

WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK

2-stündige Vorlesung für den Bachelor-Studiengang
Angewandte Informatik

Version vom Wintersemester 2018/19

Inhaltsverzeichnis

0	Einleitung	7
1	Grundbegriffe der Wahrscheinlichkeitsrechnung	11
1.1	Grundlegende Begriffe	11
1.2	Laplace-Experimente; Berechnung von Wahrscheinlichkeiten durch kombinatorische Überlegungen	16
1.3	Mehrstufige Zufallsexperimente	20
1.4	Bedingte Wahrscheinlichkeiten, der Satz von Bayes und Beispiele	23
2	Diskrete Zufallsvariable. Wichtige diskrete Wahrscheinlichkeitsverteilungen	33
2.1	Diskrete Zufallsvariable	33
2.2	Erwartungswert und Varianz diskreter Zufallsvariablen	35
2.3	Wichtige diskrete Wahrscheinlichkeitsverteilungen	39
2.3.1	Die Binomialverteilung	39
2.3.2	Die geometrische Verteilung	42
2.3.3	Die negative Binomial-Verteilung	43
2.3.4	Die hypergeometrische Verteilung	45
2.3.5	Die Poisson-Verteilung	47
2.4	Schwaches Gesetz großer Zahlen	48
3	Stetige Zufallsvariable	51
3.1	Dichte und Verteilungsfunktion	51
3.2	Erwartungswert und Varianz stetiger Zufallsvariablen	53
3.3	Wichtige stetige Wahrscheinlichkeitsverteilungen	53
3.3.1	Gamma-Verteilung und Exponentialverteilung	53
3.3.2	Die Normal-Verteilung	54
3.3.3	Die χ^2 -Verteilung	59
3.3.4	Die Student-Verteilung	60
4	Grenzwertsätze	61
4.1	Der Satz von de Moivre-Laplace	62
4.2	Der zentrale Grenzwertsatz	66
5	Statistik	69
5.1	Elemente der deskriptiven Statistik	69
5.2	Schätzprobleme	73

5.3	Konfidenzintervalle	80
5.4	Hypothesentests	85
	Grundlegendes aus der Analysis	91

Vorwort

Die Vorlesung “Wahrscheinlichkeitsrechnung und Statistik” wurde bis zum WS 2014/15 von Gerhard Freiling, Hans-Bernd Knoop und Frank Müller betreut. Seit dem Wintersemester 2015/16 ist mir die Vorlesung anvertraut worden.

Grundlage der aktuellen Version des Skripts ist das Skript meiner Vorgänger Gerhard Freiling und Hans-Bernd Knoop. Das Kapitel “Grundlegendes aus der Analysis” wurde von Frank Müller ergänzt. Die von mir vorgenommenen Änderungen bestehen im Wesentlichen aus Kürzungen des ursprünglichen Skripts und einer Umstellung in der Reihenfolge der behandelten Themen.

Ich möchte mich bei meinen Vorgängern für die Zurverfügungstellung Ihrer Materialien herzlich bedanken. Ebenfalls möchte ich mich bei Frau Mirjam Rieser für die Hilfe bei der Bearbeitung des Skriptes bedanken.

Ebenfalls habe ich mich von dem Lehrbuch von Hartmann [1] inspirieren lassen. Dieses kann ich Studierenden der Informatik als Lektüre empfehlen, es ist über die Bibliotheksseite der Universität auch online verfügbar.

Ursula Ludwig, Universität Duisburg-Essen, WS 2018/19.

Kapitel 0

Einleitung

Im Zusammenhang mit der Mathematik des Zufalls hört man die Begriffe *Wahrscheinlichkeitsrechnung*, *Wahrscheinlichkeitstheorie*, (*mathematische*) *Statistik* und *Stochastik*. Die beiden ersten Begriffe werden synonym verwandt, das Wort *Stochastik* wird dagegen als Sammelbegriff für die *Wahrscheinlichkeitsrechnung* und die *Statistik* benutzt. Diesen beiden Teilgebieten der *Stochastik* wollen wir uns in dieser Vorlesung zuwenden.

Dass beim Umgang mit Zufallsphänomenen selbst Fachleute - sprich Mathematiker - Irrtümern unterliegen können, zeigt das sog. "Ziegenproblem" aus dem Jahr 1991, das längere Zeit "heiß" diskutiert wurde. Es geht um folgendes "Glückspiel":

Ein Teilnehmer einer (amerikanischen) Fernsehshow erhält die Gelegenheit, ein Auto zu gewinnen. Dazu sind auf der Bühne drei geschlossene Türen aufgebaut. Hinter einer dieser Türen befindet sich ein Auto, hinter den beiden anderen eine Ziege, sozusagen als Niete. Der Kandidat wählt nun eine der Türen aus, die aber zunächst verschlossen bleibt. Der Spielleiter, der genau weiß, hinter welcher Tür das Auto steht, zeigt dem Kandidaten durch das Öffnen einer der beiden anderen Türen eine Ziege. Daraufhin erhält der Kandidat die Möglichkeit, bei seiner Wahl zu bleiben oder aber die andere noch verschlossene Tür zu wählen.

Soll der Kandidat umwählen oder nicht? Hierüber entbrannte eine heiße Diskussion. Jeder glaubte seine Empfehlung richtig begründen zu können. Hier eine Auswahl der "Begründungen":

1. Die Chance, das Auto zu gewinnen ist $\frac{1}{3}$. Also ist es gleichgültig, ob umgewählt wird oder nicht.
2. Die Wahrscheinlichkeit, bei der ersten Wahl das Auto zu treffen, ist geringer als die, eine Ziege zu treffen. Wenn aber eine Ziegentür geöffnet ist, stehen die Chancen für Auto und Ziege 50:50. Also ist es besser, umzuwählen, denn so verbessert man die Gewinnchance.
3. Am Anfang erwischt man mit einer Wahrscheinlichkeit von $\frac{2}{3}$ eine Ziege. Man hat also im ersten Anlauf eher eine Niete. Deshalb sollte man umwählen. Das wäre logischer.

4. Wählt man grundsätzlich um, so gewinnt man das Auto nicht, wenn man schon bei der ersten Wahl die Autotür getroffen hatte.

In der Tat erhöht das Umwählen die Gewinnchancen auf das Auto. Wer nämlich im ersten Durchgang eine Ziegentür gewählt hat - und die Wahrscheinlichkeit dafür beträgt $\frac{2}{3}$ -, gewinnt durch Umwählen das Auto. Wer dagegen im ersten Durchgang die Tür mit dem Auto gewählt hat - und die Wahrscheinlichkeit dafür beträgt "nur" $\frac{1}{3}$ -, bekommt durch Umwählen eine Ziege und hat damit Pech. Umwählen erhöht also die Gewinnchance von $\frac{1}{3}$ auf $\frac{2}{3}$. Wir werden später noch einmal ausführlich auf die "theoretischen" Hintergründe für diesen Sachverhalt eingehen.

Das **zweite** Beispiel bezieht sich eher auf die Statistik. Nachdem im Wintersemester 2007/08 an der Universität Duisburg-Essen der Bachelor-Studiengang "Angewandte Informatik" den Diplom-Studiengang abgelöst hat, schließen am Ende des Sommersemesters 2010 die ersten 10 und am Ende des Wintersemesters 2010/11 weitere 10 Studierende das Bachelor-Studium ab. Daraus wird der Schluss gezogen, dass die durchschnittliche Studiendauer in dem neuen Studiengang bei 6,5 Semestern liegt und damit fast bei der Regelstudienzeit. Also beschleunigen die neuen Studiengänge das Studium.

Wer sonst sollte denn nach 6 Semestern - einmal abgesehen von Seiteneinsteigern, die an anderen Hochschulen schon zu einem früheren Termin das Bachelor-Studium beginnen konnten, und von Wechslern aus dem Diplom-Studiengang, die in ein höheres Semester im Bachelor-Studiengang eingestuft wurden, - das Studium beenden? Das müssen die Anfänger aus dem WS 2007/08 sein, und zwar diejenigen, die nach 6 Semestern alle notwendigen Prüfungen abgelegt haben. Dies sind die sehr guten und guten Studierenden aus dem genannten Anfangssemester. Aus den Zahlen alleine kann man gar keinen Rückschluss auf die durchschnittliche Studiendauer ziehen. Das geht erst, wenn der Studiengang mehrere Jahre studiert wurde. In den ersten Jahren ist z.B. der Median ein aussagekräftigerer Wert als der Mittelwert. Aber auch später muss der Mittelwert die tatsächliche Studiendauer nicht richtig wiedergeben. Nehmen wir mal an, dass im Sommersemester 2015 insgesamt 30 Studierende das Bachelor-Studium abschließen; davon haben 10 nach 6 Semestern, 5 nach 7 Semestern, 5 nach 8 Semestern, 5 nach 10 Semestern, 3 nach 12 Semestern und 2 nach 16 Semestern das Studium abgeschlossen. Das ergäbe als Mittelwert

$$d = \frac{1}{30} (10 \cdot 6 + 5 \cdot 7 + 5 \cdot 8 + 5 \cdot 10 + 3 \cdot 12 + 2 \cdot 16) \approx 8,4 ;$$

schreibt man dagegen die Semesterzahlen der Größe nach auf, so ergibt sich der Median wegen der folgenden Rangplätzeauflistung

6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 10, 10, 10, 10, 10, 12, 12, 12, 16, 16

als arithmetisches Mittel der 15-ten und 16-ten Zahl in dieser Auflistung, d.h. 7,5. Die wenigen "Ausreißer" mit hohen Semesterzahlen heben den Schnitt nach oben, während der Median aussagt, dass die Hälfte der Absolventen im Sommersemester 2015 höchstens 7 Semester studiert hat. Um Zahlen zu interpretieren muss das "Umfeld" betrachtet werden, in dem diese Zahlen erhoben wurden.

Als **weiteres** Beispiel betrachten wir die Aussage:

”Die Wahrscheinlichkeit, dass ein (zufällig ausgewähltes) neugeborenes Kind ein Junge ist, beträgt 0,515.”

Mit naturwissenschaftlichen Argumenten wird man diese Aussage nicht begründen können. Dieses Ergebnis ergibt sich aus einer statistischen Datenauswertung, d.h. als relative Häufigkeit aus einer (großen) Gruppe von Kindern, bei der geschlechtsspezifisch abgezählt wurde.

Es hat viele Erklärungsansätze gegeben, etwa den, dass auf Grund der Tatsache, dass bei kriegerischen Auseinandersetzungen häufig Männer zu Tode kommen, aus Art-erhaltenden Gründen mehr Jungen als Mädchen geboren werden müssen.

Man kann sagen, dass die ”Wahrscheinlichkeitstheorie” Methoden zur Verfügung stellt, mit denen man sozusagen einen Gewissheitsgrad für das Eintreten eines Ereignisses berechnen kann.

Die ”Statistik” stellt Methoden bereit, um aus empirisch gewonnenen Daten Rückschlüsse auf eine der Untersuchung nicht zugängliche weitaus größere Datenmenge zu ziehen, wie im 3. Beispiel oder bei Hochrechnungen im Zusammenhang mit Wahlen.

Kapitel 1

Grundbegriffe der Wahrscheinlichkeitsrechnung

1.1 Grundlegende Begriffe

Der Begriff "wahrscheinlich" wird im Alltag in verschiedenen Situationen verwendet, hat dabei auch unterschiedliche Bedeutung. Der mathematischen Interpretation des Begriffs kommt man sehr nahe, wenn man folgende Aussage betrachtet:

"Die Wahrscheinlichkeit, mit einem (nicht gezinkten) Würfel eine 5 zu würfeln, beträgt $1/6$."

Es gibt 6 Möglichkeiten als Ergebnis eines Wurfes; wenn der Würfel nicht gezinkt ist, wird jedes Ergebnis gleichwahrscheinlich sein. Der Statistiker wird diese Definition für einen nicht gezinkten Würfel zum Anlass nehmen, um durch eine häufige Wiederholung des Experimentes festzustellen, ob der Würfel nicht gezinkt ist. Wenn er einen Würfel 100-mal wirft und dabei folgende Ergebnisse erzielt

k	1	2	3	4	5	6
h_k	15	16	18	17	16	18

(dabei sei k die Augenzahl und h_k die Häufigkeit, mit der diese Augenzahl gewürfelt wurde), so wird er wahrscheinlich zu dem Ergebnis kommen, dass dieser Würfel nicht gezinkt ist. Es ist nämlich die relative Häufigkeit für das Auftreten der Augenzahl k gleich $\frac{h_k}{100}$, also $r_1 = 0,15$, $r_2 = 0,16$, $r_3 = 0,18$ usw. im Vergleich zu $\frac{1}{6} = 0,1\bar{6}$.

Es stellt sich die Frage, wie groß die Abweichung von r_k zu $\frac{1}{6}$ sein darf, damit man noch von einem nicht gezinkten Würfel sprechen kann?

Stochastische Verfahren werden in den unterschiedlichsten Bereichen unseres täglichen Lebens eingesetzt, z.B.

Medizin:	Wirksamkeit von verschiedenen Medikamenten Modelle für die Ausbreitung von Epidemien
Verkehrswesen:	Überbuchung von Flugzeugen
Versicherungswesen:	Kalkulation von Prämien
Meinungsforschung:	Hochrechnungen aufgrund repräsentativer Stichproben
Informatik:	Analyse von Algorithmen oder Netzwerken

Wir wollen nun unsere Überlegungen idealisieren.

Definition 1.1. Wir nennen einen Vorgang ein (*ideales*) *Zufallsexperiment*, wenn folgende Gegebenheiten vorliegen:

1. Das Experiment wird unter genau festgelegten *Versuchsbedingungen* durchgeführt, z.B. mit denselben Würfeln auf einem bestimmten Tisch.
2. Das Experiment hat verschiedene mögliche *Ergebnisse*,
 - die alle vor der Durchführung bekannt sind und
 - von denen jeweils genau eines eintritt.
3. Das Experiment ist nicht *determiniert*, d.h. dass vor Beendigung des Experiments das Ergebnis ungewiss ist.
4. Das Experiment ist (zumindest in der Vorstellung) beliebig oft unter den gleichen Bedingungen durchzuführen.

Definition 1.2. Die Menge aller Ergebnisse eines Zufallsexperiments bezeichnet man üblicherweise mit Ω . Ω heißt *Ergebnisraum* oder *Ergebnismenge* oder *Grundraum* oder *Raum der Elementarereignisse*.

Jedes $\omega \in \Omega$ heißt ein *Elementarereignis*. Ein *Ereignis* ist dann eine Teilmenge von Ω .

Im obigen Würfel-Beispiel kann z.B. $\Omega = \{\omega_1, \dots, \omega_6\}$ gewählt werden, wobei ω_k bedeutet, dass das Ergebnis beim Würfeln die Augenzahl k ergeben hat.

Z.B. ist $E = \{\omega_1, \omega_3, \omega_5\}$ das Ereignis, dass sich beim Würfeln eine ungerade Augenzahl ergibt. Wir sagen, dass ein Ereignis $A \subset \Omega$ eintritt, wenn ein Elementarereignis $\omega \in A$ eintritt.

Die Menge Ω kann endlich, abzählbar unendlich oder auch überabzählbar unendlich, z.B. eine Teilmenge der reellen Zahlen, sein.

Alle möglichen Ereignisse fasst man zusammen zu einer Menge, die üblicherweise mit \mathcal{A} bezeichnet wird. Diese Menge \mathcal{A} muss bestimmten Anforderungen genügen, damit man damit "mathematisch" arbeiten kann:

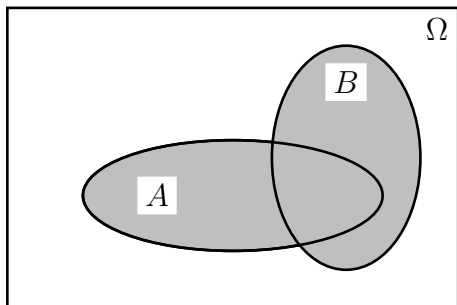
Definition 1.3. Ein System \mathcal{A} von Teilmengen von Ω heißt eine *Ereignisalgebra* oder *σ -Algebra* auf Ω , wenn folgende Bedingungen erfüllt sind:

- (i) Ist $A \in \mathcal{A}$, so ist auch $\bar{A} = \Omega \setminus A \in \mathcal{A}$.
- (ii) Es ist stets $\Omega \in \mathcal{A}$.
- (iii) Sind $A_1, A_2, \dots \in \mathcal{A}$, so ist auch $\bigcup_k A_k \in \mathcal{A}$. Dabei kann die Vereinigung eine endliche oder abzählbar unendliche Vereinigung sein.

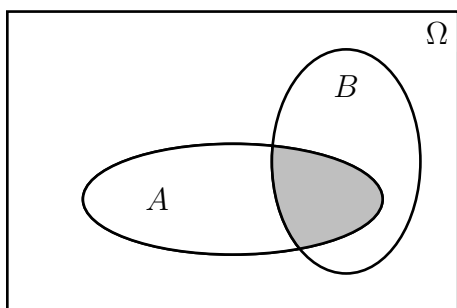
Sprechweisen:

- \bar{A} : A tritt nicht ein; $A \in \mathcal{A}$
- $A \cup B$: A oder B treten ein; $A, B \in \mathcal{A}$
- $A \cap B$: A und B treten ein; $A, B \in \mathcal{A}$.

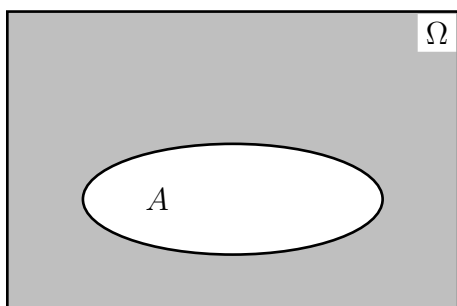
$A \cup B$ □

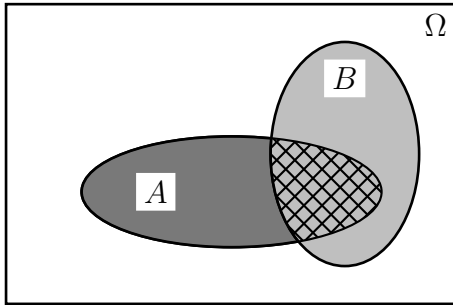


$A \cap B$ □



\bar{A} □





Elementare Eigenschaften von σ -Algebren werden in den Übungen behandelt.

Bemerkung 1.4. (i) Ist Ω eine endliche Menge mit n Elementen, so bildet die Potenzmenge $\mathfrak{P}(\Omega)$ von Ω eine Ereignisalgebra. $\mathfrak{P}(\Omega)$ besitzt 2^n Elemente; dies zeigt man induktiv.

Betrachten wir den Würfel mit $\Omega = \{\omega_1, \dots, \omega_6\}$, so bedeutet z.B. $A = \{\omega_1, \omega_4\}$, dass das Ergebnis die Augenzahl 1 oder 4 ist. Das komplementäre Ereignis ist dann $\bar{A} = \{\omega_2, \omega_3, \omega_5, \omega_6\}$.

(ii) Ist Ω abzählbar unendlich, etwa $\Omega = \mathbb{N}$, so können wir ebenfalls als Ereignisalgebra die Potenzmenge von Ω wählen. Allgemeiner ist für $\Omega \neq \emptyset$ stets $\mathfrak{P}(\Omega)$ eine Ereignisalgebra.

(iii) Seien $\Omega \neq \emptyset, \mathcal{F} \subset \mathfrak{P}(\Omega)$.

Die kleinste σ -Algebra, die \mathcal{F} enthält, ist gegeben durch

$$\mathcal{A}(\mathcal{F}) := \{A \in \mathfrak{P}(\Omega) \mid \text{für jede } \sigma\text{-Algebra } \mathcal{A} \text{ mit } \mathcal{F} \subset \mathcal{A} \text{ gilt } A \in \mathcal{A}\}.$$

$\mathcal{A}(\mathcal{F})$ heißt die von \mathcal{F} erzeugte σ -Algebra.

(iv) Oft zur Modellierung benötigt: $\Omega = \mathbb{R}^n$.

Wähle \mathcal{F} als Menge aller nach links halboffenen Intervalle

$$(a, b] = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid a_i < x_i \leq b_i, 1 \leq i \leq n\}$$

für $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in \mathbb{R}^n$.

Die zugehörige σ -Algebra $\mathcal{A}(\mathcal{F}) =: \mathcal{B}^n$ heißt Borel'sche σ -Algebra; die Elemente von \mathcal{B}^n heißen Borelmengen.

Bemerkung 1.5. Alle offenen und abgeschlossenen Teilmengen liegen in \mathcal{B}^n , aber es gilt $\mathcal{B}^n \neq \mathfrak{P}(\mathbb{R}^n)$.

\mathcal{B}^n wird auch vom System aller offenen (oder abgeschlossenen) Teilmengen des \mathbb{R}^n erzeugt.

Nun wollen wir jedem Ereignis, d.h. jedem $A \in \mathcal{A}$ eine Zahl zuordnen, die Wahrscheinlichkeit dafür, dass dieses Ereignis eintritt. Diese Zuordnung soll ebenfalls bestimmten Bedingungen genügen.

Definition 1.6. a) Es sei \mathcal{A} eine Ereignisalgebra auf Ω ; eine Abbildung (oder: eine Mengenfunktion) $P : \mathcal{A} \rightarrow [0, 1]$ heißt ein *Wahrscheinlichkeitsmaß* oder eine *Wahrscheinlichkeitsverteilung*, wenn P folgende Bedingungen erfüllt:

- (i) $P(\Omega) = 1$
- (ii) $P\left(\bigcup_k A_k\right) = \sum_k P(A_k)$ (σ -Additivität)

für höchstens abzählbar viele, paarweise disjunkte Ereignisse A_1, A_2, \dots aus \mathcal{A} . Für endlich viele Mengen A_k steht auf der rechten Seite eine endliche Summe; ist dagegen die Anzahl der Mengen A_k abzählbar unendlich, so steht auf der rechten Seite eine unendliche Reihe. Das Tripel (Ω, \mathcal{A}, P) heißt *Wahrscheinlichkeitsraum*.

Ist Ω höchstens abzählbar, so heißt (Ω, \mathcal{A}, P) *diskreter Wahrscheinlichkeitsraum*.

- b) Allgemeiner nennt man eine auf einer σ -Algebra $\mathcal{A} \subset \mathfrak{P}(\Omega)$ definierte nicht-negative, σ -additive Mengenfunktion μ ein *Maß* und das Tripel $(\Omega, \mathcal{A}, \mu)$ ein *Maßraum*; das Paar (Ω, \mathcal{A}) heißt ein *messbarer Raum* (oder Messraum) und die Elemente $A \in \mathcal{A}$ heißen *messbare Mengen*.

Satz 1.7. Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B, A_1, A_2, \dots \in \mathcal{A}$. Dann gilt:

- (1) $P(\emptyset) = 0$ und $0 \leq P(A) \leq 1$ für alle $A \in \mathcal{A}$.
- (2) $P(\bar{A}) = 1 - P(A)$.
- (3) $A \subset B \Rightarrow P(A) \leq P(B)$, i.e. p ist monoton.
- (4) $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$.
- (5) $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$.
- (6) $A \subset B \Rightarrow P(B \setminus A) = P(B) - P(A)$.
- (7) (Siebformel von Sylvester-Poincare)

$$\begin{aligned}
 P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\
 &\quad \pm \dots + (-1)^{n-2} \sum_{1 \leq i_1 < \dots < i_{n-1} \leq n} P(A_{i_1} \cap \dots \cap A_{i_{n-1}}) \\
 &\quad + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n).
 \end{aligned}$$

(8) (Bonferroni-Ungleichung)

$$\sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \leq P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

Weitere obere bzw. untere Schranken ergeben sich durch Abbruch in (7) nach + bzw. - Zeichen.

Beweis. Wir beweisen hier nur (1)-(4) und (6).

Zu (2): Aus Definition 1.6 (a), (i) und (ii) folgt direkt wegen $A \cap \bar{A} = \emptyset$ und $A \cup \bar{A} = \Omega$

$$P(\bar{A}) = P(\Omega) - P(A) = 1 - P(A).$$

Zu (1): Sofort aus (2) mit $A = \Omega$ wegen $P(\Omega) = 1$.

Zu (4): Es ist

$$A \cup B = A \cup (\bar{A} \cap B),$$

wobei $A \cap (\bar{A} \cap B) = \emptyset$ ist; also gilt nach Definition 1.6 (a), (ii)

$$\begin{aligned} P(A \cup B) &= P(A) + P(\bar{A} \cap B) \\ &= P(A) + P(\bar{A} \cap B) + P(A \cap B) - P(A \cap B) \\ &= P(A) + P((\bar{A} \cap B) \cup (A \cap B)) - P(A \cap B) \\ &= P(A) + P((\bar{A} \cup A) \cap B) - P(A \cap B) \\ &= P(A) + P(\Omega \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \leq P(A) + P(B). \end{aligned}$$

Zu (3), (6): Mit $B = A \cup (B \setminus A)$ und $A \cap (B \setminus A) = \emptyset$ folgt aus Definition 1.6 (a), (ii):

$$P(B) = P(A) + P(B \setminus A) \geq P(A).$$

□

1.2 Laplace-Experimente; Berechnung von Wahrscheinlichkeiten durch kombinatorische Überlegungen

Bei einem fairen Würfel haben wir angenommen, dass alle Ergebnisse gleichwahrscheinlich sind. Diese Situation trifft bei vielen Zufallsexperimenten zu. Dies führt auf die

Definition 1.8. Ist (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum mit endlichem Grundraum Ω und ist $P(\omega)$ für alle Elementarereignisse $\omega \in \Omega$ gleich groß, so heißt (Ω, \mathcal{A}, P) ein *Laplace-Raum*. In diesem Fall ist $P(\omega) := P(\{\omega\}) = \frac{1}{|\Omega|}$, wobei $|\Omega| = \#\Omega$ die Anzahl der Elemente von Ω bedeutet. Man nennt dann das Zufallsexperiment, das zu dem Ergebnisraum Ω führt, auch ein *Laplace-Experiment*.

Bemerkungen 1.9. a) Ist (Ω, \mathcal{A}, P) ein Laplace-Raum und A ein (beliebiges) Ereignis, so gilt aufgrund der Eigenschaften von P :

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

Häufig wird dies auch folgendermaßen ausgedrückt:

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}.$$

b) Wir betrachten das Werfen mit zwei Würfeln und fragen nach der Wahrscheinlichkeit, dass die Augensumme eine der Zahlen zwischen 2 und 12 ist. Als Grundraum ist die Menge

$$\Omega = \{(i, j) \mid 1 \leq i, j \leq 6\}$$

geeignet. Ω enthält 36 Elemente. Die Wahrscheinlichkeit, ein Elementarereignis zu würfeln ist $\frac{1}{36}$. Nun müssen wir zählen, auf wieviele Arten die Summe k mit $2 \leq k \leq 12$ gewürfelt werden kann. Für $k = 2$ und $k = 12$ ergibt sich jeweils genau eine Möglichkeit; also ist die Wahrscheinlichkeit jeweils $\frac{1}{36}$ dafür, dass die Augensumme 2 oder 12 ist. Für $k = 3$ und $k = 11$ gibt es dagegen jeweils zwei Möglichkeiten, nämlich durch das Würfeln des Tupels $(1, 2)$ oder $(2, 1)$ bzw. durch $(5, 6)$ oder $(6, 5)$. Die Augensumme $k = 6$ ergibt sich durch Würfeln von $(1, 5)$, $(2, 4)$, $(3, 3)$, $(4, 2)$ oder $(5, 1)$; also ist die Wahrscheinlichkeit, die Augensumme $k = 6$ zu erhalten: $\frac{5}{36}$.

Zur Berechnung von Wahrscheinlichkeiten bei Laplace-Experimenten können kombinatorische Überlegungen herangezogen werden:

Wir beginnen mit der Betrachtung von r Mengen A_1, A_2, \dots, A_r mit n_k Elementen für $1 \leq k \leq r$. Dann ist die Anzahl der geordneten r -Tupel (a_1, \dots, a_r) mit $a_k \in A_k$ für $1 \leq k \leq r$ das Produkt

$$\prod_{k=1}^r n_k$$

(vgl. das Abzähltheorem 1.16 aus Mathematik für Informatiker I); speziell für $n_1 = \dots = n_r =: n$ erhalten wir als Anzahl

$$n^r.$$

Ist also A eine Menge mit n Elementen, so ist die Anzahl der geordneten r -Tupel (a_1, \dots, a_r) mit (nicht notwendig paarweise verschiedenen) Elementen $a_1, \dots, a_r \in A$ gerade n^r ; diese r -Tupel nennt man auch eine r -Permutation der Menge A mit Wiederholung; für diese Permutationen schreiben wir auch $Per_r(A)$; es gilt also $|Per_r(A)| = n^r$. Ist A eine Menge mit n Elementen, so ist die Anzahl der geordneten r -Tupel (a_1, \dots, a_r) mit paarweise verschiedenen a_k gleich

$$n(n-1) \cdot \dots \cdot (n-(r-1)) = \frac{n!}{(n-r)!} = \binom{n}{r} r! =: (n)_r.$$

Dies sind die r -Permutationen der Menge A ohne Wiederholung. All diese Permutationen fassen wir in der Menge $Per_r^*(A)$ zusammen; es gilt $|Per_r^*(A)| = (n)_r$. Dies macht nur

dann Sinn, wenn $r \leq n$ ist. Die n -Permutationen der Menge A (mit n Elementen) ohne Wiederholung heißen auch *Permutationen der Menge A* ; es gilt

$$\text{Per}_n^*(A) = (n)_n = n! \quad \text{für } |A| = n.$$

Man beachte, dass jede dieser Permutationen genau einer bijektiven Abbildung $f : A \rightarrow A$ entspricht.

Für die Anzahl der r -elementigen Teilmengen von A mit $|A| = n$ ergibt sich nach dem oben Gesagten der Binomialkoeffizient

$$\binom{n}{r} = \frac{n!}{(n-r)!r!},$$

weil sich aus einer r -elementigen Menge genau $r!$ verschiedene r -Tupel mit paarweise verschiedenen Einträgen bilden lassen.

Sei nun $A = \{1, \dots, n\}$. Für die Menge der r -Permutationen der Menge $\{1, \dots, n\}$ mit bzw. ohne Wiederholung schreiben wir auch kurz $\text{Per}_r(n)$ bzw. $\text{Per}_r^*(n)$.

Wir definieren noch die Menge der r -Kombinationen mit und ohne Wiederholung der Menge $\{1, \dots, n\}$. Sei $B \subset \{1, \dots, n\}$ eine r -elementige Teilmenge. Wir können die Elemente von $B = \{b_1, \dots, b_r\}$ der Größe nach sortieren und erhalten $a_1 < a_2 < \dots < a_r$ (mit $a_j \in B$). Ein solches r -Tupel (a_1, \dots, a_r) heißt *r -Kombination aus $\{1, \dots, n\}$ ohne Wiederholung*. Wir schreiben

$$\text{Kom}_r^*(n) := \{(a_1, \dots, a_r) \mid 1 \leq a_1 < \dots < a_r \leq n\}.$$

Wir betrachten die Menge

$$\text{Kom}_r(n) := \{(a_1, \dots, a_r) \mid 1 \leq a_1 \leq \dots \leq a_r \leq n\};$$

dies ist die Menge der *r -Kombinationen aus $\{1, \dots, n\}$ mit Wiederholung*. Die Anzahl der Elemente ist

$$|\text{Kom}_r^*(n)| = \binom{n}{r}, \quad |\text{Kom}_r(n)| = \binom{n+r-1}{r}.$$

Um letztere Beziehung einzusehen, bedient man sich des folgenden Tricks: Sei (a_1, \dots, a_r) ein beliebiges Element aus $\text{Kom}_r(n)$, so transformieren wir die Elemente dieser Kombination durch

$$(*) \quad b_j := a_j + j - 1.$$

Dann gilt $1 \leq b_1 < b_2 < \dots < b_r \leq n + r - 1$; also ist $b = (b_1, \dots, b_r)$ ein Element aus $\text{Kom}_r^*(n + r - 1)$. Da durch $(*)$ verschiedene a 's auf verschiedene b 's abgebildet werden und da umgekehrt durch

$$a_j := b_j - j + 1$$

jedes $b \in \text{Kom}_r^*(n + r - 1)$ in genau ein $a \in \text{Kom}_r(n)$ transformiert wird, gilt

$$|\text{Kom}_r(n)| = |\text{Kom}_r^*(n + r - 1)| = \binom{n+r-1}{r}.$$

Seien schließlich Zahlen r_1, \dots, r_s aus \mathbb{N} gegeben mit $\sum_{k=1}^s r_k = n$. Dann ist die Anzahl der möglichen Zerlegungen von A in Teilmengen A_1, \dots, A_s mit jeweils r_k Elementen, d.h. $\bigcup_{k=1}^s A_k = A$ und $|A_k| = r_k$ gleich

$$\frac{n!}{r_1! \cdot \dots \cdot r_s!}.$$

In der Tat: Aus der Menge A können wir $\binom{n}{r_1}$ Teilmengen von A mit r_1 Elementen auswählen. Aus dem Rest $A \setminus A_1$ mit $n - r_1$ Elementen können wir dann $\binom{n - r_1}{r_2}$ Teilmengen A_2 mit r_2 Elementen bilden usw. Somit erhalten wir insgesamt

$$\begin{aligned} & \binom{n}{r_1} \cdot \binom{n - r_1}{r_2} \cdot \dots \cdot \binom{n - r_1 - r_2 - \dots - r_{s-1}}{r_s} \\ &= \frac{n!}{(n - r_1)! \cdot r_1!} \cdot \frac{(n - r_1)!}{(n - r_1 - r_2)! \cdot r_2!} \cdot \dots \cdot \frac{(n - r_1 - r_2 - \dots - r_{s-1})!}{\underbrace{(n - r_1 - \dots - r_s)!}_{=1} \cdot r_s!}. \end{aligned}$$

Beispiel 1.10 (Das Urnenmodell). (a) Zur Veranschaulichung der r -Permutationen und r -Kombinationen der Menge $\{1, \dots, n\}$ mit bzw. ohne Wiederholung können wir das Urnenmodell heranziehen.

Dabei denken wir uns eine Urne gefüllt mit $n \in \mathbb{N}$ durchnummerierten Kugeln. Nun ziehen wir r -mal hintereinander je eine Kugel aus der Urne (ohne hinzusehen) und notieren nach jeder Ziehung die Nummer der gezogenen Kugel. Legen wir nach jeder Ziehung die gezogene Kugel in die Urne zurück, so entspricht das Ergebnis dieses Zufallsexperiments (unter Beachtung der Reihenfolge) genau einer r -Permutation einer n -elementigen Menge mit Wiederholung.

Dagegen entspricht das r -malige Ziehen einer Kugel ohne Zurücklegen (unter Beachtung der Reihenfolge) genau einer r -Permutation ohne Wiederholung. Man beachte, daß wir in beiden Fällen (d.h. mit/ohne Zurücklegen) die Reihenfolge der gezogenen Kugeln berücksichtigen müssen. Berücksichtigen wir die Reihenfolge nicht, so erhalten wir ein Modell für die r -Kombinationen.

Zusammenfassend können wir für unser Urnenmodell folgende Aussagen treffen:

Die Anzahl der möglichen Ausgänge beim r -maligen Ziehen einer Kugel aus einer Urne mit n durchnummerierten Kugeln ist

	ohne Zurücklegen	mit Zurücklegen
ohne Berücksichtigung der Reihenfolge	$ Kom_r^*(n) = \binom{n}{r}$	$ Kom_r(n) = \binom{n+r-1}{r}$
mit Berücksichtigung der Reihenfolge	$ Per_r^*(n) = (n)_r$	$ Per_r(n) = n^r$

- (b) Wie groß ist die Wahrscheinlichkeit beim Lotto (6 aus 49) einen Fünfer mit bzw. ohne Zusatzzahl zu tippen? Wir betrachten das Zahlenlotto als Laplace-Experiment. Wir können uns die 49 Kugeln zerlegt denken in 6 Glückszahlen und 43 Restzahlen, bestehend aus 42 Nieten und einer Zusatzzahl. Berechnen wir die Wahrscheinlichkeit als relative Häufigkeit, so müssen wir zunächst die Anzahl der Zerlegungen der 49 Zahlen in eine 43- und eine 6- elementige Menge berechnen, dies ist die Anzahl der verschiedenen Möglichkeiten, den Lottoschein auszufüllen. Diese Anzahl ist nach den Vorüberlegungen

$$\frac{49!}{6! \cdot 43!} = \binom{49}{6}.$$

Wie viele Möglichkeiten gibt es, dass 5 Zahlen mit denen aus der 6-elementigen "Glücksmenge" übereinstimmen? Es gibt $6 = \binom{6}{5}$ Möglichkeiten.

Beim "Fünfer mit Zusatzzahl" muss die 6. Zahl mit der Zusatzzahl übereinstimmen, also ist

$$p(\text{"Fünfer mit Zusatzzahl"}) = \frac{6}{\binom{49}{6}} \approx 4.291 \cdot 10^{-7}.$$

(In ca. 43 von 100 Millionen Fällen.)

Beim "Fünfer ohne Zusatzzahl" muss genau eine der sechs Zahlen mit einer der 42 Nieten übereinstimmen; dafür gibt es je 42 Möglichkeiten, also ist

$$p(\text{"Fünfer ohne Zusatzzahl"}) = \frac{6 \cdot 42}{\binom{49}{6}} \approx 1.802 \cdot 10^{-5}.$$

(In ca. 18 von 1 Million Fällen.)

1.3 Mehrstufige Zufallsexperimente

Häufig besteht ein Zufallsexperiment aus mehreren Teilexperimenten. So können wir z.B. das Ziehen der ersten 6 Zahlen beim 6-er Zahlenlotto als 6-stufiges Zufallsexperiment auffassen. Als Ergebnismenge beim Ziehen der ersten Zahl können wir die Menge $\Omega_1 = \{1, 2, \dots, 49\}$ der ersten 49 natürlichen Zahlen betrachten. Wird die Zahl n_1 gezogen, so

verändert sich beim Ziehen der zweiten Zahl die Ergebnismenge auf $\Omega_2 = \Omega_1 \setminus \{n_1\}$ usw. Als Ergebnismenge betrachten wir dann $\Omega = \Omega_1 \times \dots \times \Omega_6$.

Wir wollen das Wesentliche bei mehrstufigen Zufallsexperimenten allerdings an einem einfacheren Beispiel verdeutlichen.

Beispiel 1.11. Wir betrachten eine Urne mit 2 roten und 3 schwarzen Kugeln. Es wird rein zufällig eine Kugel aus der Urne gezogen; ihre Farbe wird notiert und anschließend werden diese Kugel und eine weitere Kugel derselben Farbe in die Urne zurückgelegt. Nach gutem Durchmischen wird wiederum eine Kugel aus der Urne gezogen. Mit welcher Wahrscheinlichkeit ist diese Kugel rot?

Schreiben wir sozusagen zur Abkürzung für das Ziehen einer roten Kugel eine 1 und für das Ziehen einer schwarzen Kugel eine 0, so können wir als Ergebnismenge für das 1. und das 2. Ziehen jeweils $\Omega_1 = \Omega_2 = \{0, 1\}$ wählen. Wir suchen dann die Wahrscheinlichkeit für das Ereignis $A = \{(1, 1), (0, 1)\}$. Wie müssen wir auf der Potenzmenge von $\Omega = \Omega_1 \times \Omega_2$ ein Wahrscheinlichkeitsmaß gemäß Definition 1.6 festlegen?

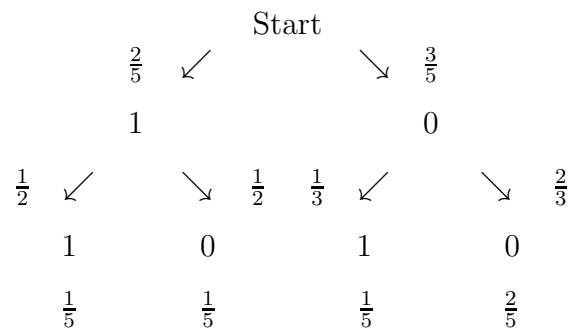
Wir würden bei einer oftmaligen Wiederholung des Experiments beim ersten Zug mit Wahrscheinlichkeit $p_1 = \frac{2}{5}$ eine rote und mit Wahrscheinlichkeit $p_2 = \frac{3}{5}$ eine schwarze Kugel erwarten. Wird beim 1. Zug eine rote Kugel gezogen, d.h. ist $a_1 = 1$, so befinden sich vor dem 2. Zug in der Urne 3 rote und 3 schwarze Kugeln. Ist dagegen $a_1 = 0$, so befinden sich vor dem 2. Zug 2 rote und 4 schwarze Kugeln in der Urne. Im ersten Fall wird man dann in der Hälfte aller Fälle als zweite Kugel eine rote Kugel erwarten. Im zweiten Fall erwartet man dagegen in 2 von 6 Ziehungen eine rote Kugel. Also ist es sinnvoll, die Wahrscheinlichkeiten auf Ω folgendermaßen festzulegen:

$$P(1, 1) = \frac{2}{5} \cdot \frac{1}{2}, \quad P(1, 0) = \frac{2}{5} \cdot \frac{1}{2}, \quad P(0, 1) = \frac{3}{5} \cdot \frac{1}{3} \quad \text{und} \quad P(0, 0) = \frac{3}{5} \cdot \frac{2}{3}.$$

Offenbar ist z.B. die Wahrscheinlichkeit dafür, dass die zweite gezogene Kugel rot ist:

$$P(1, 1) + P(0, 1) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}.$$

Wir können mehrstufige Zufallsexperimente an einem *Baumdiagramm* veranschaulichen.



Definition 1.12. Es sei Ω_j endlich oder abzählbar unendlich für $j = 1, \dots, n$; ferner sei P_1 eine Wahrscheinlichkeitsverteilung auf Ω_1 . Für jedes $a_1 \in \Omega_1$ sei ein *System*

von Übergangswahrscheinlichkeiten $P(a_2|a_1) \geq 0$ mit

$$\sum_{a_2 \in \Omega_2} P_2(a_2|a_1) = 1 \quad \text{für alle } a_1 \in \Omega_1$$

gegeben. Wir betrachten als Grundraum eines n -stufigen Zufallsexperiments die Menge $\Omega = \Omega_1 \times \dots \times \Omega_n$ und definieren unter der Voraussetzung, dass für die Modellierung des Übergangs vom $(j-1)$ -ten zum j -ten Telexperiment ($2 \leq j \leq n$) ein System von Übergangswahrscheinlichkeiten

$$P_j(a_j|a_1, \dots, a_{j-1}) \geq 0 \quad \text{für jedes } (a_1, \dots, a_{j-1}) \in \Omega_1 \times \dots \times \Omega_{j-1}$$

mit

$$\sum_{a_j \in \Omega_j} P_j(a_j|a_1, \dots, a_{j-1}) = 1 \quad \text{für alle } (a_1, \dots, a_{j-1}) \in \Omega_1 \times \dots \times \Omega_{j-1}$$

gegeben ist, die Wahrscheinlichkeit $P(\omega)$ für $\omega = (a_1, \dots, a_n)$ nach der *ersten Pfadregel* durch

$$P(\omega) := P_1(a_1) \cdot P_2(a_2|a_1) \cdot \dots \cdot P_n(a_n|a_1, \dots, a_{n-1}).$$

Dann ist $(\Omega, \mathfrak{P}(\Omega), P)$ ein Wahrscheinlichkeitsraum, wenn wir für eine Teilmenge $A \subset \Omega$ definieren

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

Die Wahrscheinlichkeit eines Ereignisses A berechnet sich als Summe aller Wahrscheinlichkeiten, die in der n -ten Stufe zu einem $\omega \in A$ führen. Diese Eigenschaft wird auch *zweite Pfadregel* genannt.

Häufig kommt es vor, dass die Telexperimente von dem Ablauf der vorhergehenden Telexperimente unabhängig sind. Wir präzisieren dies in der folgenden

Definition 1.13. Ist ein n -stufiges Zufallsexperiment gegeben und gilt für das System der Übergangswahrscheinlichkeiten

$$P_j(a_j|a_1, \dots, a_{j-1}) = P_j(a_j)$$

für alle $a_j \in \Omega_j, a_1 \in \Omega_1, \dots, a_{j-1} \in \Omega_{j-1}$, dann ist $(\Omega, \mathfrak{P}(\Omega), P)$ mit

$$P(\omega) = \prod_{j=1}^n P_j(a_j) \quad \text{für } \omega = (a_1, \dots, a_n)$$

ein Wahrscheinlichkeitsraum. Man nennt solche mehrstufigen Zufallsexperimente auch *Produktexperimente* (vgl. auch Definition 1.21).

1.4 Bedingte Wahrscheinlichkeiten, der Satz von Bayes und Beispiele

Wir wollen uns nun damit beschäftigen, wie man Vor- und Zusatzinformationen verarbeiten kann. Dazu betrachten wir das folgende

Beispiel 1.14. Drei Maschinen produzieren denselben Artikel, allerdings mit unterschiedlicher Qualität. Aus langer Erfahrung weiß man, dass Maschine 1 nur 2 % Ausschuss produziert, Maschine 2 dagegen 10 % und Maschine 3 schließlich 4 %. Die Anteile der drei Maschinen an der Gesamtproduktion betragen 30 %, 50 % bzw. 20 %. Von der Gesamtproduktion wird ein Artikel zufällig ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass der Artikel unbrauchbar ist?

Die Wahrscheinlichkeit, dass ein Artikel, der von Maschine 1 produziert wurde, unbrauchbar ist, ist 0.02. Wir schreiben dafür

$$P(U|M_1) = P_{M_1}(U) = 0.02$$

und meinen dabei mit U das Ereignis, dass ein Artikel unbrauchbar ist und mit M_1 , dass der Artikel von Maschine 1 produziert wurde. Entsprechend ist

$$P(U|M_2) = P_{M_2}(U) = 0.1$$

und

$$P(U|M_3) = P_{M_3}(U) = 0.04.$$

Definition 1.15. Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Sind A und B zwei Ereignisse, d.h. $A, B \in \mathcal{A}$, mit $P(B) > 0$, so ist die *bedingte Wahrscheinlichkeit* $P(A|B) = P_B(A)$ ("Wahrscheinlichkeit von A unter der Bedingung B ") definiert durch

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

Betrachten wir zwei Ereignisse A und B und ihre "Gegenereignisse" \bar{A} und \bar{B} , so können wir folgende "Tafel" aufstellen, wobei jeweils der Ergebnisraum als disjunkte Vereinigung dargestellt wird.

	B	\bar{B}
A	$A \cap B$	$A \cap \bar{B}$
\bar{A}	$\bar{A} \cap B$	$\bar{A} \cap \bar{B}$

Unter den Ereignissen, die die 'Eigenschaft' B erfüllen, betrachten wir noch die Ereignisse, die zusätzlich A erfüllen. Das ergibt die bedingte Wahrscheinlichkeit $P_B(A)$. Die obige Tafel nennt man auch eine *Vierfeldertafel*.

Als Beispiel blicken wir auf das statistische Jahrbuch der Bundesrepublik Deutschland aus dem Jahre 2000, das unter den ungefähr 82,2 Millionen Einwohnern der BRD folgende Altersstruktur aufzeigt:

	mindestens 70 Jahre alt	unter 70 Jahre alt	Summe
Männer	3,19 Mill.	36,90 Mill.	40,09 Mill.
Frauen	6,15 Mill.	35,93 Mill.	42,08 Mill.
Summe	9,34 Mill.	72,83 Mill.	82,17 Mill.

Wir schreiben für das Merkmal "Mann" kurz M und für das Merkmal "mindestens 70 Jahre alt" kurz S . Die Wahrscheinlichkeit dafür, dass eine beliebig ausgewählte Person ein mindestens 70 Jahre alter männlicher Einwohner ist, ist

$$P(M \cap S) = \frac{3,19}{82,17} \approx 0,038.$$

Die Wahrscheinlichkeit dafür, dass ein männlicher Einwohner mindestens 70 Jahre alt ist, ist

$$P_M(S) = \frac{3,19}{40,09} \approx 0,08.$$

Die Wahrscheinlichkeit dafür, dass ein mindestens 70 jähriger Einwohner ein Mann ist, ist

$$P_S(M) = \frac{3,19}{9,34} \approx 0,342.$$

Als Folgerung aus Definition 1.15 erhalten wir

Bemerkung 1.16 (Multiplikationssatz). (a) Sind $A, B \in \mathcal{A}$ mit $P(A) > 0$ und $P(B) > 0$, so folgt

$$P(A \cap B) = P_B(A) \cdot P(B) = P_A(B) \cdot P(A).$$

Induktiv ergibt sich

(b) Sind $A_1, A_2, \dots, A_n \in \mathcal{A}$ mit $n \geq 2$, so dass $P(A_1 \cap A_2 \cap \dots \cap A_{\nu-1}) > 0$ gilt für alle $\nu = 2, \dots, n$. Dann ist

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \prod_{\nu=2}^n P(A_\nu | A_1 \cap \dots \cap A_{\nu-1}).$$

Satz 1.17 (Satz von der totalen Wahrscheinlichkeit). Es sei $\Omega = \bigcup_{k=1}^n B_k$ eine dis-

junkte Zerlegung mit $P(B_k) > 0$ für $1 \leq k \leq n$. Dann gilt für beliebiges $A \in \mathcal{A}$:

$$P(A) = \sum_{k=1}^n P(B_k) \cdot P_{B_k}(A).$$

Beweis. Aus

$$A = A \cap \Omega = A \cap \left(\bigcup_{k=1}^n B_k \right) = \bigcup_{k=1}^n (A \cap B_k)$$

und der paarweisen Disjunktheit von $A \cap B_k$ folgt

$$P(A) = \sum_{k=1}^n P(A \cap B_k) = \sum_{k=1}^n P_{B_k}(A) \cdot P(B_k).$$

□

Wir kehren zu unserem **Beispiel 1.14** zurück: Die Gesamtmenge der produzierten Artikel können wir als disjunkte Vereinigung $M_1 \cup M_2 \cup M_3$ auffassen. Dann gilt nach Satz 1.17 für die Wahrscheinlichkeit, dass ein beliebig herausgegriffener Artikel unbrauchbar ist:

$$\begin{aligned} P(U) &= P_{M_1}(U) \cdot P(M_1) + P_{M_2}(U) \cdot P(M_2) + P_{M_3}(U) \cdot P(M_3) \\ &= 0.02 \cdot 0.3 + 0.1 \cdot 0.5 + 0.04 \cdot 0.2 = 0.064. \end{aligned}$$

Es wird also 6,4 % Ausschuss produziert.

Folgerung 1.18. Es sei $B \in \mathcal{A}$ mit $0 < P(B) < 1$, d.h. auch $0 < P(\bar{B}) < 1$. Dann ist für beliebiges $A \in \mathcal{A}$:

$$P(A) = P_B(A) \cdot P(B) + P_{\bar{B}}(A) \cdot P(\bar{B}).$$

Bei der Berechnung bedingter Wahrscheinlichkeiten können manchmal erstaunliche Phänomene auftreten. Wir konstruieren ein Beispiel über die Zulassungszahlen in 2 Studiengängen an einer Universität:

	Frauen		Männer	
	Bewerbungen	zugelassen	Bewerbungen	zugelassen
Fach 1	900	720 (80 %)	200	180 (90 %)
Fach 2	100	20 (20 %)	800	240 (30 %)
Summe	1000	740 (74 %)	1000	420 (42 %)

Obwohl die Zulassungsquoten der Männer diejenigen der Frauen in jedem Fach übertreffen, erscheint die Universität insgesamt auf den ersten Blick männerfeindlich. Das liegt daran, dass die globale Zulassungsquote als gewichtetes Mittel der Zulassungsquoten in den einzelnen Fächern berechnet wird; es ist

$$0,74 = 0,9 \cdot 0,8 + 0,1 \cdot 0,2 \quad \text{und} \quad 0,42 = 0,2 \cdot 0,9 + 0,8 \cdot 0,3$$

Hieran sieht man, dass durch Verschweigen gewisser Aspekte nicht mehr der Realität entsprechende Schlussfolgerungen gezogen werden können.

Der historische Ursprung für die Entdeckung des oben geschilderten als *Simpson-Paradoxon* bekannten Phänomens war wohl eine Statistik aus dem Jahre 1910 über Tbc-Todesfälle in New York und Richmond, aufgegliedert für Weiße und Farbige. Dabei waren die Einzeltodesraten für Weiße und Farbige in New York höher, die Gesamttodesrate war aber in Richmond höher.

Satz 1.19 (Satz von Bayes). *Es sei $\Omega = \bigcup_{k=1}^n B_k$ eine disjunkte Zerlegung mit $P(B_k) > 0$ für $1 \leq k \leq n$. Ist $A \in \mathcal{A}$ beliebig mit $P(A) > 0$, so gilt*

$$P_A(B_k) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k \cap A)}{\sum_{j=1}^n P(B_j) \cdot P_{B_j}(A)}.$$

Ist $n = 2$, so erhalten wir mit $B_1 = B$ und $B_2 = \bar{B}$:

$$P_A(B) = \frac{P(A \cap B)}{P(B) \cdot P_B(A) + P(\bar{B}) \cdot P_{\bar{B}}(A)}.$$

Beispiel 1.20. Wir betrachten einen medizinischen Labortest zur Erkennung von Krankheiten. Es können bei einem solchen Test zwei Arten von Fehlern auftreten:

1. Der Patient hat die Krankheit; sie wird aber durch den Test nicht erkannt.
2. Der Patient ist gesund, wird aber aufgrund des Tests als krank diagnostiziert.

Konkret beziehen wir uns auf folgendes Beispiel:

In der BRD waren 1975 etwa 0,5 % der Bevölkerung an Tbc erkrankt. Man weiß aufgrund langjähriger Erfahrung, dass durch eine spezielle Tbc-Röntgenuntersuchung 90 % der Kranken und 99 % der Gesunden richtig diagnostiziert werden.

Wie groß ist die Wahrscheinlichkeit, dass eine als krank diagnostizierte Person wirklich an Tbc erkrankt ist und wie groß ist die Wahrscheinlichkeit, dass eine als gesund diagnostizierte Person wirklich gesund ist.

Wir betrachten zur Vereinfachung das Ereignis K : "Die Person ist krank" und das Ereignis N : "Der Test fällt negativ aus", d.h. die Person wird als gesund diagnostiziert.

Dann ergeben die obigen Daten:

$$P(K) = 0.005, \quad P_K(\bar{N}) = 0.9, \quad P_{\bar{K}}(N) = 0.99$$

und somit

$$P(\overline{K}) = 0.995, \quad P_K(N) = 0,1, \quad P_{\overline{K}}(\overline{N}) = 0.01.$$

Wegen

$$P_{\overline{N}}(K) = \frac{P(\overline{N} \cap K)}{P(K) \cdot P_{\overline{K}}(\overline{N}) + P(\overline{K}) \cdot P_{\overline{K}}(\overline{N})} = \frac{P(K) \cdot P_{\overline{K}}(\overline{N})}{P(K) \cdot P_{\overline{K}}(\overline{N}) + P(\overline{K}) \cdot P_{\overline{K}}(\overline{N})}$$

und

$$P_N(\overline{K}) = \frac{P(\overline{K}) \cdot P_{\overline{K}}(N)}{P(K) \cdot P_{\overline{K}}(N) + P(\overline{K}) \cdot P_{\overline{K}}(N)}$$

erhalten wir daraus

$$P_{\overline{N}}(K) \approx 0.3114 \quad \text{sowie} \quad P_N(\overline{K}) \approx 0.9995.$$

Definition 1.21. (i) Zwei Ereignisse A und B heißen (*stochastisch*) *unabhängig*, wenn gilt

$$P(A \cap B) = P(A) \cdot P(B);$$

andernfalls heißen sie *abhängig*. n Ereignisse A_1, \dots, A_n (mit $n \geq 2$) heißen (*stochastisch*) *unabhängig*, falls für alle mindestens zweielementigen Teilmengen $T \subset \{1, 2, \dots, n\}$ gilt

$$P\left(\bigcap_{j \in T} A_j\right) = \prod_{j \in T} P(A_j).$$

(ii) Eine Familie von Ereignissen $(A_i)_{i \in I} \subset \mathcal{A}$ heißt *paarweise (stochastisch) unabhängig*, falls

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j), \quad \forall i, j \in I, i \neq j.$$

Beispiel 1.22 (Ziegenproblem).

Situation:

- 3 Türen, dahinter 1 Auto und 2 Ziegen
- Kandidat wählt eine Tür
- Eingriff des Moderators: Dieser öffnet eine der nicht-gewählten Türen, hinter der kein Auto steht
- Kandidat darf Wahl ändern

Frage: Ist die Änderung der Entscheidung sinnvoll?

Konkret: O.b.d.A. wegen Symmetrie wählt Kandidat Tür 1, Quizmaster öffnet Tür 3. Soll Kandidat bei Tür 1 bleiben oder zu 2 wechseln?

Sei, ohne den Wahrscheinlichkeitsraum zu spezifizieren:

$$\begin{aligned}
 A_i &\hat{=} \text{ Auto ist hinter Tür } i \\
 P(A_i) &= \frac{1}{3}, \quad i \in \{1, 2, 3\} \\
 K_i &\hat{=} \text{ Kandidat wählt Tür } i \\
 P(K_i) &= \frac{1}{3}, \quad i \in \{1, 2, 3\} \\
 A_i, K_j &\text{ stochastisch unabhängig, } 1 \leq i, j \leq 3 \\
 Q_i &\hat{=} \text{ Quizmaster öffnet Tür } i \text{ (nicht unabhängig von } A_i, K_j) \\
 \Rightarrow \underbrace{P(A_1 | K_1 \cap Q_3)}_{\text{Bleibestrategie}} &= \frac{P(A_1 \cap K_1 \cap Q_3)}{P(K_1 \cap Q_3)} \\
 &= \frac{P(Q_3 | A_1 \cap K_1)}{P(K_1 \cap Q_3)} \cdot P(A_1 \cap K_1).
 \end{aligned}$$

Es ist

$$\begin{aligned}
 P(K_1 \cap Q_3) &= P(K_1 \cap Q_3 \cap A_1) + P(K_1 \cap Q_3 \cap A_2) + \underbrace{P(K_1 \cap Q_3 \cap A_3)}_{=0} \\
 &= \underbrace{P(Q_3 | A_1 \cap K_1)}_{=\frac{1}{2}} \cdot \underbrace{P(A_1 \cap K_1)}_{=\frac{1}{9}} + \underbrace{P(Q_3 | A_2 \cap K_1)}_{=1} \cdot \underbrace{P(A_2 \cap K_1)}_{=\frac{1}{9}} \\
 &= \frac{1}{2} \cdot \frac{1}{9} + 1 \cdot \frac{1}{9} = \frac{1}{6}.
 \end{aligned}$$

Eingesetzt in die Bleibestrategie:

$$P(A_1 | K_1 \cap Q_3) = \frac{P(Q_3 | A_1 \cap K_1)}{P(K_1 \cap Q_3)} \cdot P(A_1 \cap K_1) = \frac{\frac{1}{2}}{\frac{1}{6}} \cdot \frac{1}{9} = \frac{1}{3}.$$

Analog folgt

$$P(A_2 | K_1 \cap Q_3) = \frac{P(Q_3 | A_2 \cap K_1)}{P(K_1 \cap Q_3)} \cdot P(A_2 \cap K_1) = \frac{1}{\frac{1}{6}} \cdot \frac{1}{9} = \frac{2}{3},$$

d.h. die Änderung der Entscheidung verdoppelt die Gewinnwahrscheinlichkeit.

Lemma 1.23. 1. Mit A, B sind auch A, \bar{B} und \bar{A}, \bar{B} stochastisch unabhängig.

2. Ist $P(B) > 0$, so gilt:

$$A, B \text{ stochastisch unabhängig} \iff P(A | B) = P(A).$$

3. Ist A eine sogenannte Nullmenge, d.h. $P(A) = 0$, so sind A, B stochastisch unabhängig für alle $B \in \mathcal{A}$.

Beweis. (Übung) □

Beispiel 1.24. Werfen von 2 fairen Würfeln, d.h. Laplaceverteilung über $\Omega = \{(i, j) \mid i, j \in \{1, \dots, 6\}\}$.

Wir bezeichnen mit

$$\begin{aligned} A_1 &\hat{=} \text{Würfel 1 zeigt eine gerade Zahl,} \\ A_2 &\hat{=} \text{Würfel 2 zeigt eine gerade Zahl,} \\ A_3 &\hat{=} \text{die Augensumme ist gerade.} \end{aligned}$$

Es gilt

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = \frac{1}{4},$$

d.h. A_1 und A_2 sind stochastisch unabhängig. Weiterhin gilt

$$P(A_3 \mid A_1) = P(A_3) = \frac{1}{2} \text{ und } P(A_3 \mid A_2) = P(A_3) = \frac{1}{2}.$$

Nach dem in Lemma 1.23 bewiesenen Kriterium sind also auch A_2 und A_3 stochastisch unabhängig; A_1 und A_3 sind ebenfalls stochastisch unabhängig. Also wissen wir, dass $\{A_1, A_2, A_3\}$ paarweise stochastisch unabhängig sind.

Nun gilt

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2) = \frac{1}{4} \text{ und } P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{8},$$

und damit

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3),$$

d.h. A_1, A_2, A_3 sind NICHT stochastisch unabhängig.

Beispiel 1.25 (Sortieren (Rencontre-Problem)). Betrachte Felder der Länge n von vergleichbaren, verschiedenen Elementen. Alle Anordnungen seien gleichwahrscheinlich. Modelliere die Situation wie folgt:

$$\Omega = \{\text{Permutationen von } \{1, \dots, n\}\}$$

$$\mathcal{A} = \mathfrak{P}(\Omega)$$

$$P(\{\omega\}) = \frac{1}{n!} \quad \forall \omega \in \Omega.$$

Teil a: Bestimme die Wahrscheinlichkeit, dass mindestens ein Element an der richtigen Stelle ist (vorsortiert). Definiere dazu Ereignismengen A_j , welche jeweils alle Ergebnisse beinhalten, bei denen Element j an der j -ten Stelle ist, wie folgt:

$$A_j = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega \mid \omega_j = j\}.$$

Gesucht ist dann $P(A_1 \cup \dots \cup A_n) = P\left(\bigcup_{j=1}^n A_j\right)$. Da die A_j sicher nicht paarweise disjunkt sind, erfolgt die Berechnung mit Hilfe der Siebformel:

Sei $1 \leq i_1 < \dots < i_l \leq n$ und $l \leq n$. Dann ist

$$\bigcap_{j=1}^l A_{i_j} = \{\omega \in \Omega \mid \omega_{i_j} = i_j, j = 1, \dots, l\}$$

die Menge aller Permutationen, bei denen die Elemente i_1, \dots, i_l sich an der richtigen Stelle befinden. Die Mächtigkeit dieser Menge ist

$$\left| \bigcap_{j=1}^l A_{i_j} \right| = (n-l)!,$$

weil man die gegebenen l Elemente auf die richtigen Positionen verteilen muss, dann aber die verbliebenen $(n-l)$ Elemente beliebig auf die restlichen $(n-l)$ Positionen verteilen darf. Die Wahrscheinlichkeit dafür, dass sich die l Elemente i_1, \dots, i_l auf den richtigen Positionen befinden, ist deshalb

$$P\left(\bigcap_{j=1}^l A_{i_j}\right) = \frac{(n-l)!}{n!} = \frac{1}{\binom{n}{l} l!}, \quad l = 1, \dots, n.$$

Außerdem ist die Mächtigkeit der Menge aller l -elementigen Teilmengen von n , also die Menge aller Möglichkeiten, zunächst l Elemente aus den vorhandenen n auszuwählen

$$|\{(i_1, \dots, i_l) \mid 1 \leq i_1 < \dots < i_l \leq n\}| = \binom{n}{l}.$$

Insgesamt ergibt sich für die Wahrscheinlichkeit, dass mindestens ein Element sich an der richtigen Position befindet

$$\begin{aligned} P\left(\bigcup_{j=1}^n A_j\right) &\stackrel{\text{Siebformel}}{=} \sum_{j=1}^n P(A_j) - \sum_{i < j} P(A_i \cap A_j) + \dots + (-1)^{n+1} P\left(\bigcap_{j=1}^n A_j\right) \\ &= \frac{n}{n} - \binom{n}{2} \frac{1}{\binom{n}{2} 2!} + \dots + (-1)^{n+1} \binom{n}{n} \frac{1}{\binom{n}{n} n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + \frac{(-1)^{n+1}}{n!} \\ &= 1 - \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}\right) \\ &\xrightarrow{n \rightarrow \infty} 1 - e^{-1} \approx 0,6321. \end{aligned}$$

Erstaunlicherweise konvergiert die Wahrscheinlichkeit gegen einen festen Wert. Das bedeutet, dass die Wahrscheinlichkeit dafür, ob mindestens ein Element in einem Feld der Länge n an der richtigen Position ist, für große n fast unabhängig von n ist.

Teil b: Eine Abschätzung dafür, dass mindestens k Elemente vorsortiert sind

$$\begin{aligned} P\left(\bigcup_{1 \leq i_1 < \dots < i_k \leq n} \bigcap_{l=1}^k A_{i_l}\right) &\stackrel{\text{Satz 1.7}}{\leq} \sum_{1 \leq i_1 < \dots < i_k \leq n} P\left(\bigcap_{l=1}^k A_{i_l}\right) \\ &= \binom{n}{k} \cdot \frac{1}{\binom{n}{k} k!} = \frac{1}{k!}. \end{aligned}$$

Teil c: Die Wahrscheinlichkeit dafür, dass *genau* k Elemente vorsortiert sind.

Nach Teil a betragt die Wahrscheinlichkeit dafur, dass in einem Feld der Lange $n - k$ kein Element vorsortiert ist

$$1 - \frac{1}{1!} + \frac{1}{2!} - \dots + \frac{(-1)^{n-k}}{(n-k)!}.$$

Daher ist die Anzahl der Anordnungen, bei denen kein Element vorsortiert ist,

$$(n-k)! \cdot \left(1 - \frac{1}{1!} + \frac{1}{2!} - \dots + \frac{(-1)^{n-k}}{(n-k)!} \right).$$

Nun gibt es noch $\binom{n}{k}$ Moglichkeiten, ein Feld der Lange n in eines der Lange k und eines der Lange $n - k$ aufzuteilen. Somit ergibt sich fur die Wahrscheinlichkeit, dass k Elemente vorsortiert sind

$$\begin{aligned} \frac{1}{n!} \binom{n}{k} (n-k)! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \dots + \frac{(-1)^{n-k}}{(n-k)!} \right) \\ = \frac{1}{k!} \left(1 - \frac{1}{1!} + \frac{1}{2!} - \dots + \frac{(-1)^{n-k}}{(n-k)!} \right). \end{aligned}$$

Wie werden Wahrscheinlichkeiten bestimmt, wenn schon bekannt ist, dass das Ergebnis in einer bestimmten Teilmenge liegt?

Sei beispielsweise Ω eine Menge von Chips, die von zwei verschiedenen Firmen stammen und $|\Omega| = 5000$. Von Firma A stammen $|A| = 1000$ Chips und von Firma B $|B| = 4000$ Chips. Unter den 5000 Chips sind insgesamt $|D| = 300$ defekt, von denen $|A \cap D| = 100$ von Firma A und $|B \cap D| = 200$ von Firma B stammen. Von Firma A sind also 10% aller Chips defekt und von Firma B 5%. Es werde nun zufallig ein Chip gezogen (Laplace Modell). Die Wahrscheinlichkeit dafur, dass der Chip defekt ist, falls er von Firma A stammt ist

$$P(D | A) = \frac{|D \cap A|}{|A|} = \frac{\frac{|D \cap A|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{P(D \cap A)}{P(A)} = \frac{\frac{100}{5000}}{\frac{1000}{5000}} = \frac{1}{10}.$$

Die Wahrscheinlichkeit, dass ein Chip von Firma A stammt, falls er defekt ist, ist

$$P(A | D) = \frac{|A \cap D|}{|D|} = \frac{\frac{|A \cap D|}{|\Omega|}}{\frac{|D|}{|\Omega|}} = \frac{1}{3}.$$

Kapitel 2

Diskrete Zufallsvariable. Wichtige diskrete Wahrscheinlichkeitsverteilungen

Häufig ist es so, dass den Ausgängen eines Zufallexperiments, d.h. den Elementen der Ereignisalgebra, eine Zahl zugeordnet wird. Das wollen wir etwas mathematischer fassen.

2.1 Diskrete Zufallsvariable

Definition 2.1. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Eine Funktion

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

mit endlichem oder abzählbarem Wertebereich

$$X(\Omega) = \{x_1, x_2, \dots\}$$

heißt *diskrete Zufallsvariable* auf Ω .

Beispiel 2.2. Wir betrachten einen Laplace-Würfel; wir werfen ihn dreimal. X ordne jedem (dreimaligen) Wurf die Augensumme zu; dann ist $X(\Omega) = \{3, 4, \dots, 18\}$. Wir berechnen z. B.

$$X^{-1}(\{4\}) = \{(1, 1, 2), (1, 2, 1), (2, 1, 1)\}$$

oder

$$\begin{aligned} X^{-1}([2.5, 4.8]) &= X^{-1}(\{3\}) \cup X^{-1}(\{4\}) \\ &= \{(1, 1, 1)\} \cup X^{-1}(\{4\}). \end{aligned}$$

Betrachten wir dann die Wahrscheinlichkeit $P(X^{-1}(\{4\}))$, so erhalten wir

$$P(X^{-1}(\{4\})) = \frac{3}{216} = \frac{1}{72}.$$

Man schreibt auch $P(X = 4) = \frac{1}{72}$. Damit haben wir Teilmengen von \mathbb{R} über den Wahrscheinlichkeitsraum eine Wahrscheinlichkeit zugeordnet.

Wir bezeichnen mit

$$(X = x) := \{\omega \in \Omega \mid X(\omega) = x\} \quad \text{bzw.} \quad (X \leq x) := \{\omega \in \Omega \mid X(\omega) \leq x\}.$$

Definition 2.3. (a) Die Funktion

$$V = V_X : \begin{array}{ll} X(\Omega) & \longrightarrow [0, 1] \\ x & \mapsto V(x) = P(X = x). \end{array}$$

heißt *Verteilung* der Zufallsvariablen X .

(b) Die Funktion $F = F_X : \mathbb{R} \rightarrow [0, 1]$ mit

$$F(x) = P(X \leq x)$$

heißt *Verteilungsfunktion* von X .

Beispiel 2.4. Ein Laplace-Würfel wird dreimal geworfen. Die Zufallsvariable X bezeichne die Anzahl der ungeraden Zahlen, die dabei geworfen wird. Es ist $X(\Omega) = \{0, 1, 2, 3\}$. Bezeichnet G das Ergebnis, dass eine gerade Augenzahl gewürfelt wird und U das Ereignis, dass sich eine ungerade Augenzahl ergibt, so erhalten wir (wegen der Unabhängigkeit der Ereignisse)

$$P(X = 0) = P(GGG) = \left(\frac{1}{2}\right)^3 = 0.125,$$

$$\begin{aligned} P(X = 1) &= P(UGG) + P(GUG) + P(GGU) \\ &= 3 \cdot \left(\frac{1}{2}\right)^3 = 0.375, \end{aligned}$$

$$\begin{aligned} P(X = 2) &= P(UUG) + P(UGU) + P(GUU) \\ &= 3 \cdot \left(\frac{1}{2}\right)^3 = 0.375, \end{aligned}$$

und

$$P(X = 3) = P(UUU) = \left(\frac{1}{2}\right)^3 = 0.125.$$

Die Verteilungsfunktion $F = F_X$ von X ist dann eine Treppenfunktion mit

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.125 & \text{für } 0 \leq x < 1 \\ 0.5 & \text{für } 1 \leq x < 2 \\ 0.875 & \text{für } 2 \leq x < 3 \\ 1 & \text{für } 3 \leq x. \end{cases}$$

Satz 2.5. Ist X eine diskrete Zufallsvariable mit Wertemenge $X(\Omega) = \{x_i \mid i \in \mathbb{N}\} \subset \mathbb{R}$, so gilt für die zugehörige Verteilungsfunktion F :

$$(a) \text{ Für } x \in \mathbb{R} \text{ ist } F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i).$$

(b) Ist $x < y$, so gilt $F(x) \leq F(y)$. (Monotonie)

(c) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$.

Bemerkung 2.6. In Teil (a) von Satz 2.5 benutzen wir die Diskretheit der Zufallsvariablen, und zwar um die Summe auf der rechten Seite bilden zu können. Warum diese unendliche Summe im mathematischen Sinne existiert wird in der Vorlesung “Mathematik 1” begründet. Ebenfalls in der Vorlesung “Mathematik 1” wird das Symbol $\lim_{x \rightarrow \infty}$ und der Begriff des (uneigentlichen) Grenzwerts von Funktionen erklärt. (Siehe auch das Kapitel “Grundlegendes aus der Analysis” in diesem Skript).

Satz 2.7. Ist F die Verteilungsfunktion einer diskreten Zufallsvariablen X , so gilt für alle reellen Zahlen $a < b$:

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

(b) $P(X > a) = 1 - F(a)$.

Teil (a) des Satzes 2.7 besagt: Für $a < b$ gibt

$$F(b) - F(a) = \sum_{x_j \leq b} P(X = x_j) - \sum_{x_j \leq a} P(X = x_j) = \sum_{a < x_j \leq b} P(X = x_j)$$

die Wahrscheinlichkeit dafür an, dass die Zufallsvariable X einen Wert im Intervall $]a, b]$ annimmt.

Definition 2.8. Zwei diskrete Zufallsvariable X und Y auf einem Wahrscheinlichkeitsraum Ω heißen *unabhängig*, wenn die Ereignisse $X = x$ und $Y = y$ für jedes beliebige Tupel (x, y) mit $x \in X(\Omega)$ und $y \in Y(\Omega)$ unabhängig sind, d.h. wenn gilt:

$$P((X = x) \cap (Y = y)) = P(X = x) \cdot P(Y = y).$$

Anderenfalls heißen X und Y *abhängig*.

2.2 Erwartungswert und Varianz diskreter Zufallsvariablen

Welche Augenzahl erwarten wir im Mittel beim Werfen eines Würfels; ein Maß wäre die Summe aller Möglichkeiten dividiert durch die Mindestanzahl, mit der man dies erreichen kann:

$$\frac{1}{6}(1 + 2 + \dots + 6) = 3,5.$$

Diese Zahl ergibt sich bei keinem Wurf als Ergebnis, ist also mehr eine ”theoretische” Zahl.

Definition 2.9. Sei X eine diskrete Zufallsvariable. Falls die Reihe $\sum_{\omega \in \Omega} |X(\omega)|P(\omega)$ konvergiert, so heißt

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) \quad (2.1)$$

Erwartungswert von X . Der Erwartungswert wird oft mit $E(X) = \mu$ bezeichnet.

Bemerkung 2.10. Eine äquivalente Formel für den Erwartungswert ist:

$$E(X) = \sum_{x_i \in X(\Omega)} x_i \cdot P(X = x_i).$$

Beispiel 2.11. Wir betrachten das wiederholte Würfeln mit einem fairen Würfel. Wie lange muss man im Mittel auf die erste Sechs warten. Wir haben es hier mit einem Experiment mit zwei möglichen Ergebnissen zu tun, nämlich mit Erfolg (eine Sechs) oder Misserfolg (keine Sechs). Bei einem fairen Würfel tritt der Erfolg mit der Wahrscheinlichkeit $p = \frac{1}{6}$ und der Misserfolg mit der Wahrscheinlichkeit $q = 1 - p = \frac{5}{6}$ ein. Gibt die Zufallsvariable X die Anzahl der Versuche bis zum 1. Auftreten des Erfolgs an, so ist

$$p_k = P(X = k) = q^{k-1}p.$$

Dann gilt (geometrische Reihe!):

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

Für den Erwartungswert erhalten wir

$$E(X) = \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1}.$$

Um den Reihenwert zu bestimmen, betrachten wir die Potenzreihe $\sum_{k=0}^{\infty} x^k$, die für $|x| < 1$ konvergiert; wir dürfen gemäß Kapitel 1 aus Math. für Inf. 2 die Potenzreihe differenzieren, indem wir gliedweise differenzieren; wir erhalten so

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{d}{dx} \left(\sum_{k=0}^{\infty} x^k \right) = \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{1}{(1-x)^2}.$$

In unserem Beispiel ergibt sich daher

$$E(X) = p \frac{1}{(1-q)^2} = \frac{1}{p}.$$

Der Erwartungswert beim Würfeln mit $p = \frac{1}{6}$ ist damit $E(X) = 6$, d.h. dass man im Durchschnitt 6 Würfe benötigt, um eine Sechs zu würfeln.

Satz 2.12. Sind X, Y zwei diskrete Zufallsvariablen mit existierenden Erwartungswerten, so gelten folgende Aussagen:

(i) *Linearität des Erwartungswerts:* Für beliebige Konstante $a, b \in \mathbb{R}$ gilt:

$$E(aX + b) = aE(X) + b. \quad (2.2)$$

Zudem gilt:

$$E(X + Y) = E(X) + E(Y). \quad (2.3)$$

(ii) *Für unabhängige Zufallsvariable X und Y gilt:*

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

(iii) *Monotonie des Erwartungswerts:* Gilt $X \leq Y$, d.h. $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$, so folgt

$$E(X) \leq E(Y).$$

Beweis. Wir beweisen hier nur Teil (i). Zu (2.2): Wir halten etwas allgemeiner fest: Ist X eine diskrete Zufallsvariable mit $X(\Omega) = \{x_1, x_2, \dots\}$ und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine bijektive Funktion, so ist auch $g \circ X$ eine diskrete Zufallsvariable, und es gilt

$$E(g \circ X) = \sum_i g(x_i)P(g \circ X = g(x_i)) = \sum_i g(x_i)P(X = x_i),$$

falls die Reihe $\sum_i |g(x_i)|P(X = x_i)$ konvergiert. Speziell für $g(x) = ax + b$ mit zwei Konstanten $a, b \in \mathbb{R}$ erhalten wir

$$E(aX + b) = aE(X) + b.$$

(Für $a = 0$ folgt $E(b) = b$ sofort aus der Definition 2.9.)

Wir beweisen nun (2.3). Es gilt

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

Im letzten Schritt haben wir die Grenzwertsätze für konvergente Reihen aus Mathe 1 benutzt. \square

Während der Erwartungswert eine Maßzahl für den "Schwerpunkt" einer Verteilung ist, ist die Varianz eine Maßzahl für die "Streuung" um diesen Schwerpunkt.

Definition 2.13. (a) Ist X eine diskrete Zufallsvariable und existiert $E(X^2)$, so heißt

$$\text{Var}(X) := E((X - E(X))^2) = \sum_i (x_i - E(X))^2 p_i$$

die Varianz von X .

(b)

$$\sigma := \sigma_X := \sqrt{\text{Var}(X)}$$

heißt *Standardabweichung* von X .

(c) Ist Y eine weitere Zufallsvariable, für die $E(Y^2)$ existiert, so heißt

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

die *Covarianz* von X und Y und

$$\rho_{XY} := \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

der *Korrelationskoeffizient* von X und Y . X und Y heißen *unkorreliert*, wenn die Covarianz $\text{Cov}(X, Y) = 0$ ist.

Beispiel 2.14. Die Zufallsvariable gebe die höchste Augenzahl beim zweimaligen Würfeln an. Ist $\Omega := \{(i, j) \mid 1 \leq i, j \leq 6\}$, so ist $X(\omega) := \max(i, j)$ für $\omega = (i, j)$. Definieren wir für ein Elementarereignis $P(\omega) = \frac{1}{36}$, so erhalten wir:

$$\begin{aligned} P(X = 1) &= \frac{1}{36}, & P(X = 2) &= \frac{3}{36}, & P(X = 3) &= \frac{5}{36}, \\ P(X = 4) &= \frac{7}{36}, & P(X = 5) &= \frac{9}{36}, & \text{und } P(X = 6) &= \frac{11}{36}. \end{aligned}$$

Für den Erwartungswert ergibt sich somit

$$E(X) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} = \frac{161}{36} = 4 \frac{17}{36}.$$

Nun berechnen wir die Varianz

$$\text{Var}(X) = \sum_{i=1}^6 \left(i - \frac{161}{36} \right)^2 \cdot \frac{2i-1}{36} = \frac{2555}{1296} \approx 1.97.$$

Unmittelbar aus der Definition der Kovarianz und aus Satz 2.12 (ii) folgt

Satz 2.15. Sind zwei diskrete Zufallsvariable X und Y unabhängig, so sind sie auch unkorreliert.

Satz 2.16 (Rechenregeln für die Varianz). Sind $X, X_i, i = 1, \dots, n$ Zufallsvariable, für die $E(X^2)$ und $E(X_i^2)$ existieren, so gilt:

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

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

(iii) Sind X_1, \dots, X_n paarweise unabhängig, so gilt

$$\text{Var}(X_1 + \dots + X_n) = \sum_{k=1}^n \text{Var}(X_k).$$

2.3 Wichtige diskrete Wahrscheinlichkeitsverteilungen

Wir betrachten zunächst einige Verteilungsfunktionen für Produktexperimente.

2.3.1 Die Binomialverteilung

Wir betrachten ein Zufallsexperiment (zum Beispiel das Werfen einer Münze), bei dem nur zwei Ereignisse eintreten können, nämlich das Ereignis A mit der Wahrscheinlichkeit

$$p := P(A) > 0$$

und das Ereignis \bar{A} mit der Wahrscheinlichkeit

$$P(\bar{A}) = 1 - p > 0.$$

Ein Experiment dieser Form nennt man ein **Bernoulli-Experiment** und p heißt **Bernoulli-Verteilung**.

Wir führen das Experiment n -mal (unter gleichen Bedingungen) hintereinander durch und fragen, wie groß die Wahrscheinlichkeit ist, dass dabei k -mal das Ereignis A eintritt. Eine von insgesamt $\binom{n}{k}$ Möglichkeiten (vgl. Abschnitt 1.2) für das k -malige Auftreten des Ereignisses A wäre z.B. die Reihenfolge

$$\underbrace{AA \dots A}_{k\text{-mal}} \quad \underbrace{\bar{A} \dots \bar{A}}_{(n-k)\text{-mal}} .$$

Bezeichnen wir das Ereignis A mit 1 und das Ereignis \bar{A} mit 0, so ist der Grundraum für das n -stufige Zufallsexperiment die Menge

$$\Omega = \{0, 1\}^n = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \{0, 1\}, 1 \leq i \leq n\}$$

mit

$$P(\omega) = p^k (1 - p)^{n-k}, \text{ wenn } k \text{ die Anzahl der Einsen in } \omega \text{ ist.}$$

Das Ereignis, dass insgesamt k -mal das Ereignis A eintritt, lässt sich in der Form

$$E_k = \left\{ \omega \in \Omega \mid \sum_{i=1}^n \omega_i = k \right\}$$

beschreiben. Dann gilt also:

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

Erklären wir S_n als die Zufallsvariable, die zählt, wie oft das Ereignis A eintritt, so erhalten wir

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

Definition 2.17. Eine diskrete Zufallsvariable X heißt *binomial-verteilt* mit den Parametern n und p , wenn für $0 \leq k \leq n$ gilt

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} =: B_{n,p}(k).$$

Wir sagen kurz: X ist $B(n, p)$ -verteilt.

Die so definierte Binomialverteilung erfüllt die Kriterien einer diskreten Wahrscheinlichkeitsverteilung, insbesondere gilt:

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n B_{n,p}(k) = 1. \quad (2.4)$$

Die Identität (2.4) folgt mithilfe der Binomischen Formel:

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

Satz 2.18. (a) Es gilt für $0 \leq k \leq n$

$$B_{n,p}(k) = B_{n,1-p}(n-k).$$

(b) Sei X die Zufallsvariable, die in einem n -stufigen Bernoulli-Experiment mit Wahrscheinlichkeitsparameter p die Anzahl der Erfolge misst. Sei \bar{X} die Zufallsvariable, die die Anzahl der Misserfolge misst. Dann gilt für $0 \leq m \leq n$:

$$P(X \leq m) = 1 - P(\bar{X} \leq n - m - 1).$$

Beweis. Die Behauptung in (a) folgt aus $\binom{n}{k} = \binom{n}{n-k}$. Die Behauptung in (b) ergibt sich dann wie folgt:

$$\begin{aligned} P(X \leq m) &= \sum_{k=0}^m B_{n,p}(k) = \sum_{k=0}^m B_{n,1-p}(n-k) = \sum_{\ell=n-m}^n B_{n,1-p}(\ell) \\ &= \sum_{\ell=n-m}^n P(\bar{X} = \ell) = P(\bar{X} \geq n-m) = 1 - P(\bar{X} < n-m). \end{aligned}$$

□

Beispiel 2.19. Aus einer Urne mit genau 30 Kugeln, nämlich 12 weißen und 18 roten werden (blind) nacheinander und mit Zurücklegen genau 50 Kugeln entnommen. Wir berechnen die Wahrscheinlichkeit dafür, dass von den entnommenen Kugeln höchstens die Hälfte rot ist.

Mit X bezeichnen wir die Zufallsvariable, die die Anzahl der roten unter den entnommenen Kugeln angibt. Dann ist X binomial-verteilt mit $n = 50$ und $p = \frac{18}{30} = 0.6$. Wir müssen $P(X \leq 25)$ berechnen, d.h.

$$P(X \leq 25) = \sum_{k=0}^{25} P(X = k) = \sum_{k=0}^{25} \binom{50}{k} 0.6^k 0.4^{50-k}.$$

Diesen Wert können wir z.B. mit dem Taschenrechner berechnen. Es gibt jedoch für die Binomial-Verteilung auch Tabellen, aus denen man die Werte für die sogenannten *kumulativen Wahrscheinlichkeiten*

$$P(X \leq m) = \sum_{k=0}^m B_{n,p}(k), \quad 0 \leq m \leq n,$$

in Abhängigkeit von p und n ablesen kann. Diese *Binomial-Tabellen* beinhalten aber nur Werte für $0 < p \leq 0.5$ und in der Regel $n = 10, 20, 50, 100$. Wir erhalten in unserem Beispiel nach Satz 2.18 wegen $1 - p = 0.4$ und $n - m - 1 = 24$:

$$P(X \leq 25) = 1 - P(\bar{X} \leq 24) \approx 1 - 0.902 = 0.098.$$

Um die Binomial-Tabellen zu erstellen, verwendet man eine Rekursionsformel. Es gilt für $0 < p < 1$ und $1 \leq k \leq n$:

$$\begin{aligned} B_{n,p}(k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{(n-k+1)p}{k(1-p)} \binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} \\ &= \frac{(n-k+1)p}{k(1-p)} B_{n,p}(k-1). \end{aligned}$$

Satz 2.20. Ist X $B(n, p)$ -verteilt, so gilt für den Erwartungswert $E(X) = np$ und für die Varianz $\text{Var}(X) = np(1-p)$.

Beweis. Nach Definition des Erwartungswerts

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

Nach Definition der Varianz und Wegen $k^2 = k(k-1) + k$ gilt:

$$\begin{aligned}
 \text{Var}(X) &= \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k (1-p)^{n-k} \\
 &= n(n-1) \sum_{k=2}^n \frac{k(k-1)}{n(n-1)} \binom{n}{k} p^k (1-p)^{n-k} + n \sum_{k=1}^n \frac{k}{n} \binom{n}{k} p^k (1-p)^{n-k} \\
 &\quad - 2n^2 p \sum_{k=1}^n \frac{k}{n} \binom{n}{k} p^k (1-p)^{n-k} + n^2 p^2 \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \\
 &= n(n-1)p^2 + np - 2n^2 p^2 + n^2 p^2 = np(1-p).
 \end{aligned}$$

□

2.3.2 Die geometrische Verteilung

Definition 2.21. Die diskrete Zufallsvariable X heißt *geometrisch-verteilt* mit dem Parameter $0 < p < 1$, wenn für $k \in \mathbb{N}_0$ gilt

$$P(X = k) = (1-p)^k p.$$

Wir sagen kurz: X ist $G(p)$ -verteilt.

Beachte, dass die geometrische Verteilung die Kriterien einer diskreten Wahrscheinlichkeitsverteilung erfüllt, denn mithilfe der geometrischen Reihe berechnen wir

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} (1-p)^k p = p \frac{1}{1-(1-p)} = 1. \quad (2.5)$$

Als *Modellbeispiel* dient auch hier ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $P(A)$, welches wiederholt und bei gleichen Bedingungen durchgeführt wird. Die Zufallsvariable X , die die Anzahl der Misserfolge \bar{A} vor dem ersten Erfolg A zählt, ist dann $G(p)$ -verteilt (vgl. auch Beispiel 2.11; die dortige Zufallsvariable zählte aber, wann der Erfolg das erste Mal eintrat).

Satz 2.22. Ist X $G(p)$ -verteilt, so gilt für den Erwartungswert $E(X) = \frac{1-p}{p} = \frac{1}{p} - 1$ und für die Varianz $\text{Var}(X) = \frac{1-p}{p^2}$.

Beweis. Der oben erwähnte Vergleich mit Beispiel 2.11 erklärt, weshalb der Erwartungswert hier um 1 kleiner als in Beispiel 2.11 ist; dies lässt sich natürlich auch analog der dortigen Rechnung nachrechnen.

Für die Varianz ergibt sich wegen

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

durch zweimalige Differentiation der geometrischen Reihe

$$\begin{aligned} E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1)(1-p)^k p \\ &= p(1-p)^2 \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &= p(1-p)^2 \frac{2}{(1-(1-p))^3} = \frac{2(1-p)^2}{p^2} \end{aligned}$$

und damit

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

□

2.3.3 Die negative Binomial-Verteilung

Beispiel 2.23. Wir betrachten n Bernoulli-Experimente mit Erfolgswahrscheinlichkeit p und stellen die Frage nach der Wahrscheinlichkeit dafür, dass das n -te Experiment das r -te erfolgreiche Experiment ist und dass genau k Misserfolge dem r -ten Erfolg vorangegangen sind. Es ist also $n = r + k$. Ein solches Ereignis können wir durch ein n -Tupel $(\omega_1, \dots, \omega_n)$ aus k Nullen und r Einsen beschreiben, wobei die letzte Komponente eine 1 ist. Es gibt insgesamt $\binom{n-1}{k} = \binom{k+r-1}{k}$ Möglichkeiten k Nullen auf die ersten $n-1$ Komponenten zu verteilen. Jedes solche n -Tupel hat die Wahrscheinlichkeit $p^r \cdot (1-p)^k$. Die gesuchte Wahrscheinlichkeit ist damit

$$\binom{k+r-1}{k} p^r (1-p)^k.$$

Definition 2.24. Die diskrete Zufallsvariable X besitzt eine *negative Binomialverteilung* mit den Parametern $r \in \mathbb{N}$ und $0 < p < 1$, wenn für $k \in \mathbb{N}_0$ gilt

$$P(X = k) = \binom{k+r-1}{k} p^r (1-p)^k.$$

Wir sagen kurz: X ist $Nb(r, p)$ -verteilt.

Bemerkungen 2.25. Definiert man für negative ganze Zahlen m und $k \in \mathbb{N}_0$ den Binomialkoeffizienten durch

$$\binom{m}{k} := \frac{m \cdot (m-1) \cdot \dots \cdot (m-k+1)}{k!},$$

so gilt für den Binomialkoeffizienten aus Definition 2.24:

$$\binom{k+r-1}{k} = \frac{(k+r-1) \cdot \dots \cdot (k+r-1-k+1)}{k!} = \frac{(k+r-1) \cdot \dots \cdot r}{k!}$$

$$= (-1)^k \frac{(-r) \cdot (-r-1) \cdot \dots \cdot (-r-k+1)}{k!} = (-1)^k \binom{-r}{k}.$$

Dies erklärt die Namensgebung für die Negative Binomial-Verteilung. Wir prüfen wieder, dass die Kriterien einer diskreten Wahrscheinlichkeitsverteilung erfüllt sind:

$$\begin{aligned} \sum_{k=0}^{\infty} \binom{k+r-1}{k} p^r (1-p)^k &= p^r \sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} (1-p)^k \\ &= p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-(1-p))^k = 1. \end{aligned}$$

Die zuletzt angegebene unendliche Reihe stellt nämlich die *Binomialreihe* mit Exponent $-r$ dar; daher gilt

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

Satz 2.26. *Ist die Zufallsvariable X $Nb(r, p)$ -verteilt, so gilt*

$$E(X) = r \frac{1-p}{p} \quad \text{und} \quad \text{Var}(X) = r \frac{1-p}{p^2}.$$

Beweis. Nach Definition gilt (verwende hier Bemerkung 2.25):

$$\begin{aligned} E(X) &= p^r (1-p) r \sum_{k=1}^{\infty} \frac{k}{r} \binom{k+r-1}{k} (1-p)^{k-1} \\ &= p^r (1-p) r \sum_{k=1}^{\infty} \binom{k+r-1}{k-1} (1-p)^{k-1} \\ &= \frac{p^r (1-p) r}{p^{r+1}} \sum_{k=0}^{\infty} \binom{k+(r+1)-1}{k} p^{r+1} (1-p)^k = \frac{(1-p)r}{p} \end{aligned}$$

und entsprechend

$$E(X(X-1)) = \frac{r(r+1)(1-p)^2}{p^2},$$

woraus wegen

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

folgt

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

□

Beispiel 2.27. Als Variante von Beispiel 2.23 fragen wir nach der Wahrscheinlichkeit, dass das r -te erfolgreiche Experiment im j -ten Versuch (mit $j \geq r$) auftritt. Also müssen unter den ersten $j - 1$ Experimenten $r - 1$ Erfolge und $j - r$ Misserfolge aufgetreten sein. Das j -te Experiment muss wieder erfolgreich sein. Wir identifizieren den Erfolg wieder mit der Zahl 1 und den Misserfolg mit der 0. Da jedes j -Tupel aus $j - r$ Nullen und r Einsen mit der Wahrscheinlichkeit

$$(1 - p)^{j-r} p^r$$

auftritt und da es $\binom{j-1}{r-1}$ Möglichkeiten gibt, ein $(j-1)$ -Tupel mit $r-1$ Einsen (und $(j-r)$ Nullen) zu bilden, erhalten wir als Wahrscheinlichkeit $p_{r,j}$ für den r -ten Erfolg im j -ten Experiment

$$p_{r,j} = \binom{j-1}{r-1} (1-p)^{j-r} p^r \quad \text{für } j = r, r+1, \dots$$

Mit der Substitution $k = j - r$ erhalten wir

$$p_{r,r+k} = \binom{k+r-1}{r-1} (1-p)^{j-r} p^r = \binom{k+r-1}{k} (1-p)^k p^r$$

für $k = 0, 1, 2, \dots$. Dies ist die Wahrscheinlichkeit dafür, dass im $(r+k)$ -ten Experiment der r -te Erfolg eintritt.

2.3.4 Die hypergeometrische Verteilung

Beispiel 2.28. Gegeben sind N verschiedene Objekte, darunter S mit der Eigenschaft A ; es wird eine Stichprobe vom Umfang n (ohne Zurücklegen) gezogen. Mit welcher Wahrscheinlichkeit sind darunter genau k Stück mit der Eigenschaft A ? Nach Abschnitt 1.2 gibt es $\binom{N}{n}$ Stichproben vom Umfang n ; dann müssen k Elemente der Stichprobe die Eigenschaft A besitzen und $n-k$ die Eigenschaft \bar{A} ; dafür gibt es $\binom{S}{k}$ bzw. $\binom{N-S}{n-k}$ Möglichkeiten. Also ist die Wahrscheinlichkeit dafür, dass in der Stichprobe genau k Objekte die Eigenschaft A besitzen

$$P(X = k) = \frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}},$$

wobei X die Anzahl der gezogenen Objekte mit der Eigenschaft A zähle.

Definition 2.29. Eine diskrete Zufallsvariable X heißt *hypergeometrisch-verteilt* mit den Parametern N, S und n , wenn für $0 \leq k \leq n$ gilt

$$P(X = k) = \frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}}.$$

Wir schreiben kurz: X ist $H(N, S, n)$ -verteilt.

Bemerkungen 2.30. a) Dass die Summe der Wahrscheinlichkeiten gleich 1 ist, ergibt sich aus der sog. *Vandermondeschen Faltungsformel*

$$\sum_{m=0}^n \binom{S}{m} \binom{N-S}{n-m} = \binom{N}{n}$$

b) Verwenden wir die Schreibweise aus Abschnitt 1.2, so lässt sich die Wahrscheinlichkeit aus Definition 2.29 auch folgendermaßen darstellen:

$$P(X = k) = \binom{n}{k} \frac{(S)_k (N - S)_{n-k}}{(N)_n}, \quad (S)_k := \binom{S}{k} k! = \frac{S!}{(S - k)!}.$$

Satz 2.31. *Ist X $H(N, S, n)$ -verteilt, so gilt für den Erwartungswert*

$$E(X) = n \frac{S}{N}$$

und für die Varianz

$$\text{Var}(X) = n \frac{S}{N} \left(1 - \frac{S}{N}\right) \frac{N - n}{N - 1}.$$

Beweis. Es ist

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} \frac{(S)_k (N - S)_{n-k}}{(N)_n} \\ &= n \frac{S}{N} \sum_{k=1}^n \binom{n-1}{k-1} \frac{(S-1)_{k-1} (N-1-(S-1))_{n-1-(k-1)}}{(N-1)_{n-1}} \\ &= n \frac{S}{N} \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{(S-1)_k (N-1-(S-1))_{n-1-k}}{(N-1)_{n-1}} = n \frac{S}{N}. \end{aligned}$$

Wegen $\text{Var}(X) = E(X(X-1)) + E(X) - E(X)^2$ folgt für die Varianz

$$\text{Var}(X) = n(n-1) \frac{S(S-1)}{N(N-1)} + n \frac{S}{N} - n^2 \frac{S^2}{N^2} = n \frac{S}{N} \left(1 - \frac{S}{N}\right) \frac{N-n}{N-1}.$$

□

Bemerkungen 2.32. Die Formel für die Varianz sieht übersichtlicher aus, wenn man die Wahrscheinlichkeit, im ersten Zug ein Element mit der Eigenschaft A zu ziehen, mit $p := \frac{S}{N}$ bezeichnet; dann erhalten wir

$$\text{Var}(X) = np(1-p) \frac{N-n}{N-1}.$$

Man kann sich übrigens überlegen, dass die Wahrscheinlichkeit, im k -ten Zug ein Element mit der Eigenschaft A zu ziehen, ebenfalls $p = \frac{S}{N}$ ist.

Wenn wir für die Binomialverteilung als Modell das Ziehen von Kugeln aus einer Urne mit Zurücklegen wählen und als Modell für die hypergeometrische Verteilung das Ziehen von Kugeln aus einer Urne ohne Zurücklegen, so liefert der Vergleich der Varianzen

$$\text{Var}(X_1) = np(1-p) \quad \text{und} \quad \text{Var}(X_2) = np(1-p) \left(1 - \frac{n-1}{N-1}\right),$$

also beim zweiten Modell eine kleinere Varianz als beim ersten Modell. Der Grund hierfür ist der Informationsgewinn beim Ziehen ohne Zurücklegen.

2.3.5 Die Poisson-Verteilung

Definition 2.33. Eine diskrete Zufallsvariable X heißt *Poisson-verteilt* mit dem Parameter $\lambda > 0$, wenn für $k \in \mathbb{N}_0$ gilt:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Wir sagen kurz: X ist $P(\lambda)$ -verteilt.

Satz 2.34. Ist X $P(\lambda)$ -verteilt, so gilt für den Erwartungswert bzw. die Varianz

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

wobei wir $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, $x \in \mathbb{R}$, verwendet haben. Wegen $k^2 = k(k-1) + k$ gilt weiter

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^{\infty} (k - \lambda)^2 \frac{\lambda^k}{k!} \cdot e^{-\lambda} \\ &= e^{-\lambda} \left(\sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} - 2\lambda \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} + \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) \\ &= e^{-\lambda} (\lambda^2 e^{\lambda} + \lambda e^{\lambda} - 2\lambda^2 e^{\lambda} + \lambda^2 e^{\lambda}) = \lambda. \end{aligned}$$

Bemerkungen 2.35. Für große n und kleine p gibt es eine Möglichkeit, die $B_{n,p}$ -Verteilung durch die Poisson-Verteilung zu ersetzen. Dazu betrachten wir mit $\mu = np$, d.h. $p = \frac{\mu}{n}$, folgende Beziehung:

$$\begin{aligned} B_{n,p}(k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!} \cdot \frac{\mu^k}{n^k} \cdot \frac{\left(1 - \frac{\mu}{n}\right)^n}{\left(1 - \frac{\mu}{n}\right)^k} \\ &= \frac{1 \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\mu}{n}\right)^k} \cdot \frac{\mu^k}{k!} \cdot \left(1 - \frac{\mu}{n}\right)^n \\ &= \frac{1}{1 - \frac{\mu}{n}} \cdot \frac{1 - \frac{1}{n}}{1 - \frac{\mu}{n}} \cdot \frac{1 - \frac{2}{n}}{1 - \frac{\mu}{n}} \cdots \frac{1 - \frac{k-1}{n}}{1 - \frac{\mu}{n}} \cdot \frac{\mu^k}{k!} \cdot \left(1 - \frac{\mu}{n}\right)^n \end{aligned}$$

Für $n \rightarrow \infty$ konvergieren die ersten k Faktoren gegen 1, wenn wir μ als fest auffassen. Wegen $e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$, $x \in \mathbb{R}$, konvergiert der letzte Faktor gegen $e^{-\mu}$. Daraus ergibt

sich die *Poisson-Näherung*

$$B_{n,p}(k) \approx e^{-\mu} \frac{\mu^k}{k!} .$$

Faustregel: Im Allgemeinen kann man die Poisson-Näherung für $p \leq 0.1$ und $n \geq 100$ gebrauchen.

Beispiel 2.36. In einen Teig werden 250 Rosinen geknetet und dann daraus 200 Hörnchen gebacken. Wir wählen ein Hörnchen beliebig aus. Mit welcher Wahrscheinlichkeit enthält es genau 2 Rosinen?

Wir gehen davon aus, dass für jede der 250 Rosinen die gleiche Wahrscheinlichkeit $\frac{1}{200}$ besteht, in das ausgewählte Hörnchen zu gelangen. Das Ereignis A tritt ein, wenn sich die "Rosine Nr. i " im ausgewählten Hörnchen befindet. Dann lässt sich das Geschehen durch eine Binomial-Verteilung mit $n = 250$ und $p = \frac{1}{200}$ beschreiben. Die gesuchte Wahrscheinlichkeit ist

$$B_{250, \frac{1}{200}}(2) = \binom{250}{2} \left(\frac{1}{200}\right)^2 \left(\frac{199}{200}\right)^{248} \approx 0.22448 ;$$

verwenden wir die Poisson-Näherung, so ergibt sich

$$B_{250, \frac{1}{200}}(2) \approx e^{-1.25} \frac{1.25^2}{2!} \approx 0.22383 .$$

2.4 Schwaches Gesetz großer Zahlen

Als Vorbereitung zum schwachen Gesetz für große Zahlen zeigen wir

Satz 2.37 (Tschebyscheffsche Ungleichung). *Es seien $(\Omega, \mathfrak{P}(\Omega), P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable mit endlicher Varianz. Dann gilt für jedes $\varepsilon > 0$:*

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

Beweis. Sei $Z = X - E(X)$. Wir definieren eine neue Zufallsvariable $Y : \Omega \rightarrow \mathbb{R}$ durch

$$Y(\omega) = \begin{cases} 0 & \text{für } \omega \text{ mit } |Z(\omega)| < \varepsilon \\ \varepsilon^2 & \text{für } \omega \text{ mit } |Z(\omega)| \geq \varepsilon \end{cases} .$$

Dann ist $Y \leq Z^2$, also

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

□

Hieraus folgern wir

Satz 2.38 (Schwaches Gesetz der großen Zahlen für unabhängige Zufallsvariable mit beschränkter Varianz). *Seien X_1, \dots, X_n paarweise unabhängige Zufallsvariable mit gleichem Erwartungswert und endlicher Varianz $\text{Var}(X_k) \leq M$ für $1 \leq k \leq n$. Dann gilt für alle $\varepsilon > 0$:*

$$P\left(\left|\frac{1}{n}(X_1 + \dots + X_n) - E(X_1)\right| \geq \varepsilon\right) \leq \frac{M}{n\varepsilon^2}.$$

Beweis. Es sei $X := \frac{1}{n}(X_1 + \dots + X_n)$. Dann ist $E(X) = E(X_1)$ und nach Satz 2.16 (ii) und (v) folgt:

$$\text{Var}(X) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) \leq \frac{M}{n}.$$

Die direkte Anwendung der Tschebyscheffschen Ungleichung liefert die Behauptung. \square

Bemerkung 2.39. a) Sind Y_1, Y_2, \dots Zufallsvariable, die auf einem gemeinsamen Wahrscheinlichkeitsraum definiert sind, und ist a eine reelle Zahl mit der Eigenschaft

$$\lim_{n \rightarrow \infty} P(|Y_n - a| \geq \varepsilon) = 0 \text{ für jedes } \varepsilon > 0,$$

so sagt man, dass die Folge (Y_n) **stochastisch gegen a konvergiert**.

b) Das schwache Gesetz der großen Zahlen besagt, dass die Folge der arithmetischen Mittel von unabhängigen Zufallsvariablen mit gleichem Erwartungswert μ und beschränkter Varianz stochastisch gegen μ konvergiert. In diesem Sinne wird die intuitive Vorstellung des Erwartungswertes als ein bei häufiger Durchführung des Experimentes erhaltener durchschnittlicher Wert präzisiert.

Kapitel 3

Stetige Zufallsvariable

3.1 Dichte und Verteilungsfunktion

Manchmal ist es hilfreich häufig auftretende Verteilungen durch stetige Verteilungen anzunähern. Daher wollen wir uns in diesem Kapitel auch mit stetigen Zufallsvariablen beschäftigen.

Definition 3.1. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt *stetige Zufallsvariable*, falls es eine integrierbare, nicht negative reelle Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

gibt mit der Eigenschaft

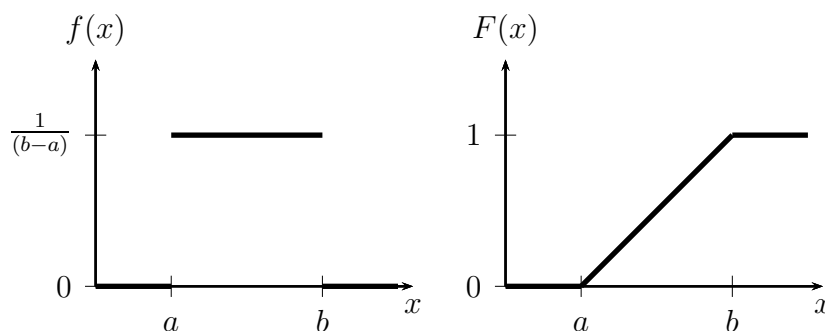
$$P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Die Funktion

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto P(X \leq x) = \int_{-\infty}^x f(t) dt \end{aligned}$$

heißt *Verteilungsfunktion* von X , die Funktion f heißt *Dichte* der Zufallsvariablen X .

Anbei ein Beispiel:



Ohne Beweis geben wir die folgenden beiden Sätze an, die den Sätzen 2.5 bzw. 2.7 für diskrete Zufallsvariable entsprechen:

Satz 3.2. Ist X eine stetige Zufallsvariable mit Verteilungsfunktion F und Dichte f , so gilt:

(a) Ist $x < y$, so gilt $F(x) \leq F(y)$.

(b) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$.

Satz 3.3. Ist F die Verteilungsfunktion einer stetigen Zufallsvariablen X mit Dichte f , so gilt für alle reellen Zahlen $a < b$:

(a) $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t) dt$.

(b) $P(X > a) = 1 - F(a) = \int_a^\infty f(t) dt$.

Beispiel 3.4. Es sei $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ für $x \in \mathbb{R}$; dann ist f nichtnegativ, stetig, und es gilt nach [4, Satz 11.28]:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1.$$

Wir definieren nun für ein Intervall $I =]a, b]$

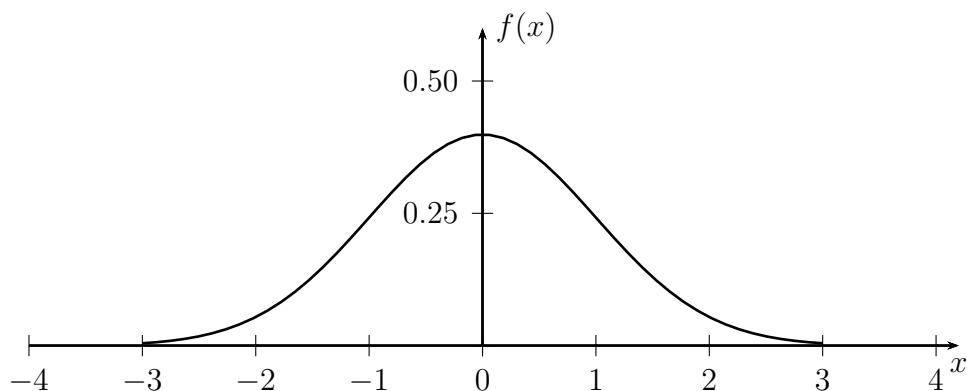
$$P(I) := F(b) - F(a) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx,$$

wobei durch

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

die Verteilungsfunktion zu der Wahrscheinlichkeitsdichte f gegeben ist. $P(I)$ gibt also den Flächeninhalt an, der unter der "Flächenkurve" f zwischen $x = a$ und $x = b$ liegt.

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



Definition 3.5. Zwei stetige Zufallsvariable X und Y heißen *unabhängig*, wenn die Ereignisse $(X \leq x)$ und $(Y \leq y)$ für beliebige $(x, y) \in \mathbb{R}^2$ unabhängig sind, d.h. wenn

$$P((X \leq x) \cap (Y \leq y)) = P(X \leq x) \cdot P(Y \leq y)$$

gilt. Sonst heißen X und Y *abhängig*.

3.2 Erwartungswert und Varianz stetiger Zufallsvariabler

Definition 3.6. (a) Ist X eine stetige Zufallsvariable mit der Dichtefunktion f , so heißt

$$E(X) := \int_{-\infty}^{\infty} xf(x)dx$$

der *Erwartungswert* von X .

(b) Ist X eine stetige Zufallsvariable mit der Dichtefunktion f derart, dass $E(X^2)$ existiert, so definiert man die *Varianz* durch

$$Var(X) := \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx.$$

$\sigma = \sigma_X = \sqrt{Var(X)}$ heißt *Standardabweichung* von X .

Die in Satz 2.12 sowie Satz 2.16 studierten Eigenschaften des Erwartungswerts und der Varianz diskreter Zufallsvariablen gelten auch für stetige Zufallsvariablen. Ebenfalls läßt sich die Ungleichung von Tschebyscheff und das Schwache Gesetz großer Zahlen aus Abschnitt 2.4 auf stetige Zufallsvariable übertragen.

3.3 Wichtige stetige Wahrscheinlichkeitsverteilungen

3.3.1 Gamma-Verteilung und Exponentialverteilung

Zahlreiche Zufallsvariable können nur nichtnegative Werte annehmen. Ihre empirische Wahrscheinlichkeitsdichte ist oft unimodal und asymmetrisch. Als Beispiel sei die Zeitspanne zwischen Funktionsstörungen bei Flugzeugmotoren, die Dauer von Routineuntersuchungen bei Flugzeugen oder Autos erwähnt. Derartige Situationen lassen sich nur schlecht mit einer Normalverteilung modellieren, da deren Dichte einerseits symmetrisch um den Erwartungswert ist, andererseits auch negativen Werten positive Wahrscheinlichkeitsdichten zugewiesen werden. In solchen Fällen ist oft der Einsatz der Gamma-Verteilung sinnvoll.

Definition 3.7. Eine stetige Zufallsvariable X besitzt eine *Gamma-Verteilung* mit

den Parametern $\alpha > 0$ und $\lambda > 0$, wenn ihre Dichte gegeben ist durch

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Die Gamma-Funktion $\Gamma(\alpha)$ ist dabei definiert durch

$$\Gamma(\alpha) := \int_0^\infty u^{\alpha-1} e^{-u} du, \quad \alpha > 0.$$

Die Form der Verteilungsfunktion hängt stark von α ab. Man nennt daher α den Formparameter der Gamma-Verteilung. Für $\alpha = 1$ erhält man die *Exponentialverteilung*.

Der Parameter λ heißt Skalierungsparameter. Multipliziert man nämlich eine Gamma-verteilte Zufallsvariable X mit einer Konstanten β , so erhält man wieder eine Gamma-verteilte Zufallsvariable mit gleichem α , der Parameter λ wird durch $\frac{\lambda}{\beta}$ ersetzt. Einer Änderung von λ entspricht also die Änderung der Maßeinheit beim zugrundeliegenden Zufallsexperiment.

Die Gamma-Funktion ist eine Verallgemeinerung der Fakultät: Direkte Integration zeigt

$$\Gamma(1) = 1,$$

und mit partieller Integration verifiziert man

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha > 0,$$

woraus insbesondere

$$\Gamma(n) = (n - 1)!, \quad n \in \mathbb{N},$$

folgt. Ferner gilt: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Für $\alpha \notin \mathbb{N}$ ist es nicht möglich, einen geschlossenen Ausdruck für die Wahrscheinlichkeit $P(a < X \leq b)$ anzugeben.

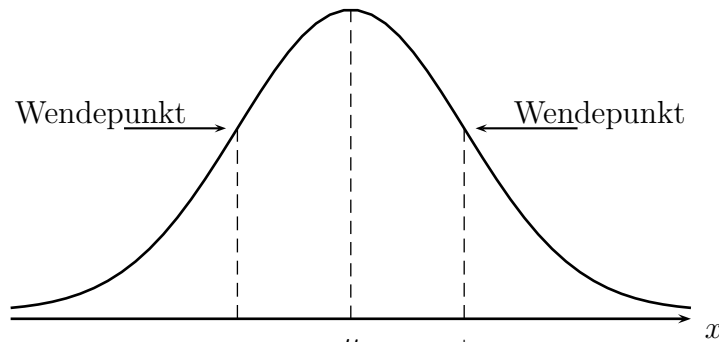
3.3.2 Die Normal-Verteilung

Definition 3.8. Eine stetige Zufallsvariable heißt *normal-verteilt* mit den Parametern μ und σ (kurz: $X \sim N(\mu, \sigma^2)$), wenn die Dichtefunktion folgende Gestalt hat:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right).$$

Wir sagen kurz: X ist $N(\mu, \sigma^2)$ -verteilt. Die Dichtefunktion aus Beispiel 3.4 ergibt sich mit $\sigma = 1$ und $\mu = 0$; die $N(0, 1)$ -Verteilung heißt *Standardnormalverteilung*.

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



Satz 3.9. Ist $X \sim N(\mu, \sigma^2)$ -verteilt, so erhalten wir mit der Substitution $\frac{x-\mu}{\sigma} = t$

$$\begin{aligned} E(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) \exp\left(-\frac{1}{2}t^2\right) \sigma dt \\ &= \frac{1}{\sqrt{2\pi}} \mu \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}t^2\right) dt = \mu \end{aligned}$$

und

$$\begin{aligned} \text{Var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 t^2 \exp\left(-\frac{1}{2}t^2\right) \sigma dt \\ &= \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} \sigma^2 t^2 \exp\left(-\frac{1}{2}t^2\right) dt = \frac{2}{\sqrt{\pi}} \sigma^2 \int_0^{\infty} \sqrt{u} \exp(-u) du \\ &= \frac{2}{\sqrt{\pi}} \sigma^2 \Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}} \sigma^2 \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \sigma^2. \end{aligned}$$

mit der Gammafunktion $\Gamma(x) := \int_0^{\infty} u^{x-1} \exp(-u) du$, $x > 0$, die $\Gamma(x+1) = x \cdot \Gamma(x)$ sowie $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ erfüllt.

Bemerkungen 3.10. Ist $X \sim N(0, 1)$ -verteilt ($X \sim N(0, 1)$), so bezeichnet man üblicherweise die zugehörige Verteilungsfunktion mit Φ , d.h.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2}t^2\right) dt.$$

Die Wahrscheinlichkeit, dass X Werte im Intervall $]a, b]$ mit $a < b$ annimmt, ist dann $P(a < X \leq b) = \phi(b) - \phi(a)$. Da man ϕ nicht mit elementaren Funktionen darstellen kann, gibt es Tabellen, in denen viele Werte von ϕ eingetragen sind. Um diese Tabellen zu erzeugen, entwickelt man den Integranden in eine Potenzreihe, integriert gliedweise und wertet die entstehende Reihe numerisch aus.

In den Tabellen sind allerdings nur Werte $\phi(x)$ für $x \geq 0$ angegeben. Wegen der Symmetrie der Dichtefunktion ergibt sich für $x > 0$

$$\phi(-x) = 1 - \phi(x) .$$

Ist X $N(\mu, \sigma^2)$ -verteilt, so erhalten wir

$$P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt .$$

Die Substitution $t - \mu = \sigma u$ liefert

$$P(X \leq x) = \phi\left(\frac{x-\mu}{\sigma}\right)$$

und damit

$$P(a < X \leq b) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right) .$$

Also kann man mit den Tabellen für die $N(0, 1)$ -Normalverteilung auch die Verteilungsfunktion für die $N(\mu, \sigma^2)$ -Verteilung näherungsweise berechnen.

Übrigens: Aufgrund der Stetigkeit von ϕ (bzw. f) ist es unerheblich, ob die Intervallgrenzen a, b zum Intervall von X dazu gehören.

Bemerkungen 3.11. Wir wollen ein paar Bemerkungen zur Gewinnung der Tabellen für die Normalverteilung machen. Man wählt als Ersatz für den exakten Funktionswert $\phi(x)$ die Näherung

$$1 - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) (a_1 s + a_2 s^2 + a_3 s^3) \quad \text{mit } s = \frac{1}{1 + bx}$$

und

$$b = 0.33267, \quad a_1 = 0.4361836, \quad a_2 = -0.1201676 \quad \text{und} \quad a_3 = 0.937298 .$$

Beispiel 3.12. Ein Werkstück soll eine Bohrung erhalten mit einem Durchmesser von 50 mm. Die Toleranzgrenzen sind $t_u = 49.97$ mm und $t_o = 50.04$ mm. Es sei bekannt, dass die von den Bohrautomaten erstellten Bohrungen $N(\mu, \sigma^2)$ -verteilt sind, wobei $\mu = 50$ mm und $\sigma = 0.02$ mm gelten soll. Ein Werkstück ist Ausschuss, wenn der Durchmesser größer als t_o ausfällt. Ist der Durchmesser kleiner als t_u , so muss eine Nachbohrung durchgeführt werden.

- a) Wie groß ist die Wahrscheinlichkeit, dass die Bohrung sofort qualitätsgerecht ausfällt?

- b) Wie groß ist die Wahrscheinlichkeit dafür, dass das Werkstück nachgebessert werden muss?
- c) Wie groß ist die Ausschusswahrscheinlichkeit?

Um die erste Frage zu beantworten, gilt mit den Bemerkungen 3.10 und einer Tabelle

$$\begin{aligned} P(t_u \leq X \leq t_o) &= \phi\left(\frac{50.04 - 50}{0.02}\right) - \phi\left(\frac{49.97 - 50}{0.02}\right) = \phi(2) - \phi(-1.5) \\ &= \phi(2) - (1 - \phi(1.5)) \approx 0.9772 - (1 - 0.9332) = 0.9104. \end{aligned}$$

Also ist bei 91,04 % der Werkstücke von einer qualitätsgerechten Bohrung auszugehen. Analog zur Berechnung für Teil a) erhalten wir zur Beantwortung von Teil b):

$$P(X < t_u) = \phi\left(\frac{49.97 - 50}{0.02}\right) = \phi(-1.5) \approx 0.0668.$$

Also ist mit einer Wahrscheinlichkeit von ungefähr 6,68 % von einer Nachbesserung auszugehen. Für die Ausschusswahrscheinlichkeit gilt

$$P(X > t_o) = 1 - P(X \leq t_o) = 1 - \phi\left(\frac{50.04 - 50}{0.02}\right) = 1 - \phi(2) \approx 0.0228.$$

Mit einer Wahrscheinlichkeit von 2,28 % ist der Durchmesser zu groß.

Bemerkung 3.13. Häufig ist man an der Wahrscheinlichkeit interessiert, dass die $N(\mu, \sigma^2)$ -verteilte Zufallsvariable Werte in einem zu μ symmetrischen Intervall $[\mu - k\sigma, \mu + k\sigma]$ mit $k \in \mathbb{N}$ annimmt. Es ist üblich, die Abweichung von μ in Einheiten von σ anzugeben. Deshalb spricht man vom $k\sigma$ -Intervall. Wir erhalten aus den Bemerkungen 3.10:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = \phi(k) - \phi(-k) = 2\phi(k) - 1.$$

Speziell für $k = 1, 2, 3$ ergeben sich folgende Werte

$$P(\mu - 1 \cdot \sigma \leq X \leq \mu + 1 \cdot \sigma) = 2\phi(1) - 1 \approx 0.6826,$$

$$P(\mu - 2 \cdot \sigma \leq X \leq \mu + 2 \cdot \sigma) = 2\phi(2) - 1 \approx 0.9544,$$

$$P(\mu - 3 \cdot \sigma \leq X \leq \mu + 3 \cdot \sigma) = 2\phi(3) - 1 \approx 0.9974.$$

Also liegen ca. 68 % der beobachteten Werte bei einer $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen zwischen $\mu - \sigma$ und $\mu + \sigma$, ca. 95 % liegen zwischen $\mu - 2\sigma$ und $\mu + 2\sigma$ und ca. 99.7 % liegen zwischen $\mu - 3\sigma$ und $\mu + 3\sigma$.

Tabelle mit den Werten der Verteilungsfunktion ϕ der Standardnormalverteilung

x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$
0,00	0,5000	0,53	0,7019	1,06	0,8554	1,59	0,9441	2,12	0,9830	2,65	0,9960	3,18	0,99926
0,01	0,5040	0,54	0,7054	1,07	0,8577	1,60	0,9452	2,13	0,9834	2,66	0,9961	3,19	0,99929
0,02	0,5080	0,55	0,7088	1,08	0,8599	1,61	0,9463	2,14	0,9838	2,67	0,9962	3,20	0,99931
0,03	0,5120	0,56	0,7123	1,09	0,8621	1,62	0,9474	2,15	0,9842	2,68	0,9963	3,21	0,99934
0,04	0,5160	0,57	0,7157	1,10	0,8643	1,63	0,9485	2,16	0,9846	2,69	0,9964	3,22	0,99936
0,05	0,5199	0,58	0,7190	1,11	0,8665	1,64	0,9495	2,17	0,9850	2,70	0,9965	3,23	0,99938
0,06	0,5239	0,59	0,7224	1,12	0,8686	1,65	0,9505	2,18	0,9854	2,71	0,9966	3,24	0,99940
0,07	0,5279	0,60	0,7258	1,13	0,8708	1,66	0,9515	2,19	0,9857	2,72	0,9967	3,25	0,99942
0,08	0,5319	0,61	0,7291	1,14	0,8729	1,67	0,9525	2,20	0,9861	2,73	0,9968	3,26	0,99944
0,09	0,5359	0,62	0,7324	1,15	0,8749	1,68	0,9535	2,21	0,9865	2,74	0,9969	3,27	0,99946
0,10	0,5398	0,63	0,7357	1,16	0,8770	1,69	0,9545	2,22	0,9868	2,75	0,9970	3,28	0,99948
0,11	0,5438	0,64	0,7389	1,17	0,8790	1,70	0,9554	2,23	0,9871	2,76	0,99711	3,29	0,99950
0,12	0,5478	0,65	0,7422	1,18	0,8810	1,71	0,9564	2,24	0,9875	2,77	0,99720	3,30	0,99952
0,13	0,5517	0,66	0,7454	1,19	0,8830	1,72	0,9573	2,25	0,9878	2,78	0,99728	3,31	0,99953
0,14	0,5557	0,67	0,7486	1,20	0,8849	1,73	0,9582	2,26	0,9881	2,79	0,99736	3,32	0,99955
0,15	0,5596	0,68	0,7518	1,21	0,8869	1,74	0,9591	2,27	0,9884	2,80	0,99744	3,33	0,99957
0,16	0,5636	0,69	0,7549	1,22	0,8888	1,75	0,9599	2,28	0,9887	2,81	0,99752	3,34	0,99958
0,17	0,5675	0,70	0,7580	1,23	0,8907	1,76	0,9608	2,29	0,9890	2,82	0,99760	3,35	0,99960
0,18	0,5714	0,71	0,7612	1,24	0,8925	1,77	0,9616	2,30	0,9893	2,83	0,99767	3,36	0,99961
0,19	0,5754	0,72	0,7642	1,25	0,8944	1,78	0,9625	2,31	0,9896	2,84	0,99774	3,37	0,99962
0,20	0,5793	0,73	0,7673	1,26	0,8962	1,79	0,9633	2,32	0,9898	2,85	0,99781	3,38	0,99964
0,21	0,5832	0,74	0,7704	1,27	0,8980	1,80	0,9641	2,33	0,9901	2,86	0,99788	3,39	0,99965
0,22	0,5871	0,75	0,7734	1,28	0,8997	1,81	0,9649	2,34	0,9904	2,87	0,99795	3,40	0,99966
0,23	0,5910	0,76	0,7764	1,29	0,9015	1,82	0,9656	2,35	0,9906	2,88	0,99801	3,41	0,99968
0,24	0,5948	0,77	0,7794	1,30	0,9032	1,83	0,9664	2,36	0,9909	2,89	0,99807	3,42	0,99969
0,25	0,5987	0,78	0,7823	1,31	0,9049	1,84	0,9671	2,37	0,9911	2,90	0,99813	3,43	0,99970
0,26	0,6026	0,79	0,7852	1,32	0,9066	1,85	0,9678	2,38	0,9913	2,91	0,99819	3,44	0,99971
0,27	0,6064	0,80	0,7881	1,33	0,9082	1,86	0,9686	2,39	0,9916	2,92	0,99825	3,45	0,99972
0,28	0,6103	0,81	0,7910	1,34	0,9099	1,87	0,9693	2,40	0,9918	2,93	0,99831	3,46	0,99973
0,29	0,6141	0,82	0,7939	1,35	0,9115	1,88	0,9700	2,41	0,9920	2,94	0,99836	3,47	0,99974
0,30	0,6179	0,83	0,7967	1,36	0,9131	1,89	0,9706	2,42	0,9922	2,95	0,99841	3,48	0,99975
0,31	0,6217	0,84	0,7996	1,37	0,9147	1,90	0,9713	2,43	0,9925	2,96	0,99846	3,49	0,99976
0,32	0,6255	0,85	0,8023	1,38	0,9162	1,91	0,9719	2,44	0,9927	2,97	0,99851	3,50	0,99977
0,33	0,6293	0,86	0,8051	1,39	0,9177	1,92	0,9726	2,45	0,9929	2,98	0,99856	3,51	0,99978
0,34	0,6331	0,87	0,8079	1,40	0,9192	1,93	0,9732	2,46	0,9931	2,99	0,99861	3,52	0,99978
0,35	0,6368	0,88	0,8106	1,41	0,9207	1,94	0,9738	2,47	0,9932	3,00	0,99865	3,53	0,99979
0,36	0,6406	0,89	0,8133	1,42	0,9222	1,95	0,9744	2,48	0,9934	3,01	0,99869	3,54	0,99980
0,37	0,6443	0,90	0,8159	1,43	0,9236	1,96	0,9750	2,49	0,9936	3,02	0,99874	3,55	0,99981
0,38	0,6480	0,91	0,8186	1,44	0,9251	1,97	0,9756	2,50	0,9938	3,03	0,99878	3,56	0,99981
0,39	0,6517	0,92	0,8212	1,45	0,9265	1,98	0,9762	2,51	0,9940	3,04	0,99882	3,57	0,99982
0,40	0,6554	0,93	0,8238	1,46	0,9279	1,99	0,9767	2,52	0,9941	3,05	0,99886	3,58	0,99983
0,41	0,6591	0,94	0,8264	1,47	0,9292	2,00	0,9773	2,53	0,9943	3,06	0,99889	3,59	0,99983
0,42	0,6628	0,95	0,8289	1,48	0,9306	2,01	0,9778	2,54	0,9945	3,07	0,99893	3,60	0,99984
0,43	0,6664	0,96	0,8315	1,49	0,9319	2,02	0,9783	2,55	0,9946	3,08	0,99896	3,61	0,99985
0,44	0,6700	0,97	0,8340	1,50	0,9332	2,03	0,9788	2,56	0,9948	3,09	0,99900	3,62	0,99985
0,45	0,6736	0,98	0,8365	1,51	0,9345	2,04	0,9793	2,57	0,9949	3,10	0,99903	3,63	0,99986
0,46	0,6772	0,99	0,8389	1,52	0,9357	2,05	0,9798	2,58	0,9951	3,11	0,99906	3,64	0,99986
0,47	0,6808	1,00	0,8413	1,53	0,9370	2,06	0,9803	2,59	0,9952	3,12	0,99910	3,65	0,99987
0,48	0,6844	1,01	0,8438	1,54	0,9382	2,07	0,9808	2,60	0,9953	3,13	0,99913	3,66	0,99987
0,49	0,6879	1,02	0,8461	1,55	0,9394	2,08	0,9812	2,61	0,9955	3,14	0,99916	3,67	0,99988
0,50	0,6915	1,03	0,8485	1,56	0,9406	2,09	0,9817	2,62	0,9956	3,15	0,99918		
0,51	0,6950	1,04	0,8508	1,57	0,9418	2,10	0,9821	2,63	0,9957	3,16	0,99921		
0,52	0,6985	1,05	0,8531	1,58	0,9430	2,11	0,9826	2,64	0,9959	3,17	0,99924		

3.3.3 Die χ^2 -Verteilung

Definition 3.14. Es sei X eine standardnormal-verteilte Zufallsvariable. Man nennt die Verteilung von $Z = X^2$ χ^2_1 -Verteilung mit einem Freiheitsgrad.

Ist X also standardnormal-verteilt, so erhalten wir für die Verteilung von $Z = X^2$ (m.H. der Substitution $s = t^2$):

$$\begin{aligned} F_Z(x) &= P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} \exp\left(-\frac{t^2}{2}\right) dt = \frac{1}{\sqrt{2\pi}} \int_0^x \frac{1}{\sqrt{s}} e^{-\frac{s}{2}} ds. \end{aligned}$$

Wir haben in Definition 3.14 festgelegt, dass das Quadrat einer standardnormal-verteilten Zufallsvariablen χ^2_1 -verteilt (1 Freiheitsgrad) ist und deren Verteilung ausgerechnet. Ein Vergleich zeigt, dass die Dichte der χ^2_1 -Verteilung durch die Dichte der Gamma-Verteilung mit $\alpha = \lambda = \frac{1}{2}$ gegeben ist. Wir betrachten nun eine etwas allgemeinere Situation.

Definition 3.15. Es seien U_1, \dots, U_n unabhängige χ^2_1 -verteilte Zufallsvariable. Die Verteilung von $V = U_1 + \dots + U_n$ heißt χ^2_n -Verteilung mit n Freiheitsgraden und wird χ^2_n bezeichnet.

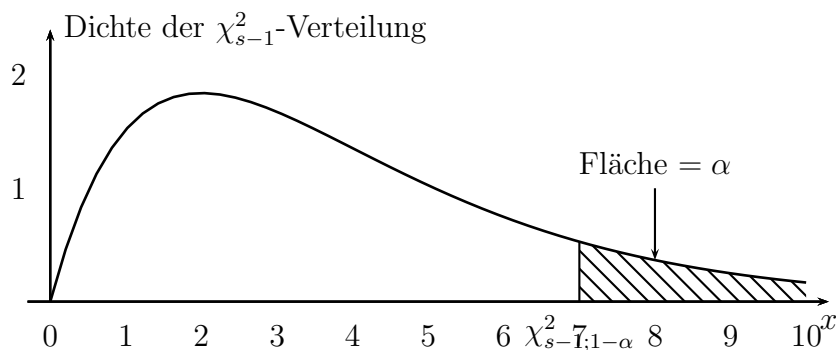
Da die Zufallsvariablen U_i Gamma-verteilt mit den Parametern $(\alpha, \lambda) = (\frac{1}{2}, \frac{1}{2})$ sind, folgt, dass V ebenfalls Gamma-verteilt und zwar mit den Parametern $(\frac{n}{2}, \frac{1}{2})$ ist. Somit erhalten wir für die Wahrscheinlichkeitsdichte der χ^2_n -Verteilung

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(n/2)} x^{n/2-1} e^{-x/2}, \quad x \geq 0.$$

Es gilt

$$E(V) = n, \quad \sigma_V^2 = 2n.$$

Die χ^2_n -Verteilung besitzt folgende bemerkenswerte Eigenschaft: wenn U einer χ^2_n -Verteilung, V einer χ^2_m -Verteilung folgt, dann besitzt $U + V$ eine χ^2_{n+m} -Verteilung.



3.3.4 Die Student-Verteilung

Definition 3.16. Es sei X standardnormal-verteilt und U folge einer χ_n^2 -Verteilung, $n \in \mathbb{N}$. Die Verteilung von $T = \frac{X}{\sqrt{U/n}}$ heißt t -Verteilung mit n Freiheitsgraden, kurz t_n -Verteilung.

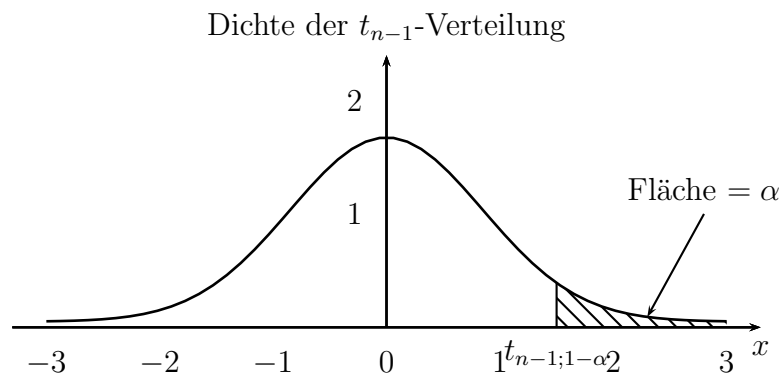
Bemerkung 3.17 (Die t -Verteilung). Die t -Verteilung wurde von W. S. Gosset eingeführt, der im Department of Agriculture in Dublin arbeitete und unter dem Pseudonym "Student" veröffentlichte. Die t -Verteilung heißt daher auch **Student-Verteilung**.

Auf solche Verteilungen stößt man, wenn man aus einer Stichprobe mit unbekannter Streuung Aussagen über den Erwartungswert machen soll.

Für die Dichte der t_n -Verteilung erhält man nach längerer Rechnung

$$f_T(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad t \in \mathbb{R}.$$

Die Dichte ist eine gerade Funktion und besitzt eine große Ähnlichkeit zu der Dichte der Standardnormalverteilung.



Kapitel 4

Grenzwertsätze

Bevor wir uns mit den eigentlichen Grenzwertsätzen beschäftigen, stellen wir noch ein paar Begriffe zur Verfügung.

Definition 4.1. a) Ist X eine Zufallsvariable mit dem Erwartungswert $\mu = E(X)$ und der Varianz $\sigma^2 = Var(X) > 0$, so gilt für die Zufallsvariable

$$X^* := \frac{X - \mu}{\sigma}$$

nach den Sätzen 2.12 und 2.16

$$E(X^*) = 0 \quad \text{und} \quad Var(X^*) = 1 .$$

X^* heißt die zu X gehörende *standardisierte Variable*.

b) Zwei Zufallsvariable X und Y auf demselben Wahrscheinlichkeitsraum (Ω, \mathcal{A}, p) heißen *gleichverteilt* oder *identisch verteilt*, wenn ihre Verteilungsfunktionen übereinstimmen.

Beispiel 4.2. Wir betrachten einen einfachen Würfelwurf; dann ist $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathfrak{P}(\Omega)$ und $p(\omega) = \frac{1}{6}$ für alle $\omega \in \Omega$. Wird eine gerade Zahl gewürfelt, so wird ein Euro ausgezahlt; sonst verliert der Spieler den Einsatz von einem Euro. Die Zufallsvariable X beschreibe dieses Spiel. Wird eine Primzahl gewürfelt, so gewinnt der Spieler einen Euro; sonst verliert er seinen Einsatz. Dieses Spiel wird durch Y beschrieben. Wir erhalten

ω	1	2	3	4	5	6
$X(\omega)$	-1	1	-1	1	-1	1
$Y(\omega)$	-1	1	1	-1	1	-1

und damit die Verteilungsfunktionen F_X und F_Y mit

$$F_X(x) = F_Y(x) = \begin{cases} 0 & \text{für } x < -1 \\ \frac{1}{2} & \text{für } -1 \leq x < 1 \\ 1 & \text{für } 1 \leq x \end{cases} .$$

Verschiedene Zufallsvariable können also dieselbe Verteilungsfunktion besitzen. Hier sind X und Y identisch verteilt.

4.1 Der Satz von de Moivre-Laplace

Wir wollen nun motivieren, warum wir die Binomialverteilung für große n durch die Normalverteilung ersetzen dürfen.

Es sei X $B(n, p)$ -verteilt; dann besitzt die Verteilungsfunktion Sprungstellen in den Punkten $0, 1, \dots, n$. Betrachten wir die zu X gehörende standardisierte Variable X^* , so hat die zu X^* gehörende Verteilungsfunktion Sprungstellen in den Punkten

$$u_k^{(n)} = \frac{k - \mu}{\sigma} \quad \text{für } k = 0, \dots, n.$$

Z.B. hat die Verteilungsfunktion X für $n = 12$ und $p = 0.25$, d.h. $\mu = np = 3$ und $\sigma = \sqrt{np(1-p)} = \frac{3}{2}$ die Sprungstellen $0, \dots, 12$. Die zu $X^* = \frac{2}{3}(X - 3)$ gehörende Verteilungsfunktion hat die Sprungstellen

$$-2, -\frac{4}{3}, -\frac{2}{3}, 0, \frac{2}{3}, \frac{4}{3}, \dots, 6.$$

Auf dem Intervall $[k, k + 1[$ gilt für die zu X gehörende Dichte f_X :

$$f_X(x) = B_{n,p}(k), \quad x \in [k, k + 1[.$$

Für die zu X^* gehörende Dichte f_{X^*} gilt

$$f_{X^*}(z) = \sigma B_{n,p}(k), \quad \text{für } z \in [u_k^{(n)}, u_{k+1}^{(n)}[.$$

Wir ersetzen nun die stückweise konstante Dichtefunktion f_{X^*} durch einen Polygonzug m_n . Dazu verbinden wir die Punkte $(u_k^{(n)} + \frac{1}{2\sigma}, f_{X^*}(u_k^{(n)}))$ und $(u_{k+1}^{(n)} + \frac{1}{2\sigma}, f_{X^*}(u_{k+1}^{(n)}))$. Wir nehmen nun an, dass dieser Polygonzug für $n \rightarrow \infty$ gegen eine auf \mathbb{R} differenzierbare Funktion φ konvergiert, und werden zeigen, dass sich als Grenzwert die Dichtefunktion der Normalverteilung ergibt.

Wir berechnen bei fest vorgegebenem $u \in \mathbb{R}$ zunächst k in Abhängigkeit von n so, dass

$$u \in [u_k^{(n)}, u_{k+1}^{(n)}[$$

gilt, vorausgesetzt u liegt zwischen $u_0^{(n)}$ und $u_n^{(n)}$. Nun ist

$$u_k^{(n)} = \frac{k - \mu}{\sigma} \leq u < u_{k+1}^{(n)} = \frac{k + 1 - \mu}{\sigma}$$

genau dann, wenn

$$\sigma u + \mu - 1 < k \leq \sigma u + \mu,$$

d.h. wenn

$$u\sqrt{np(1-p)} + np - 1 < k \leq u\sqrt{np(1-p)} + np$$

gilt. Wir betrachten nun den Polygonzug m_n an der Stelle u . Die Ableitung von m_n an der Stelle u ergibt sich durch die Steigung des Polygonzuges, d.h.

$$m'_n(u) = \frac{\sigma B_{n,p}(k+1) - \sigma B_{n,p}(k)}{\frac{1}{\sigma}}.$$

Aus der Rekursionsformel in Beispiel 2.19 folgt

$$m'_n(u) = \sigma^2 B_{n,p}(k) \left(\frac{(n-k)p}{(k+1)(1-p)} - 1 \right) = \sigma^2 B_{n,p}(k) \frac{np - k - (1-p)}{(k+1)(1-p)}.$$

Ersetzen wir $\sigma B_{n,p}(k)$ durch $m_n(u_k^{(n)})$ und k durch $\sigma u_k^{(n)} + \mu$, so erhalten wir wegen $\mu = np$ und $\mu(1-p) = \sigma^2$:

$$\begin{aligned} m'_n(u) &= \sigma m_n(u_k^{(n)}) \frac{np - \sigma u_k^{(n)} - \mu - (1-p)}{(\sigma u_k^{(n)} + \mu + 1)(1-p)} \\ &= m_n(u_k^{(n)}) \frac{-(\sigma^2 u_k^{(n)} + \sigma(1-p))}{\sigma u_k^{(n)}(1-p) + \sigma^2 + (1-p)} \\ &= m_n(u_k^{(n)}) \frac{-(u_k^{(n)} + \frac{1-p}{\sigma})}{\frac{u_k^{(n)}(1-p)}{\sigma} + 1 + \frac{1-p}{\sigma^2}}. \end{aligned}$$

Wegen

$$|u - u_k^{(n)}| \leq |u_{k+1}^{(n)} - u_k^{(n)}| \leq \frac{1}{\sigma} = \frac{1}{\sqrt{np(1-p)}}$$

konvergiert $u_k^{(n)}$ für $n \rightarrow \infty$ gegen u . Damit erhalten wir durch Grenzübergang für die Grenzfunktion φ die Beziehung

$$\varphi'(u) = -u \cdot \varphi(u).$$

Dies ist eine homogene lineare Differentialgleichung 1. Ordnung für φ , deren Lösung sich zu

$$\varphi(u) = C \exp\left(-\frac{1}{2}u^2\right)$$

ergibt, wobei C so zu bestimmen ist, dass

$$\int_{-\infty}^{\infty} \varphi(u) du = 1$$

gilt. Daraus folgt $C = \frac{1}{\sqrt{2\pi}}$; also ist φ die Dichtefunktion der Normalverteilung.

Satz 4.3 (Satz von de Moivre-Laplace). *Es sei $0 < p < 1$ und X_n $B(n, p)$ -verteilt sowie X_n^* die zu X_n gehörende standardisierte Zufallsvariable. Dann gilt für alle $a < b$:*

$$\lim_{n \rightarrow \infty} P(a \leq X_n^* \leq b) = \phi(b) - \phi(a),$$

wobei ϕ (wie üblich) die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Bemerkung 4.4. Numerische Experimente zeigen, dass eine "gute" Näherung der Binomialverteilung durch die Normalverteilung vorliegt, wenn die **Faustregel**

$$np(1-p) > 9$$

erfüllt ist. Wir erhalten dann für die $B(n, p)$ -Verteilung

$$\begin{aligned} P(X = k) &= B_{n,p}(k) = \frac{f_{X^*}\left(\frac{k-\mu}{\sigma}\right)}{\sigma} \approx \frac{1}{\sigma} \varphi\left(\frac{k-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2\right) \end{aligned}$$

und

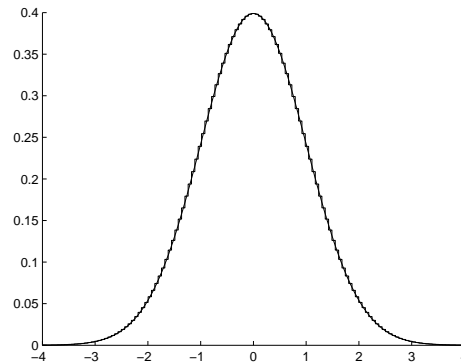
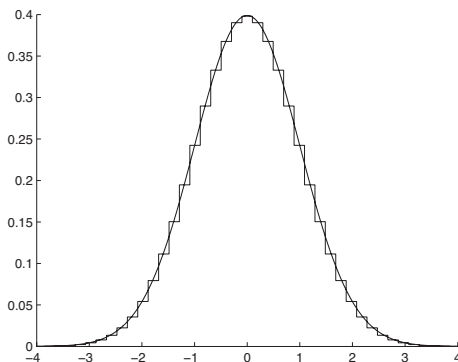
$$P(X \leq x) \approx \phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

sowie die (durch eine sog. Stetigkeitskorrektur) verbesserten Näherungen

$$P(X \leq x) \approx \phi\left(\frac{x - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

und

$$P(a \leq X \leq b) \approx \phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) - \phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{np(1-p)}}\right).$$



Wir wollen die letzten Überlegungen an einem Beispiel demonstrieren:

Beispiel 4.5. Für viele Sportveranstaltungen in Stadien werden häufig Freikarten vergeben. Erfahrungsgemäß nutzen nur 85% der auf diese Weise geladenen Gäste ihre Freikarten.

- a) Es werden 200 Freikarten ausgegeben; wir berechnen die Wahrscheinlichkeit dafür, dass genau 170 Ehrenplätze belegt werden.
- b) Für eine Veranstaltung werden 200 Freikarten vergeben. Wie viele Ehrenplätze sind mindestens bereitzustellen, damit die Wahrscheinlichkeit, dass alle ins Stadion kommenden geladenen Ehrengäste jeweils noch einen freien Ehrenplatz vorfinden, mindestens 97.5% beträgt?

Bezeichnen wir mit X die zufällige Anzahl der belegten Ehrenplätze (d.h. der genutzten Freikarten), so ist X $B(n, p)$ -verteilt mit $n = 200$ und $p = 0.85$. Es ist $\mu = np = 170$. Wegen $np(1-p) = 25.5 > 9$ ist die Faustregel aus Bemerkung 4.4 erfüllt. Um Teil a) zu beantworten, erhalten wir wegen $\sigma = \sqrt{np(1-p)} \approx 5.05$

$$\begin{aligned} P(X = 170) &\approx \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{170-170}{\sigma}\right)^2\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp(0) \approx \frac{1}{5.05} \cdot 0.399 \approx 0.079. \end{aligned}$$

(Dabei können wir die Werte $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$ entweder mit einem Taschenrechner berechnen oder aber einer Tabelle entnehmen.)

Die Wahrscheinlichkeit, dass genau 170 Ehrenplätze belegt werden, beträgt also ungefähr 7.9 %.

Um die zweite Frage zu beantworten, berechnen wir k so, dass

$$P(X \leq k) \geq 0.975$$

gilt. Wir ersetzen die Wahrscheinlichkeit $P(X \leq k)$ durch

$$\Phi\left(\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

und erhalten die Bedingung

$$\Phi\left(\frac{k - 170 + \frac{1}{2}}{5.05}\right) \geq 0.975,$$

woraus sich mit Hilfe der Tabelle für die Normalverteilung

$$\frac{k - 170 + \frac{1}{2}}{5.05} \geq 1.96$$

ergibt. Daraus erhalten wir $k > 179.3$ bzw. $k \geq 180$. Also benötigt man mindestens 180 Ehrenplätze.

4.2 Der zentrale Grenzwertsatz

Bei den **Zentralen Grenzwertsätzen** (ZGWS) handelt es sich um eine Familie schwacher Konvergenzaussagen aus der Wahrscheinlichkeitstheorie. Allen gemeinsam ist die Aussage, dass die (normierte und zentrierte) Summe einer großen Zahl von unabhängigen, identisch verteilten Zufallsvariablen annähernd (standard) normal verteilt ist. Dies erklärt auch die Sonderstellung der Normalverteilung.

Die wichtigste und bekannteste Aussage wird auch einfach als *Der Zentrale Grenzwertsatz* bezeichnet und befasst sich mit unabhängigen, identisch verteilten Zufallsvariablen, deren Erwartungswert und Varianz endlich sind; er wird auch als Grenzwertsatz von Lindeberg-Lewy bezeichnet, siehe Satz 4.6 unten.

Es existieren verschiedene Verallgemeinerungen, für die eine identische Verteilung keine notwendige Voraussetzung ist. Stattdessen wird dann eine andere Voraussetzung gefordert, die sicherstellt, dass keine der Variablen zu großen Einfluss auf das Ergebnis erhält. Beispiele sind die Lindeberg-Bedingung und die Ljapunow-Bedingung. Darüber hinausgehende Verallgemeinerungen gestatten sogar "schwache" Abhängigkeit der Zufallsvariablen.

Die Bezeichnung geht auf G. Polya's Arbeit *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem* von 1920 zurück.

Satz 4.6 (Grenzwertsatz von Lindeberg-Lewy). *Es sei X_1, X_2, \dots eine Folge (stochastisch) unabhängiger und identisch verteilter (kurz: u.i.v. oder i.i.d.) Zufallsvariablen mit $\sigma^2 = \text{Var}(X_1) > 0$. Setzen wir $\mu := E(X_1)$ und $S_n := X_1 + \dots + X_n$, so gilt:*

$$\lim_{n \rightarrow \infty} P \left(a \leq \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq b \right) = \phi(b) - \phi(a).$$

Der Beweis ist mathematisch etwas anspruchsvoller, deshalb verzichten wir hier darauf.

Folgerung 4.7. Betrachten wir in Satz 4.6 das symmetrische Intervall $[-k, k]$, so erhalten wir wegen $E(S_n) = n \cdot E(X_1) = n \cdot \mu$ und $\text{Var}(S_n) = n \cdot \text{Var}(X_1) = n\sigma^2$ die Beziehung

$$\lim_{n \rightarrow \infty} P \left(-k \leq \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \leq k \right) = \phi(k) - \phi(-k) = 2 \cdot \phi(k) - 1,$$

d.h.

$$\lim_{n \rightarrow \infty} P \left(E(S_n) - k\sqrt{\text{Var}(S_n)} \leq S_n \leq E(S_n) + k\sqrt{\text{Var}(S_n)} \right) = 2 \cdot \phi(k) - 1.$$

Mit Bemerkung 3.13 erhalten wir also, dass die Summe von n unabhängigen und identisch verteilten Zufallsvariablen (als Faustregel für große n) mit einer ungefähren Wahrscheinlichkeit von

- 0.6826 in den Grenzen $E(S_n) \pm 1 \cdot \sqrt{\text{Var}(S_n)}$
- 0.9544 in den Grenzen $E(S_n) \pm 2 \cdot \sqrt{\text{Var}(S_n)}$
- 0.9974 in den Grenzen $E(S_n) \pm 3 \cdot \sqrt{\text{Var}(S_n)}$

liegt.

Beispiel 4.8. Wir betrachten einen fairen Würfel, der n -mal geworfen wird. Die Zufallsvariable X_i gebe das Ergebnis des i -ten Wurfs an. Wir können davon ausgehen, dass die Würfe unabhängig voneinander und unter gleichen Bedingungen stattfinden. Also können wir die X_i als unabhängig und identisch verteilt ansehen. In Abschnitt 2.2 haben wir den Erwartungswert $E(X_1) = 3,5$ berechnet. Für die Varianz erhalten wir

$$\text{Var}(X_1) = \sum_{i=1}^6 \left(i - \frac{7}{2}\right)^2 \cdot \frac{1}{6} = \frac{35}{12} \approx 2.917.$$

Gehen wir einmal von 100 Würfeln aus, so ergibt sich mit der obigen Faustregel wegen $\sqrt{291.7} \approx 17.1$, dass die Augensumme dann mit einer Wahrscheinlichkeit von

- 0.6826 in den Grenzen $350 \pm 1 \cdot 17.1$, also zwischen 332 und 368,
- 0.9544 in den Grenzen $350 \pm 2 \cdot 17.1$, also zwischen 315 und 385
- 0.9974 in den Grenzen $350 \pm 3 \cdot 17.1$, also zwischen 298 und 402

liegt.

Kapitel 5

Statistik

Die Statistik wird üblicherweise in die *deskriptive* (beschreibende) und die *induktive* (beurteilende, schließende) Statistik eingeteilt. Diese Einteilung ist insofern irreführend, da der Eindruck erweckt wird, dass die deskriptive Statistik von subjektiven Einschätzungen frei sei. Das ist aber häufig nicht der Fall. Die Hauptaufgabe der deskriptiven Statistik ist zwar in erster Linie eine übersichtliche graphische und/oder tabellarische Darstellung der erhobenen Daten; es wird aber oft durch die Art der Präsentation (z.B. bzgl. der Umsatzentwicklung eines Unternehmens) eine Beeinflussung z.B. von potentiellen Geldgebern (Banken, Aktionäre usw.) versucht.

5.1 Elemente der deskriptiven Statistik

Wir stellen ein paar Begriffe zusammen, die uns eine leichtere Beschreibung von erhobenen oder gemessenen Daten ermöglichen und die uns aufmerksam machen auf häufig voreilig aus erhobenen Daten gezogene Schlüsse. Dabei gehen wir davon aus, dass die erhobenen Daten in Form von Zahlen $x_i \in \mathbb{R}$ vorliegen.

Definition 5.1. a) Die Zahl

$$\bar{x} := \bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i$$

heißt (*Stichproben-*) oder *arithmetisches Mittel* bzw. kurz *Mittelwert* der Daten x_1, \dots, x_n .

b) Die Zahl

$$s^2 := s_x^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

heißt (*Stichproben-*) oder *empirische Varianz* der Daten x_1, \dots, x_n .

Die Zahl $s_x = \sqrt{s_x^2}$ heißt (*Stichproben-*) oder *empirische Standardabweichung* von x_1, \dots, x_n .

Bemerkungen 5.2. a) Werden die Daten x_1, \dots, x_n linear transformiert zu

$$y_i = a \cdot x_i + b \quad \text{mit } a \neq 0,$$

so gilt (vgl. Sätze 2.12 und 2.16)

$$\bar{y} = a \cdot \bar{x} + b,$$

$$s_y^2 = a^2 \cdot s_x^2,$$

$$s_y = |a| \cdot s_x.$$

b) Die Größen \bar{x} , s_x^2 und s_x sind *ausreißerempfindlich*, d.h. dass eine Abänderung eines einzigen Wertes den Mittelwert, die Varianz und die Standardabweichung beliebig klein bzw. groß werden lassen kann.

Definition 5.3. a) Sortieren wir die Daten x_1, \dots, x_n der Größe nach, d.h. bilden wir $x_{(1)} = \min_{1 \leq i \leq n} x_i$ bis $x_{(n)} = \max_{1 \leq i \leq n} x_i$, so nennen wir

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

die *geordnete Stichprobe* zu x_1, \dots, x_n . Die Differenz $x_{(n)} - x_{(1)}$ heißt *Stichprobenspannweite*.

b) Der (*empirische*) *Median* oder *Zentralwert* der Stichprobe x_1, \dots, x_n ist definiert durch

$$x_{1/2} := \begin{cases} x_{(\frac{n+1}{2})} & \text{für ungerades } n \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{für gerades } n \end{cases}.$$

Der Median von $|x_1 - x_{1/2}|, |x_2 - x_{1/2}|, \dots, |x_n - x_{1/2}|$ heißt *Median-Abweichung* von x_1, \dots, x_n .

c) Ist $0 < p < 1$, so heißt die Zahl

$$x_p := \begin{cases} x_{(\lfloor np \rfloor + 1)} & \text{falls } n \cdot p \notin \mathbb{N} \\ \frac{1}{2}(x_{(np)} + x_{(np+1)}) & \text{falls } n \cdot p \in \mathbb{N} \end{cases}.$$

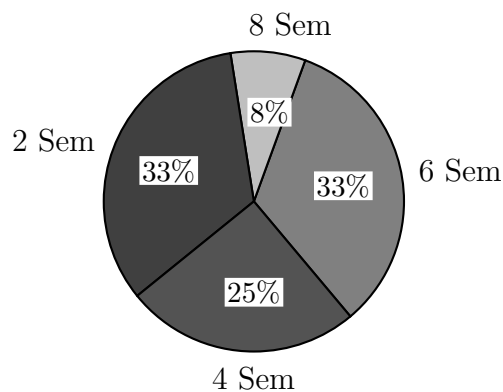
empirisches p-Quantil von x_1, \dots, x_n . (Dabei ist $\lfloor y \rfloor$ die größte ganze Zahl kleiner oder gleich $y \in \mathbb{R}$.)

d) Die Quantile $x_{3/4}$ und $x_{1/4}$ heißen *oberes* bzw. *unteres Quartil*. Die Differenz $x_{3/4} - x_{1/4}$ heißt *Quartilsabstand* der x_1, \dots, x_n .

Bemerkungen 5.4. Das p -Quantil x_p teilt die geordnete Stichprobe im Verhältnis $p : (1-p)$ auf. Links vom p -Quantil liegen $100 \cdot p$ Prozent aller Beobachtungswerte bzw. $100 \cdot p$ Prozent der Gesamtzahl der Zufallswerte. Rechts davon liegen $100 \cdot (1-p)$ Prozent aller Beobachtungswerte bzw. $100 \cdot (1-p)$ Prozent der Gesamtzahl der Zufallswerte. p heißt auch der Unterschreitungsanteil.

Wir wollen uns nun graphischen Darstellungsmöglichkeiten für Stichproben zuwenden. Zu erwähnen sind das *Stab-* und das *Kreisdiagramm*, das *Histogramm* und der *Box-Plot*. Beim Stabdiagramm werden die absoluten bzw. die relativen Häufigkeiten, mit denen die Daten auftreten, als Stäbchen gezeichnet, deren Länge die absolute Häufigkeit bzw. die relative Häufigkeit ist.

Beim Kreisdiagramm wird eine Kreisfläche in Sektoren aufgeteilt, deren Flächen proportional zu den (absoluten oder relativen) Häufigkeiten für das Auftreten verschiedener Daten sind. Beide Diagramme werden z.B. bei Ergebnissen von Wahlen verwendet.



Definition 5.5. Wir betrachten die Daten x_1, \dots, x_n . Wir teilen diese Daten in s disjunkte Klassen auf, indem wir s halboffene Intervalle

$$[a_1, a_2[, [a_2, a_3[, \dots, [a_s, a_{s+1}[$$

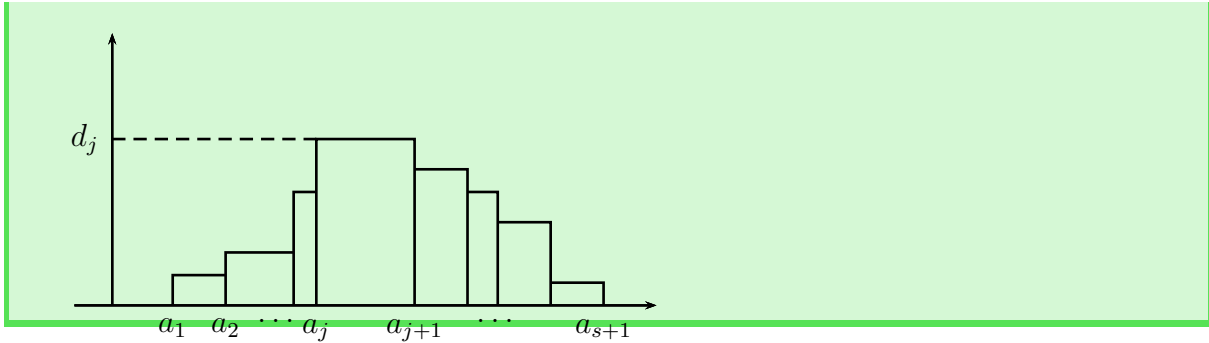
mit $a_1 < a_2 < \dots < a_{s+1}$ betrachten, in denen alle Daten liegen. Nun bilden wir über jedem Teilintervall $[a_i, a_{i+1}[$ ein Rechteck der Höhe d_i , wobei

$$d_i \cdot (a_{i+1} - a_i) = k_i \quad \text{für } 1 \leq i \leq s$$

ist mit

$$k_i = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{a_i \leq x_j < a_{i+1}\}}$$

Dadurch entsteht ein *Histogramm*. Die Gesamtfläche aller Rechtecke ist 1. Die einzelnen Rechteckflächen sind proportional zur relativen Häufigkeit des Auftretens der Daten. (Dabei ist $\mathbf{1}$ die Indikatorfunktion.)



Bemerkungen 5.6. Die Breite der Teilintervalle ist willkürlich; treten einige Daten nur selten auf, so sollten die entsprechenden Klassen (sprich Intervalle) weggelassen werden. Ist die Länge aller Teilintervalle gleich groß, so ist auch die Höhe der Rechtecke proportional zur sog. Klassenhäufigkeit.

Definition 5.7. Der *Box-Plot* wird häufig beim Vergleich verschiedener Stichproben verwendet. Er benutzt Quantile zur graphischen Darstellung von Lage und Streuung der Daten. Außerdem werden potentielle Ausreißer hervorgehoben.

Zur Anfertigung des Box-Plot wird ein senkrecht oder waagrecht Rechteck (eine *Kiste*) gezeichnet, die vom unteren bis zum oberen Quartil geht und beim Median unterteilt wird. Die Breite des Rechtecks wird meist nach ästhetischen Gesichtspunkten gewählt. Nach oben und unten bzw. links und rechts wird die Kiste durch zwei Stäbe verlängert, wobei der Endpunkt des nach oben aufgesetzten Stabes kleiner ist als das obere Quartil plus das 1,5-fache des Quartilsabstandes, also kleiner als

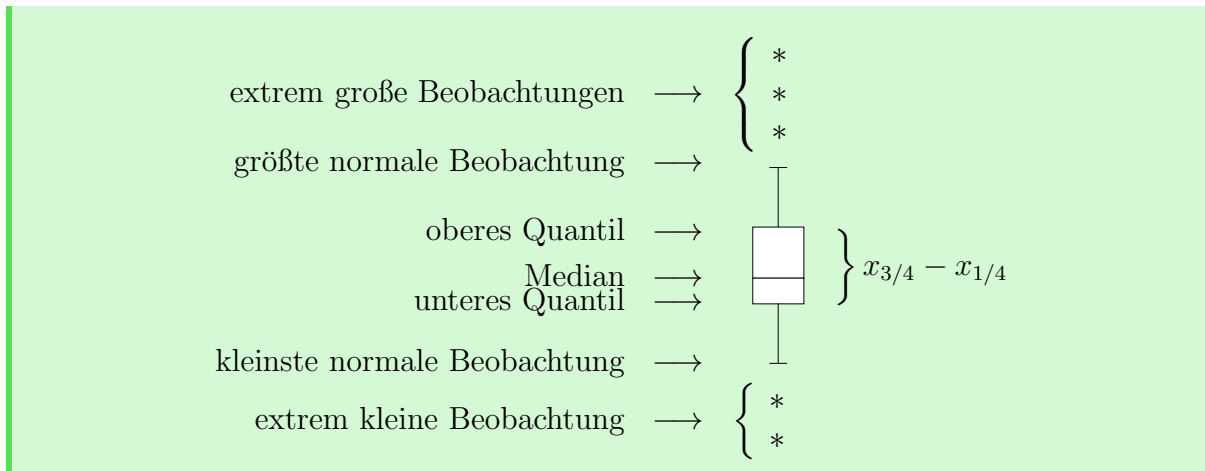
$$x_{3/4} + 1,5 \cdot (x_{3/4} - x_{1/4})$$

ist, die sog. *größte normale Beobachtung*. Der Endpunkt des nach unten aufgesetzten Stabes ist größer als

$$x_{1/4} - 1,5 \cdot (x_{3/4} - x_{1/4}),$$

die sog. *kleinste normale Beobachtung*. Extrem große Beobachtungen sind Daten, die oberhalb von $x_{3/4} + 1,5 \cdot (x_{3/4} - x_{1/4})$ liegen, extrem kleine Beobachtungen sind Daten, die unterhalb von $x_{1/4} - 1,5 \cdot (x_{3/4} - x_{1/4})$ liegen. Die sog. Ausreißer nach oben und unten werden durch einen Stern oder Punkt gekennzeichnet.

Wegen des Rechtecks in der Mitte des Diagramms spricht man auch von einem *Kisten-Diagramm*.



Eine derartige Präsentation von Daten findet man z.B. im Bericht "PISA 2000 - Die Länder der Bundesrepublik Deutschland im Vergleich", herausgegeben vom Deutschen PISA-Konsortium bei Leske und Budrich, Opladen 2002. Beim Box-Plot ist allerdings statt der aufgesetzten Stäbe das Rechteck nach oben und unten verlängert worden. Außerdem wurde auch bei dem Rechteck zwischen dem unteren und oberen Quartil um den Mittelwert das sog. Konfidenzintervall angegeben. Es wurden farblich getrennt nach unten noch das 10- und das 5 %-Perzentil und nach oben das 90- und das 95 %-Perzentil angefügt.

5.2 Schätzprobleme

Bisher sind wir davon ausgegangen, dass die Parameter der Wahrscheinlichkeitsverteilungen, beispielsweise μ und σ in der Normalverteilung, bekannt sind. Bei praktischen Anwendungen ist dies jedoch selten der Fall. Die schließende Statistik stellt Methoden bereit, mit deren Hilfe man aus Stichproben Informationen über die interessierenden Parameter gewinnen kann. Da naturgemäß eine Stichprobe nur einen kleinen Teil der Grundpopulation umfasst, birgt diese Information stets ein bestimmtes Maß an Unsicherheit in sich. Absolut zuverlässige Information wäre nur bei Erfassen der gesamten Grundpopulation zu erzielen. Es ist auch Aufgabe der schließenden Statistik, das Ausmaß an Unsicherheit zu quantifizieren.

Eine wesentliche Voraussetzung für die Anwendung statistischer Methoden ist die Zufälligkeit der Auswahl der Elemente in der Stichprobe: für jedes Individuum der Grundpopulation muss die Wahrscheinlichkeit, in die Stichprobe aufgenommen zu werden, gleich sein. Nur dadurch ist gewährleistet, dass die Stichprobe das Verhalten der Grundpopulation ausreichend widerspiegelt. Auf die Methoden der statistisch korrekten Entnahme einer Stichprobe, ein durchaus schwieriges Problem, kann hier nicht eingegangen werden.

Beispiel 5.8. a) Ein Teich enthalte eine unbekannte Anzahl N von Fischen, die geschätzt werden soll. Dazu werden W Fische gefangen, mit einem weißen Punkt markiert und wieder im Teich ausgesetzt. Nach einer gewissen Zeit werden bei einem zweiten Fang n Fische gefangen, und es wird die Anzahl x der mit einem weißen Punkt gekennzeichneten Fische ermittelt.

Eine plausible Schätzung $\hat{N}(x)$ für N ergibt sich aus folgender Überlegung: Wenn x nicht zu klein ist, sollte der Anteil x/n der markierten Fische am zweiten Fang ungefähr dem Anteil W/N der markierten Fische an der Gesamt-Population sein. Wenn also $\hat{N}(x)$ eine gute Schätzung für N ist, müsste gelten:

$$\frac{x}{n} \approx \frac{W}{\hat{N}(x)} \quad \text{oder} \quad \hat{N}(x) \approx W \cdot \frac{n}{x}.$$

Dabei wähle man als Näherung die zu der eventuell nichtganzzahligen Zahl $W \cdot \frac{n}{x}$ nächstgelegene ganze Zahl. Für kleine x ist diese Schätzung nicht sehr zuverlässig.

Diese heuristische Schätzung ergibt sich auch aus einem anderen Ansatz. Dazu betrachten wir Beispiel 2.28: Ist N die Anzahl aller Fische im Teich, W die Anzahl der markierten und $N - W$ die Anzahl der nicht-markierten Fische im Teich, so ist die Wahrscheinlichkeit $P_N(x)$, dass von n gefangenen Fischen beim zweiten Fang genau x markiert sind:

$$P_N(x) = \frac{\binom{W}{x} \binom{N-W}{n-x}}{\binom{N}{n}} \quad \text{mit } 0 \leq x \leq n.$$

Hier ist N ein unbekannter Wert, den es zu schätzen gilt. Wir nehmen als Schätzung N so, dass $P_N(x)$ für das beobachtete x maximal wird. Dazu betrachten wir

$$\frac{P_N(x)}{P_{N-1}(x)} = \frac{\binom{W}{x} \binom{N-W}{n-x} \binom{N-1}{n}}{\binom{N}{n} \binom{W}{x} \binom{N-W-1}{n-x}} = \frac{(N-W)(N-n)}{N(N-W-n+x)};$$

es ist $P_N(x) > P_{N-1}(x)$ genau dann, wenn $nW > Nx$ gilt; entsprechend folgt $P_N(x) < P_{N-1}(x)$ genau dann, wenn $nW < Nx$ gilt und $P_N(x) = P_{N-1}(x)$ genau für $nW = Nx$. Also ist $P_N(x)$ (als Funktion von N betrachtet) maximal für

$$\hat{N}(x) = \left\lceil \frac{nW}{x} \right\rceil.$$

Ist $\frac{nW}{x}$ keine ganze Zahl, so ist $\hat{N}(x)$ eindeutig bestimmt. Ist dagegen $\frac{nW}{x} \in \mathbb{Z}$, so sind $\frac{nW}{x}$ und $\frac{nW}{x} - 1$ Werte von N , für die $P_N(x)$ maximal ist.

- b) In n Bernoulli-Experimenten soll die Erfolgswahrscheinlichkeit p aus der Zahl x der Erfolge geschätzt werden (siehe Abschnitt 2.3.1, Binomialverteilung). Hierzu suchen wir $p \in [0, 1]$ so, dass

$$L_x(p) = \binom{n}{x} p^x (1-p)^{n-x}$$

maximal wird. Mit Hilfe der Produktregel erhalten wir

$$\frac{d}{dp} L_x(p) = \binom{n}{x} p^{x-1} (1-p)^{n-x-1} \cdot (x(1-p) - (n-x)p),$$

woraus sich aus der notwendigen Bedingung für das Vorliegen eines Extremwertes $\frac{d}{dp} L_x(p) = 0$ für den Fall $0 < p < 1$ die Nullstelle $\hat{p}(x) = \frac{x}{n}$ ergibt. Da die Ableitung

von L_x für $p < \hat{p}(x)$ positiv und für $p > \hat{p}(x)$ negativ ist, liegt in $\hat{p}(x)$ ein Maximum vor. Die relative Häufigkeit $\hat{p}(x) = \frac{x}{n}$ ist also eine Schätzung für die Erfolgswahrscheinlichkeit p .

Der Begriff der Stichprobe, der bereits in vorhergehenden Abschnitten gelegentlich (ohne nähere Begründung) verwendet wurde, ist ein Grundbegriff der mathematischen Statistik. In engem Zusammenhang damit steht der Begriff der Grundgesamtheit.

Definition 5.9. Sei (Ω, \mathcal{A}, p) ein Wahrscheinlichkeitsraum und X eine Zufallsgröße über diesem Wahrscheinlichkeitsraum, die ein gewisses Merkmal beschreibt. Dann nennt man X (in diesem Zusammenhang) eine *Grundgesamtheit*.

Zur Gewinnung von Informationen über die Wahrscheinlichkeitsverteilung von X wird ein Versuch n -mal unabhängig voneinander wiederholt. Damit erhält man als Realisierungen der Zufallsgröße X die Zahlen x_1, x_2, \dots, x_n . Betrachtet man nun die Zahl x_k (d.h. die Realisierung von X im k -ten Versuch) als Realisierung einer Zufallsgröße X_k , so sind die n Zufallsgrößen X_1, X_2, \dots, X_n unabhängig und identisch wie X verteilt (n Exemplare der Zufallsgröße X).

Definition 5.10. Sei X eine Zufallsgröße mit der Verteilungsfunktion F . Dann heißt der Zufallsvektor (X_1, \dots, X_n) dessen Koordinaten X_k unabhängig und identisch wie X verteilt sind, eine *mathematische Stichprobe* vom Umfang n aus der Grundgesamtheit X mit der Verteilungsfunktion F . Die Zufallsgrößen X_1, \dots, X_n heißen (in diesem Zusammenhang) *Stichprobenvariable*.

Jede Realisierung (x_1, \dots, x_n) des Zufallsvektors (X_1, \dots, X_n) heißt *konkrete Stichprobe* vom Umfang n aus der Grundgesamtheit X mit der Verteilungsfunktion F . Jede einzelne Realisierung x_k heißt *Element* der Stichprobe.

Während wir also wie bisher Zufallsgrößen (Merkmale, Eigenschaften eines Untersuchungsobjektes) mit Großbuchstaben X, Y, Z, \dots bezeichnen, werden Realisierungen dieser Zufallsgrößen (Merkmalswerte, Messwerte) mit den entsprechenden Kleinbuchstaben x, y, z, \dots bezeichnet.

Unter den Voraussetzungen von Def. 5.10 ist die Verteilungsfunktion des Zufallsvektors (X_1, \dots, X_n) gegeben durch

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{k=1}^n F(x_k), \quad (x_1, \dots, x_n) \in \mathbb{R}^n. \quad (5.1)$$

In der Praxis ist die Verteilungsfunktion oft nur bis auf einen noch zu bestimmenden Parameter ϑ bekannt; um dies hervorzuheben, schreiben wir dann $F_{X_1, \dots, X_n}(x_1, \dots, x_n; \vartheta)$, $F(x_k; \vartheta)$ etc.

Mittels der Schätztheorie soll bei einer **Punktschätzung** unter Verwendung einer mathematischen Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X eine zur Schätzung von ϑ geeignete Stichprobenfunktion $\hat{\theta}(X_1, \dots, X_n)$ angegeben werden. Liegt

dann eine konkrete Stichprobe (x_1, \dots, x_n) vor, so betrachtet man die Zahl $\hat{\theta}(x_1, \dots, x_n)$ einerseits als Realisierung der Zufallsgröße $\hat{\theta}(X_1, \dots, X_n)$, andererseits als Schätzwert für den Parameter ϑ . Die Stichprobenfunktion $\hat{\vartheta} = \hat{\theta}(X_1, \dots, X_n)$ heißt *Schätzung* (oder auch: *Schätzer*, *Schätzfunktion*) für den Parameter ϑ .

Zur konkreten Ermittlung von geeigneten Schätzern gibt es mehrere Verfahren - wir beschränken uns hier auf das bekannteste, nämlich die *Maximum-Likelihood-Methode*, die auf R.A. Fisher zurückgeht.

Seien X eine Grundgesamtheit, deren Verteilung von einem Parameter $\vartheta \in \Theta$ abhängen soll, und (X_1, \dots, X_n) eine mathematische Stichprobe aus X . Ist X stetig verteilt, so ergibt sich für die Verteilungsdichte des Zufallsvektors (X_1, \dots, X_n)

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \vartheta) = \prod_{k=1}^n f(x_k; \vartheta), \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

wobei $f(x_k; \vartheta)$ den Wert der Verteilungsdichte der Zufallsgröße X an der Stelle x_k in Abhängigkeit vom Parameter ϑ bedeutet. Ist X diskret verteilt, so ergibt sich für die Verteilung des Zufallsvektors (X_1, \dots, X_n) entsprechend

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{k=1}^n P(X_k = x_k; \vartheta).$$

Definition 5.11. Es sei $(x_1, \dots, x_n) \in \mathbb{R}^n$ eine konkrete Stichprobe vom Umfang n aus der Grundgesamtheit X , die von einem Parameter ϑ abhängt.

(i) Ist X stetig verteilt mit der Verteilungsdichte f , so heißt

$$L_x(\vartheta) = L(x_1, \dots, x_n; \vartheta) = \prod_{k=1}^n f(x_k; \vartheta), \quad (5.2)$$

Likelihood-Funktion der konkreten Stichprobe $x = (x_1, \dots, x_n)$.

(ii) Ist X diskret verteilt mit der Verteilung $P(X = x_i, \vartheta)$, $i = 1, 2, \dots, n$, so heißt

$$L_x(\vartheta) = L(x_1, \dots, x_n; \vartheta) = \prod_{k=1}^n P(X_k = x_k; \vartheta) \quad (5.3)$$

Likelihood-Funktion der konkreten Stichprobe $x = (x_1, \dots, x_n)$.

Definition 5.12. Nimmt die Likelihood-Funktion L_x mit $L_x(\vartheta) := L(x_1, \dots, x_n; \vartheta)$ in $\hat{\vartheta}(x)$ ein Maximum an, d.h. gilt

$$L_x(\hat{\vartheta}(x)) := \sup\{L_x(\vartheta) \mid \vartheta \in \Theta\},$$

so heißt $\hat{\vartheta}(x)$ eine *Maximum-Likelihood-Schätzung* von ϑ .

- Bemerkungen 5.13.** a) Der Parameter $\hat{\vartheta}(x)$ ist die „beste Erklärung“ für die Beobachtung $X_1 = x_1, \dots, X_n = x_n$, weil er dem beobachteten Ereignis im diskreten Fall die größte Wahrscheinlichkeit des Eintretens (bei stetigen Zufallsvariablen: Wahrscheinlichkeitsdichte) verleiht.
- b) In den meisten Fällen gibt es einen eindeutig bestimmten Maximum-Likelihood-Schätzer, und er ist auch ein „guter“ Schätzer.
- c) In vielen Fällen ist Θ ein reelles Intervall, und eine Maximum-Likelihood-Schätzung kann mit Hilfe der Differentiation gefunden werden. Häufig ist es zweckmäßig, statt der Funktion L_x die Funktion \mathcal{L}_x mit

$$\mathcal{L}_x = \ln L_x$$

zu betrachten. Wegen der strengen Monotonie der Logarithmus-Funktion haben (im Fall der Existenz) beide Funktionen das Maximum an der gleichen Stelle.

Beispiel 5.14. X sei Poisson-verteilt zum Parameter ϑ , also

$$P_\vartheta(x) = P(X = x; \vartheta) = \frac{\vartheta^x}{x!} e^{-\vartheta}, \quad x \in \mathbb{N}.$$

Dann ergibt sich die Likelihood-Funktion zu

$$L_x(\vartheta) = L(x_1, \dots, x_n; \vartheta) = \prod_{k=1}^n P_\vartheta(x_k), \quad x_1, \dots, x_n \in \mathbb{N}.$$

In diesem Beispiel gilt also

$$L_x(\vartheta) = \frac{\vartheta^{\sum_{k=1}^n x_k}}{\prod_{k=1}^n (x_k!)} e^{-n\vartheta}, \quad x_1, \dots, x_n \in \mathbb{N}.$$

Da beim Schätzproblem der gesuchte Parameter ϑ nicht bekannt ist, soll ein „guter“ Schätzer $\hat{\vartheta}$ zumindest im Mittel den richtigen Wert liefern. Das führt zu

Definition 5.15. Sei $\hat{\theta}_n := \hat{\theta}(X_1, \dots, X_n)$ eine Schätzfunktion für den Parameter θ .

- (a) Die Schätzfunktion $\hat{\theta}_n$ heißt erwartungstreue Schätzfunktion für θ (*engl: unbiased*), wenn gilt

$$E_\theta(\hat{\theta}(X_1, \dots, X_n)) = \theta \text{ für alle } \vartheta \in \Theta.$$

- (b) Man nennt

$$B_\theta(\hat{\theta}_n) = E_\theta(\hat{\theta}(X_1, \dots, X_n)) - \theta$$

den systematischen Fehler (*engl. Bias*).

Bemerkung 5.16. a) Definition 5.15 besagt, dass der Schwerpunkt der Verteilung des zufälligen Schätzwertes gerade der zu schätzende Parameter ϑ ist, und zwar unabhängig davon, welches ϑ der wahre Parameter ist.

b) Für erwartungstreue Schätzer ist der Bias = 0.

Satz 5.17. Es seien X_1, \dots, X_n unabhängig mit $E_{\vartheta}(X_i) = \mu$ und $\text{Var}_{\vartheta}(X_i) = \sigma^2$ für $1 \leq i \leq n$ und alle $\vartheta \in \Theta$; dann gilt:

a) Der Mittelwert $\bar{X} (= \bar{X}(X_1, \dots, X_n)) = \frac{1}{n} \sum_{i=1}^n X_i$ ist ein erwartungstreuer Schätzer für μ .

b) Die empirische Varianz $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ist ein erwartungstreuer Schätzer für σ^2 .

c) $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ ist kein erwartungstreuer Schätzer für σ^2 .

Beweis. Es gilt

$$E_{\vartheta}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E_{\vartheta}(X_i) = \mu$$

und

$$E_{\vartheta}(S^2) = \frac{1}{n-1} \sum_{i=1}^n E_{\vartheta}((X_i - \bar{X})^2)$$

mit

$$\begin{aligned} E_{\vartheta}((X_i - \bar{X})^2) &= E_{\vartheta}([(X_i - \mu) - (\bar{X} - \mu)]^2) \\ &= E_{\vartheta}((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2) \\ &= E_{\vartheta}((X_i - \mu)^2) - \frac{2}{n} \sum_{j=1}^n E_{\vartheta}((X_i - \mu)(X_j - \mu)) \\ &\quad + E_{\vartheta}((\bar{X} - \mu)^2). \end{aligned}$$

Wegen der Unabhängigkeit der X_1, \dots, X_n ist

$$E_{\vartheta}((\bar{X} - \mu)^2) = \text{Var}_{\vartheta}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}_{\vartheta}(X_i) = \frac{\sigma^2}{n}$$

und

$$E_{\vartheta}((X_i - \mu)(X_j - \mu)) = 0 \quad \text{für } i \neq j.$$

Damit folgt

$$E_{\vartheta}(S^2) = \frac{1}{n-1} \sum_{i=1}^n \left(\sigma^2 - \frac{2}{n} \sigma^2 + \frac{\sigma^2}{n} \right) = \sigma^2.$$

Also gelten b) und c). □

Beispiel 5.18. a) Sind X_1, \dots, X_n unabhängig und $N(\mu, \sigma^2)$ -verteilt, so ist

$$L_x(\mu, \sigma) = \left(\frac{1}{\sigma \cdot \sqrt{2\pi}} \right)^n \exp \left(- \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

zu maximieren. Gemäß Bemerkung 5.13 c) gehen wir zum Logarithmus über und erhalten

$$\mathcal{L}_x(\mu, \sigma) = -n \ln(\sigma \cdot \sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

als zu maximierende Funktion für die Parameter μ und σ . Eigentlich sind hier 3 Fälle zu unterscheiden:

- (i) $\mu = \mu_0$ ist bekannt, aber σ nicht. Wir setzen diesen Wert für μ ein und maximieren bzgl. σ . Das notwendige Kriterium für das Vorliegen von Extremwerten liefert:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$$

An dieser Stelle liegt ein Maximum vor.

- (ii) $\sigma^2 = \sigma_0^2$ ist bekannt, aber μ nicht. Durch Differentiation nach μ erhalten wir aus der notwendigen Bedingung für Extremwerte

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

- (iii) μ und σ^2 sind unbekannt. Hier könnte man Kriterien für Extremwerte von Funktionen mehrerer Veränderlicher anwenden (vgl. Math. 2 für Inf.). Man kann aber auch einfacher verfahren: Der obige Ausdruck wird maximal, wenn $\sum_{i=1}^n (x_i - \mu)^2$ minimal wird; das ist der Fall, wenn

$$\sum_{i=1}^n (x_i - \mu) = 0 \quad \text{d.h.} \quad \hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

ist. Dann erhalten wir

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

- b) X_1, \dots, X_n seien unabhängig und P_λ -verteilt mit $\lambda > 0$, d.h. wir haben die Situation aus Beispiel 5.14 mit $\vartheta = \lambda$. Gesucht ist ein Schätzer für den Parameter λ . Wir maximieren die Funktion

$$\mathcal{L}_x(\lambda) = \ln \left(\frac{\lambda^{\sum_{i=1}^n x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^n (x_i)!} \right) = -n\lambda + \ln \lambda \cdot \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!) .$$

Das notwendige Kriterium liefert

$$\frac{d}{d\lambda} \mathcal{L}_x(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i = 0$$

genau für

$$\lambda = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Wegen

$$\frac{d^2}{d\lambda^2} \mathcal{L}_x(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^n x_i < 0$$

liegt an der Stelle $\lambda = \bar{x}$ ein Maximum vor.

5.3 Konfidenzintervalle

Bisher haben wir nur Punktschätzer betrachtet: die Stichprobe wurde benutzt, um einen Schätzwert für einen Verteilungsparameter zu berechnen. Ein **Intervallschätzer** dagegen berechnet aus der Stichprobe ein Intervall $[\hat{\vartheta}_u, \hat{\vartheta}_o]$, welches den Zielparameter ϑ mit einer vorgegebenen Wahrscheinlichkeit enthält:

$$P_{\vartheta}(\hat{\vartheta}_u \leq \vartheta \leq \hat{\vartheta}_o) = 1 - \alpha.$$

Die Intervallgrenzen $\hat{\vartheta}_u, \hat{\vartheta}_o$ hängen neben α auch von der jeweiligen Stichprobe ab und variieren daher in zufälliger Weise von Stichprobe zu Stichprobe.

Schätzt man z.B. die Erfolgswahrscheinlichkeit p eines Bernoulli-Experiments aus den Ergebnissen einer Stichprobe $X = (X_1, \dots, X_n)$, so geht man davon aus, dass die Schätzung umso näher an p liegt, je größer die Stichprobe ist. Ziel ist die Angabe eines Intervalls (das von der Beobachtung x abhängt) derart, dass die Wahrscheinlichkeit dafür, dass p in dem Intervall liegt, einen möglichst großen Wert hat, etwa 0,95.

Etwas allgemeiner haben wir die folgende Situation: Sei X_1, \dots, X_n die mathematische Stichprobe (die X_i 's sind stochastisch unabhängige, identisch verteilte Zufallsvariable!). Es sei θ der zu schätzende Parameter. Bei der Methode "Konfidenzintervall" benötigen wir *zwei Schätzfunktionen*

$$\hat{\theta}_u = \hat{\theta}_u(X_1, \dots, X_n) \text{ und } \hat{\theta}_o = \hat{\theta}_o(X_1, \dots, X_n).$$

Definition 5.19. Ist θ ein zu schätzender Parameter und sind $\hat{\theta}_u$ und $\hat{\theta}_o$ Schätzfunktionen mit der Eigenschaft

$$P_{\theta}(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) = 1 - \alpha \text{ (oder auch } \geq 1 - \alpha)$$

- so liefert jede konkrete Stichprobe $x = (x_1, \dots, x_n)$ ein *Konfidenzintervall* $[\hat{\theta}_u(x), \hat{\theta}_o(x)]$.
- Die Zahl $1 - \alpha$ heißt *Konfidenzniveau*. Die Zahl α heißt *Irrtumswahrscheinlichkeit*.

Häufig ist man in der Situation, die *unteren Konfidenz- oder Vertrauensgrenze* und die *obere Konfidenz- oder Vertrauensgrenze* für einen Parameter θ einzeln zu bestimmen:

- $\hat{\theta}_u$ heißt *untere Konfidenzgrenze* für den Parameter θ zum Konfidenzniveau $1 - \beta$, wenn

$$P_\theta(\hat{\theta}_u \leq \theta) \geq 1 - \beta.$$

- $\hat{\theta}_o$ heißt *obere Konfidenzgrenze* für den Parameter θ zum Konfidenzniveau $1 - \beta$, wenn

$$P_\theta(\theta \leq \hat{\theta}_o) \geq 1 - \beta.$$

Bemerkung: Aus $P_\theta(\hat{\theta}_u \leq \theta) \geq 1 - \frac{\alpha}{2}$ und $P_\theta(\theta \leq \hat{\theta}_o) \geq 1 - \frac{\alpha}{2}$ folgt

$$P_\theta(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) \geq 1 - \alpha.$$

Die Notation P_θ bedeutet, dass wir die Wahrscheinlichkeit unter der Annahme, dass der Parameter θ ist, berechnen.

Beispiel 5.20. In einer Bernoulli-Kette S_n vom Umfang n (das bedeutet: n identische und unabhängige Bernoulli-Experimente) seien k Treffer aufgetreten. Wir wollen eine untere und eine obere Vertrauensgrenze $\hat{p}_u(k)$ bzw. $\hat{p}_o(k)$ für die Erfolgswahrscheinlichkeit $p \in [0, 1]$ bestimmen. Die Idee besteht darin, solche p 's auszuschließen, für die die Wahrscheinlichkeit für höchstens k bzw. mindestens k Treffer in n Versuchen hinreichend klein wird.

Im Fall $k = 0$ ist $\hat{p}_u(0) = 0$ zu setzen. Setzen wir $\hat{p}_o(0) = 1 - \beta^{1/n}$, so gilt für alle $p \geq \hat{p}_o(0)$

$$(1 - p)^n = \binom{n}{0} \cdot p^0 \cdot (1 - p)^{n-0} \leq \beta;$$

der Term $(1 - p)^n$ gibt für das unbekannte p die Wahrscheinlichkeit an, dass in n Versuchen kein Treffer erzielt wird.

Im Fall $k = n$ ergibt sich mit entsprechenden Überlegungen $\hat{p}_o(n) = 1$ und $\hat{p}_u(n) = \beta^{1/n}$. Im Fall $1 \leq k \leq n - 1$ ergibt sich $\hat{p}_o(k)$ als eindeutig bestimmte Lösung p der Gleichung

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

und $\hat{p}_u(k)$ als eindeutig bestimmte Lösung p der Gleichung

$$\sum_{j=k}^n \binom{n}{j} \cdot p^j \cdot (1 - p)^{n-j} = \beta.$$

Die Lösung der obigen Gleichung zu bestimmen ist recht kompliziert. Meist benutzt man dazu numerische Verfahren.

Deswegen wollen wir einen anderen Weg einschlagen, um ein Konfidenzintervall für p zu bestimmen. Die Strategie besteht darin, die $B(n, p)$ -Verteilung mithilfe des Satzes von

Moivre-Laplace mit einer geeigneten Normalverteilung zu approximieren. Damit können wir dann Näherungen für die Konfidenzgrenzen $\hat{p}_u(k)$ und $\hat{p}_o(k)$ berechnen. Die Voraussetzung für die Anwendbarkeit des Satzes von Moivre-Laplace ist, dass der Umfang n der Stichprobe genügend groß ist. Wegen

$$P_p(S_n \leq k) = P_p\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{k - np}{\sqrt{np(1-p)}}\right) \approx \phi\left(\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

erhalten wir eine Näherung für $\hat{p}_o(k)$ aus der Gleichung

$$\phi\left(\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) = \beta$$

durch Auflösen nach p , d.h. aus der Gleichung

$$\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}} = \phi^{-1}(\beta).$$

Entsprechend erhalten wir eine Näherung für $\hat{p}_u(k)$ durch Auflösen der Gleichung

$$\frac{k - np - \frac{1}{2}}{\sqrt{np(1-p)}} = \phi^{-1}(1 - \beta)$$

nach p . Setzen wir $c := \phi^{-1}(1 - \beta)$, so gilt aus Symmetriegründen $\phi^{-1}(\beta) = -c$. Damit ergeben sich als Näherungen

$$\hat{p}_o(k) \approx \frac{k + \frac{1}{2} + \frac{c^2}{2} + c \cdot \sqrt{k + \frac{1}{2} - \frac{1}{n}(k + \frac{1}{2})^2 + \frac{c^2}{4}}}{n + c^2}$$

und

$$\hat{p}_u(k) \approx \frac{k - \frac{1}{2} + \frac{c^2}{2} - c \cdot \sqrt{k - \frac{1}{2} - \frac{1}{n}(k - \frac{1}{2})^2 + \frac{c^2}{4}}}{n + c^2}.$$

Die Werte c können in Abhängigkeit von β aus der Tabelle für die Standardnormalverteilung (Seite 58) abgelesen werden. Oft findet man die benötigten Angaben auch in speziellen Tabellen mit den wichtigsten Quantilen der Standardnormalverteilung:

$1 - \beta$	0.900	0.950	0.975	0.990	0.995
$c = \phi^{-1}(1 - \beta)$	1.282	1.645	1.960	2.326	2.576

Damit erhält man für den Fall $\beta = 0.025$, also $c = 1.96$, $n = 50$ und $k = 10$ die Näherungen $\hat{p}_o(10) \approx 0.341$ und $\hat{p}_u(10) \approx 0.105$ im Vergleich zu den exakten Werten $\hat{p}_o = 0.337$ und $\hat{p}_u = 0.100$.

Beispiel 5.21. Seien X_1, \dots, X_n stochastisch unabhängige und identisch verteilte Zufallsvariable $X_i \sim N(\mu, \sigma^2)$. Der Parameter σ^2 (die Varianz) sei fest vorgegeben und

bekannt, μ sei zu schätzen. Bekanntlich ist dann das arithmetische Mittel \bar{X} normal verteilt mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$. (Dies folgt aus den Sätzen 2.12 und 2.16.) Also folgt

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Bestimmt wird nun $u \in \mathbb{R}$ so, dass

$$P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \leq u\right) = 1 - \alpha.$$

Das Prinzip hierfür wird anhand der Dichte der $N(0, 1)$ -Verteilung erläutert.

$u = u_{1-\frac{\alpha}{2}} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ ist dabei das $(1 - \frac{\alpha}{2})$ -Quantil der $N(0, 1)$ -Verteilung (siehe Tabelle unten). Auflösen nach μ ergibt die folgende Gleichung

$$\begin{aligned} P\left(|\bar{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right) &= P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right) \\ &= 1 - \alpha. \end{aligned}$$

Also ist

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right]$$

ein $(1 - \alpha)$ -Konfidenzintervall für μ . Wie erwartet fällt die Länge des Konfidenzintervalls mit wachsendem n und wächst mit wachsendem Niveau $1 - \alpha$.

α	0,1	0,05	0,025	0,01	0,005	0,0025
$1 - \alpha$	0,9	0,95	0,975	0,99	0,995	0,9975
$u_{1-\alpha}$	1,282	1,645	1,960	2,326	2,576	2,807

Einige Quantile der Standardnormalverteilung sind in obiger Tabelle angegeben. Ausführliche Tabellen finden sich in der Literatur.

Satz 5.22. Seien X_1, \dots, X_n stochastisch unabhängig und identisch verteilt, $X_i \sim N(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ und $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Es gilt

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \quad \text{und} \quad \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$$

wobei t_{n-1} bzw. χ_{n-1}^2 die t -Verteilung bzw. die χ^2 -Verteilung mit $n-1$ Freiheitsgraden bezeichnen. (Siehe Definition 3.15 und 3.16).

Es gilt sogar, dass \bar{X} und S^2 stochastisch unabhängig sind. Der Beweis dieser Aussage und des Satzes findet sich in Casella, Berger auf Seite 226.

Man beachte folgende Extremfälle für den Parameter $m = n - 1$ und die t -Verteilung:

- Für $m = 1$ ergibt sich die Dichte der *Cauchy-Verteilung*.

- Für $Y_m \sim t_m$, $m \in \mathbb{N}$, gilt $Y_m \stackrel{as}{\sim} N(0, 1)$ mit $m \rightarrow \infty$. t_m -Verteilungen konvergieren also mit $m \rightarrow \infty$ gegen die Standardnormalverteilung.

Beispiel 5.23. Seien X_1, \dots, X_n i.i.d. $\sim N(\mu, \sigma^2)$. Bestimmt wird ein Konfidenzintervall für μ bei unbekanntem σ^2 . Die Idee ist, in Beispiel 5.21 σ^2 durch den erwartungstreuen Schätzer S^2 zu ersetzen und dann die Quantile der t -Verteilung zu benutzen.

Bestimmt wird $v \in \mathbb{R}$ so, dass

$$P\left(\left|\frac{\bar{X} - \mu}{S/\sqrt{n}}\right| \leq v\right) = 1 - \alpha.$$

Analog zu Beispiel 5.21 wird $v = v_{1-\frac{\alpha}{2}}$ als das $(1-\frac{\alpha}{2})$ -Quantil der t_{n-1} -Verteilung gewählt. Auflösen obiger Gleichung nach μ liefert

$$P\left(|\bar{X} - \mu| \leq v_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) = P\left(\bar{X} - v_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + v_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right).$$

Also ist

$$\left[\bar{X} - v_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + v_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$

ein $1 - \alpha$ -Konfidenzintervall für μ bei ungekannter Varianz σ^2 .

Beispiel 5.24. Von einem Automaten werden Produkte hergestellt, bei denen ein Merkmal (Länge, Gewicht, Festigkeit usw.) untersucht wird. Die Abweichungen vom vorgegebenen Nennmaß (etwa als μ_W mm gegeben) kann man als Realisierungen einer normalverteilten Zufallsgröße X auffassen. Der Erwartungswert $E(X)$ ist von der Einstellung des Automaten abhängig und daher nicht bekannt. Aus Prüfungen der Funktionsgenauigkeit des Automaten ist die empirische Varianz $S^2 = 225$ bekannt. Für den Erwartungswert $\mu = E(X)$ ist eine konkrete Konfidenzschätzung mit dem Konfidenzniveau $1 - \alpha = 0,95$ zu ermitteln.

Aus einer konkreten Stichprobe vom Umfang $n = 25$ sei das arithmetische Mittel $\bar{x}_n = 48$ bekannt. Für eine Irrtumswahrscheinlichkeit $\alpha = 0,05$ folgt aus obiger Tabelle $u_{1-\frac{\alpha}{2}} = u_{0,975} \approx 1,96$. Damit erhält man als konkretes Konfidenzintervall für den Parameter μ :

$$\bar{x}_n - u_{1-\alpha/2} \frac{S}{\sqrt{n}} = 48 - 1,96 \frac{15}{5} < \mu < 48 + 1,96 \frac{15}{5}$$

und damit

$$42,12 < \mu < 53,88.$$

Bei Verkleinerung der Irrtumswahrscheinlichkeit vergrößert sich naturgemäß das Konfidenzintervall. Für $\alpha = 0,01$ folgt aus obiger Tabelle der Wert $u_{1-\frac{\alpha}{2}} = u_{0,995} \approx 2,576$ und damit das konkrete Konfidenzintervall für den Parameter μ :

$$40,26 < \mu < 55,74.$$

5.4 Hypothesentests

Wir beginnen mit einem in der Literatur bekannten Beispiel (vgl. z.B. N. Henze, S. 234 ff. und U. Krengel, S. 92 ff.):

Beispiel 5.25 ("tea-testing lady"). Eine englische Lady trinkt regelmäßig ihren 5-Uhr-Tee mit Milch. Eines Tages überrascht sie ihre Teerunde mit der Behauptung, dass sie am Geschmack des Tees feststellen könne, ob zuerst die Milch und anschließend der Tee eingegossen wurde oder umgekehrt. Sie sei zwar nicht unfehlbar, würde aber in der Mehrzahl der Fälle die richtige Reihenfolge feststellen.

Um der Lady die Möglichkeit zu geben, ihre Fähigkeit unter Beweis zu stellen, wird folgendes Verfahren vorgeschlagen: der Lady werden n -mal hintereinander jeweils zwei Tassen Tee gereicht, jeweils eine vom Typ "Milch vor Tee" und eine vom Typ "Tee vor Milch". In welcher Reihenfolge die Lady die beiden Tees probiert, wird durch den Wurf einer Münze festgelegt. Die Pausen zwischen den Geschmacksproben sollen so groß sein, dass die Lady unbeeinflusst von der vorhergehenden Probe urteilen kann.

Wir können diese Versuchsanordnung interpretieren als n unabhängige Versuche, wobei mit unbekannter Trefferwahrscheinlichkeit p die richtige Reihenfolge genannt wird. Bezeichnen wir mit S_n die Anzahl der richtig angegebenen Reihenfolgen, so wären wir im Fall $n = 20$ bei 10 richtigen Reihenfolgen nicht von den Fähigkeiten der Lady überzeugt, denn durch "blindes" Raten könnte sich auch diese Anzahl ergeben. Wie groß muss die Anzahl der richtig angegebenen Reihenfolgen sein, damit man der Lady die Fähigkeit attestieren kann, die richtige Reihenfolge beim Eingießen durch eine Geschmacksprobe zu erkennen. Sind wir bei 13 " Treffern" von den Fähigkeiten der Lady überzeugt oder erst bei 14 Treffern. Wie groß ist die Wahrscheinlichkeit durch bloßes Raten 13 "Treffer" zu erzielen?

Definition 5.26. Wir zerlegen den Parameterbereich Θ in zwei nichtleere disjunkte Teile

$$\Theta = \Theta_0 \cup \Theta_1 .$$

Ein *statistischer Test* oder *Hypothesentest* ist eine Entscheidungsregel, die innerhalb des vorgegebenen Modell-Rahmens für jede mögliche Stichprobe x festlegt, ob man sich für die

$$\text{Nullhypothese } H_0 : \text{ es gilt } \vartheta \in \Theta_0$$

oder für die

$$\text{Gegenhypothese (Alternative) } H_1 : \text{ es gilt } \vartheta \in \Theta_1$$

entscheidet. Die Zerlegung von Θ impliziert eine Zerlegung

$$\mathcal{K}_0 \cup \mathcal{K}_1$$

des Stichprobenraums und damit folgende Entscheidungsregel:

$$\text{"Ist } x \in \mathcal{K}_0, \text{ so entscheide für } H_0."$$

"Ist $x \in \mathcal{K}_1$, so entscheide für H_1 ."

Man sagt auch: "Zu testen ist die Hypothese H_0 gegen die Alternative H_1 ". \mathcal{K}_0 heißt *Annahmehereich* des Tests und \mathcal{K}_1 heißt *Verwerfungsbereich* oder auch *kritischer Bereich*. Die Hypothese H_0 wird häufig auch *Nullhypothese* genannt.

Beispiel (Fortsetzung von Beispiel 5.25). Bei der "tea-testing lady" bietet sich als Θ das Intervall $[\frac{1}{2}, 1]$ an mit der Zerlegung in $\Theta_0 = \{\frac{1}{2}\}$ und $\Theta_1 =]\frac{1}{2}, 1]$. Die Hypothese H_0 lautet: "Die Lady hat keine besondere Gabe, die Eingießreihenfolge am Geschmack zu erkennen." Überprüfen wir die Fähigkeiten mit 20 Tassenpaaren und unterteilen wir den Stichprobenraum in den kritischen Bereich

$$\mathcal{K}_1 = \{14, 15, \dots, 20\}$$

und den Annahmehereich $\mathcal{K}_0 = \{0, 1, \dots, 13\}$ für die Hypothese H_0 , so ist die Wahrscheinlichkeit, dass die Lady durch bloßes Raten mindestens 14-mal die richtige Reihenfolge trifft,

$$p_{\frac{1}{2}}(S_{20} \geq 14) = \sum_{j=14}^{20} \binom{20}{j} \left(\frac{1}{2}\right)^j \cdot \left(1 - \frac{1}{2}\right)^{20-j} \approx 0,058.$$

Wenn diese Wahrscheinlichkeit, durch blindes Raten mindestens 14 Treffer zu erzielen, noch zu groß ist, der kann statt der 14 auch mindestens 15 Treffer fordern; dann gilt

$$p_{\frac{1}{2}}(S_{20} \geq 15) \approx 0,0207.$$

Bei jedem Test muss man den Annahmehereich und damit auch den Verwerfungsbereich festlegen. Nun gibt es bei der Festlegung mit z.B. mindestens 14 Treffern folgende Möglichkeiten:

- Die Lady besitzt nicht die o.g. Fähigkeit und sie errät auch nur bei 10 Tassenpaaren die richtige Reihenfolge. Man entscheidet sich für die Nullhypothese.
- Die Lady besitzt nicht die o.g. Fähigkeit, durch "Zufall" kommt sie aber auf 15 Treffer. Man entscheidet sich für die Gegenhypothese.
- Die Lady besitzt die o.g. Fähigkeit und sie hat auch 16 Treffer. Man entscheidet sich für die Gegenhypothese.
- Die Lady besitzt die o.g. Fähigkeit, sie hat aber nur bei 13 Tassenpaaren recht. Man entscheidet sich für die Nullhypothese.

In den Fällen b) und d) macht man einen Fehler.

Definition 5.27. Wir übernehmen die Bezeichnungen aus Definition 5.26. Gilt (in Wirklichkeit) $\vartheta \in \Theta_0$ und man entscheidet sich für die Gegenhypothese, so spricht man von einem *Fehler erster Art*. Gilt dagegen $\vartheta \in \Theta_1$ und man entscheidet sich für die Nullhypothese, so spricht man von einem *Fehler zweiter Art*. Man kann dies folgendermaßen zusammenfassen:

		Wirklichkeit	
		$H_0 : \vartheta \in \Theta_0$	$H_1 : \vartheta \in \Theta_1$
Entscheidung	für H_0	richtige Entscheidung	Fehler 2. Art
	für H_1	Fehler 1. Art	richtige Entscheidung

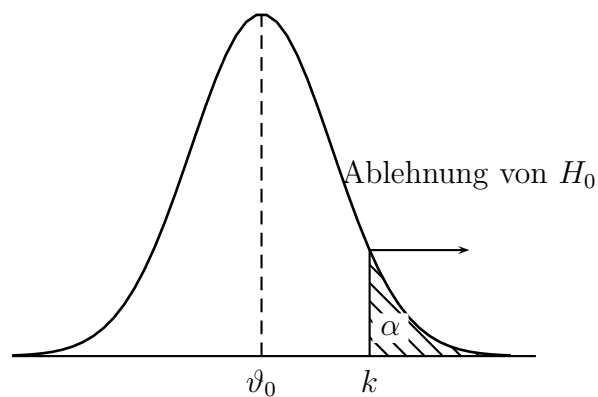
Um die Wahrscheinlichkeit für eine falsche Entscheidung möglichst klein zu halten, ist eine sog. *Gütefunktion* g mit kleinen Werten auf Θ_0 und großen Werten auf Θ_1 wünschenswert. Wir definieren $g : \Theta \rightarrow [0, 1]$ durch

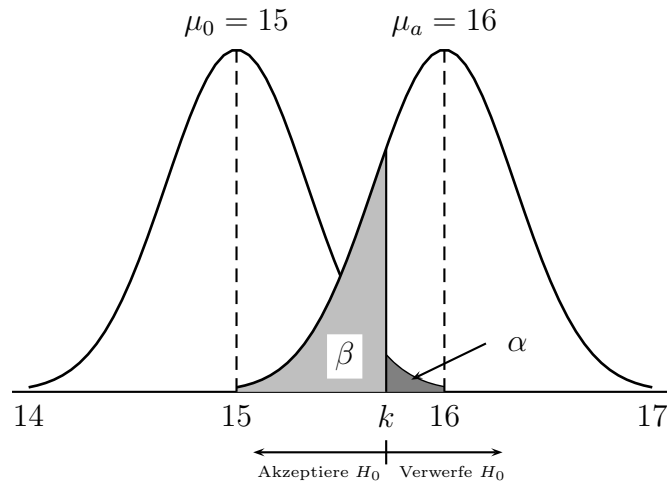
$$g(\vartheta) := p_\vartheta(X \in \mathcal{K}_1), \quad \vartheta \in \Theta.$$

g ordnet jedem ϑ die sog. *Verwerfungswahrscheinlichkeit der Hypothese H_0 unter p_ϑ* zu. Man gibt eine obere Schranke $\alpha \in]0, 1[$ für die Wahrscheinlichkeit des Fehlers erster Art vor und legt \mathcal{K}_1 so fest, dass

$$g(\vartheta) \leq \alpha \quad \text{für alle } \vartheta \in \Theta_0$$

gilt. Ein solcher Test heißt (*Signifikanz-*)*Test zum (Signifikanz-)Niveau α* . Dabei sind für α Werte aus dem Intervall $[0.01, 0.1]$ üblich.





Beispiel 5.28. Ein Produzent von Überraschungseiern versichert, dass in mindestens 14 % der Eier Figuren aus einem beliebigen Fantasy-Abenteuer stecken. Eine Verbraucherorganisation ist misstrauisch und möchte diese Aussage überprüfen. Aus diesem Grund untersucht sie 1000 zufällig ausgewählte Überraschungseier, findet darin aber nur 130 Fantasy-Figuren. Genügt dieses Ergebnis, um die Behauptung des Herstellers zu widerlegen, oder sollte die Verbraucherorganisation lieber schweigen, da eine gewisse Abweichung vom Erwartungswert 140 in der zufällig ausgewählten Stichprobe immer möglich ist? Es sei p der wahre, aber unbekannte Anteil der Eier mit Fantasy-Figuren. Die Hypothese des Herstellers lautet dann

$$H_0 : p \geq 0.14 ,$$

die Hypothese der Verbraucherorganisation

$$H_1 : p < 0.14 .$$

Wir bezeichnen mit X die Zufallsvariable, die die Anzahl der Figuren in der Stichprobe vom Umfang $n = 1000$ zählt. Wir können davon ausgehen, dass X Binomial-verteilt ist. Wenn H_0 richtig ist, wird X nicht so klein ausfallen; also gibt es eine noch zu bestimmende Zahl c derart, dass $X > c$ ist. Wenn $X \leq c$ ausfällt, ist H_0 wahrscheinlich nicht richtig. Die Wertemenge von X zerfällt also in

$$\mathcal{K}_0 = \{c + 1, \dots, 1000\}, \quad \mathcal{K}_1 = \{0, \dots, c\}.$$

Der Wert c ist so zu wählen, dass die Wahrscheinlichkeit α , die richtige Hypothese H_0 zu verwerfen, möglichst klein ist, d.h.

$$p(X \leq c \text{ und es ist } H_0 \text{ richtig}) \leq \alpha .$$

Wir haben bei Vorgabe der sogenannten *Irrtums-Wahrscheinlichkeit* α und des daraus bestimmten c insgesamt vier Entscheidungssituationen:

- a) H_0 trifft zu, und es ist $X > c$; wir entscheiden uns für H_0 .
- b) H_0 trifft zu, und es ist $X \leq c$; wir entscheiden uns für H_1 .

c) H_1 trifft zu, und es ist $X > c$; wir entscheiden uns für H_0 .

d) H_1 trifft zu, und es ist $X \leq c$; wir entscheiden uns für H_1 .

Ersetzen wir die Binomial-Verteilung durch die Gaußsche Glockenkurve (Satz 4.3 von de Moivre-Laplace), so erhalten wir die Größe c z.B. für $\alpha = 0.05$ aus der Beziehung

$$\begin{aligned} 0.05 &= P(X \leq c \text{ und es ist } H_0 \text{ richtig}) \\ &\approx \phi\left(\frac{c - np}{\sqrt{np(1-p)}}\right) = \phi\left(\frac{c - 140}{\sqrt{120.4}}\right), \end{aligned}$$

wobei wir ohne Korrekturterm arbeiten. Wegen der Symmetrie der Normalverteilung erhalten wir daraus:

$$0.95 = 1 - 0.05 = 1 - \alpha = \phi\left(\frac{140 - c}{\sqrt{120.4}}\right)$$

und damit aus einer Tabelle

$$\frac{140 - c}{\sqrt{120.4}} \approx 1.65,$$

woraus

$$c \approx 121.9$$

folgt. Würden wir mit Korrekturterm rechnen, d.h. mit

$$\phi\left(\frac{c - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) = \phi\left(\frac{c - 139.5}{\sqrt{120.4}}\right),$$

so ergäbe sich statt 121.9 die Zahl 121.4, was zu derselben nächstgelegenen, kleineren ganzen Zahl führt. Also erhalten wir die sog. *Testvorschrift*, die Hypothese H_0 erst dann abzulehnen, wenn in der Stichprobe von 1000 Eiern die Anzahl der Figuren 121 oder weniger beträgt. Es wurden aber 130 Figuren gefunden. Also ist die Behauptung des Herstellers durch dieses Testergebnis nicht zu beanstanden. Berechnen wir für $c = 121$ die Wahrscheinlichkeit für den Fehler 1. Art, so ergibt sich

$$P(X \leq 121 \text{ und } H_0 \text{ ist richtig}) \approx \phi\left(\frac{121 - 140}{\sqrt{120.4}}\right) \approx 0.0418$$

bzw.

$$P(X \leq 121 \text{ und } H_0 \text{ ist richtig}) \approx \phi\left(\frac{121 - 139.5}{\sqrt{120.4}}\right) \approx 0.0465.$$

(Wir können hier nicht exakt den Wert 0.05 erzielen, da c nur die Menge der natürlichen Zahlen durchläuft.) Das Ergebnis ist folgendermaßen zu interpretieren: Wenn wir viele Stichproben im Umfang von jeweils 1000 Überraschungseiern ziehen würden, in denen höchstens 121 Fantasy-Figuren stecken, so würden wir in 4.18 % (bzw. 4.65 %) der gezogenen Stichproben dem Hersteller zu Unrecht vorwerfen, dass weniger als 14 % Fantasy-Figuren in seinem Produkt seien.

Zum Schluss ein paar Bemerkungen, die man bei der Durchführung eines Hypothesentests immer berücksichtigen soll:

Bemerkungen 5.29. Man sollte wissen, dass

- a) Hypothesen und Gegenhypothesen nie bewiesen werden können,
- b) das Nichtverwerfen einer Hypothese H_0 im Allgemeinen nur bedeutet, dass die vorliegende Datenbasis zu gering ist, um einen signifikanten Widerspruch zu H_0 herbeizuführen.
- c) Hypothesen, die anhand von Daten gebildet werden, nie anhand derselben Daten getestet werden dürfen.

Grundlegendes aus der Analysis

Wir fassen kurz einige grundlegende Begriffe und Aussagen aus der Analysis zusammen, die wir im Laufe der Vorlesung benötigen werden.

1. Abbildungen ([4, Abschnitt 0.3])

Eine *Abbildung* oder *Funktion* ist eine Zuordnungsvorschrift zwischen zwei Mengen X und Y , bei der jedem Element $x \in X$ genau ein Element $y \in Y$ zugeordnet wird. Wir schreiben $f : X \rightarrow Y$, $x \mapsto y = f(x)$. Man nennt X den *Definitions-* oder *Urbildbereich* und Y den *Werte-* oder *Bildbereich* von f .

Beispiele. • $X = \mathbb{R}$ (= Menge der reellen Zahlen), $Y = \mathbb{R}$, $f(x) := x^2$.

- $X = \{A \in \mathcal{P}(\mathbb{N}) \mid A \text{ hat endlich viele Elemente } \sharp A < \infty\}$,
 $Y = \mathbb{N} = \{1, 2, 3, \dots\}$, $f(A) := \sharp A$.

2. Folgen und Reihen ([4, Kapitel 5 und 6])

Eine Abbildung $f : \mathbb{N} \rightarrow X$ nennen wir eine *Folge* (in X); das Bild $f(n)$ heißt *n-tes Glied der Folge*. Eine Folge in X ordnet also jeder natürlichen Zahl genau ein Element aus X zu. Man schreibt für die Folge $(x_n)_{n \in \mathbb{N}}$ oder (x_1, x_2, \dots) mit $x_n := f(n)$, $n \in \mathbb{N}$. Im Falle von $X = \mathbb{R}$ spricht man auch von (*reellen*) *Punktfolgen*. Statt \mathbb{N} kann man als Urbild auch $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ oder eine unendliche Teilmenge von \mathbb{N} verwenden.

Beispiele. • $X = \mathbb{R}$, $x_n = f(n) = n^2$, also $(1, 4, 9, 16, \dots)$.

- $X =$ Menge der Seiten eines Spielwürfels $\cong \{1, 2, \dots, 6\}$,
 $f(n) :=$ Ergebnis des n -ten Wurfes eines Würfelexperiments,
also z.B. $(6, 3, 2, 5, \dots)$.

Sei nun $(x_n)_{n \in \mathbb{N}}$ eine reelle Punktfolge. Wir definieren

$$s_n := x_1 + x_2 + \dots + x_n =: \sum_{k=1}^n x_k, \quad n \in \mathbb{N}.$$

Die so erklärte Folge $(s_n)_{n \in \mathbb{N}}$ heißt *Folge der Partialsummen* (von $(x_n)_{n \in \mathbb{N}}$); falls die Indexmenge \mathbb{N} durch \mathbb{N}_0 oder eine unendliche Teilmenge von \mathbb{N} ersetzt wird, definiert man s_n analog.

Beispiel (Geometrische Reihe). $X = \mathbb{R}$, $x_n = q^n$, $n \in \mathbb{N}_0$, mit einem $q \in \mathbb{R}$.

- Für $q = 1$ ergibt sich $x_n = 1$ und somit $s_n = \sum_{k=0}^{n-1} x_k = n$, $n \in \mathbb{N}$.
- Für $q \in \mathbb{R} \setminus \{1\}$ ergibt sich $s_n = \sum_{k=0}^{n-1} x_k = \sum_{k=0}^{n-1} q^k = \frac{1 - q^n}{1 - q}$, $n \in \mathbb{N}$.

(Beweis der *Summenformel* durch vollständige Induktion oder m.H. von Teleskopsummen \rightarrow *Übung*.)

Was geschieht, wenn man nicht nur die ersten n Glieder sondern alle (unendlich vielen) Glieder der Folge $(x_n)_{n \in \mathbb{N}}$ aufsummiert? Bei dem so entstehenden Objekt spricht man von einer (*unendlichen*) *Reihe* und schreibt

$$s = \sum_{k=1}^{\infty} x_k.$$

Im Falle des obigen Beispiels $x_n = q^n$ führt diese Betrachtung auf die wohl wichtigste Reihe, nämlich die *geometrische Reihe* $\sum_{k=0}^{\infty} q^k$. Für die Untersuchung von Reihen benötigt man den Begriff des Grenzwerts, der für die gesamte Analysis grundlegend ist:

3. Grenzwerte von Folgen, Reihenwerte ([4, Kapitel 5 und 6])

Zu Folgen: Sei $(x_n)_{n \in \mathbb{N}}$ eine reelle Punktfolge. Gibt es ein $a \in \mathbb{R}$, so dass für beliebige $\varepsilon > 0$ ein $N(\varepsilon) \in \mathbb{N}$ so existiert, dass gilt

$$|x_n - a| < \varepsilon \quad \text{für alle } n \geq N(\varepsilon),$$

so heißt a *Grenzwert von* $(x_n)_{n \in \mathbb{N}}$ und wir schreiben

$$a = \lim_{n \rightarrow \infty} x_n \quad \text{oder} \quad x_n \rightarrow a \quad (n \rightarrow \infty).$$

Man sagt auch: $(x_n)_{n \in \mathbb{N}}$ *konvergiert (gegen a)*.

Beispiele. • $x_n = \frac{n^2 - 7}{2n^2 + 5n}$. M.H. von $\frac{1}{n} \rightarrow 0$, $\frac{1}{n^2} \rightarrow 0$ ($n \rightarrow \infty$) berechnen wir

$$x_n = \frac{n^2 - 7}{2n^2 + 5n} = \frac{1 - \frac{7}{n^2}}{2 + \frac{5}{n}} \rightarrow \frac{1}{2} \quad (n \rightarrow \infty).$$

- $x_n = q^n$, $q \in \mathbb{R}$.
 - ◊ $q = 1$, also $(x_n)_{n \in \mathbb{N}} = (1, 1, 1, \dots)$. Diese *konstante Folge* hat trivialerweise den Grenzwert $\lim_{n \rightarrow \infty} x_n = 1$.
 - ◊ $q = -1$, also $(x_n)_{n \in \mathbb{N}} = (-1, 1, -1, 1, \dots)$. Da der Abstand je zweier aufeinander folgender Glieder = 2 ist, kann $((-1)^n)_{n \in \mathbb{N}}$ nicht konvergieren.

◇ $|q| < 1$, dann gilt $q^n \rightarrow 0$ ($n \rightarrow \infty$). Ist nämlich $\varepsilon > 0$ beliebig, so gilt

$$\begin{aligned} |q^n - 0| < \varepsilon &\Leftrightarrow |q|^n < \varepsilon &\Leftrightarrow e^{n \ln |q|} < \varepsilon \\ &\Leftrightarrow n \ln |q| < \ln \varepsilon &\Leftrightarrow n > \frac{\ln \varepsilon}{\ln |q|} \end{aligned}$$

Wählen wir also eine natürliche Zahl $N(\varepsilon) > \frac{\ln \varepsilon}{\ln |q|}$, so folgt

$$|q^n - 0| < \varepsilon \quad \text{für alle } n \geq N(\varepsilon).$$

◇ $|q| > 1$. Die Folge $(q^n)_{n \in \mathbb{N}}$ konvergiert nicht, da sie unbeschränkt ist in folgendem Sinne:

Eine Folge $(x_n)_{n \in \mathbb{N}}$ heißt *unbeschränkt*, wenn zu jedem $M > 0$ ein $n \in \mathbb{N}$ so existiert, dass $|x_n| > M$ gilt. Wir erwähnen noch die zwei folgenden Spezialfälle:

- Existiert zu jedem $M > 0$ ein $N(M) \in \mathbb{N}$ mit $x_n > M$ für alle $n \geq N(M)$, so heißt $(x_n)_{n \in \mathbb{N}}$ *uneigentlich konvergent (gegen $+\infty$)*. Schreibweise: $x_n \rightarrow +\infty$ ($n \rightarrow \infty$).
- Existiert zu jedem $M > 0$ ein $N(M) \in \mathbb{N}$ mit $x_n < -M$ für alle $n \geq N(M)$, so heißt $(x_n)_{n \in \mathbb{N}}$ *uneigentlich konvergent (gegen $-\infty$)*. Schreibweise: $x_n \rightarrow -\infty$ ($n \rightarrow \infty$).

In der Vorstellung streben uneigentlich konvergente Folgen beliebig weit nach rechts bzw. links auf der Zahlengeraden.

Beispiele. Die Folgen $(n)_{n \in \mathbb{N}}$, $(2^n)_{n \in \mathbb{N}}$ sind uneigentlich konvergent gegen $+\infty$, $(-n^2)_{n \in \mathbb{N}}$ ist z.B. uneigentlich konvergent gegen $-\infty$. Hingegen ist $((-3)^n)_{n \in \mathbb{N}}$ zwar unbeschränkt, aber nicht uneigentlich konvergent.

Zu Reihen: Eine Reihe $\sum_{k=1}^{\infty} x_k$ heißt *konvergent*, wenn die Folge der Partialsummen $s_n = \sum_{k=1}^n x_k$, $n \in \mathbb{N}$, konvergiert. Der Grenzwert heißt dann *Wert der Reihe* und wird ebenfalls mit $\sum_{k=1}^{\infty} x_k$ bezeichnet.

Beispiel. Die geometrische Reihe $\sum_{k=0}^{\infty} q^k$ konvergiert für $|q| < 1$ und divergiert für $|q| \geq 1$. Im letzteren Fall ist nämlich der Abstand zweier aufeinander folgender Partialsummen ≥ 1 . Und im Fall $|q| < 1$ gilt nach den obigen Beispielen:

$$s_n = \sum_{k=0}^{n-1} q^k = \frac{1 - q^n}{1 - q} \rightarrow \frac{1}{1 - q} \quad (n \rightarrow \infty).$$

Für den Wert der geometrischen Reihe haben wir also

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1 - q}, \quad \text{falls } |q| < 1.$$

4. Funktionsgrenzwerte im Unendlichen ([4, Kapitel 9])

Bei der Untersuchung einer Funktion $f = f(x) : \mathbb{R} \rightarrow \mathbb{R}$ ist häufig deren Verhalten bzw. Aussehen für beliebig große bzw. beliebig kleine Werte von x interessant. Man bestimmt also – wenn (uneigentlich) existent – die folgenden Grenzwerte

$$\lim_{x \rightarrow +\infty} f(x) = a \quad \text{bzw.} \quad \lim_{x \rightarrow -\infty} f(x) = b,$$

die folgendermaßen zu verstehen sind:

- Es existiert ein $a \in \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$, so dass gilt: Für beliebige Punktfolgen $(x_n)_{n \in \mathbb{N}}$ mit $\lim_{n \rightarrow \infty} x_n = +\infty$ gilt $\lim_{n \rightarrow \infty} f(x_n) = a$.
- Es existiert ein $b \in \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$, so dass gilt: Für beliebige Punktfolgen $(x_n)_{n \in \mathbb{N}}$ mit $\lim_{n \rightarrow \infty} x_n = -\infty$ gilt $\lim_{n \rightarrow \infty} f(x_n) = b$.

Beispiele. • $f(x) = e^x$. Es gilt $\lim_{x \rightarrow +\infty} e^x = +\infty$, $\lim_{x \rightarrow -\infty} e^x = 0$.

- $f(x) = \sin x$. Dann gilt z.B. $\lim_{n \rightarrow \infty} \sin(n\pi) = 0$ und $\lim_{n \rightarrow \infty} \sin(\frac{\pi}{2} + 2n\pi) = 1$; der Grenzwert $\lim_{x \rightarrow +\infty} \sin x$ existiert also nicht. Analoges gilt für den Grenzwert $\lim_{x \rightarrow -\infty} \sin x$.

5. Stammfunktionen und Integrale ([4, Kapitel 11])

Aus der Schule kennt man das Konzept der Stammfunktion: Ist $f : \mathbb{R} \rightarrow \mathbb{R}$ gegeben, so heißt eine Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ *Stammfunktion* von f , wenn gilt:

$$\frac{dF}{dx}(x) = f(x) \quad \text{für alle } x \in \mathbb{R};$$

dabei bezeichnet $\frac{dF}{dx} = F'$ die erste Ableitung von F . Wir nennen die Stammfunktion auch *unbestimmtes Integral* und schreiben

$$F = \int f(x) dx.$$

Sind $a, b \in \mathbb{R}$ mit $a < b$ beliebig, so besagt der *Fundamentalsatz der Differential- und Integralrechnung*:

$$\int_a^b f(x) dx = F(b) - F(a) =: F(x) \Big|_{x=a}^{x=b},$$

wobei die linke Seite das *bestimmte Integral von f (zwischen a und b)* bezeichnet.

Geometrisch misst $\int_a^b f(x) dx$ die Fläche zwischen dem Graphen von f und der x -Achse, die von den senkrechten Geraden $x = a$ und $x = b$ begrenzt wird. Dabei werden Flächenstücke unterhalb der x -Achse negativ bewertet.

Beispiel. $f(x) = \cos x$. Eine Stammfunktion ist $F(x) = \sin x$ (auch $F(x) = \sin x + c$ mit beliebiger Konstante $c \in \mathbb{R}$ ist Stammfunktion). Für das bestimmte Integral folgt:

$$\int_a^b \cos x \, dx = \sin b - \sin a.$$

6. Uneigentliche Integrale ([4, Abschnitt 11.5])

Das bestimmte Integral kann durch einen Grenzübergang wie in 4. genutzt werden, um auch Flächeninhalte über unendlich ausgedehnte Abschnitte der x -Achse zu berechnen:

Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ mit Stammfunktion $F : \mathbb{R} \rightarrow \mathbb{R}$ gegeben und sei $a \in \mathbb{R}$ fixiert. Dann heißt

$$\int_a^{+\infty} f(x) \, dx := \lim_{b \rightarrow +\infty} \int_a^b f(x) \, dx = \lim_{b \rightarrow +\infty} F(b) - F(a)$$

das *uneigentliche Integral von f (zwischen a und $+\infty$)*, falls der angegebene Funktionsgrenzwert existiert. Analog definiert man - bei Existenz - für beliebiges $b \in \mathbb{R}$:

$$\int_{-\infty}^b f(x) \, dx := \lim_{a \rightarrow -\infty} \int_a^b f(x) \, dx = F(b) - \lim_{a \rightarrow -\infty} F(a).$$

Und man setzt schließlich noch

$$\int_{-\infty}^{+\infty} f(x) \, dx := \lim_{b \rightarrow +\infty} F(b) - \lim_{a \rightarrow -\infty} F(a),$$

wenn beide Grenzwerte existieren.

Beispiele. • $f(x) = \frac{1}{x^2}$. Eine Stammfunktion ist $F(x) = -\frac{1}{x}$. Damit existiert z.B. das uneigentliche Integral

$$\int_1^{+\infty} \frac{1}{x^2} \, dx = \lim_{b \rightarrow +\infty} \left(-\frac{1}{b} \right) + 1 = 1.$$

Dass f nur für $x \neq 0$ definiert ist, macht uns keine Probleme; dann sind die Stammfunktion und entsprechend die uneigentlichen Integrale nur auf dieser Teilmenge sinnvoll.

• $f(x) = \cos x$. Wie in 5. folgt

$$\int_a^b \cos x \, dx = \sin b - \sin a;$$

aber z.B. $\lim_{b \rightarrow +\infty} \sin b$ existiert nicht, wie in 4. gesehen; also existiert das uneigentliche Integral $\int_a^{+\infty} \cos x \, dx$ nicht.

- Für uns wird der Fall $f(x) = \exp(-\frac{x^2}{2})$ interessant sein. Mit einiger Arbeit kann man zeigen, dass das uneigentliche Integral von f zwischen $-\infty$ und $+\infty$ existiert und dass gilt:

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi};$$

siehe [4, Satz 11.39].

Literaturverzeichnis

- [1] Peter Hartmann, *Mathematik für Informatiker*, Springer Vieweg, 2012.
- [2] Lothar Sachs, Jürgen Hedderich, *Angewandte Statistik. Methodensammlung mit R.*, Springer Spektrum, 2015.
- [3] Norbert Henze, *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*, Springer Spektrum, 2013.
- [4] *Skript zur Vorlesung Mathematik für Studierende der Informatik 1, Duisburg-Essen*, Version SS 17.

Index

- $B(n, p)$ -Verteilung, 40
- $N(\mu, \sigma^2)$ -Verteilung, 57
- $P(\lambda)$ -Verteilung, 47

- Bayes, Satz von, 26
- bedingte Wahrscheinlichkeit, 23
- Bernoulli-Experiment, 39
- Bernoulli-Verteilung, 39
- Binomial-Verteilung, 40

- Covarianz, 38

- Ereignis, 12
- Erwartungswert, 36, 53

- geometrische Verteilung, 42
- Grenzwertsätze, 61

- Hypergeometrische-Verteilung, 45

- identisch verteilt, 61
- Irrtums-Wahrscheinlichkeit, 88

- Kombinatorik, 16
- Korrelationskoeffizient, 38

- Maß, 15
- Maßraum, 15
- messbare Menge, 15
- messbarer Raum, 15
- Multiplikationssatz, 24

- negative Binomialverteilung, 43
- Normal-Verteilung, 54

- Poisson-Näherung, 48
- Poisson-Verteilung, 47

- Raum der Elementarereignisse, 12

- Satz von der totalen Wahrscheinlichkeit, 24
- Standardabweichung, 38, 53
- standardisierte Variable, 61

- unabhängige Ereignisse, 27
- unabhängige Zufallsvariable, 35, 53
- unkorrelierte Zufallsvariable, 38

- Vandermondesche Faltungsformel, 45
- Varianz, 38
- Verteilungsfunktion, 34

- Wahrscheinlichkeitsmaß, 15
- Wahrscheinlichkeitsraum, 15

- Zufallsvariable, 33