)^2\right) \\
&= \frac{1}{n\cdot\frac{1}{2}}\cdot\left(\left(X_0 - n\cdot\frac{1}{2}\right)^2 + \left(n\cdot\frac{1}{2} - X_0\right)^2\right) \\
&= \frac{2}{n\cdot\frac{1}{2}}\cdot\left(X_0 - n\cdot\frac{1}{2}\right)^2 \\
&= \frac{\left(X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}\cdot(1-\frac{1}{2})}
\end{aligned}$$

Wenn die Null-Hypothese zutrifft, dann ist die Zufalls-Variable X_0 binomial verteilt mit dem Erwartungswert $\mu = n\cdot\frac{1}{2}$ und der Varianz $\sigma^2 = n\cdot\frac{1}{2}\cdot(1-\frac{1}{2})$. Für große Werte von n kann diese Verteilung durch eine Normal-Verteilung $N_{\mu,\sigma}$ mit Mittelwert μ und Varianz σ approximiert werden. Brauchbare Werte erhalten wir, sobald $\sigma^2 > 9$ gilt. Für $n = 100$ haben wir

$$\sigma^2 = n\cdot\frac{1}{2}\cdot\left(1-\frac{1}{2}\right) = 100\cdot\frac{1}{4} = 25,$$

so dass diese Bedingung erfüllt ist. Wenn X_0 aber ein Normal-Verteilung mit Mittelwert $n\cdot\frac{1}{2}$ und Varianz $n\cdot\frac{1}{2}\cdot(1-\frac{1}{2})$ hat, dann hat die Zufalls-Variable

$$Y := \frac{X_0 - n\cdot\frac{1}{2}}{\sqrt{n\cdot\frac{1}{2}\cdot(1-\frac{1}{2})}}$$

eine Standard-Normal-Verteilung, es gilt also $Y \sim \mathcal{N}(0,1)$. Im letzten Kapitel haben wir gesehen, dass dann Y^2 eine χ^2-Verteilung mit dem Freiheitsgrad 1 genügt. Die Zufalls-Variable Y^2 ist aber gleich C, so dass wir damit die Verteilung von C gefunden haben.

$$C = Y^2 \sim \chi^2(1).$$

Daher können wir jetzt die Wahrscheinlichkeit, dass C einen Wert größer oder gleich 25 annimmt, berechnen. Für die Wahrscheinlichkeits-Dichte einer χ^2-Verteilung mit dem Freiheitsgrad 1 hatten wir im letzten Kapitel den Ausdruck

$$p_C(x) = \frac{1}{A_1} \cdot \frac{1}{\sqrt{x}} \cdot \exp\left(-\frac{x}{2}\right) \quad \text{mit } A_1 = \sqrt{2} \cdot \Gamma\left(\frac{1}{2}\right) = \sqrt{2 \cdot \pi}$$

gefunden. Wir setzen zur Abkürzung $z = 25$ und rechnen wie folgt:

$$\begin{array}{lcl}
P(C \geq 25) & = & 1 - P(C < 25) \\
& \approx & 1 - \displaystyle\int_0^z p_C(x)\ \mathrm{d}x \\
& = & 1 - \displaystyle\int_0^z \frac{1}{\sqrt{2 \cdot \pi \cdot x}} \cdot \exp\left(-\frac{x}{2}\right)\ \mathrm{d}x \\
& & \text{Substitution } x = y^2\text{, also } \mathrm{d}x = 2 \cdot y\ \mathrm{d}y \\
& = & 1 - \displaystyle\int_0^{\sqrt{z}} \frac{1}{\sqrt{2 \cdot \pi} \cdot y} \cdot \exp\left(-y^2/2\right) \cdot 2 \cdot y\ \mathrm{d}y \\
& = & 1 - 2 \cdot \dfrac{1}{\sqrt{2 \cdot \pi}} \cdot \displaystyle\int_0^{\sqrt{z}} \exp\left(-y^2/2\right)\ \mathrm{d}y \\
& = & 1 - 2 \cdot \left(\dfrac{1}{\sqrt{2 \cdot \pi}} \cdot \displaystyle\int_{-\infty}^{\sqrt{z}} \exp\left(-y^2/2\right)\ \mathrm{d}y - \dfrac{1}{\sqrt{2 \cdot \pi}} \cdot \displaystyle\int_{-\infty}^{0} \exp\left(-y^2/2\right)\ \mathrm{d}y\right) \\
& = & 1 - 2 \cdot \left(\dfrac{1}{\sqrt{2 \cdot \pi}} \cdot \displaystyle\int_{-\infty}^{\sqrt{z}} \exp\left(-y^2/2\right)\ \mathrm{d}y - \dfrac{1}{\sqrt{2 \cdot \pi}} \cdot \dfrac{1}{2} \cdot \displaystyle\int_{-\infty}^{\infty} \exp\left(-y^2/2\right)\ \mathrm{d}y\right) \\
& = & 1 - 2 \cdot \left(\dfrac{1}{\sqrt{2 \cdot \pi}} \cdot \displaystyle\int_{-\infty}^{\sqrt{z}} \exp\left(-y^2/2\right)\ \mathrm{d}y - \dfrac{1}{2} \cdot 1\right) \\
& = & 2 - 2 \cdot \dfrac{1}{\sqrt{2 \cdot \pi}} \cdot \displaystyle\int_{-\infty}^{\sqrt{z}} \exp\left(-y^2/2\right)\ \mathrm{d}y \\
& = & 2 - 2 \cdot \Phi\big(\sqrt{z}\big) \\
& = & 2 - 2 \cdot \big(1 - \Phi\big(-\sqrt{z}\big)\big) \quad \text{denn } \Phi(x) + \Phi(-x) = 1 \\
& = & 2 \cdot \Phi\big(-\sqrt{z}\big) = 2 \cdot \Phi(-5) \approx 5.7 \cdot 10^{-7}
\end{array}$$

Wir können den Wert des gaußschen Fehlerintegrals $\Phi(x)$ mit SetlX berechnen, denn es gilt

$\Phi(x) =$ `stat_normalCDF(x, 0, 1)`.

Wenn die Null-Hypothese zutrifft, dann ist die Wahrscheinlichkeit, dass die Zufalls-Variable C einen Wert größer oder gleich 25 annimmt, verschwindend gering. Daher werden wir die Null-Hypothese verwerfen. ◇

Bemerkung: Eventuell fragen Sie sich, warum ich oben den Ausdruck $2-2\cdot\Phi\bigl(\sqrt{z}\bigr)$ in den äquivalenten Ausdruck $2\cdot\Phi\bigl(-\sqrt{z}\bigr)$ umgeformt habe. Der Grund ist, dass der erste Ausdruck numerisch instabil ist: Für große Werte von $\sqrt{z}$ hat $\Phi\bigl(\sqrt{z}\bigr)$ annähernd den Wert 1 und die Subtraktion führt dann zu einer Auslöschung. Setzen wir beispielsweise $z := 100$, so liefert SetlX für den Ausdruck $\Phi(10)$ das Ergebnis 1.0, so dass $2-2\cdot\Phi(10)$ dann den Wert `0.0` hat. Dem gegenüber liefert der Ausdruck $2\cdot\Phi(-10)$ das Ergebnis `1.523970604832111E-23`, das zwar sehr klein, aber eben nicht exakt 0 ist. ⋄

Aufgabe 47: Berechnen Sie die Wahrscheinlichkeit, dass beim Werfen einer fairen Münze das Ergebnis "Wappen" mindestens 75 mal. Benutzen Sie zur Lösung dieser Aufgabe die Binomial-Verteilung und berechnen Sie die Summe, die bei der Lösung dieser Aufgabe auftritt, mit einem geeigneten Programm. Das Ergebnis, das Sie erhalten werden, ist etwa halb so groß wie der Wert, den wir mit Hilfe der χ^2-Quadrat-Verteilung erhalten haben. Der Grund ist, dass die Normal-Verteilung nur für solche Werte, die nicht zu weit vom Mittelwert entfernt sind, eine gute Näherung für die Binomial-Verteilung ist. ⋄

Beispiel: Wir verallgemeinern das letzte Beispiel und betrachten nun eine Anwendung der Statistik aus dem wirklichen Leben: Die Würfel, die an der DHBW zur Notenfindung eingesetzt werden, werden jedes Jahr in einem aufwendigen Prozess[1] geeicht. Dazu wird mit dem zu eichenden Würfel 100 mal gewürfelt und die Häufigkeit der einzelnen Ziffern wird notiert. Da an der DHBW nur Noten von 1 bis 5 vergeben werden, handelt es sich bei den verwendeten Würfeln um kostspielige Spezialanfertigungen, mit denen sich nur Zahlen aus der Menge $\{1,\cdots,5\}$ würfeln lassen. Bei meinem Würfel ergab sich die dabei die folgenden Verteilung der Häufigkeiten:

Zahl	1	2	3	4	5
Häufigkeit	10	16	14	35	25

Wir würden erwarten, dass jede Zahl mit einer durchschnittlichen Häufigkeit von $100\cdot\frac{1}{5}=20$ in der Tabelle erscheint, aber es ist auch klar, dass es statistische Schwankungen dieser Häufigkeiten geben wird. Für $i=1,\cdots,5$ bezeichnen wir die Häufigkeit des Auftretens der Zahl i mit X_i. Als ein Maß für die Abweichung der oben gezeigten Häufigkeiten von dem Erwartungswert verwenden wir die Zufalls-Variable C, die wir in Analogie zu dem letzten Beispiel als

$$C=\sum_{i=1}^{5}\frac{(X_i-20)^2}{100\cdot\frac{1}{5}}=\frac{1}{20}\cdot\sum_{i=1}^{5}(X_i-20)^2$$

definieren. Für meinen Würfel ergibt sich dabei ganz konkret der Wert

$$C=\frac{1}{20}\cdot(10^2+4^2+6^2+15^2+5^2)=\frac{402}{20}=20{,}1.$$

Wir fassen C nun als Zufalls-Variable auf und überlegen uns, wie wahrscheinlich es ist, dass die Zufalls-Variable C einen Wert größer oder gleich $20{,}1$ annimmt, wir berechnen also die Wahrscheinlichkeit

$$P(C\geq 20{,}1).$$

Um diese Wahrscheinlichkeit berechnen zu können, müssen wir zunächst eine Annahme machen, die uns zeigt, wie die Zufalls-Variablen X_i verteilt sind. Wir machen die Annahme, dass es sich bei dem Würfel um einen Laplace-Würfel handelt, dass also die Wahrscheinlichkeit für das Auftreten jeder einzelnen Ziffer den Wert $\frac{1}{5}$ hat. Diese Annahme ist jetzt die zu prüfende Null-Hypothese.

Die Zufalls-Variablen X_i sind dann binomial verteilt und es gilt

$$P\bigl(X_i=k\bigr)=\binom{100}{k}\cdot\left(\frac{1}{5}\right)^k\cdot\left(\frac{4}{5}\right)^{100-k}.$$

Da $100\cdot\frac{1}{5}\cdot\frac{4}{5}=16>9$ ist, können wir diese Verteilung durch eine Normalverteilung approximiert. Wir setzen also

[1] Unter den Studenten ist dieser Prozess als die sogenannte T2000-Prüfung gefürchtet.

$$\mu = 100 \cdot \frac{1}{5} = 20 \quad \text{und} \quad \sigma^2 = 100 \cdot \frac{1}{5} \cdot \frac{4}{5} = 16.$$

und haben dann für die Verteilung der Zufalls-Variablen X_i die Approximation

$$P\big(X_i \leq k\big) \approx \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{k} \exp\left(-\frac{(k-\mu)^2}{2 \cdot \sigma^2}\right)\, \mathrm{d}x$$

In Analogie zur letzten Aufgabe kann gezeigt werden, dass die Zufalls-Variable C einer χ^2-Verteilung mit $5-1$ Freiheitsgraden genügt. Die 1, die wir von der 5 abziehen, hat ihre Ursache in der Tatsache, dass die fünf Zufalls-Variablen $X_1, \cdots, X_5$ nicht unabhängig sind, denn wenn wir X_1, X_2, X_3 und X_4 kennen, dann können wir X_5 über die Formel $X_1 + X_2 + X_3 + X_4 + X_5 = 100$ berechnen. Dies erklärt, dass die χ^2-Verteilung nicht fünf, sondern nur vier Freiheitsgrade hat. Wir wollen dieses Ergebnis hier nicht beweisen sondern nur darauf hinweisen, dass die Situation ganz ähnlich ist wie im Satz 53. Also gilt

$$\begin{aligned}
P(C \geq 20{,}1) &= 1 - P(C < 20{,}1) \\
&= 1 - P\left(C \leq z\right) \quad \text{mit } z := 20{,}1 \\
&= 1 - \int_0^z \frac{1}{A_4} \cdot x \cdot \exp\left(-\frac{x}{2}\right)\, \mathrm{d}x \\
& \quad\ \text{mit} \quad A_4 = 2^2 \cdot \Gamma(2) = 4
\end{aligned}$$

Wir berechnen das Integral separat:

$$\begin{aligned}
& \int_0^z \frac{1}{4} \cdot x \cdot \exp\left(-\frac{x}{2}\right)\, \mathrm{d}x \quad \text{mit der Substitution } y = \frac{x}{2} \text{ wird daraus} \\
= \ & \int_0^{z/2} y \cdot \exp\left(-y\right)\, \mathrm{d}y \\
& \text{partielle Integration: } u(y) = y,\ v'(y) = \exp(-y) \\
& \text{also:} \qquad\qquad u'(y) = 1,\ v(y) = -\exp(-y) \\
= \ & -y \cdot \exp(-y)\Big|_0^{z/2} + \int_0^{z/2} \exp\left(-y\right)\, \mathrm{d}y \\
= \ & -\frac{z}{2} \cdot \exp\left(-\frac{z}{2}\right) + \int_0^{z/2} \exp\left(-y\right)\, \mathrm{d}y \\
= \ & -\frac{z}{2} \cdot \exp\left(-\frac{z}{2}\right) + \left(-\exp(-y)\Big|_0^{z/2}\right) \\
= \ & -\frac{z}{2} \cdot \exp\left(-\frac{z}{2}\right) - \exp\left(-\frac{z}{2}\right) + 1 \\
= \ & 1 - \left(\frac{z}{2} + 1\right) \cdot \exp\left(-\frac{z}{2}\right)
\end{aligned}$$

Damit finden wir für die gesuchte Wahrscheinlichkeit $P(C \geq 20{,}1)$

$$\begin{aligned}
P(C \geq 20{,}1) &= 1 - \left(1 - \left(\frac{z}{2} + 1\right) \cdot \exp\left(-\frac{z}{2}\right)\right) \\
&= \left(\frac{z}{2} + 1\right) \cdot \exp\left(-\frac{z}{2}\right) \\
&\approx 4.8 \cdot 10^{-4}
\end{aligned}$$

Diese Wahrscheinlichkeit ist sehr klein. Dies deutet mit hoher Sicherheit darauf hin, dass die Null-Hypothese falsch ist. Wir verwerfen daher die Null-Hypothese und sehen, dass sich die Wahrscheinlichkeit der einzelnen Zahlen bei meinem Würfel deutet von dem Wert $\frac{1}{5}$ unterscheidet. ◇

Bemerkung: Das in dem obigen Beispiel skizzierte Verfahren wird als χ^2-Test bezeichnet. Die nächste Aufgabe zeigt, wie dieser Test in der Medizin angewendet wird.

Aufgabe 48: Zur Untersuchung der Frage, ob ein Zusammenhang zwischen Zigaretten-Rauchen und dem Auftreten von Lungenkrebs besteht, wurde eine Gruppe von $30\,000$ Rauchern über einen Zeitraum von 10 Jahren beobachtet. Am Ende des Tests wurde für jede einzelne Person überprüft, ob die Person während des Tests an Lungenkrebs erkrankt ist. Dabei ergaben sich die folgenden Ergebnisse:

	Raucher	Nichtraucher	Gesamt
Lungenkrebs	62	14	76
kein Lungenkrebs	9 938	19 986	29 924
Gesamt	10 000	20 000	30 000

Die zu testende Null-Hypothese lautet in diesem Fall, dass es zwischen dem Rauchen und dem Auftreten von Lungenkrebs keinen Zusammenhang gibt. Folglich hätte die Wahrscheinlichkeit p, dass eine Person an Lungenkrebs erkrankt, in beiden Fällen den selben Wert. Im Gegensatz zu dem letzten Beispiel kennen wir den Wert von p nicht und müssen ihn daher schätzen. Als Schätzwert wählen wir das Verhältnis der Gesamtzahl der Lungenkrebs-Erkrankungen zu der Zahl aller Test-Personen und erhalten

$$p = \frac{76}{30\,000} = 0.25\overline{3}\,\%.$$

Überprüfen Sie die Null-Hypothese mit Hilfe des χ^2-Tests.

Bemerkung: Selbstverständlich handelt es sich bei allen Zahlen, die ich den Beispielen verwende um Fakten. Bei den obigen Zahlen handelt es sich sogar um Fakten, die ich mir nicht selber ausgedacht habe. Sie können diese Zahlen in dem Buch von Sheldon M. Ross [Ros04] wiederfinden.

Lösung: Wir bezeichnen die Anzahl der an Lungenkrebs erkrankten Rauchern mit $X_{1,1}$, die Anzahl der Raucher, die nicht erkrankt sind mit $X_{2,1}$, die Anzahl der erkrankten Nichtraucher mit $X_{1,2}$ und die Anzahl der Nichtraucher, die nicht erkrankt sind, mit $X_{2,2}$. Weiter sei N_1 die Anzahl der Raucher und N_2 sei die Anzahl der Nichtraucher, es gilt also

$$N_1 = 10\,000 \quad \text{und} \quad N_2 = 20\,000.$$

In Analogie zu den vorigen Beispielen definieren wir die Zufalls-Variable C als

$$C = \frac{\big(X_{1,1} - N_1 \cdot p\big)^2}{N_1 \cdot p} + \frac{\big(X_{2,1} - N_1 \cdot (1-p)\big)^2}{N_1 \cdot (1-p)} + \frac{\big(X_{1,2} - N_2 \cdot p\big)^2}{N_2 \cdot p} + \frac{\big(X_{2,2} - N_2 \cdot (1-p)\big)^2}{N_2 \cdot (1-p)}$$

Setzen wir die konkreten Werte ein, so erhalten wir

$$C \approx 79{,}81$$

Wie in den vorherigen Beispielen auch hat C eine χ^2-Verteilung, wir müssen aber noch die Anzahl der Freiheitsgrade bestimmen.

1. Wir haben vier Zufalls-Variablen $X_{1,1}$, $X_{2,1}$, $X_{1,2}$ und $X_{2,2}$.
2. Zwischen diesen vier Zufalls-Variablen gibt es drei verschiedene Beziehungen:
   (a) $X_{1,1} + X_{2,1} = N_1$
   (b) $X_{1,2} + X_{2,2} = N_2$
   (c) $\dfrac{X_{1,1} + X_{1,2}}{N_1 + N_2} = p$

Daher hat die χ^2-Verteilung nur $4 - 3 = 1$ Freiheitsgrad! Setzen wir $z := 79{,}81$ so gilt also

$$
\begin{aligned}
P(C \geq 79{,}81) &= P(C \geq z) \\
&= 1 - P(C < z) \\
&\approx 1 - \int_0^z p_C(x)\, \mathrm{d}x \\
&= 1 - \int_0^z \frac{1}{\sqrt{2\pi \cdot x}} \cdot \exp\left(-\frac{x}{2}\right)\, \mathrm{d}x
\end{aligned}
$$

Das Integral, das hier auftritt, haben wir bereits zu Beginn dieses Abschnitts berechnet. Wir hatten damals die Formel

$$P(C \geq z) = 1 - \int_0^z \frac{1}{\sqrt{2\pi \cdot x}} \cdot \exp\left(-\frac{x}{2}\right)\, \mathrm{d}x = 2 \cdot \Phi\big(-\sqrt{z}\big)$$

gefunden. Mit $z = 79{,}81$ gilt

$$P(C \geq z) \approx 2 \cdot \Phi\big(-\sqrt{79{,}81}\big) \approx \texttt{4,1220121535273025E-19}.$$

Dieser Wert ist so klein, dass der Zufall als Ursache für die höhere Zahl der Lungenkrebs-Erkrankungen bei den Rauchern praktisch ausgeschlossen werden kann. Das heißt natürlich noch lange nicht, dass zwischen dem Rauchen und dem Auftreten von Lungenkrebs eine kausale Beziehung besteht, denn die beobachtete Korrelation könnte auch andere Ursachen haben:

1. Nehmen wir einmal an, dass Rauchen vor Aids schützt[2]. Wenn dann die Nichtraucher zu einem nennenswerten Prozentsatz an Aids wegsterben würden, noch bevor Sie ihre Ansprüche auf einen vollwertigen Lungenkrebs realisieren könnten, so würde dies die beobachteten Unterschiede erklären.

2. Es wäre theoretisch möglich, dass es ein Gen gibt, dass einerseits das Auftreten von Lungenkrebs begünstigt, andererseits aber auch den gesunden Menschenverstand dahingehend beeinträchtigt, dass Personen, die dieses Gen besitzen, verstärkt zu Rauchern werden. Wenn dies der Fall wäre, so würde es einem Raucher wenig nützen, wenn er das Rauchen aufgibt, denn einerseits würde sich dadurch sein Risiko, an Lungenkrebs zu erkanken, nicht ändern und andererseits würde sich auch sein Intelligenz-Quotient nicht erhöhen. Das einzige was sich erhöhen würde, wäre das Risiko an Aids zu erkanken.

   Einer der größten Statistiker des letzten Jahrhunderts, Sir Ronald Aylmer Fisher (1890 - 1962), war tatsächlich der Ansicht, dass die Korrelation zwischen dem Rauchen und dem Auftreten von Lungenkrebs genetisch bedingt ist. Der Fairness halber soll erwähnt werden, dass zu dem Zeitpunkt, an dem Sir Ronald diese steile Hypothese aufstellte, der medizinische Wissensstand noch keine klare Aussage zuließ. Der Vollständigkeit halber soll allerdings auch erwähnt werden, dass Sir Ronald bei der Tabak-Industrie unter Vertrag stand und selber ein leidenschaftlicher Zigarren-Raucher war. Um die Geschichte abzuschließen, bemerken wir noch, dass Sir Ronald dann auch nicht an Lungenkrebs gestorben ist, sondern an Kehlkopfkrebs. Dies ist die bevorzugte Krebsart bei Zigarren-Rauchern und hat in etwa den selben Spaßfaktor wie Lungenkrebs. ◇

Wir fassen die bisherigen Beispiele in einem Satz zusammen, der 1900 von Karl Pearson (1857–1936) bewiesen wurde. Leider liegt ein Beweis dieses Satzes außerhalb unserer Möglichkeiten.

Satz 54 (χ^2-Test) *Es seien n verschiedene Zufalls-Variablen X_1, $\cdots$, X_n gegeben. Weiter sei eine Null-Hypothese H_0 gegeben. Wenn diese Hypothese zutrifft, dann gelte:*

1. *Die Zufalls-Variablen genügen (zumindest näherungsweise) einer Normalverteilung.*

2. *Zwischen den Zufalls-Variablen bestehen k unabhängige Beziehungen.*

   Mit unabhängig ist hier gemeint, dass keine der Beziehungen aus den restlichen $k-1$ Beziehungen gefolgert werden kann.

[2] Diese Annahme ist keinesweg so absurd wie sie zunächst erscheinen mag, denn bekanntlich macht Rauchen häßlich.

Dann genügt die Zufalls-Variable

$$C = \sum_{i=1}^{n} \frac{(X_i - E[X_i])^2}{E[X_i]}$$

einer χ^2*-Verteilung mit* $n - k$ *Freiheitsgraden:*

$$C \sim \chi^2(n-k).$$

Nimmt die Zufalls-Variable also einen bestimmen Wert z *an, so gilt für die Wahrscheinlichkeit, dass* $C \geq z$ *ist*

$$P(C \geq z) = 1 - \int_0^z p_C(x)\ dx.$$

Ist diese Wahrscheinlichkeit kleiner als ein vorgegebener Wert α *(in der Praxis nimmt man oft* $\alpha = 0.05$*,* $\alpha = 0.01$ *oder* $\alpha = 0.001$*), so sagen wir, dass die Null-Hypothese auf dem Signifikanz-Niveau* α *verworfen werden kann.* □

Aufgabe 49: Besuchen Sie im Internet die Seite

`http://www.virtualgrave.eu/html_core/ksiega_zmarlych/`

Bei dieser Seite handelt es sich um einen virtuellen Friedhof. Notieren Sie die Wochentage des Ablebens der Verblichenen. Überprüfen Sie die Hypothese, dass für den Todestag eines Menschen alle Wochentage dieselbe Wahrscheinlichkeit haben mit Hilfe des χ^2-Tests und interpretieren Sie das Ergebnis. ⋄

Literaturverzeichnis

[BH14] Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability*. CRC Press, 2014.

[Bul79] Michael G. Bulmer. *Principles of Statistics*. Dover Publications, 1979.

[FHK+16] Ludwig Fahrmeir, Christian Heumann, Rita Künstler, Iris Pigeot, and Gerhard Tutz. *Statistik — Der Weg zur Datenanalyse*. Springer, 8te Auflage, 2016.

[GH85] Donald Gross and Carl M. Harris. *Fundamentals of Queueing Theory*. Wiley, 1985.

[Gil14] Peter Gill. *Misleading DNA Evidence*. Academic Press, 2014.

[GS97] Charles M. Grinstead and Laurie J. Snell. *Introduction to Probability*. American Mathematical Society, 1997.

[Lew03] Michael Lewis. *Moneyball: The Art of Winning an Unfair Game*. Norton & Company, 2003.

[NIS12] NIST. *Engineering Statistics*. National Institute of Standards and Technology, 2012. http://www.itl.nist.gov/div898/handbook/index.htm.

[Ros04] Sheldon M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists*. Elsevier Academic Press, 2004.

[SS00] Murray R. Spiegel and Larry J. Stephens. *Probability and Statistics*. McGraw-Hill; New York, 2000.