

Stochastik

Dr. Jana Bielagk
Humboldt-Universität zu Berlin

Wintersemester 2019/20

Notation

\emptyset	die leere Menge bzw. das unmögliche Ereignis
$\{a, b, c, \dots\}$	ungeordnete Menge, z.B. $\{1, 2, 3\} = \{3, 1, 2\}$
(a, b, c, \dots)	geordnete Menge, Tupel, Vektor
\mathbb{N}	natürliche Zahlen $\{1, 2, 3, \dots\}$
\mathbb{Z}	ganze Zahlen $\{\dots, -2, -1, 0, 1, 2, \dots\}$
\mathbb{Q}	rationale Zahlen
\mathbb{R}	reelle Zahlen
\mathbb{C}	komplexe Zahlen
$A \times B$	kartesisches Produkt von A und B , z.B. $\{1, 2, 3\} \times \{4, 5\} = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$
$A \cap B$	Schnittmenge von A und B , z.B. $\{0, 1\} \cap \{1, 2\} = \{1\}$
$A \cup B$	Vereinigung von Mengen A und B , z.B. $\{0, 1\} \cup \{1, 2, 3\} = \{0, 1, 2, 3\}$
$A \subseteq B$	Menge A ist in Menge B enthalten
$A \subset B$	Menge A ist echt in Menge B enthalten
$A \setminus B$	Menge von Elementen, die in A , aber nicht in B sind
A^c	Komplement der Menge / des Ereignisses A
$a \in A$	a ist ein Element der Menge A , z.B. $5 \in \mathbb{N}$
\forall	für alle, z.B. $\forall n \in \mathbb{N} \dots$
\exists	es existiert, z.B. $\forall n \in \mathbb{N} \exists m \in \mathbb{N} : m > n$
$\exists!$	es existiert genau ein – Quantor kombiniert Aussagen zu Existenz und Eindeutigkeit
$f: V \rightarrow W$	eine Funktion f von V nach W , V ist Definitions- und W Wertebereich
$f \circ g$	Verkettung der Abbildungen f und g , d.h. $(f \circ g)(x) := f(g(x))$
$a \mapsto b$	Element a wird auf b abgebildet, z.B. $f(x) = x^2$ entspricht $x \mapsto x^2$
$A := B$	definiere A als B , z.B. $A := \{1, 2, 3\}$
$A \Rightarrow B$	Implikation (aus der Aussage A folgt Aussage B)
$A \Leftrightarrow B$	Aussage A ist äquivalent zu B
$\det(A)$	Determinante der (quadratischen) Matrix A , wird auch mit $ A $ bezeichnet, wenn keine Verwechslung mit dem Betrag möglich ist
$f^{-1}(M)$	Urbild der Menge M unter f , d.h. für $f: V \rightarrow W$ ist $f^{-1}(M) = \{v \in V \mid f(v) \in M\}$
f^{-1}	die Umkehrfunktion der Abbildung f
A^{-1}	die inverse Matrix zu A
A^T	die Transponierte der Matrix A
$(x_k)_{k \in \mathbb{N}}$	Folge x_1, x_2, x_3, \dots
$x \rightarrow a$	Grenzwertbetrachtung, lasse x gegen a gehen
$x \rightarrow a_+$	einseitige Grenzwertbetrachtung von rechts
$x \rightarrow a_-$	einseitige Grenzwertbetrachtung von links
$f(x_+)$	Grenzwert $\lim_{x \rightarrow x_+} f(x)$
$[a, b]$	abgeschlossenes Intervall von a bis b
(a, b)	offenes Intervall von a bis b
$x \vee y$	$\max\{x, y\}$
$x \wedge y$	$\min\{x, y\}$
f^+	Positivteil einer Funktion in der Zerlegung $f = f^+ - f^-$; $f^+(x) = f(x) \vee 0$
$\operatorname{sgn}(x)$	Vorzeichen von $x \in \mathbb{R}$, d.h. $\operatorname{sgn}(x) \in \{-1, 0, 1\}$.

Inhaltsverzeichnis

1 Einführung	1
1.1 Zufallsexperimente und Ereignisse	1
1.2 Wahrscheinlichkeitsräume	4
1.2.1 Sigma-Algebren und die Potenzmenge	5
1.2.2 Wahrscheinlichkeitsmaße und -räume	6
1.3 Laplace-Experimente	9
1.4 Kombinatorik – Die Kunst des Zählens	10
1.4.1 Das Grundprinzip des Zählens – Produktregel	11
1.4.2 Permutationen	12
1.4.3 Variationen und Kombinationen	14
1.4.4 Ununterscheidbare Objekte	18
1.4.5 Laplace-Experimente II	20
1.5 Relative Häufigkeiten und Wahrscheinlichkeiten	23
2 Bedingte Wahrscheinlichkeiten	25
2.1 Definition bedingter Wahrscheinlichkeiten	25
2.2 Satz von Bayes und Gesetz der totalen Wahrscheinlichkeit	30
2.3 Stochastische Unabhängigkeit	33
2.4 Versuchsfolgen	36
2.5 Stetigkeit des Wahrscheinlichkeitsmaßes	38
3 Diskrete Zufallsvariablen und Verteilungen	41
3.1 Was sind Zufallsvariablen?	41
3.2 Die Massenfunktion diskret verteilter Zufallsvariablen	43
3.3 Die wichtigsten diskreten Verteilungen	46
3.3.1 Bernoulli- und Binomialverteilung	46
3.3.2 Poisson-Verteilung	48
3.3.3 Geometrische Verteilung	50
3.3.4 Hypergeometrische Verteilung	51
3.3.5 Diskrete Gleichverteilung	53
3.4 Erwartungswert und Varianz	54
3.5 Verteilungsfunktionen von Zufallsvariablen	63

4	Absolutstetige Zufallsvariablen und Verteilungen	67
4.1	Wann heißt eine Verteilung absolutstetig?	67
4.2	Erwartungswert und Varianz	72
4.3	Die wichtigsten absolutstetigen Verteilungen	76
4.3.1	Gleichverteilung	76
4.3.2	Exponentialverteilung	78
4.3.3	Normalverteilung	80
4.4	Der Satz von de Moivre-Laplace	87
5	Gemeinsame Verteilungen von Zufallsvariablen	95
5.1	Gemeinsame Verteilung	95
5.2	Stochastische Unabhängigkeit von Zufallsvariablen	101
5.3	Summen unabhängiger Zufallsvariablen	104
5.3.1	Diskrete Zufallsvariablen	104
5.3.2	Absolutstetige Zufallsvariablen	106
5.4	Erwartungswert von Funktionen von Zufallsvariablen	112
5.4.1	Kovarianz und Korrelation	115
5.4.2	Anwendung: Lineare Regression	121
6	Grenzwertsätze	125
6.1	Zwei wichtige Ungleichungen	125
6.2	Gesetze der großen Zahlen	127
6.2.1	Schwaches Gesetz der großen Zahlen	127
6.2.2	Exkurs: Konvergenz von Folgen von Zufallsvariablen	128
6.3	Zentraler Grenzwertsatz	129
7	Elemente der Statistik	133
7.1	Schätzen von unbekanntem Parametern	133
7.2	Konfidenzintervalle	138
7.3	Hypothesentests	140
7.3.1	Typen von Tests	144
7.4	Beschreibende Statistik	161
7.4.1	Merkmale, Ausprägungen und Histogramme	161
7.4.2	Lage- und Streuungsparameter	163
7.4.3	Schreibweise durch Datenvektoren und empirische lineare Regression	165
7.4.4	Box-Plots	168
8	Modellierung in der Stochastik – zwei Beispiele	169
8.1	Das Ziegenproblem	169
8.2	Das Bertrand-Paradoxon	170
A	Relevante Inhalte der Analysis	173
A.1	Fakultät und Binomialkoeffizient	173
A.2	Folgen, Reihen, Funktionenfolgen und Konvergenzbegriffe	174
A.3	Stetigkeit, Differenzierbarkeit	178
A.4	Integrationsregeln	180

B	Kontroll- und Ergänzungsfragen	185
B.1	Zu Kapitel 1	185
B.2	Zu Kapitel 2	186
B.3	Zu Kapitel 3	187
B.4	Zu Kapitel 4	188
B.5	Zu Kapitel 5	189
C	Python Programme	191
D	Kartenspiele	193

Einleitung

Was ist Stochastik? Stochastik ist dem Wort nach die *Kunst des Vermutens*. Sie beinhaltet sowohl die Wahrscheinlichkeitstheorie, welche sich mit dem Formalisieren von Zufallsereignissen und der Analyse von Modellen befasst, als auch die mathematische Statistik, welche auf Grund beobachteter zufälliger Ereignisse ein geeignetes Modell auswählt, was zu den Beobachtungen passt und damit ggf. Aussagen über zukünftige zufällige Ereignisse zu ermöglichen.

Was ist Zufall? Diese Frage beantwortet nicht die Mathematik, sondern eher die Philosophie. Für uns genügt die Vorstellung, dass auch unter augenscheinlich gleichen Ausgangsbedingungen verschiedene Ergebnisse auftreten können – dann ist das Ergebnis zufällig.

Was ist die Besonderheit des Faches Stochastik? In den ersten Semestern eines Mathematikstudiums lernt man in den Fächern Analysis und Lineare Algebra und Analytische Geometrie das Vorgehen Definition – Satz – Beweis kennen. Mit anderen Worten: Neue Begriffe werden eingeführt, deren Eigenschaften postuliert und anschließend bewiesen. Im Gegensatz dazu ist eine Besonderheit der Stochastik, dass wir von Anfang an von *Modellen* sprechen. Ist ein Problem gegeben, muss zunächst ein mathematisches (und insbesondere mathematisch verwertbares) Modell aufgestellt werden, dann erst wird das Problem mit den Begriffen und Methoden, die wir im Laufe des Kurses kennen lernen werden, gelöst um am Schluss eine Antwort auf die Fragestellung geben zu können. Damit erfordert die Stochastik in besonderem Maße, dass in (in unserem Fall) deutscher Sprache verfasste Probleme in die Sprache der Mathematik übersetzt werden. Daher werden Beispiele und das Modellieren auch im Skript eine besondere Rolle spielen.

Was sind die Ziele dieses Kurses? Neben dem Erlernen des fachspezifischen Vokabulars, sowie der Eigenschaften und Beziehungen der Begriffe, der Beweis- und Problemlösungsmethoden ist es ein erklärtes Ziel in der Stochastik, das Aufstellen von Modellen zu erlernen, Annahmen zu treffen und kritisch zu hinterfragen, sowie aus einer Auswahl von Methoden und Modellen bewusst entscheiden zu können, welches in einer gegebenen Situation angemessen und zielführend ist.

Wie können Sie mit diesem Skript arbeiten? Da Mathematik grundsätzlich eine aktive Auseinandersetzung mit dem Stoff erfordert, gibt es Abschnitte in denen Sie bereits in der Vorlesung mitdenken und mitmachen sollen. Diese sind seitlich durch das Symbol  gekennzeichnet. Kommentare, die zu wichtig sind um sie übersehen zu dürfen, sind durch das Symbol  gekennzeichnet. Ein Bezug zu Hausaufgaben ist durch  erkennbar. Zudem gibt es im Anhang B Kontrollfragen zu jedem Kapitel. Diese sollen Sie selbstständig beantworten sobald der Stoff behandelt wurde. Zusätzlich gibt es Übungsaufgaben in den Präsenzübungen und Hausaufgaben, die zeigen sollen, ob Sie mit den gelernten Begriffen umgehen können.

Was gibt es in diesem Skript Besonderes? Neben den inhaltlichen Kapiteln enthält dieses Skript einen Anhang mit Zusatzmaterial. In Anhang A sind wichtige Inhalte der Analysis zur Referenz gesammelt; Anhang B enthält die bereits erwähnten Ergänzungsfragen; Anhang C enthält Python-Codes, die im Laufe der Vorlesung vorkommen; Anhang D enthält einen Überblick über ein 52-Karten-Spiel zur Referenz für die in der Stochastik obligatorischen Aufgaben, die sich auf Kartenspiele beziehen.

Was für Fragen werden Sie am Ende des Semesters beantworten können? Im Folgenden ist eine Reihe von Fragestellungen, die jeweils am Ende eines Kapitels beantwortet werden kann.

1. Ein Lehrer sammelt die Abgaben seiner 25 Schüler ein, mischt sie gut durch und verteilt sie wieder an die Schüler. Wie wahrscheinlich ist es, dass kein Schüler seine eigene Abgabe bekommt?
2. Nehmen wir an, dass auf ewig jede Woche sechs Lottozahlen nacheinander gezogen werden. Wie wahrscheinlich ist es, dass irgendwann die Kombination (1, 2, 3, 4, 5, 6) gezogen wird?

Kapitel 1

Einführung – Modellieren von Zufallsexperimenten

Ziele

- Mengenoperationen und die Siebformel kennen
- Die Bestandteile eines Wahrscheinlichkeitsraumes mit ihren Eigenschaften kennen
- Wahrscheinlichkeitsräume zu Zufallsexperimenten aufstellen können
- Laplace-Experimente (er-)kennen
- Die wichtigsten Formeln und Rechenregeln der Kombinatorik kennen und anwenden können

1.1 Zufallsexperimente und Ereignisse

In Experimenten in Physik oder Chemie wird in der Regel *Reproduzierbarkeit* gefordert, d.h. wenn ein Versuch unter den gleichen Voraussetzungen mehrfach wiederholt wird, sollte das gleiche Ergebnis beobachtet werden können.

Befindet man sich auf der Erdoberfläche, hält einen Stein in die Höhe und lässt ihn los, so fällt er zum Boden. Man kann den Stein auch durch einen Apfel oder einen nicht aufgeblasenen Luftballon ersetzen – das Ergebnis bleibt das Gleiche. Hält man jedoch einen prall gefüllten Luftballon in die Höhe und lässt ihn los, so wird er möglicherweise das Weite suchen – ob steil nach oben oder eher parallel zum Boden hängt von diversen Faktoren ab, u.a. dem Gas, mit dem der Luftballon gefüllt ist, sowie dem Wetter zum Zeitpunkt des Experiments.

Dieses Beispiel widerspricht nicht der Reproduzierbarkeit, denn der Füllzustand des Luftballons und die herrschenden Wetterbedingungen sind Voraussetzungen, die einen Einfluss auf den Ausgang des Experiments haben – hält man sämtliche Einflussfaktoren konstant, so sollte der Ausgang des Experiments vorhersagbar sein.

Auch bei *Zufallsexperimenten* müssen die Rahmenbedingungen für das Experiment festgelegt werden, jedoch kann der Ausgang des Experiments nicht mit Sicherheit vorhergesagt werden. Es ist zwar bekannt, welche Ausgänge möglich sind, jedoch sind auch unter (weitestgehend²) gleichen Voraussetzungen die Ergebnisse des Versuchs nicht immer gleich.

²Messungenauigkeiten und Ungenauigkeiten bei der Ausführung des Experiments können in der Regel zwar nicht ganz verhindert werden, aber ihr Einfluss auf das Ergebnis ist bei deterministischen Experimenten vernachlässigbar, während sie bei Zufallsexperimenten den Zufall mit verursachen können.

Zur mathematischen Beschreibung von Zufallsexperimenten benötigen wir geeignete Modelle.

Definition 1.1

Jeder mögliche Ausgang eines Zufallsexperiments wird als *Ergebnis* oder *Elementarereignis* bezeichnet. Die Menge aller Ergebnisse bezeichnen wir als *Grundraum* oder *Ergebnisraum* und notieren ihn als Ω . Teilmengen des Grundraumes bezeichnen wir als *Ereignisse*.

- Das Ereignis $A = \emptyset$ (leere Menge) wird als *unmögliches Ereignis* bezeichnet.
- Das Ereignis $A = \Omega$ wird als *sicheres Ereignis* bezeichnet.

Beispiel 1.2. Zufallsexperimente und mögliche Grundräume zur Modellierung:

i) Ergebnis eines Würfelwurfes: $\Omega = \{1, 2, 3, 4, 5, 6\}$

ii) Zweifacher Münzwurf: $\Omega = \{K, Z\}^2 = \{(K, K), (K, Z), (Z, K), (Z, Z)\}$

iii) Reihenfolge beim Zieleinlauf von n Sprintern beim 100-Meter-Lauf:

$$\Omega = \{(a_1, \dots, a_n) \in \{1, \dots, n\}^n \mid a_i \neq a_j \text{ für } i \neq j\}$$

iv) Lebensdauer einer Glühlampe in Stunden: $\Omega = [0, \infty)$



Handelt es sich bei folgenden Ereignissen um Elementarereignisse? Wie schreibt beschreibt man sie mathematisch?

	ja	nein	Ereignis
Der Würfel zeigt eine 3.			
Der Würfel zeigt eine gerade Zahl.			
Beim zweifachen Münzwurf erscheint zuerst Kopf.			
Beim zweifachen Münzwurf erscheint zweimal Zahl.			
Sprinter Nummer 7 siegt.			
Die Glühlampe leuchtet mindestens 1000 Stunden.			

Mengenoperationen werden unverändert auch auf Ereignisse angewandt.

Vereinigung zweier Ereignisse: $A \cup B = \{\omega \in \Omega \mid \omega \in A \vee \omega \in B\}$

Schnitt zweier Ereignisse: $A \cap B = \{\omega \in \Omega \mid \omega \in A \wedge \omega \in B\}$

Komplementär-/Gegenereignis: $A^c = \{\omega \in \Omega \mid \omega \notin A\}$

Bemerkung 1.3. Wir wählen bewusst nicht die in der Literatur ebenfalls zu findende Schreibweise \bar{A} für das Komplementärereignis zu A , da diese Schreibweise auch für den Abschluss einer Menge verwendet wird.

Notation

Ist $(A_k)_{k \in \mathbb{N}}$ eine Folge von Ereignissen, so setzt man:

$$\begin{aligned} \bigcup_{k=1}^{\infty} A_k &= \{\omega \in \Omega \mid \exists n \in \mathbb{N}: \omega \in A_n\} \\ &= \{\omega \in \Omega \mid \omega \in A_n \text{ für mindestens ein } n \in \mathbb{N}\}, \\ \bigcap_{n=1}^{\infty} A_n &= \{\omega \in \Omega \mid \forall n \in \mathbb{N}: \omega \in A_n\} \\ &= \{\omega \in \Omega \mid \omega \in A_n \text{ für alle } n \in \mathbb{N}\}. \end{aligned}$$

Zur Vollständigkeit der Darstellung wiederholen wir kurz die Rechenregeln für Mengen bzw. Ereignisse.

Satz 1.4

Seien $E, F, G, A_1, \dots, A_n \subseteq \Omega$ Ereignisse. Dann gelten folgende Rechenregeln:

Kommutativität $E \cup F = F \cup E$ und $E \cap F = F \cap E$

Assoziativität: $(E \cup F) \cup G = E \cup (F \cup G)$ und $(E \cap F) \cap G = E \cap (F \cap G)$

Distributivität: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ und $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$

De Morgan'sche Regeln: $(\bigcup_{k=1}^n A_k)^c = \bigcap_{k=1}^n A_k^c$ und $(\bigcap_{k=1}^n A_k)^c = \bigcup_{k=1}^n A_k^c$

Beweis der De Morgan'schen Regeln. Die erste Gleichung zeigt man dadurch, dass ein beliebiges Element ω genau dann in der einen Menge enthalten ist, wenn es in der anderen Menge enthalten ist.

$$\begin{aligned} \omega \in \left(\bigcup_{k=1}^n A_k \right)^c &\iff \omega \notin \bigcup_{k=1}^n A_k \\ &\iff \omega \notin A_k, \forall k \in \{1, \dots, n\} \\ &\iff \omega \in A_k^c, \forall k \in \{1, \dots, n\} \iff \omega \in \bigcap_{k=1}^n A_k^c \end{aligned}$$

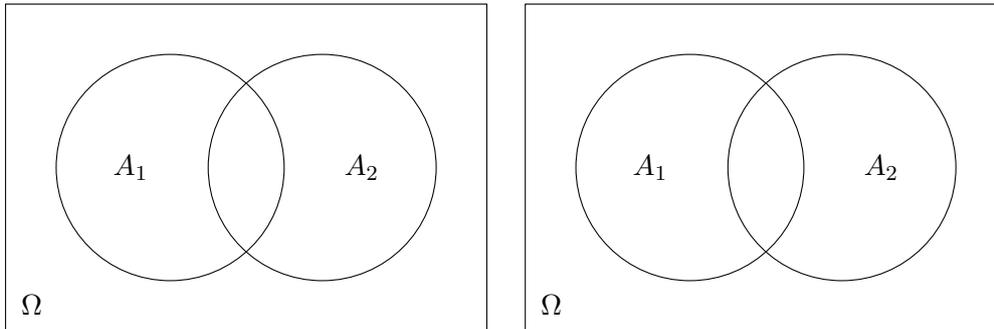
Die zweite Aussage beweist man mit Hilfe der ersten. Zunächst einmal gilt mit der bereits bewiesenen Aussage

$$\left(\bigcup_{k=1}^n A_k^c \right)^c \stackrel{1. \text{ Aussage}}{=} \bigcap_{k=1}^n (A_k^c)^c = \bigcap_{k=1}^n A_k.$$

Bildet man das Komplement auf beiden Seiten, so erhält man die zu zeigende Aussage. \square

Man kann sich die DeMorgan'schen Regeln auch mit Hilfe von Venn-Diagrammen verdeutlichen, zum Beispiel gilt im einfachen Fall $n = 2$:



**Definition 1.5**

Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge von Ereignissen. Die Ereignisse heißen *paarweise disjunkt*, falls Folgendes gilt:

$$A_k \cap A_j = \emptyset, \quad \forall k \neq j.$$

Notation

Sind $A, B \subseteq \Omega$ unvereinbar, d.h. disjunkte Ereignisse, so schreibt man für deren (disjunkte) Vereinigung $A \dot{\cup} B$.

1.2 Wahrscheinlichkeitsräume

Die Wahrscheinlichkeit eines Ereignisses ist eine Zahl zwischen 0 und 1 und desto größer die Zahl ist, desto wahrscheinlicher ist das Eintreten des Ereignisses.

Überlegen wir uns anhand eines Beispiels, welche Eigenschaften eine Abbildung haben sollte, die Ereignissen ihre Wahrscheinlichkeit zuordnet.

Beispiel 1.6. Wir betrachten das Zufallsexperiment, in dem ein fairer Würfel geworfen wird, d.h. wir setzen $\Omega = \{1, 2, 3, 4, 5, 6\}$ und alle sechs Elementarereignisse sind gleich wahrscheinlich.

- Werfen wir einen Würfel ohne dass dieser verschwindet, so muss irgendetwas passieren, d.h. Ω soll die Wahrscheinlichkeit 1 haben. Gleichzeitig wird nicht nichts passieren, d.h. das unmögliche Ereignis \emptyset muss die Wahrscheinlichkeit 0 haben.
- Da alle Elementarereignisse gleich wahrscheinlich sein sollen, geben wir ihnen intuitiv die Wahrscheinlichkeit $\frac{1}{6}$.
- Das Ereignis $\{1, 2\}$ ist doppelt so wahrscheinlich wie $\{1\}$, jedoch genau so wahrscheinlich wie $\{3, 4\}$. Die Wahrscheinlichkeit eines Ereignisses sollte also dadurch bestimmt sein, wie viele der Elementarereignisse darin enthalten sind.
- Es tritt entweder das Ereignis $\{1, 2\}$ oder das Ereignis $\{1, 2\}^c = \{3, 4, 5, 6\}$ ein – die Summe beider Wahrscheinlichkeiten muss also 1 sein.

Wir müssen für die Abbildung, die jedem Ereignis eine Wahrscheinlichkeit zuordnet, Definitions- und Wertebereich festlegen. Der Wertebereich ist klar $[0, 1]$ – für den Definitionsbereich brauchen wir jedoch, wie wir gerade gesehen haben, mehr als nur die Elementarereignisse.

1.2.1 Sigma-Algebren und die Potenzmenge

Gehen wir zunächst den einfachen Weg und betrachten die Menge aller Teilmengen des Grundraumes Ω .

Definition 1.7: Potenzmenge

Die Potenzmenge $\mathcal{P}(\Omega)$ einer Menge Ω ist die Menge der Teilmengen von Ω , d.h.

$$\mathcal{P}(\Omega) := \{ A \mid A \subseteq \Omega \}.$$

Wie lauten die Potenzmengen folgender Mengen?

Ω	$\mathcal{P}(\Omega)$
\emptyset	
$\{1\}$	
$\{0, 1\}$	



Wir können also die Potenzmenge des Grundraumes als Definitionsbereich setzen. In manchen Fällen kann es jedoch sinnvoll sein, weniger als die ganze Potenzmenge als Definitionsbereich zu wählen, da wir nicht für sämtliche Ereignisse Wahrscheinlichkeiten bestimmen können. Sehen wir uns dazu ein Beispiel an.

Beispiel 1.8. Wir betrachten das Geburtsgewicht eines Babys. Aus einer Statistik¹ leiten wir ab, dass die Wahrscheinlichkeit, dass das Gewicht zwischen 3000 und 3500 Gramm liegt, bei 0,37 liegt und mit einer Wahrscheinlichkeit von 0,3 liegt es zwischen 3500 und 4000 Gramm. Setzen wir $\Omega = [0, \infty)$, so können wir dennoch nicht die Wahrscheinlichkeit angeben, dass ein Neugeborenes unter 3000 Gramm wiegt, obwohl $[0, 3000) \subseteq \Omega$ ist. Hätten wir die Zusatzinformation, wie wahrscheinlich ein Neugeborenes über 4000 Gramm wiegt, wäre das jedoch möglich.

Wenn wir nicht die Potenzmenge wählen, was ist dann die Minimalanforderung an das System von Mengen, dem wir Wahrscheinlichkeiten zuordnen wollen?

Wie wir in Beispiel 1.6 gesehen haben, sollen \emptyset und Ω immer eine Wahrscheinlichkeit zugeordnet bekommen. Zudem soll zu disjunkten Ereignissen $A, B \subseteq \Omega$, deren Wahrscheinlichkeiten bekannt sind, auch die Wahrscheinlichkeit von $A \cup B$ bekannt sein – auch bei einer ganzen Folge paarweise disjunkter Ereignisse soll das zutreffen. Und zu einem beliebigen Ereignis kennen wir auch stets die Wahrscheinlichkeit des Gegenereignisses. Ein Mengensystem, welches all die genannten Mengen enthält, ist die σ -Algebra².

¹Wir werden noch sehen, was relative Häufigkeiten sind und warum wir diese mit Wahrscheinlichkeiten gleich setzen.

²Die Pluralform von Algebra ist Algebren. Das σ signalisiert, dass nicht nur Vereinigungen endlich vieler, sondern sogar abzählbar vieler Mengen wieder in dem Mengensystem enthalten sein sollen.

Definition 1.9: σ -Algebra

Ein Mengensystem $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, d.h. eine Menge von Teilmengen von Ω , heißt σ -Algebra, falls sie folgende Bedingungen erfüllt:

1. $\Omega \in \mathcal{A}$.
2. Ist $A \in \mathcal{A}$, so ist auch das Komplement $A^c \in \mathcal{A}$.
3. Sind $A_1, A_2, \dots \in \mathcal{A}$, so ist auch deren Vereinigung $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.

Die kleinste σ -Algebra, die ein gegebenes Mengensystem $\mathcal{M} \subseteq \mathcal{P}(\Omega)$ enthält, bezeichnen wir mit $\sigma(\mathcal{M})$ und wir sprechen von der *durch \mathcal{M} erzeugten σ -Algebra*.



Seien $A, B \subseteq \Omega$ Mengen. Ergänzen Sie die angegebenen Mengensysteme so, dass σ -Algebren entstehen.

\mathcal{M}	kleinste σ -Algebra, die \mathcal{M} enthält
$\{\Omega\}$	
$\{A\}$	
$\{A, B\}$	
$\mathcal{P}(\Omega)$	



Wann immer Ω endlich oder abzählbar unendlich ist, wählen wir $\mathcal{A} = \mathcal{P}(\Omega)$. Ist Ω jedoch überabzählbar unendlich, so verwenden wir i.d.R. die so genannte *Borel'sche σ -Algebra*.

Definition 1.10

Ist $\Omega = \mathbb{R}$, so bezeichnet $\mathcal{B}(\mathbb{R})$ die *Borel'sche σ -Algebra*, welche erzeugt wird durch die offenen Mengen in \mathbb{R} .

Bemerkung 1.11. Die Borel'sche σ -Algebra enthält nicht nur alle offenen Intervalle, sondern auch alle abgeschlossenen und halboffenen Intervalle. Man kann statt $\mathcal{B}(\mathbb{R})$ auch die Borel'sche σ -Algebra über einer Teilmenge von \mathbb{R} bilden, beispielsweise $\mathcal{B}(\mathbb{R}_+)$ oder $\mathcal{B}([0, 1])$.

1.2.2 Wahrscheinlichkeitsmaße und -räume

Nun kommen wir zur Definition einer Abbildung, die Ereignissen ihre Wahrscheinlichkeiten zuordnet.

Definition 1.12

Auf einem Grundraum Ω sei $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine σ -Algebra. Eine Abbildung $\mathbb{P}: \mathcal{A} \rightarrow [0, 1]$ heißt *Wahrscheinlichkeitsmaß*, falls sie folgende Eigenschaften erfüllt.

Normierung: $\mathbb{P}(\Omega) = 1$.

σ -Additivität: Für jede Folge paarweise disjunkter Ereignisse $(A_k)_{k \in \mathbb{N}} \subseteq \mathcal{A}$ gilt

$$\mathbb{P}\left(\dot{\bigcup}_{k \in \mathbb{N}} A_k\right) = \sum_{k \in \mathbb{N}} \mathbb{P}(A_k).$$

Beispiel 1.13. Für einen Würfelwurf (vgl. Beispiel 1.6) mit $\Omega = \{1, 2, 3, 4, 5, 6\}$ und σ -Algebra $\mathcal{P}(\Omega)$ ist ein geeignetes Wahrscheinlichkeitsmaß gegeben durch

$$\mathbb{P}(\omega) = \frac{1}{6}, \quad \forall \omega \in \Omega.$$

In der Tat gilt wie in Beispiel 1.6 gefordert $\mathbb{P}(\{1, 2\}) = 2 \cdot \mathbb{P}(\{1\}) = \frac{2}{6} = \frac{1}{3} = \mathbb{P}(\{3, 4\})$, was aus der σ -Additivität von \mathbb{P} folgt. Zudem können wir aus der σ -Additivität und Normierung Folgendes schlussfolgern:

$$\mathbb{P}(\{1, 2\}^c) + \mathbb{P}(\{1, 2\}) \stackrel{\text{disjunkt}}{=} \mathbb{P}(\Omega) = 1 \implies \mathbb{P}(\{1, 2\}^c) = 1 - \mathbb{P}(\{1, 2\}) = \frac{2}{3} = \mathbb{P}(\{3, 4, 5, 6\}).$$

Definition 1.14

Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$, bestehend aus einer nichtleeren Menge Ω , einer σ -Algebra $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ und einem Wahrscheinlichkeitsmaß $\mathbb{P}: \mathcal{A} \rightarrow [0, 1]$.

Der folgende Satz bietet eine Reihe nützlicher Eigenschaften von Wahrscheinlichkeitsmaßen, die es einem analog zu Beispiel 1.13 ermöglichen, aus den W 'keiten von Elementarereignissen auf die W 'keiten beliebiger Ereignisse zu schlussfolgern.

Satz 1.15

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Dann hat \mathbb{P} folgende Eigenschaften:

1. $\mathbb{P}(\emptyset) = 0$;
2. $\mathbb{P}\left(\bigcup_{k \in \{1, \dots, n\}} A_k\right) = \sum_{k=1}^n \mathbb{P}(A_k)$ für endlich viele paarweise disjunkte Ereignisse $A_1, \dots, A_n \in \mathcal{A}$;
3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ für alle $A \in \mathcal{A}$;
4. Ist $A \subseteq B$ für Ereignisse $A, B \in \mathcal{A}$, so gilt $\mathbb{P}(A) \leq \mathbb{P}(B)$;
5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ für alle $A, B \in \mathcal{A}$;
6. $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ für alle $A, B \in \mathcal{A}$.

Beweis.

1. Wählt man $A_k = \emptyset$ für alle $k \in \mathbb{N}$, so ist $(A_k)_{k \in \mathbb{N}}$ eine Folge paarweise disjunkter Ereignisse und es gilt wegen $\emptyset = \bigcup_{k \in \mathbb{N}} A_k$

$$\mathbb{P}(\emptyset) = \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right) \stackrel{\text{disjunkt}}{=} \sum_{k \in \mathbb{N}} \mathbb{P}(A_k) = \sum_{k \in \mathbb{N}} \mathbb{P}(\emptyset). \quad (1.2.1)$$

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

2. Seien $A_{n+1} = A_{n+2} = \dots = \emptyset$. Dann ist wieder $(A_k)_{k \in \mathbb{N}}$ eine Folge paarweise disjunkter Ereignisse und es folgt

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \stackrel{\sigma\text{-Add.}}{=} \sum_{k=1}^{\infty} \mathbb{P}(A_k) \stackrel{!}{=} \sum_{k=1}^n \mathbb{P}(A_k).$$

3. Es gilt $A \dot{\cup} A^c = \Omega$, so dass aus 2. direkt

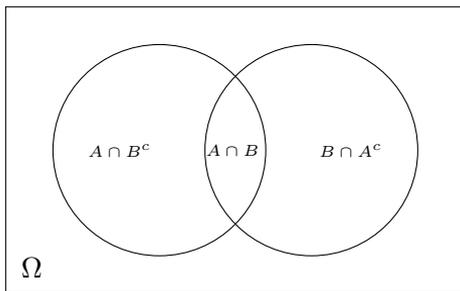
$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\Omega) \stackrel{\text{Normierung}}{=} 1$$

und somit die Behauptung folgt.

4. Ist $A \subseteq B$, so gilt die disjunkte Zerlegung $B = A \dot{\cup} (B \cap A^c)$. Somit können wir wieder 2. verwenden und erhalten

$$\mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \cap A^c)}_{\geq 0} \geq \mathbb{P}(A).$$

5. Wir zerlegen $A \cup B$ wieder in disjunkte Teilmengen. Dazu hilft eine Skizze.



$$\begin{aligned} A &= (A \cap B) \dot{\cup} (A \cap B^c) \\ B &= (B \cap A) \dot{\cup} (B \cap A^c) \\ A \cup B &= (A \cap B^c) \dot{\cup} (B \cap A^c) \dot{\cup} (A \cap B) \end{aligned}$$

Aus den disjunkten Zerlegungen folgt

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(B \cap A^c) + \mathbb{P}(A \cap B) \\ &= [\mathbb{P}(A) - \mathbb{P}(A \cap B)] + [\mathbb{P}(B) - \mathbb{P}(A \cap B)] + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \end{aligned}$$

6. Diese Aussage folgt sofort aus 5.:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \underbrace{\mathbb{P}(A \cap B)}_{\geq 0} \leq \mathbb{P}(A) + \mathbb{P}(B).$$

□

Die 5. Aussage lässt sich auf eine beliebige (endliche) Anzahl von Ereignissen verallgemeinern. Die Formel ist dann bekannt als *Einschluss-Ausschluss-Formel*, *Prinzip von Inklusion und Exklusion* oder als *Siebformel von Poincaré und Sylvester*.¹

Satz 1.16: Siebformel von Poincaré und Sylvester

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Für $A_1, \dots, A_n \in \mathcal{A}$ gilt dann

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) &= \sum_{k=1}^n \mathbb{P}(A_k) - \sum_{1 \leq j < k \leq n} \mathbb{P}(A_j \cap A_k) + \sum_{1 \leq i < j < k \leq n} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots \\ &\quad + (-1)^{n+1} \mathbb{P}\left(\bigcap_{k=1}^n A_k\right). \end{aligned}$$

¹Jules Henri Poincaré (1854-1912) war ein französischer Mathematiker, Physiker und Astronom; James Joseph Sylvester (1814-1897) ein britischer Mathematiker.

Beweis. Der Beweis wird über vollständige Induktion geführt. (ÜA) □

1.3 Laplace-Experimente

Definition 1.17

Sei $\Omega = \{\omega_1, \dots, \omega_n\}$ ein endlicher Grundraum. Ein Zufallsexperiment wird *Laplace-Experiment* genannt, falls alle Versuchsausgänge, d.h. alle Elementarereignisse, die gleiche Wahrscheinlichkeit besitzen. Es gilt in diesem Fall

$$\mathbb{P}(\omega_k) = \frac{1}{n} \quad \text{für } k = 1, \dots, n.$$

Bemerkung 1.18. *Ob ein Zufallsexperiment sinnvoll als Laplace-Experiment modelliert werden kann, muss von Fall zu Fall entschieden werden. Das Werfen eines fairen Würfels oder einer fairen Münze sind gewissermaßen die Prototypen für Laplace-Experimente, da das Wort fair gerade dafür steht, dass alle Möglichkeiten gleich wahrscheinlich sind. Allerdings nehmen wir in beiden Fällen (ohne dies jedes Mal zu erwähnen) an, dass Würfel oder Münze nicht auf einer Kante liegen bleiben.*

Ein Wahrscheinlichkeitsmaß auf einem diskreten Grundraum Ω ist durch Angabe der Wahrscheinlichkeiten der Elementarereignisse eindeutig bestimmt. 

Speziell gilt im Falle eines Laplace-Experimentes folgende Faustformel:

$$\text{Wahrscheinlichkeit des Ereignisses} = \frac{\text{Anzahl günstiger Ergebnisse}}{\text{Anzahl möglicher Ergebnisse}}$$

Um dies mathematisch formulieren zu können, benötigen wir eine Schreibweise für die Anzahl der Elemente einer Menge.

Notation

Die Anzahl der Elemente einer Menge M bezeichnen wir mit $|M|$ oder $\#M$.

Damit können wir die Wahrscheinlichkeit eines beliebigen Ereignisses bei einem Laplace-Experiment angeben.

Korollar 1.19

Beschreibt ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein Laplace-Experiment, so ist die Wahrscheinlichkeit eines beliebigen Ereignisses $A \in \mathcal{P}(\Omega)$ gegeben durch

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

Beweis. Sei $\Omega = \{\omega_1, \dots, \omega_n\}$. Dann gilt nach Eigenschaft 2 von Satz 1.15

$$\mathbb{P}(A) = \sum_{k: \omega_k \in A} \mathbb{P}(\{\omega_k\}) = \frac{1}{n} \cdot |\{k \in \{1, \dots, n\} \mid \omega_k \in A\}| = \frac{|A|}{|\Omega|}.$$

□

Beispiel 1.20. Ein fairer Würfel wird zweimal geworfen. Wie groß ist die Wahrscheinlichkeit, dass die Augensumme 7 beträgt?

- Wahrscheinlichkeitsraum: $\Omega = \{1, \dots, 6\}^2$ und $\mathcal{A} = \mathcal{P}(\Omega)$, \mathbb{P} ist gegeben durch $\mathbb{P}(\{(k, j)\}) = \frac{1}{36}$, da $|\Omega| = 36$ ist.
- Ereignis: $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$ mit $|A| = 6$
- Berechnung: $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$.

Beispiel 1.21. Eine faire Münze wird zweimal geworfen.

- Wie groß ist die Wahrscheinlichkeit, dass beim zweiten Wurf Zahl geworfen wird?
- Wie groß ist die Wahrscheinlichkeit, dass mindestens einmal Zahl geworfen wird?

- Wahrscheinlichkeitsraum: $\Omega = \{K, Z\}^2$ und $\mathcal{A} = \mathcal{P}(\Omega)$. Da es wieder ein Laplace-Experiment ist, ist \mathbb{P} ist gegeben durch $\mathbb{P}(\{(i, j)\}) = \frac{1}{4}$, da $|\Omega| = 4$ ist.
- Ereignisse: Um beide Fragen beantworten zu können, definieren wir mehrere Ereignisse:

$$A = \{\text{Beim 1. Wurf fällt Zahl.}\} = \{(Z, K), (Z, Z)\}$$

$$B = \{\text{Beim 2. Wurf fällt Zahl.}\} = \{(K, Z), (Z, Z)\}$$

- Berechnung: Für die erste Frage ist $\mathbb{P}(B)$ gesucht. Dazu berechnen wir:

$$\mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{2}{4} = \frac{1}{2} = \mathbb{P}(A).$$

Damit ist die erste Frage beantwortet. Als Antwort auf die zweite Frage suchen wir $\mathbb{P}(A \cup B)$. Wir können dazu die Siebformel (mit $n = 2$) verwenden. Dazu benötigen wir noch die Wahrscheinlichkeit von $A \cap B$.

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{(Z, Z)\}) = \frac{1}{4}.$$

Damit ist die gesuchte Wahrscheinlichkeit

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Alternativ kann man über das Komplement $(A \cup B)^c$ argumentieren, denn

$$\begin{aligned} (A \cup B)^c &= \{\text{Es wird weder beim 1. noch beim 2. Wurf Zahl geworfen.}\} \\ &= \{\text{Es fällt bei beiden Würfeln Kopf.}\} = \{(K, K)\}. \end{aligned}$$

Folglich ist

$$\mathbb{P}(A \cup B) = 1 - \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(\{(K, K)\}) = 1 - \frac{1}{4} = \frac{3}{4}.$$

Um die Anzahl günstiger und möglicher Ergebnisse systematisch und schnell berechnen zu können, benötigen wir Kenntnisse der *Kombinatorik*.

1.4 Kombinatorik – Die Kunst des Zählens

Eine ausführliche Einführung in die Kombinatorik und was *Zählen* eigentlich bedeutet findet sich in Abschnitt 2.8 in [KS11]. Eine Übersicht über Resultate zur Fakultät und zum Binomialkoeffizienten sind in Abschnitt A.1 zum Nachschlagen zu finden.

1.4.1 Das Grundprinzip des Zählens – Produktregel

Stellen wir uns folgendes Mittagmenü eines Restaurants vor:

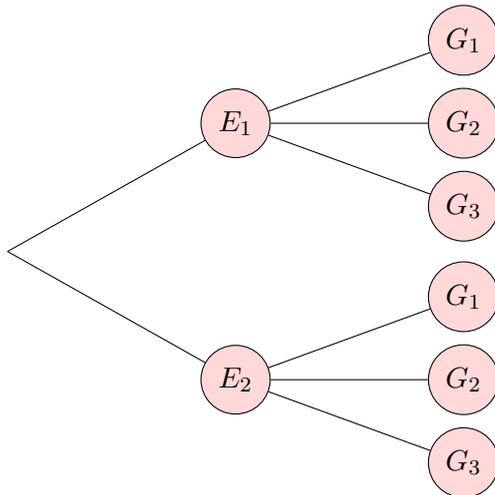
Essen

Spaghetti a la Carbonara (E_1)
Pizza Tonno (E_2)

Getränk

Mineralwasser (G_1)
Orangensaft (G_2)
Kaffee (G_3)

Wie viele Kombinationsmöglichkeiten aus Essen und Getränk gibt es? Eine Möglichkeit der Herleitung ist die Verwendung eines Baumdiagramms:



Man kann die Entscheidung als zwei verschachtelte Experimente auffassen.

1. Auswahl des Essens – 2 mögliche Ausgänge;
2. Auswahl des Getränks – 3 mgl. Ausgänge.

Fasst man beide Experimente zusammen, so gibt es insgesamt $2 \cdot 3 = 6$ mögliche Ausgänge.

Satz 1.22: Grundprinzip des Zählens

Zwei Experimente werden durchgeführt. Hat das erste m und das zweite n mögliche Ausgänge, so haben beide Experimente zusammen $m \cdot n$ mögliche Ausgänge.

Beweis. Das gemeinsame Experiment hat die Ausgänge (k, j) mit $k \in \{1, \dots, m\}$ und $j \in \{1, \dots, n\}$. Damit sind die möglichen $m \cdot n$ Ausgänge

$$\begin{array}{cccc}
 (1, 1) & (1, 2) & \cdots & (1, n) \\
 (2, 1) & (2, 2) & \cdots & (2, n) \\
 \vdots & \vdots & \ddots & \vdots \\
 (m, 1) & (m, 2) & \cdots & (m, n).
 \end{array}$$

□

Dieses Prinzip lässt sich auch auf mehr als zwei Versuche verallgemeinern.

Satz 1.23: Verallgemeinertes Zählprinzip

Besteht ein Experiment aus N Telexperimenten (mit $N \in \mathbb{N}$), die unabhängig voneinander ausgeführt werden und gibt es für das k -te Telexperiment genau n_k mögliche Versuchsausgänge ($k \in \{1, \dots, N\}$), so hat das zusammengesetzte Experiment genau $\prod_{k=1}^N n_k$ mögliche Versuchsausgänge.

Beweis. Der Beweis wird über vollständige Induktion geführt. \square

Bemerkung 1.24. Das Ergebnis aus Satz 1.23 ist auch unter den Bezeichnungen Produktregel oder Multiplikationsregel bekannt.

Beispiel 1.25. Ein Koffer ist mit einem Zifferschluss gesichert. Das Schloss hat drei Ringe, auf denen die Ziffern 1 bis 9 stehen.

Wenn man für das Probieren pro Einstellung 2 Sekunden benötigt, wie viel Zeit benötigt man maximal bis das Schloss geknackt ist?

① ② ③

Antwort: Nach dem (verallgemeinerten) Zählprinzip gibt es $9 \cdot 9 \cdot 9 = 9^3 = 729$ Möglichkeiten. Wenn das Schloss in der Ausgangsposition noch nicht offen ist, heißt das, man braucht maximal $2 \cdot 729 = 1458$ Sekunden, d.h. 24 Minuten und 18 Sekunden.



Beispiel 1.26. Beim Lotto 6 aus 49 werden aus einer Trommel mit 49 nummerierten Kugeln nacheinander 6 Kugeln gezogen.

Wie viele Möglichkeiten für Ziehungen gibt es, wenn die zeitliche Reihenfolge beim Ziehen eine Rolle spielt?

Antwort:

- | | |
|-----------|---------------|
| 1. Kugel: | Möglichkeiten |
| 2. Kugel: | Möglichkeiten |
| 3. Kugel: | Möglichkeiten |
| 4. Kugel: | Möglichkeiten |
| 5. Kugel: | Möglichkeiten |
| 6. Kugel: | Möglichkeiten |

Damit gibt es insgesamt

mögliche Ziehungen.

Um in Beispiel 1.26 die Anzahl möglicher Ziehungen zu bekommen, bei denen die Reihenfolge keine Rolle spielt, benötigen wir die Anzahl an Möglichkeiten, sechs Zahlen anzuordnen.

1.4.2 Permutationen

Definition 1.27

Sei $\Omega = (\omega_1, \dots, \omega_n)$ eine endliche Menge. Eine *Permutation* ist eine bijektive Abbildung $\pi: \Omega \rightarrow \Omega$.

Beispiel 1.28. Eine Permutation kann man beispielsweise dadurch angeben, dass man jeweils paarweise die Elemente einer Menge und deren Bilder angibt. Beispielsweise ist eine Permutation auf $(1, 2, 3, 4, 5)$ gegeben durch

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 1 & 2 \end{pmatrix}.$$

Satz 1.29

Ist M eine Menge mit n unterscheidbaren Objekten, so gibt es genau $n!$ mögliche Anordnungen aller n Elemente, d.h. es gibt $n!$ Permutationen auf M .

Beweis. Die Aussage folgt sofort aus dem Zählprinzip. Für das 1. Objekt gibt es n mögliche Positionen, für das 2. Objekt $n - 1$ mögliche Positionen u.s.w. und für das n -te Objekt gibt es nur noch eine mögliche Position. Damit gibt es insgesamt

$$n \cdot (n - 1) \cdot \dots \cdot 1 = \prod_{k=1}^n k = n!$$

Möglichkeiten. □

Beispiel 1.30. Ich habe 10 Bücher über Analysis, 4 Bücher über lineare Algebra und 13 Bücher über Stochastik und möchte diese so in einer Reihe anordnen, dass alle Bücher aus einem Fachgebiet nebeneinanderstehen. Kein Buch kommt doppelt vor und die Reihenfolge der Fachgebiete ist frei wählbar.

Wie viele Anordnungsmöglichkeiten gibt es?

Antwort: Betrachten wir zunächst die Teilerperimente, die Bücher der einzelnen Fachgebiete anzuordnen.

- Analysis: 10! mögliche Anordnungen
- Lineare Algebra: 4! mögliche Anordnungen
- Stochastik: 13! mögliche Anordnungen

Zudem gibt es 3! mögliche Anordnungen der drei Fachgebiete. Also gibt es insgesamt

$$10! \cdot 4! \cdot 13! \cdot 3!$$

mögliche Anordnungen der Bücher.

Beispiel 1.31 (Fortsetzung von Beispiel 1.26). Soll die Reihenfolge der gezogenen Zahlen im Lotto 6 aus 49 keine Rolle spielen, müssen wir die bisher berechnete Anzahl möglicher Ziehungen durch die Anzahl der Permutationen der 6 gezogenen Zahlen teilen. Damit gibt es bei vernachlässigbarer Reihenfolge insgesamt 

mögliche Ziehungen.

Permutationen mit Gruppen ununterscheidbarer Objekte

Beispiel 1.32. Wie viele Anagramme, d.h. wie viele unterscheidbare Umordnungen der Buchstaben, des Wortes **PFEFFER** gibt es?

Antwort: Zunächst einmal gibt es 7! Möglichkeiten, 7 verschiedene Buchstaben anzuordnen. Geben wir dazu den Buchstaben Indizes, d.h. das Wort lautet nun $PF_1E_1F_2F_3E_2R$. Folgende Umordnungen sind ohne die Indizes jedoch nicht vom ursprünglichen Wort zu unterscheiden:

$$PF_1E_1F_3F_2E_2R, PF_2E_1F_1F_3E_2R, PF_2E_1F_3F_1E_2R, \dots$$

Wir müssen also die Anzahlen der möglichen Permutationen der mehrfach auftretenden Buchstaben untersuchen:

- Es gibt $2!$ mögliche Permutationen der E-Buchstaben.
- Es gibt $3!$ mögliche Permutationen der F-Buchstaben.

Somit gibt es nach dem Zählprinzip $2! \cdot 3! = 12$ Permutationen der doppelten Buchstaben und somit bilden jeweils 12 Permutationen das gleiche Anagramm, wenn alle Buchstabentypen an den gleichen Stellen stehen.

Insgesamt gibt es somit

$$\frac{7!}{12} = \frac{7!}{3!2!} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{2} = 420$$

verschiedene Anagramme des Wortes **PFEFFER**.

Satz 1.33

Für n Objekte, die in r Gruppen der Größen n_1, \dots, n_r mit $n_1 + \dots + n_r = n$ zerfallen, gibt es

$$\frac{n!}{n_1!n_2! \cdots n_r!} =: \binom{n}{n_1, n_2, \dots, n_r}$$

Permutationen, bei denen die Elemente der Gruppe identifiziert werden.

Notation

Der mathematische Ausdruck

$$\binom{n}{n_1, n_2, \dots, n_r} := \frac{n!}{n_1!n_2! \cdots n_r!}$$

wird als *Multinomialkoeffizient* bezeichnet. Für $r = 2$ definiert man als Spezialfall den *Binomialkoeffizienten* für $n_1 \leq n$ via

$$\binom{n}{n_1} := \frac{n!}{n_1!(n - n_1)!}$$



Beispiel 1.34. Eine Polizeistation ist mit 10 Beamten besetzt. 5 davon sind auf Streife, 2 haben Innendienst und 3 sind auf Bereitschaft. Wie viele Aufteilungen der 10 Beamten auf die drei Gruppen sind möglich?

Antwort: Es gibt insgesamt

$$\binom{10}{5, 2, 3} = \frac{10!}{5!2!3!} = \frac{3628800}{720 \cdot 2 \cdot 6} = 2520$$

also 2520 Möglichkeiten.

1.4.3 Variationen und Kombinationen – geordnete und ungeordnete Stichproben mit und ohne Wiederholung/Zurücklegen

In diesem Abschnitt geht es um die Frage, wie viele Möglichkeiten es gibt, k Objekte aus n ($n \geq k$) verschiedenen Objekten auszuwählen. Um eine Antwort auf diese Frage geben zu können, muss die Fragestellung noch in Hinblick auf zwei Aspekte präzisiert werden:

1. Spielt die Reihenfolge, mit der die Objekte gezogen werden, eine Rolle?
2. Können Objekte mehrmals gezogen werden (d.h. sie werden nach dem Ziehen wieder zur Menge hinzugefügt) oder nur einmal?

Sehen wir uns diese Fallunterscheidung für $n = 3$ und $k = 2$ an:



	mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge
mit Zurücklegen	(1, 1) (1, 2) (1, 3) (2, 1) (2, 2) (2, 3) (3, 1) (3, 2) (3, 3)	
ohne Zurücklegen		

Die Bezeichnungen für die verschiedenen Fälle variieren in der Literatur. Zwei mögliche Bezeichnungen findet man in folgender Tabelle:

	mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge
mit Zurücklegen	Variation /	Kombination /
ohne Zurücklegen	geordnete Stichprobe	ungeordnete Stichprobe

Ob eine Variation oder Kombination mit oder ohne Zurücklegen (bzw. Wiederholung) zu verstehen ist, muss also grundsätzlich dazugesagt werden. Bisweilen wird für Variationen auch die Bezeichnung k -Permutation verwendet. Außerdem wird die oben eingeführte Permutation auch als Sonderfall der Variation betrachtet.

Kommen wir nun zu den Formeln für die Anzahlen der Variationen und Kombinationen.

Satz 1.35

Im Folgenden seien $n, k \in \mathbb{N}$ mit $k \leq n$.

Variation mit Zurücklegen: Es gibt n^k Möglichkeiten, k Objekte aus einer Menge von n verschiedenen Objekten auszuwählen, wenn jedes Objekt beliebig oft in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte beachtet wird.

Variation ohne Zurücklegen: Es gibt $\frac{n!}{(n-k)!}$ Möglichkeiten, k Objekte aus einer Menge von n verschiedenen Objekten auszuwählen, wenn jedes Objekt höchstens einmal in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte beachtet wird.

Kombination mit Zurücklegen: Es gibt $\binom{n+k-1}{k}$ Möglichkeiten, k Objekte aus einer Menge von n verschiedenen Objekten auszuwählen, wenn jedes Objekt beliebig oft in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte nicht beachtet wird.

Kombination ohne Zurücklegen: Es gibt $\binom{n}{k}$ Möglichkeiten, k Objekte aus einer Menge von n verschiedenen Objekten auszuwählen, wenn jedes Objekt höchstens einmal in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte nicht beachtet wird.



Die Formeln können in Tabellenform zusammengefasst werden:

	mit Beachtung der Reihenfolge	ohne Beachtung der Reihenfolge
mit Zurücklegen		
ohne Zurücklegen		

Beweis.

Variation mit Zurücklegen: Für jedes der k Experimente, in dem ein Objekt aus n Objekten gezogen wird, gibt es n Möglichkeiten. Nach dem (verallgemeinerten) Zählprinzip ergeben sich insgesamt also

$$\underbrace{n \cdot n \cdot \dots \cdot n}_{k \text{ Versuche}} = n^k$$

Möglichkeiten.

Variation ohne Zurücklegen: In jedem Versuch reduziert sich die Anzahl der Möglichkeiten für den folgenden Versuch um 1, so dass es insgesamt

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

Möglichkeiten gibt.

Kombination ohne Zurücklegen: Dies kann einerseits als Spezialfall von Satz 1.33 mit $r=2$, $n_1=k$ und $n_2=n-k$ aufgefasst werden. Andererseits können wir zunächst die Variation ohne Zurücklegen betrachten, d.h. es gibt $\frac{n!}{(n-k)!}$ Möglichkeiten, wenn die Reihenfolge beachtet wird. Da die Reihenfolge jedoch nicht beachtet werden soll, müssen wir durch die Anzahl möglicher Permutationen der ausgewählten k Elemente, also $k!$, teilen. Somit ergeben sich

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

Möglichkeiten.

Kombination mit Zurücklegen: Da die Elemente der Grundgesamtheit unterscheidbar sind, nummerieren wir diese und identifizieren sie mit ihrer Nummer, d.h. wir wollen k Elemente aus $\{1, 2, \dots, n\}$ auswählen ohne Beachtung der Reihenfolge und mit Zurücklegen der bereits gezogenen Zahlen. Die Menge solcher Kombinationen ist

$$K_k^n(mZ) := \left\{ (a_1, \dots, a_k) \in \mathbb{N}^k \mid 1 \leq a_1 \leq \dots \leq a_k \leq n \right\}.$$

Im Gegensatz dazu sei die Menge der Kombinationen ohne Zurücklegen

$$K_k^n(oZ) := \left\{ (a_1, \dots, a_k) \in \mathbb{N}^k \mid 1 \leq a_1 < \dots < a_k \leq n \right\}.$$

Wir wollen zeigen, dass es eine Bijektion $f: K_k^n(mZ) \rightarrow K_k^{n+k-1}(oZ)$ gibt.¹ Wenn uns das gelingt, so folgt daraus², dass beide Mengen gleich viele Elemente haben, also gerade $\binom{n+k-1}{k}$.

¹Ein alternativer Beweis über eine Äquivalenzrelation und ihre Äquivalenzklassen findet sich in [Kre05] auf S. 8.

²Bekanntes Resultat aus Analysis: Gibt es eine Bijektion zwischen zwei endlichen Mengen, so enthalten diese gleich viele Elemente.

Die bijektive Abbildung lautet wie folgt:

$$f((a_i)_{i=1,\dots,k}) := (a_i + i - 1)_{i=1,\dots,k}.$$

Setzen wir

$$\begin{aligned} b_1 &:= a_1 + 1 - 1 = a_1 \\ b_2 &:= a_2 + 2 - 1 = a_2 + 1 \\ &\vdots \\ b_k &:= a_k + k - 1, \end{aligned}$$

so ist $(b_1, b_2, \dots, b_k) \in K_k^{n+k-1}(oZ)$. Zum Nachweis der Bijektivität geben wir die Umkehrabbildung von f an:

$$f^{-1}((b_i)_{i=1,\dots,k}) := (b_i + 1 - i).$$

□

Sehen wir uns ein paar Beispiele für die genannten Formeln an.

Beispiel 1.36.

1. Wie viele Möglichkeiten gibt es, dass die ersten fünf Personen, die Ihnen in der S-Bahn begegnen, an verschiedenen Wochentagen Geburtstag haben **wenn unterschieden wird, wer an welchem Tag Geburtstag hat**?
2. Wie viele Möglichkeiten gibt es, bei einer Qualitätskontrolle zufällig aus einem Karton mit 100 Glühbirnen 4 Stück auszuwählen?
3. Wie viele Möglichkeiten gibt es für 3 nicht unterscheidbare Spatzen, sich auf 7 Bäume zu verteilen, wenn mehrere Spatzen auf dem gleichen Baum sitzen können?
4. Bei Hexadezimalzahlen werden insgesamt 16 verschiedene Ziffern (0 bis 9 und A bis F) verwendet. Wie viele verschiedene Zahlen kann man mit 3 solchen Ziffern darstellen?

Antwort:

1. Es sollen 5 **Personen** auf 7 **Wochentage** aufgeteilt werden, wobei die Reihenfolge eine Rolle spielen soll, d.h. es wird unterschieden, welche der Personen an welchem Tag Geburtstag hat. Es handelt sich also um eine Variation ohne Zurücklegen. Es gibt dafür $\frac{7!}{(7-5)!} = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 2520$ Möglichkeiten.
2. In diesem Versuch werden 4 Glühbirnen aus 100 Glühbirnen ohne Beachtung der Reihenfolge und ohne sie wieder zurückzulegen gezogen. Es handelt sich also um eine Kombination ohne Zurücklegen; es gibt dafür $\binom{100}{4}$ Möglichkeiten.
3. Es wird von $k = 3$ Spatzen einer der $n = 7$ Bäume ausgesucht, wobei die Reihenfolge der Spatzen keine Rolle spielt und die Bäume, sobald sie ausgewählt wurden, wieder „zurückgelegt“ werden in die Menge der auswählbaren Bäume. Es gibt nach der Formel für Kombinationen mit Zurücklegen also $\binom{7+3-1}{3} = \binom{9}{3}$ Möglichkeiten.
4. Es wird in $k = 3$ Versuchen jeweils eine aus $n = 16$ möglichen Ziffern ausgewählt, wobei die Reihenfolge wichtig ist und jede Ziffer nach der Auswahl wieder zur Verfügung steht. Es handelt sich also um eine Variation mit Zurücklegen; es gibt 16^3 mögliche Zahlen.



Beispiel 1.37. Ein Kommunikationssystem besteht aus n in einer Reihe aufgestellten Antennen. Es funktioniert genau dann, wenn keine zwei aufeinanderfolgenden Antennen defekt sind. Falls genau m der n Antennen ausfallen, wie wahrscheinlich ist es, dass das System funktioniert?

Antwort:

Überlegen wir uns zunächst für den Fall $n = 4$ und $m = 2$ alle Möglichkeiten.

☑	☑	☒	☒	funktioniert nicht
☑	☒	☑	☒	
☑	☒	☒	☑	
☒	☑	☑	☒	
☒	☑	☒	☑	
☒	☒	☑	☑	

Betrachten wir die Anzahl möglicher und die Anzahl funktionierender Konfigurationen und nehmen wir an, dass alle Konfigurationen gleich wahrscheinlich sind, so können wir das Experiment als Laplace-Experiment modellieren und erhalten die Wahrscheinlichkeit, dass das System aus 4 Antennen bei 2 defekten Antennen noch funktioniert als

Für den allgemeinen Fall mit n Antennen, von denen $m < n$ defekt sind, zählen wir wieder die Anzahl möglicher und (un-)günstiger Fälle.

- Wie viele Möglichkeiten gibt es, m (gleichartige) Antennen aus n (unterscheidbaren) Antennen auszuwählen, wenn die Reihenfolge des Auswählens keine Rolle spielt und jede Antenne höchstens einmal gewählt werden darf?
- Wie viele Möglichkeiten gibt es, die m defekten Antennen zwischen den $n - m$ funktionierenden Antennen anzuordnen, so dass nie zwei defekte Antennen nebeneinander sind?
 - 1 2 3 4 ... $n - m$
- Wahrscheinlichkeit einer funktionierenden Konfiguration:

1.4.4 Ununterscheidbare Objekte

Im vorherigen Abschnitt sind wir davon ausgegangen, dass wir k Objekte aus einer Menge von n unterscheidbaren Objekten auszuwählen. Nun wollen wir uns überlegen, wie viele Möglichkeiten es gibt, n ununterscheidbare Objekte in k (unterscheidbare) Gruppen aufzuteilen.

Satz 1.38

Es sollen n ununterscheidbare Objekte in $k \leq n$ unterscheidbare Gruppen aufgeteilt werden.

1. Wenn jede Gruppe mindestens ein Objekt enthalten muss, gibt es $\binom{n-1}{k-1}$ Möglichkeiten.
2. Wenn Gruppen auch null Objekte enthalten dürfen, gibt es $\binom{n+k-1}{k-1}$ Möglichkeiten.

Sehen wir uns zunächst ein Beispiel zur Illustration des Problems an.

Beispiel 1.39. Für die Lerntage haben sich 8 Helfer gefunden, die sich entweder um die Räumlichkeiten, die Verpflegung oder die Freizeitgestaltung kümmern sollen. Wer was macht, spielt keine Rolle, aber es ist ein Unterschied, ob eine Person oder zwei oder mehr Personen beispielsweise die Getränkebox besorgen. Es sollte jede Aufgabe von mindestens einer Person erledigt werden. Einige mögliche Verteilungen sind in folgender Tabelle aufgezeigt:



Räume						
Verpflegung						
Freizeit						

Systematisches Aufzählen der Möglichkeiten erfordert folgende Schritte:

- In der 1. Gruppe (Räume) können 1 bis 6 Helfer sein, damit mindestens einer für jede Gruppe bleibt.
- Sind n_1 Helfer in Gruppe 1, so sind 1 bis $7 - n_1$ Helfer in Gruppe 2 (Verpflegung).
- Sind n_1 Helfer in Gruppe 1 und n_2 Helfer in Gruppe 2, so sind notwendigerweise $8 - (n_1 + n_2)$ Helfer in Gruppe 3 (Freizeit).

Schreiben wir die Namen der 8 Helfer nacheinander auf, so ist die Frage lediglich, zwischen welchen Namen wir die Gruppengrenzen ziehen – siehe Beispiel 1.37.

1 2 3 4 5 6 7 8

Die Anzahl möglicher Aufteilungen auf die drei Gruppen ist also (gemäß der Formel für die Kombinationen ohne Zurücklegen):

$$\binom{\quad}{\quad}$$

Gibt es niemanden, der darauf achtet, dass alle Aufgabenbereiche abgedeckt sind, könnten auch leere Gruppen entstehen. Wie viele Möglichkeiten gibt es dann? Gehen wir sofort zur systematischen Zählung über analog zum ersten Fall:

- In der 1. Gruppe (Räume) können 0 bis 8 Helfer sein, damit mindestens einer für jede Gruppe bleibt.
- Sind n_1 Helfer in Gruppe 1, so sind 0 bis $8 - n_1$ Helfer in Gruppe 2 (Verpflegung).
- Sind n_1 Helfer in Gruppe 1 und n_2 Helfer in Gruppe 2, so sind notwendigerweise $8 - (n_1 + n_2)$ Helfer in Gruppe 3 (Freizeit).

Es ist egal, ob man eine Zahl aus $\{0, 1, \dots, 8\}$ oder aus $\{1, 2, \dots, 9\}$ auswählt. Somit gibt es (gemäß der Formel für die Kombinationen mit Zurücklegen)

$$\binom{\quad}{\quad} \text{ Möglichkeiten.}$$

Nachdem wir uns den Bezug zu den bekannten Formeln bereits im Beispiel hergeleitet haben, fassen wir den Beweis kurz.

Beweis von Satz 1.38.

1. Wir reihen die n Objekte zunächst auf und machen sie so unterscheidbar. Für die Gruppengrenzen gibt es genau $n - 1$ Möglichkeiten und zwischen k Gruppen gibt es genau $k - 1$ Grenzen, die auf diese $n - 1$ Positionen zu verteilen sind. Dafür gibt es genau $\binom{n-1}{k-1}$ Möglichkeiten.

2. Die Anzahl der Möglichkeiten, n ununterscheidbare Objekte auf k möglicherweise leere Gruppen zu verteilen, ist gleich der Anzahl Möglichkeiten, $n+k$ ununterscheidbare Objekte auf k Gruppen aufzuteilen. Folglich gibt es $\binom{n+k-1}{k-1}$ Möglichkeiten.

□

1.4.5 Laplace-Experimente II

Mit den Formeln der Kombinatorik können wir nun auch Laplace-Experimente auswerten, bei denen das Aufreihen und Zählen der möglichen Ausgänge zu langwierig wäre.

Beispiel 1.40. Eine Urne enthält 6 weiße und 5 schwarze Kugeln. 3 Kugeln werden zufällig und ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass 1 Kugel weiß und 2 Kugeln schwarz sind?

Antwort: Wir nummerieren die schwarzen und weißen Kugeln um sie unterscheidbar zu machen und bezeichnen sie mit w_1, \dots, w_6 und s_1, \dots, s_5 .

- Wahrscheinlichkeitsraum:

$$\Omega = \{\{w_1, w_1, w_3\}, \{w_1, w_3, w_4\}, \dots, \{w_1, s_1, s_5\}, \dots, \{s_3, s_4, s_5\}\}$$

mit $|\Omega| = \binom{11}{3}$; $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \mathcal{A}$.

- Gesuchtes Ereignis:

$$A = \{\{w_i, s_j, s_k\} \in \Omega \mid i \in \{1, \dots, 6\}, j, k \in \{1, \dots, 5\}, j \neq k\}$$

- Kombinatorische Überlegung: Die Reihenfolge der Ziehungen ist egal, daher können wir das Experiment in zwei Telexperimente aufteilen: Zunächst ziehen wir 1 der 6 weißen Kugeln und anschließend ziehen wir 2 der 5 schwarzen Kugeln. Das Produkt der Möglichkeiten für beide Telexperimente ergibt nach dem Zählprinzip $|A|$. Jedes Telexperiment entspricht einer Kombination ohne Zurücklegen, d.h. es gilt

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\binom{6}{1} \binom{5}{2}}{\binom{11}{3}} = \frac{6 \cdot 10}{11 \cdot 5 \cdot 3} = \frac{60}{165} = \frac{4}{11} \approx 0,36.$$

Bemerkung 1.41. Eine nicht-mathematische Bemerkung ist an dieser Stelle angebracht. Alle Zahlen in obigem Beispiel hätten nach dem Willen der Rechtschreibung als Wörter geschrieben werden sollen. Allerdings erschwert dies das schnelle Erfassen und damit das Verständnis der Aufgabe. Daher wählen wir auch bei kleinen Anzahlen Ziffern an Stelle von Zahlwörtern.

Bemerkung 1.42. Aufgaben mit farbigen Kugeln, die aus einer Urne gezogen werden, sind typisch für das Trainieren von Fragestellungen der Kombinatorik. Einerseits passt das gut zu den traditionellen Ziehungen der Lottozahlen, andererseits können die farbigen Kugeln durch beliebige realitätsnähere Objekte ersetzt werden. Beispielsweise greift man in einen unsortierten Besteckkasten mit Löffeln, Gabeln und (hoffentlich nicht so scharfen) Messern und kann überlegen, wie lange man braucht, bis man ein Besteckset gezogen hat. Da die Mathematik von der Umschreibung des Problems nicht beeinflusst wird, vereinfachen die Kugeln die Aufgabenformulierung. Gleichzeitig kann man beliebige innerhalb von gewissen Gruppen nicht unterscheidbare Objekte gedanklich durch farbige Kugeln ersetzen um ein auf den ersten Blick neues Problem auf ein bereits bekanntes Problem zurückzuführen.

Beispiel 1.43 (Geburtstagsproblem). *Wie viele Personen müssen in einem Raum sein, damit die Wahrscheinlichkeit, dass zwei von ihnen am gleichen Tag Geburtstag haben, mindestens 0,5 beträgt?*



Vorüberlegung: Wir wollen die beschriebene Situation als Laplace-Experiment auffassen. Welche Annahmen müssen wir dabei machen?

- *Betrachten wir das Jahr als 365 oder 366 Tage lang? Wenn wir von 365 Tagen ausgehen, könnten wir den 29. Februar als nicht zulässigen Geburtstag haben. \rightsquigarrow Nehmen wir also an, alle Personen seien im gleichen Jahr, z.B. 1997, geboren. Da dies kein Schaltjahr war, können wir Ω als die Menge aller Tage des Jahres 1997 setzen und haben damit $|\Omega| = 365$.*
- *Ist die Wahrscheinlichkeit, dass eine zufällig gewählte Person an einem festen Tag im Jahr Geburtstag hat, für alle Tage gleich groß? Die Antwort ist: Nein! Mehr Details, warum dies so ist und welchen Einfluss ungleiche Wahrscheinlichkeiten auf die Antwort unseres Problems haben findet man beispielsweise in [Hur08]. Vereinfachend nehmen wir also an, dass die Personen im Raum unabhängig voneinander mit gleicher Wahrscheinlichkeit an jedem der 365 Tage Geburtstag hat.*

Intuition: Eigene Schätzung für die Antwort auf die Frage:

Antwort: Unter Beachtung der Vorüberlegungen rechnen wir nun mit $|\Omega| = 365$. Überlegen wir uns zunächst, wie hoch die Wahrscheinlichkeit ist, dass unter n Personen keine zwei am gleichen Tag Geburtstag haben – hierbei ist natürlich $n \in \mathbb{N}$ mit $n \leq 365$ vorauszusetzen. Wir formulieren ein entsprechendes Ereignis¹:

A_n : „Unter n anwesenden Personen haben keine zwei am gleichen Tag Geburtstag.“

Die Wahrscheinlichkeit berechnen wir wie folgt:

$$|A_n| =$$

$$\mathbb{P}(A_n) = \text{_____}$$

Rechnet man diese Wahrscheinlichkeiten für die ersten natürlichen Zahlen n aus, so gibt uns die erste Wahrscheinlichkeit, die kleiner als 0,5 ist, die gesuchte Anzahl Personen.

n	$\mathbb{P}(A_n)$	n	$\mathbb{P}(A_n)$	n	$\mathbb{P}(A_n)$
1	1.0	11	0.8589	21	0.5563
2	0.9973	12	0.8330	22	0.5243
3	0.9918	13	0.8056	23	0.4927
4	0.9836	14	0.7769	24	0.4617
5	0.9729	15	0.7471	25	0.4313
6	0.9595	16	0.7164	26	0.4018
7	0.9438	17	0.6850	27	0.3731
8	0.9257	18	0.6531	28	0.3456
9	0.9054	19	0.6209	29	0.3190
10	0.8831	20	0.5886	30	0.2937

*Es müssen also mindestens
Personen
in einem Raum sein, damit mit einer Wahrscheinlichkeit von mindestens 0,5 mindestens zwei Personen den gleichen Geburtstag haben.*

¹Beschreiben wir ein Ereignis mit mathematischer Symbolik, so schreiben wir wie bisher $A = \{ \dots \}$; beschreiben wir ein Ereignis jedoch mit Worten und ohne mathematische Symbolik, so verwendet wir auch die Schreibweise A : „...“.

Bemerkung 1.44. Wer bei solchen Wahrscheinlichkeiten vorschnell den Taschenrechner zückt, kann mit einem *Math ERROR* belohnt werden. Während es klar ist, dass beispielsweise $365!$ eine Zahl ist, die dem Taschenrechner zu groß ist, dürfte $\frac{364!}{365!} = \frac{1}{365}$ eigentlich kein Problem darstellen. Allerdings ist ein Standard-Rechner nicht so programmiert, dass er Ausdrücke zuerst vereinfacht – somit muss der Benutzer eines Taschenrechners diese Arbeit für ihn ggf. übernehmen. Die hier angegebenen Werte wurden mit Python¹ erzeugt und auf vier Stellen nach dem Komma gerundet.

Bemerkung 1.45. Eine sehr umfassende Beschreibung des Geburtstagsproblem von der Herleitung eines geeigneten Modells bis zur konkreten Berechnung der Wahrscheinlichkeiten findet man in Abschnitt 1.1 in [Mil17].

Das folgende Problem ist als Hutproblem bekannt – man kann es aber auch leicht auf andere Accessoires übertragen wie Regenschirme, Rucksäcke oder Handys.



Beispiel 1.46 (Hutproblem). Die Hüte von n Personen werden gemischt und nacheinander nimmt jede Person zufällig einen Hut. Wie groß ist die Wahrscheinlichkeit, dass keine der Personen ihren eigenen Hut erhält?

Antwort: Wir stellen wieder zunächst einen Wahrscheinlichkeitsraum auf und formulieren das gesuchte Ereignis und beginnen erst dann mit dem Rechnen.

- Wahrscheinlichkeitsraum: Der Grundraum Ω ist die Menge der Permutationen der n Hüte. Somit ist

$$|\Omega| =$$

Da Ω endlich ist, setzen wir $\mathcal{A} = \mathcal{P}(\Omega)$ und wenn wir davon ausgehen, dass keine Permutation (d.h. keine Ziehung der Hüte) wahrscheinlicher ist als eine andere, gehen wir von einem Laplace-Experiment aus, d.h. für jedes Ereignis $A \in \mathcal{A}$ gilt

$$\mathbb{P}(A) =$$

- Ereignisse: Wir suchen die Wahrscheinlichkeit des Ereignisses

K : „Keine Person erhält ihren eigenen Hut.“

Um dessen Wahrscheinlichkeit zu berechnen, werden wir zusätzlich folgende Ereignisse benötigen:

E_i : „Person i erhält ihren eigenen Hut.“

- Wahrscheinlichkeiten: E_i beschreibt die Menge aller Permutationen von $(1, \dots, n)$, die i auf i abbilden. Damit gilt für $i = 1, \dots, n$:

$$|E_i| = \quad \implies \quad \mathbb{P}(E_i) =$$

Mit der gleichen Überlegung bestimmt man die Wahrscheinlichkeit, dass $k \leq n$ fest bestimmte Personen ihre Hüte bekommen. Dazu seien $1 \leq i_1 < i_2 < \dots < i_k \leq n$. Dann ist

$$|E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k}| = \quad \implies \quad \mathbb{P}(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k}) =$$

K kann man mit Hilfe von E_1, \dots, E_n folgendermaßen darstellen:

$$K = \bigcap_{i=1}^n E_i^c = \left(\quad \right)^c$$

¹Der Programmcode ist in Anhang C zu finden.

Mit Hilfe der Siebformel berechnet man

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n E_i\right) &= \sum_{i=1}^n \mathbb{P}(E_i) - \sum_{1 \leq j < k \leq n} \mathbb{P}(E_j \cap E_k) + \sum_{1 \leq i < j < k \leq n} \mathbb{P}(E_i \cap E_j \cap E_k) + \dots \\ &\quad + (-1)^{n+1} \mathbb{P}(E_1 \cap \dots \cap E_n) \\ &= \sum_{i=1}^n \mathbb{P}(E_i) - \sum_{1 \leq j < k \leq n} \mathbb{P}(E_j \cap E_k) + \sum_{1 \leq i < j < k \leq n} \mathbb{P}(E_i \cap E_j \cap E_k) + \dots + (-1)^{n+1} \mathbb{P}(E_1 \cap \dots \cap E_n). \end{aligned}$$

Somit ist

$$\begin{aligned} \mathbb{P}(K) &= 1 - \mathbb{P}(K^c) = \\ &= \\ &\xrightarrow{n \rightarrow \infty} \end{aligned}$$

1.5 Relative Häufigkeiten und Wahrscheinlichkeiten

Wie kommt man auf Wahrscheinlichkeiten? Wenn wir ein Laplace-Experiment haben, so können wir mit dem Hilfsmitteln der Kombinatorik die günstigen und möglichen Fälle zählen und der Quotient aus beiden ist die Wahrscheinlichkeit des Ereignisses. Was ist jedoch in Situationen, in denen wir kein Modell haben? Im Zweifel orientieren wir uns an der wirklichen Welt und verwenden beobachtete Ergebnisse eines Zufallsexperiments.

Beispiel 1.47. Bei der Bürgermeisterwahl in Hintertupfingen – einer Gemeinde mit 1000 Einwohnern – haben 660 Wähler dem Kandidaten Anton ihre Stimme gegeben. Wenn Sie nach dem Wahltag nach Hintertupfingen fahren und rein zufällig einen Wähler nach seiner Wahlentscheidung fragen – mit welcher Wahrscheinlichkeit wählte dieser Anton?

Wenn wir davon ausgehen, dass die Hintertupfinger nicht lügen und Ihnen alle mit gleicher Wahrscheinlichkeit begegnen könnten, werden wir die Wahrscheinlichkeit als 0,66 festlegen, denn das ist der Anteil der Wähler, die für Anton gestimmt haben.

Während die Zahl 660 die absolute Häufigkeit darstellt, ist 0,66 die relative Häufigkeit, mit der unter allen Wählern auf Anton gesetzt wurde.

Definition 1.48

Die absolute Häufigkeit $H_n(A)$ gibt an, wie oft das Ereignis A bei einem n -mal wiederholten Zufallsexperiment auftritt. Die zugehörige relative Häufigkeit ist definiert als $h_n(A) := \frac{H_n(A)}{n}$.

Bemerkung 1.49. In der beschreibenden Statistik spricht man an dieser Stelle von Merkmalsträgern (die Wähler), Merkmalen (Wahlentscheidung) und Ausprägungen (alle verfügbaren Kandidaten). Wir wollen hier jedoch vorerst bei den Bezeichnungen bleiben, die in der Wahrscheinlichkeitstheorie üblich sind.

Die folgende Gesetzmäßigkeit wurde von Jakob Bernoulli (1655 - 1705) als *theorema aureum*, also als *goldener Satz* bezeichnet.

Empirisches Gesetz der großen Zahlen

Ist A ein Ereignis eines Zufallsexperiments, so stabilisieren sich bei einer hinreichend großen Anzahl n von Durchführungen dieses Experiments die relativen Häufigkeiten $h_n(A)$.

Wenn es kein Modell gibt, aus welchem Wahrscheinlichkeiten abgeleitet werden, aber dafür empirische Daten über ein Zufallsexperiment vorliegen, so setzt man die Wahrscheinlichkeit $\mathbb{P}(A)$ gleich der beobachteten relativen Häufigkeit $h_n(A)$.

Kapitel 2

Bedingte Wahrscheinlichkeiten

Ziele

- die Formeln für bedingte Wahrscheinlichkeiten inkl. Satz von Bayes und der Regel der totalen Wahrscheinlichkeit kennen, herleiten und anwenden
- aus Texten gegebene (bedingte) Wahrscheinlichkeiten herauslesen und damit gesuchte (bedingte) Wahrscheinlichkeiten berechnen
- den Unterschied zwischen unvereinbaren und (paarweise) stochastisch unabhängigen Ereignissen kennen
- die Stetigkeit des Wahrscheinlichkeitsmaßes kennen und beweisen können

2.1 Definition bedingter Wahrscheinlichkeiten

Beispiel 2.1. Wir wollen folgende Fragen beantworten:

1. Wie wahrscheinlich ist es, dass beim Werfen zweier fairer Würfel die Augensumme 10 beträgt?
2. Wie wahrscheinlich ist es, dass beim Werfen zweier fairer Würfel der erste Würfel eine 4 zeigt?
3. Wie wahrscheinlich ist es, dass beim Werfen zweier fairer Würfel der erste Würfel eine 4 zeigt und die Augensumme 10 beträgt?
4. Von zwei fairen Würfeln wird der erste geworfen und zeigt eine 4. Wie wahrscheinlich ist es nun, dass nach dem Werfen des zweiten Würfels die Augensumme 10 beträgt?
5. Von zwei geworfenen fairen Würfeln ist nur bekannt, dass ihre Augenzahl 10 beträgt. Wie wahrscheinlich ist es, dass der erste Würfel eine 4 gezeigt hat?

Für die ersten drei Fragen werden wir immer den gleichen Grundraum als Modell verwenden, nämlich $\Omega = \{1, \dots, 6\}^2$ mit $\mathcal{A} = \mathcal{P}(\Omega)$. Davon ausgehend, dass alle Elementarereignisse gleich wahrscheinlich sind, haben wir es mit einem Laplace-Experiment zu tun, d.h. $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \mathcal{A}$, wobei $|\Omega| = 36$ gilt.

1. Wir suchen die Wahrscheinlichkeit des Ereignisses

$$A = \{\omega = (\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = 10\}.$$

Anhand einer Tabelle kann man leicht die Anzahl der Elementarereignisse ablesen:

+	◻	◻	◻	◻	◻	◻
■	2	3	4	5	6	7
■	3	4	5	6	7	8
■	4	5	6	7	8	9
■	5	6	7	8	9	10
■	6	7	8	9	10	11
■	7	8	9	10	11	12

$$|A| = 3 \implies \mathbb{P}(A) = \frac{3}{36} = \frac{1}{12}.$$

2. Das zu untersuchende Ereignis ist

$$A' = \{\omega = (\omega_1, \omega_2) \in \Omega \mid \omega_1 = 4\}.$$

In diesem Fall ist $|A'| = 6$ und somit $\mathbb{P}(A') = \frac{6}{36} = \frac{1}{6}$.

3. Das zu untersuchende Ereignis ist

$$\begin{aligned} A'' &= A \cap A' \\ &= \{\omega = (\omega_1, \omega_2) \in \Omega \mid \omega_1 = 4, \omega_1 + \omega_2 = 10\} \\ &= \{\omega = (\omega_1, \omega_2) \in \Omega \mid \omega_1 = 4, \omega_2 = 6\} \end{aligned}$$

In diesem Fall ist $|A''| = 1$ und somit $\mathbb{P}(A'') = \frac{1}{36}$.

Für die letzten zwei Fragen müssen wir das Modell jeweils verändern.

4. Sei zunächst

$$\Omega^* := \{(\omega_1, \omega_2) \in \{1, \dots, 6\}^2 \mid \omega_1 = 4\} \quad \text{und} \quad \mathcal{A}^* = \mathcal{P}(\Omega^*).$$

Es ist $|\Omega^*| = 6$ und wir setzen $\mathbb{P}^*(A) = \frac{|A|}{|\Omega^*|}$. Das zu untersuchende Ereignis ist

$$A^* = \{\omega = (\omega_1, \omega_2) \in \Omega^* \mid \omega_1 + \omega_2 = 10\}.$$

In diesem Fall ist $|A^*| = 1$ und somit $\mathbb{P}^*(A^*) = \frac{1}{6}$.

Die Wahrscheinlichkeit dafür, dass die Augensumme 10 beträgt, hat sich in dem Moment verändert, indem das Ergebnis des ersten Würfelwurfes bekannt wurde.

5. Sei nun

$$\Omega^{**} := \{(\omega_1, \omega_2) \in \{1, \dots, 6\}^2 \mid \omega_1 + \omega_2 = 10\} \quad \text{und} \quad \mathcal{A}^{**} = \mathcal{P}(\Omega^{**}).$$

Es ist $|\Omega^{**}| = 3$ und wir setzen $\mathbb{P}^{**}(A) = \frac{|A|}{|\Omega^{**}|}$. Das zu untersuchende Ereignis ist

$$A^{**} = \{\omega = (\omega_1, \omega_2) \in \Omega^{**} \mid \omega_1 = 4\}.$$

In diesem Fall ist $|A^{**}| = 1$ und somit $\mathbb{P}^{**}(A^{**}) = \frac{1}{3}$.

Bemerkung 2.2. Werden Ereignisse in Textform beschrieben, ist es wichtig, genau zu beschreiben, ob das Eintreten eines Ereignisses vorausgesetzt wird oder gleichzeitig mit einem anderen eintreten soll. Im vorherigen Beispiel sind das die Fälle 3 und 4 bzw. 3 und 5, welche je zwei vollkommen verschiedene Herangehensweisen verlangen.

Wir können in Beispiel 2.1 folgenden Zusammenhang beobachten: Die bedingte Wahrscheinlichkeit dafür, Augensumme 10 zu würfeln (A), wenn im ersten Wurf eine 4 fiel (A'), ist der Quotient $\frac{\mathbb{P}(A \cap A')}{\mathbb{P}(A')}$; die bedingte Wahrscheinlichkeit dafür, zuerst eine 4 zu würfeln (A'), wenn die Augensumme 10 betragen soll (A), ist der Quotient $\frac{\mathbb{P}(A \cap A')}{\mathbb{P}(A)}$. Verallgemeinert kann man sagen:

Ist ein Ereignis B bereits eingetreten, so tritt A genau dann ein, wenn $A \cap B$ eintritt. Somit ist die Wahrscheinlichkeit dafür, dass A eintritt unter der Voraussetzung, dass B bereits eingetreten ist, gegeben durch

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

Hierbei müssen wir natürlich voraussetzen, dass wir nicht durch Null teilen.

Definition 2.3

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A, B \in \mathcal{A}$. Die bedingte Wahrscheinlichkeit des Eintretens von A gegeben, dass B bereits eingetreten ist, ist definiert durch

$$\mathbb{P}(A|B) := \begin{cases} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, & \mathbb{P}(B) > 0, \\ 0, & \mathbb{P}(B) = 0. \end{cases}$$

In einer Übungsaufgabe wurde folgendes Resultat bereits bewiesen:

Satz 2.4

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$. Dann ist

$$\mathbb{P}_B: \mathcal{A} \rightarrow [0, 1], \quad A \mapsto \mathbb{P}_B(A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A|B)$$

ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) .

Festlegung: In allen folgenden Definitionen und Sätzen sei, sofern nichts anderes gesagt wird, $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. 

Beispiel 2.5. In einem Laden hat der Azubi 5 neue, 10 alte und 10 defekte Akkus zusammen in eine Kiste gelegt. Sie sind optisch nicht voneinander zu unterscheiden und passen alle in das neue Handy der Marke MathX. 

- Mit neuen Akkus kann man das Handy mindestens 48 Stunden benutzen.
- Mit einem gebrauchten Akku ist nach 6 Stunden Nutzung bereits Schluss.
- Mit einem defekten Akku funktioniert das Handy gar nicht.

Mit einem zufällig aus der Kiste gegriffenen Akku funktioniert das Handy. Wie groß ist die Wahrscheinlichkeit, dass es auch nach 12 Stunden Nutzung noch funktionieren wird?

Antwort: Bevor wir etwas berechnen können, müssen geeignete Ereignisse definiert und die im Text stehenden Wahrscheinlichkeiten in Formeln ausgedrückt werden.

Ereignisse: Seien N und F folgende Ereignisse:

N : „Der Akku ist neu.“

F : „Der Akku funktioniert.“

Weitere Ereignisse benötigen wir nicht, da sie mit Hilfe der bekannten Mengenoperationen durch diese Ereignisse ausgedrückt werden können.

gegebene Wahrscheinlichkeiten: Gehen wir von einem Laplace-Experiment aus, d.h. jeder der $|\Omega| = 25$ Akkus kann mit gleicher Wahrscheinlichkeit ausgewählt werden, so gilt:

$$\mathbb{P}(N) =$$

$$\mathbb{P}(F) =$$

gesuchte Wahrscheinlichkeit: Mit den gewählten Bezeichnungen der Ereignisse ist die gesuchte Wahrscheinlichkeit $\mathbb{P}(\quad)$.

Berechnung: Es gilt

$$\mathbb{P}(\quad) = \frac{\quad}{\quad} = \frac{\quad}{\quad} =$$

Bemerkung 2.6. Den Wahrscheinlichkeiten aus dem Beispiel liegt folgendes Modell zu Grunde, wobei n_k für den k -ten neuen Akku, g_k für den k -ten gebrauchten Akku und d_k für den k -ten defekten Akku steht.

$$\begin{aligned}\Omega &= \{n_1, \dots, n_5, g_1, \dots, g_{10}, d_1, \dots, d_{10}\}, \\ \mathcal{A} &= \mathcal{P}(\Omega), \\ \mathbb{P}(A) &= \frac{|A|}{|\Omega|}, \quad \forall A \in \mathcal{A}.\end{aligned}$$

Sind im Text absolute oder relative Häufigkeiten gegeben, so ist dies das zugehörige Standardmodell und wird nicht mehr explizit genannt. Sofern dies nicht explizit anders gesagt wird, genügt es, die angegebenen Häufigkeiten sofort in Wahrscheinlichkeiten zu übersetzen.



Beispiel 2.7. Eine Urne enthält 8 rote und 4 blaue Kugeln. Wir ziehen 2 Kugeln ohne Zurücklegen. Wie groß ist die Wahrscheinlichkeit, dass beide Kugeln rot sind?

1. Lösungsweg: Wir modellieren das Experiment als Laplace-Experiment.

- $\Omega = \{(\omega_1, \omega_2) \in \{r_1, \dots, r_8, b_1, \dots, b_4\}^2 \mid \omega_1 \neq \omega_2\}$ mit $|\Omega| =$
- $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \mathcal{A}$
- gesuchtes Ereignis:

$$A = \{\omega \in \Omega \mid \omega_1, \omega_2 \in \{r_1, \dots, r_8\}\} \quad \text{mit} \quad |A| =$$

- Wahrscheinlichkeit: $\mathbb{P}(A) = \frac{\quad}{\quad} =$

2. Lösungsweg: Wir berechnen bedingte Wahrscheinlichkeiten. Dazu betrachten wir folgende Ereignisse:

R_1 : „Die erste Kugel ist rot.“

R_2 : „Die zweite Kugel ist rot.“

Wir suchen $\mathbb{P}(R_1 \cap R_2)$. Kombinatorisch leicht zu bestimmen sind

$$\mathbb{P}(R_1) = \quad \text{und} \quad \mathbb{P}(R_2|R_1) =$$

Die gesuchte Wahrscheinlichkeit berechnet man indem man die Definition für die bedingte Wahrscheinlichkeit nach dem gesuchten Term umstellt.

$$\mathbb{P}(R_1 \cap R_2) =$$

Diese Möglichkeit der Berechnung des Schnittes von Ereignissen über die bedingten Wahrscheinlichkeiten lässt sich auf mehr als zwei Ereignisse verallgemeinern.

Satz 2.8: Multiplikationsregel

Seien $A_1, \dots, A_n \in \mathcal{A}$ Ereignisse. Dann gilt

$$\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Beweis. Die Aussage soll für alle $n \in \mathbb{N}$ gezeigt werden, daher beweisen wir sie über vollständige Induktion. Für $n = 1$ ist nichts zu zeigen und für $n = 2$ folgt die Aussage sofort aus der Definition der bedingten Wahrscheinlichkeit.

Nehmen wir als Induktionsvoraussetzung (IV) an, die Aussage ist für $n - 1$ Ereignisse gezeigt. Dann gilt

$$\begin{aligned} \mathbb{P}\left(\bigcap_{k=1}^n A_k\right) &= \mathbb{P}\left((A_1 \cap A_2) \cap \bigcap_{k=3}^n A_k\right) \\ &\stackrel{\text{IV}}{=} \mathbb{P}(A_1 \cap A_2) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|(A_1 \cap A_2) \cap A_3 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap A_2 \cap A_3 \cap \dots \cap A_{n-1}), \end{aligned}$$

womit gezeigt ist, dass die Aussage auch für n Ereignisse gilt. \square

Beispiel 2.9. Ein Kartenspiel mit 52 Karten (siehe Anhang D) wird zufällig in vier gleich große Haufen mit je 13 Karten aufgeteilt. Wie groß ist die Wahrscheinlichkeit, dass in jedem Haufen ein Ass ist?

Antwort: Wir definieren uns geeignete Ereignisse:

A_1 : „Kreuz-Ass ist in einem Haufen.“

A_2 : „Kreuz-Ass und Pik-Ass sind in verschiedenen Haufen.“

A_3 : „Kreuz-Ass, Pik-Ass und Karo-Ass sind in verschiedenen Haufen.“

A_4 : „Kreuz-Ass, Pik-Ass, Karo-Ass und Herz-Ass sind in verschiedenen Haufen.“

Die gesuchte Wahrscheinlichkeit $\mathbb{P}(A_4)$ berechnen wir mit Hilfe der bedingten Wahrscheinlichkeiten:

$$\begin{aligned}\mathbb{P}(A_4) &= \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \mathbb{P}(A_4|A_1 \cap A_2 \cap A_3) \\ &= 1 \cdot \frac{52-13}{52-1} \cdot \frac{52-2 \cdot 13}{52-2} \cdot \frac{52-3 \cdot 13}{52-3} \\ &= 1 \cdot \frac{39}{51} \cdot \frac{26}{50} \cdot \frac{13}{49} \approx 0,105.\end{aligned}$$

2.2 Satz von Bayes und Gesetz der totalen Wahrscheinlichkeit

Beispiel 2.10. Es gibt zwei Urnen – in der ersten sind 7 rote und 3 weiße Kugeln, in der zweiten sind 9 weiße und 1 rote Kugel. Es wird eine Kugel aus einer zufällig gewählten Urne entnommen. Mit welcher Wahrscheinlichkeit wird eine rote Kugel entnommen?

Antwort: Wir definieren folgende Ereignisse:

$U_1 :=$ „Die Kugel wird aus der 1. Urne entnommen.“

$U_2 :=$ „Die Kugel wird aus der 2. Urne entnommen.“

$R :=$ „Die entnommene Kugel ist rot.“

Hierbei gilt $U_2 = U_1^c$, jedoch behalten wir der Übersichtlichkeit halber beide Ereignisse bei. Aus dem Text können wir folgende Wahrscheinlichkeiten ablesen:

$$\mathbb{P}(U_1) = \mathbb{P}(U_2) = \frac{1}{2}, \quad \mathbb{P}(R|U_1) = \frac{7}{10} \quad \text{und} \quad \mathbb{P}(R|U_2) = \frac{1}{10}.$$

Gesucht ist die Wahrscheinlichkeit $\mathbb{P}(R)$. Das Ereignis R können wir mit Hilfe von U_1 und U_2 in zwei disjunkte Ereignisse zerlegen:

$$R = (R \cap U_1) \dot{\cup} (R \cap U_2).$$

Indem wir die Formel für die bedingte Wahrscheinlichkeit nach der Wahrscheinlichkeit des Schnittes umstellen, können wir die bekannten bedingten Wahrscheinlichkeiten verwenden:

$$\begin{aligned}\mathbb{P}(R) &= \mathbb{P}(R \cap U_1) + \mathbb{P}(R \cap U_2) = \mathbb{P}(R|U_1) \cdot \mathbb{P}(U_1) + \mathbb{P}(R|U_2) \cdot \mathbb{P}(U_2) \\ &= \frac{7}{10} \cdot \frac{1}{2} + \frac{1}{10} \cdot \frac{1}{2} = \frac{2}{5}.\end{aligned}\tag{2.2.1}$$

Die Formel, die wir in (2.2.1) für zwei komplementäre Ereignisse hergeleitet haben, können wir auch für mehr als zwei Ereignisse verallgemeinern, sofern deren disjunkte Vereinigung den Grundraum ergibt.

Satz 2.11: Gesetz der totalen Wahrscheinlichkeit

Seien $A_1, A_2, \dots \in \mathcal{A}$ paarweise disjunkte Ereignisse, deren disjunkte Vereinigung den ganzen Grundraum ergibt, d.h. $\dot{\bigcup}_{k \in \mathbb{N}} A_k = \Omega$. Dann gilt für jedes beliebige Ereignis $B \in \mathcal{A}$

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

Bemerkung 2.12. In der Literatur wird bisweilen zusätzlich verlangt, dass $\mathbb{P}(A_k) > 0$ ist für alle $k \in \mathbb{N}$. Warum können wir diese Bedingung weglassen? Wäre beispielsweise $\mathbb{P}(A_1) = 0$, so wäre auch $\mathbb{P}(B|A_1) = 0$ und der erste Summand auf der rechten Seite ist Null und ändert somit nichts an dem Wert der Reihe (oder endlichen Summe).

Beweis. B hat die Zerlegung in paarweise disjunkte Mengen

$$B = \dot{\bigcup}_{k \in \mathbb{N}} (B \cap A_k).$$

Indem wir die σ -Additivität des Wahrscheinlichkeitsmaßes ausnutzen, ergibt sich

$$\mathbb{P}(B) = \sum_{k \in \mathbb{N}} \mathbb{P}(B \cap A_k).$$

- Falls $\mathbb{P}(A_k) = 0$ ist, so ist wegen $B \cap A_k \subseteq A_k$ auch $\mathbb{P}(B \cap A_k) = 0$. Andererseits ist dann auch $0 = \mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k)$, d.h. $\mathbb{P}(B \cap A_k) = \mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k)$.
- Falls $\mathbb{P}(A_k) > 0$ ist, so gilt nach Definition der bedingten Wahrscheinlichkeit durch Umstellen nach der Wahrscheinlichkeit des Schnittes $\mathbb{P}(B \cap A_k) = \mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k)$.

Insgesamt haben wir damit die Darstellung

$$\mathbb{P}(B) = \sum_{k \in \mathbb{N}} \mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k).$$

□

Beispiel 2.13 (Fortsetzung von Beispiel 2.10). So wie in Beispiel 2.10 gibt es zwei Urnen – in der ersten sind 7 rote und 3 weiße Kugeln, in der zweiten sind 9 weiße und 1 rote Kugel. Es wird eine Kugel aus einer zufällig gewählten Urne entnommen. Wenn die entnommene Kugel rot ist, wie wahrscheinlich stammt diese aus der ersten Urne?

Antwort: Mit den Ereignissen aus Beispiel 2.10 suchen wir die Wahrscheinlichkeit $\mathbb{P}(U_1|R)$. Indem wir wieder die Definition der bedingten Wahrscheinlichkeit verwenden und die bereits bekannten Wahrscheinlichkeiten einsetzen erhalten wir die gesuchte Wahrscheinlichkeit als

$$\begin{aligned} \mathbb{P}(U_1|R) &= \frac{\mathbb{P}(U_1 \cap R)}{\mathbb{P}(R)} = \frac{\mathbb{P}(R|U_1)\mathbb{P}(U_1)}{\mathbb{P}(R)} \\ &= \frac{\frac{7}{10} \cdot \frac{1}{2}}{\frac{2}{5}} = \frac{7 \cdot 1 \cdot 5}{10 \cdot 2 \cdot 2} = \frac{7}{8}. \end{aligned} \tag{2.2.2}$$

Die in (2.2.2) hergeleitete Formel trägt den Namen von *Thomas Bayes* (1701-1761) und sie entstammt seinem posthum veröffentlichten Essay *An Essay Towards Solving a Problem in the Doctrine of Chances*.

Satz 2.14: Satz von Bayes

Seien $A, B \in \mathcal{A}$. Ist $\mathbb{P}(A) > 0$, so gilt

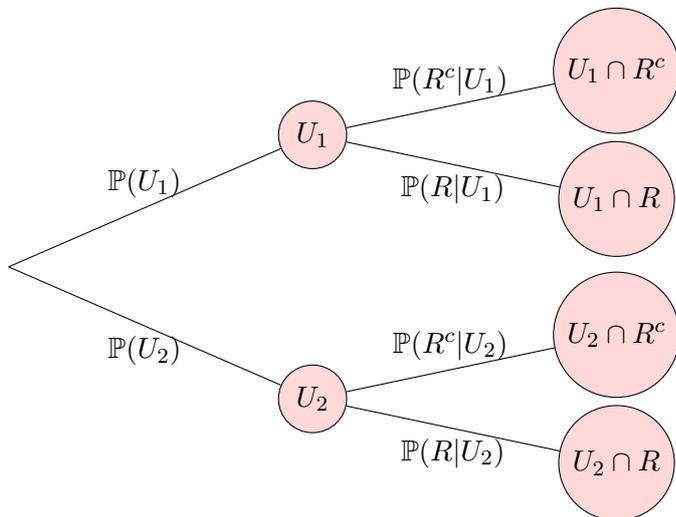
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)}.$$

Beweis. Für zwei Ereignisse $A, B \in \mathcal{A}$ gilt stets

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

Ist $\mathbb{P}(A) > 0$, so können wir alle Terme der Gleichung dadurch teilen und erhalten die Behauptung. \square

Bemerkung 2.15. Die Berechnungen in Beispiel 2.10 und 2.13 kann man auch anhand eines Baumdiagramms verdeutlichen:



Beispiel 2.16. Eine Versicherung teilt Personen in zwei Klassen ein: solche mit hohem und solche mit geringem Unfallrisiko. Gemäß ihrer internen Statistik haben Personen mit hohem Risiko im ersten Versicherungsjahr mit Wahrscheinlichkeit 0,4 (mindestens) einen Unfall, Personen mit geringem Risiko jedoch nur mit Wahrscheinlichkeit 0,2. Es werden 30% der potentiellen Versicherungsnehmer zu den Personen mit hohem Risiko gezählt.

- Wie groß ist die Wahrscheinlichkeit, dass ein neuer Versicherungsnehmer (mindestens) einen Unfall im ersten Jahr hat?
- Angenommen, der neue Versicherungsnehmer hat einen Unfall im ersten Jahr. Wie groß ist die Wahrscheinlichkeit, dass er zur Klasse der Personen mit hohem Risiko gehört?

Antwort: Zunächst formulieren wir wieder geeignete Ereignisse, notieren die gegebenen und gesuchten Wahrscheinlichkeiten und erst dann berechnen wir die gesuchten Größen.

$H :=$ „Versicherungsnehmer gehört zur Klasse mit hohem Unfallrisiko.“

$U :=$ „Versicherungsnehmer hat (mindestens) einen Unfall im ersten Jahr.“

Wahrscheinlichkeiten im Text:

$$\mathbb{P}(H) = \quad \mathbb{P}(\quad) = \quad \mathbb{P}(\quad) =$$

Gesucht werden die Wahrscheinlichkeit $\mathbb{P}(\quad)$ und $\mathbb{P}(\quad)$.

$$\mathbb{P}(\quad) =$$

$$\mathbb{P}(\quad) =$$

Korollar 2.17

Seien $A_1, \dots, A_n \in \mathcal{A}$ paarweise disjunkte Ereignisse mit $\bigcup_{k \in \{1, \dots, n\}} A_k = \Omega$. Ist $B \in \mathcal{A}$ ein weiteres Ereignis, so gilt für alle $k \in \{1, \dots, n\}$

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Beweis. Es gilt

$$\mathbb{P}(A_k|B) \stackrel{\text{Bayes}}{=} \frac{\mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k)}{\mathbb{P}(B)} \stackrel{\text{totale W.}}{=} \frac{\mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

□

Beispiel 2.18. Von drei Karten ist eine auf beiden Seiten rot, eine auf beiden Seiten weiß und eine auf einer Seite rot und auf der anderen weiß. Eine der Karten wird zufällig gezogen und von einer Seite betrachtet: Sie ist rot. Mit welcher Wahrscheinlichkeit ist die andere Seite weiß?



Antwort: Zunächst definieren wir passende Ereignisse.

- A : „Beide Seiten der gezogenen Karte sind rot.“
- B : „Beide Seiten der gezogenen Karte sind weiß.“
- C : „Die gezogene Karte hat eine weiße und eine rote Seite.“
- R : „Eine Seite der gezogenen Karte ist rot.“

Aus dem Text herauszulesende Wahrscheinlichkeiten:

$$\begin{array}{lll} \mathbb{P}(A) = & \mathbb{P}(B) = & \mathbb{P}(C) = \\ \mathbb{P}(\quad | \quad) = & \mathbb{P}(\quad | \quad) = & \mathbb{P}(\quad | \quad) = \end{array}$$

Gesucht ist die Wahrscheinlichkeit $\mathbb{P}(\quad | \quad)$.

Berechnung:

$$\begin{aligned} \mathbb{P}(\quad | \quad) &= \text{-----} \\ &= \text{-----} = \end{aligned}$$

2.3 Stochastische Unabhängigkeit

Definition 2.19

Zwei Ereignisse $A, B \in \mathcal{A}$ heißen *stochastisch unabhängig*, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ ist.

Bemerkung 2.20 (Interpretation). Ist $\mathbb{P}(B) > 0$, so ist

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

A und B sind in diesem Fall also genau dann stochastisch unabhängig, wenn die bedingte Wahrscheinlichkeit von A gegeben B gleich der unbedingten Wahrscheinlichkeit von A ist.

Beispiel 2.21. Eine Spielkarte wird rein zufällig aus einem Kartenspiel mit 52 Karten ausgewählt. Wir betrachten die Ereignisse

$$A : \text{„Die Karte ist ein Ass.“} \quad \text{und} \quad B : \text{„Die Karte hat die Farbe Pik.“}$$

Sind A und B stochastisch unabhängig?

Antwort: Wir gehen von einem Laplace-Experiment mit $|\Omega| = 52$ aus. Es gilt $|A| = 4$, $|B| = 13$ und für

$$A \cap B : \text{„Die Karte ist ein Pik-Ass.“}$$

gilt $|A \cap B| = 1$. Somit ist

$$\mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{4}{52} \cdot \frac{13}{52} = \frac{1}{52} = \mathbb{P}(A \cap B),$$

d.h. A und B sind stochastisch unabhängig.

Satz 2.22

Seien $A, B \in \mathcal{A}$ stochastisch unabhängige Ereignisse. Dann gilt:

- i) A und B^c sind stochastisch unabhängig.
- ii) A^c und B^c sind stochastisch unabhängig.

Beweis. Es gelten folgende Gleichungen:

- i) $\mathbb{P}(A) \cdot \mathbb{P}(B^c) = \mathbb{P}(A) \cdot (1 - \mathbb{P}(B)) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A \cap B^c)$.
- ii) $\mathbb{P}(A^c) \cdot \mathbb{P}(B^c) = (1 - \mathbb{P}(A)) \cdot \mathbb{P}(B^c) = \mathbb{P}(B^c) - \mathbb{P}(A \cap B^c) = \mathbb{P}(A^c \cap B^c)$.

□

Stochastische Unabhängigkeit zweier Ereignisse kann man auch an der Vierfeldertafel erkennen. In einer Tabelle trägt man dafür folgende Werte ein:

	$\mathbb{P}(B)$	$\mathbb{P}(B^c)$
$\mathbb{P}(A)$	$\mathbb{P}(A \cap B)$	$\mathbb{P}(A \cap B^c)$
$\mathbb{P}(A^c)$	$\mathbb{P}(A^c \cap B)$	$\mathbb{P}(A^c \cap B^c)$

A und B sind genau dann stochastisch unabhängig, wenn die gelb hinterlegten Einträge sich als Produkt der Wahrscheinlichkeiten in den Randspalten ergeben.

Wie kann man stochastische Unabhängigkeit von mehr als zwei Ereignissen definieren?



Beispiel 2.23. Wir werfen nacheinander zwei faire Würfel und betrachten folgende Ereignisse:

A : „Die Augensumme beider Würfel ist 7.“

B : „Der erste Würfel zeigt die Augenzahl 4.“

C : „Der zweite Würfel zeigt die Augenzahl 3.“

Wir modellieren das Zufallsexperiment als Laplace-Experiment, d.h. $\Omega = \{1, \dots, 6\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \mathcal{A}$. Damit gilt zunächst für die drei Ereignisse

$$\mathbb{P}(A) = \quad \quad \quad \mathbb{P}(B) = \quad \quad \quad \mathbb{P}(C) =$$

Für alle Schnitte von Ereignissen gilt zudem

$$\begin{aligned} A \cap B &= & \implies \mathbb{P}(A \cap B) &= \\ A \cap C &= & \implies \mathbb{P}(A \cap C) &= \\ B \cap C &= & \implies \mathbb{P}(B \cap C) &= \\ A \cap B \cap C &= & \implies \mathbb{P}(A \cap B \cap C) &= \end{aligned}$$

Die drei Ereignisse sind also paarweise stochastisch unabhängig, d.h. A und B , A und C , sowie B und C sind jeweils Paare von stochastisch unabhängigen Ereignissen. Allerdings sollten wir die drei Ereignisse nicht stochastisch unabhängig nennen, da aus B und C sofort A folgt und es gilt zudem $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$.

Definition 2.24

Sei $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A}$ eine Folge von Ereignissen.

1. Die Ereignisse $(A_n)_{n \in \mathbb{N}}$ heißen *paarweise stochastisch unabhängig*, falls für alle $k, j \in \mathbb{N}$ mit $k \neq j$ die zwei Ereignisse A_k und A_j stochastisch unabhängig sind, d.h. falls $\mathbb{P}(A_k \cap A_j) = \mathbb{P}(A_k) \cdot \mathbb{P}(A_j)$ gilt.
2. Die Ereignisse $(A_n)_{n \in \mathbb{N}}$ heißen *stochastisch unabhängig*, falls

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

für jede endliche Teilmenge $I \subset \mathbb{N}$ gilt.

Die Definition gilt analog für endliche Mengen von Ereignissen, wenn die Indexmenge \mathbb{N} durch $\{1, \dots, n\}$ ersetzt wird.

Bemerkung 2.25. Jede Menge stochastisch unabhängiger Ereignisse ist insbesondere paarweise stochastisch unabhängig. Die Umkehrung dieser Aussage gilt jedoch nicht, was bereits in Beispiel 2.23 gezeigt wurde. Das gleiche Phänomen kennen wir aus der linearen Algebra: Die Vektoren

$$u = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \quad \text{und} \quad w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

sind zwar paarweise linear unabhängig, da kein Vektor ein Vielfaches eines anderen ist, aber nicht insgesamt, da $w = u + v$ ist.

Für drei Ereignisse A , B und C müssen wir definitionsgemäß für die stochastische Unabhängigkeit folgende Eigenschaften nachweisen:

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B), & \mathbb{P}(A \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(C), & \mathbb{P}(B \cap C) &= \mathbb{P}(B) \cdot \mathbb{P}(C), \\ \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C). \end{aligned}$$

Folgendes Beispiel zeigt, dass wir auf den Nachweis der paarweisen Unabhängigkeit nicht verzichten können, wenn die letzte Eigenschaft bereits nachgewiesen wurde.



Beispiel 2.26. Eine Münze wird dreimal geworfen, d.h. um ein Laplace-Experiment zu erhalten setzen wir $\Omega = \{K, Z\}^3$, wobei K für Kopf und Z für Zahl steht. Wir betrachten folgende Ereignisse:

A : „Mindestens zwei der drei Münzwürfe zeigen Kopf.“

B : „Beim ersten Wurf erscheint Kopf.“

C : „Beim zweiten und dritten Wurf liegt die gleiche Seite der Münze oben.“

In diesem Fall gilt

$$\mathbb{P}(A) = \quad \mathbb{P}(B) = \quad \mathbb{P}(C) = \quad \mathbb{P}(A \cap B) = \quad \mathbb{P}(A \cap B \cap C) =$$

und somit

$$\mathbb{P}(A) \cdot \mathbb{P}(B) = \quad \neq \mathbb{P}(A \cap B) \quad \text{und} \quad \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \quad = \mathbb{P}(A \cap B \cap C).$$

2.4 Versuchsfolgen

Zerfällt ein Experiment in eine (endliche) Folge von Teilerperimenten, die sich gegenseitig nicht beeinflussen, so nennen wir die Teilerperimente *Versuche*. In diesem Abschnitt sehen wir uns mehrere Beispiele von Versuchsfolgen an.

Beispiel 2.27. Eine unendliche Folge von Versuchen mit zwei möglichen Ausgängen – Erfolg und Misserfolg – wird durchgeführt. Sei $p \in (0, 1)$ die Wahrscheinlichkeit für einen Erfolg in einem Versuch.¹ Wie groß ist die Wahrscheinlichkeit, dass

- mindestens ein Erfolg in den ersten n Versuchen erzielt wird?
- genau k Erfolge ($1 \leq k \leq n$) in den ersten n Versuchen erzielt werden?

Antwort: Wir definieren das Ereignis

E_k : „Erfolg im k -ten Versuch“.

- Das Ereignis $E_1^c \cap \dots \cap E_n^c$ entspricht keinem Erfolg in den ersten n Versuchen, also gerade dem Gegenereignis des gesuchten Ereignisses. Die gesuchte Wahrscheinlichkeit ist somit

$$\begin{aligned} \mathbb{P}((E_1^c \cap \dots \cap E_n^c)^c) &= 1 - \mathbb{P}(E_1^c \cap \dots \cap E_n^c) \\ &= 1 - \mathbb{P}(E_1^c) \cdot \dots \cdot \mathbb{P}(E_n^c) \\ &= 1 - \prod_{k=1}^n (1 - \mathbb{P}(E_k)) \\ &= 1 - (1 - p)^n. \end{aligned}$$

Für $n \rightarrow \infty$ konvergiert diese Wahrscheinlichkeit gegen 1.

¹Im Fall $p = \frac{1}{2}$ kann man sich unter diesem Experiment das unendlich oft wiederholte Werfen einer fairen Münze vorstellen, bei dem eine der beiden Seiten als Erfolg gewertet wird.

- b) Sei $\mathcal{I} := \{I \subseteq \{1, \dots, n\} \mid |I| = k\}$ die Menge der k -elementigen Teilmengen von $\{1, \dots, n\}$. Aus der Kombinatorik wissen wir, dass $|\mathcal{I}| = \binom{n}{k}$ ist. Wir suchen die Wahrscheinlichkeit des Ereignisses

$$A := \dot{\bigcup}_{I \in \mathcal{I}} \left(\left[\bigcap_{i \in I} E_i \right] \cap \left[\bigcap_{i \notin I} E_i^c \right] \right).$$

Es gilt

$$\begin{aligned} \mathbb{P}(A) &= \sum_{I \in \mathcal{I}} \left(\prod_{i \in I} \mathbb{P}(E_i) \cdot \prod_{i \notin I} \mathbb{P}(E_i^c) \right) \\ &= \sum_{I \in \mathcal{I}} \left(p^k \cdot (1-p)^{n-k} \right) \\ &= \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Beispiel 2.28. In einem Experiment werden beliebig oft zwei Würfel geworfen. Mit welcher Wahrscheinlichkeit wird die Augensumme (AS) 5 vor der Augensumme 7 beobachtet? 

Antwort: Wir definieren zunächst mehrere Ereignisse, die hilfreich sein können bei der Beantwortung der Frage.

A : „AS 5 erscheint vor Augensumme 7.“

A_k : „Weder AS 5 noch 7 erscheint in den ersten $k-1$ Versuchen, AS 5 im k -ten Versuch.“

B_k : „AS 5 erscheint im k -ten Versuch.“

C_k : „AS 7 erscheint im k -ten Versuch.“

D_k : „Weder 5 noch 7 erscheint als AS im k -ten Versuch.“

Die Wahrscheinlichkeiten von B_k , C_k und D_k lassen sich an Hand der Tabelle aus Beispiel 2.1 ablesen:

$$\mathbb{P}(B_k) = \quad \mathbb{P}(C_k) = \quad \mathbb{P}(D_k) =$$

1. Lösungsweg: Wir berechnen zunächst $\mathbb{P}(A_k)$:

$$\mathbb{P}(A_k) =$$

=

Damit ist

$$\mathbb{P}(A) = \mathbb{P} \left(\dot{\bigcup}_{k \in \mathbb{N}} A_k \right) =$$

=

2. Lösungsweg: Alternativ können wir das Gesetz der totalen Wahrscheinlichkeit zur Hilfe nehmen indem wir eine Fallunterscheidung nach dem ersten Schritt machen.

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|C_1)\mathbb{P}(C_1) + \mathbb{P}(A|D_1)\mathbb{P}(D_1)$$

=

Dabei verwenden wir, dass die Augenzahlen in jedem Versuch unabhängig vom vorherigen sind, so dass es für den aktuellen Versuch keine Rolle spielt, wie oft schon weder 5 noch 7 als Augenzahl aufgetaucht ist. Umstellen nach $\mathbb{P}(A)$ liefert

$$\implies \mathbb{P}(A) =$$

2.5 Stetigkeit des Wahrscheinlichkeitsmaßes

Hat man endlich viele ineinander geschachtelte Mengen bzw. Ereignisse, d.h. $A_1 \subseteq A_2 \subseteq \dots \subseteq A_{n-1} \subseteq A_n$, so gilt, wie man leicht nachprüfen kann, $\bigcup_{k=1}^n A_k = A_n$, d.h. $\mathbb{P}(\bigcup_{k=1}^n A_k) = \mathbb{P}(A_n)$. Wir benötigen in diesem Fall die Siebformel nicht.

Sind abzählbar viele Mengen gegeben, so existiert keine „größte“ Menge unter diesen, die alle Mengen enthält. Folglich müssen wir in diesem Fall mit Grenzwerten agieren.

Satz 2.29

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \dots \in \mathcal{A}$ abzählbar viele Ereignisse.

1. *Stetigkeit von unten:* Gilt $A_1 \subseteq A_2 \subseteq \dots$ und ist $A^* := \bigcup_{k \in \mathbb{N}} A_k$, so gilt

$$\lim_{k \rightarrow \infty} \mathbb{P}(A_k) = \mathbb{P}(A^*).$$

2. *Stetigkeit von oben:* Gilt $A_1 \supseteq A_2 \supseteq \dots$ und ist $A_* := \bigcap_{k \in \mathbb{N}} A_k$, so gilt

$$\lim_{k \rightarrow \infty} \mathbb{P}(A_k) = \mathbb{P}(A_*).$$

Notation

- Gilt $A_1 \subseteq A_2 \subseteq \dots$, so schreibt man auch $A_n \uparrow A^* := \bigcup_{k \in \mathbb{N}} A_k$ für $n \rightarrow \infty$.
- Gilt $A_1 \supseteq A_2 \supseteq \dots$, so schreibt man auch $A_n \downarrow A_* := \bigcap_{k \in \mathbb{N}} A_k$ für $n \rightarrow \infty$.

Beweis.

1. Um die σ -Additivität von \mathbb{P} ausnutzen zu können, müssen wir A^* als disjunkte Vereinigung schreiben. Mit der Festlegung $A_0 := \emptyset$ gilt

$$A^* = \bigcup_{k \in \mathbb{N}} A_k = \bigcup_{k \in \mathbb{N}} B_k$$

für

$$B_k :=$$

Unter Verwendung der bereits bekannten Eigenschaften eines Wahrscheinlichkeitsmaßes gilt damit

$$\begin{aligned} \mathbb{P}(A^*) &= \mathbb{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{P}(B_k) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=1}^n B_k\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(A_n). \end{aligned}$$



2. Aus $A_n \downarrow A_*$ folgt $A_n^c \uparrow (A_*)^c$ und mit dem 1. Teil folgt

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} (1 - \mathbb{P}(A_n^c)) \stackrel{1.}{=} 1 - \mathbb{P}((A_*)^c) = \mathbb{P}(A_*).$$

□

Mit Hilfe der Stetigkeit des Wahrscheinlichkeitsmaßes können wir Aussage 6 von Satz 1.15 auf abzählbar viele Ereignisse übertragen.

Korollar 2.30

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $(A_k)_{k \in \mathbb{N}} \subseteq \mathcal{A}$ eine Folge von Ereignissen. Dann gilt die Abschätzung

$$\mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right) \leq \sum_{k \in \mathbb{N}} \mathbb{P}(A_k).$$

Beweis. Für $n \in \mathbb{N}$ sei $B_n := \bigcup_{k=1}^n A_k$. Die Folge von Ereignissen $(B_n)_{n \in \mathbb{N}}$ ist aufsteigend, d.h. $B_n \subseteq B_{n+1}$ für alle $n \in \mathbb{N}$. Nach Eigenschaft 1 des vorherigen Satzes (Stetigkeit von unten) gilt also

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right) &= \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) \stackrel{\text{Satz 1.15}}{\leq} \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{P}(A_k) \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}(A_k). \end{aligned}$$

□

Kapitel 3

Diskrete Zufallsvariablen und Verteilungen

Ziele

- verstehen, wie der Zufall in die Zufallsvariable kommt
- Erwartungswerte, Varianzen und andere Kenngrößen als Formeln kennen und konkret berechnen können
- die wichtigsten diskreten Verteilungen kennen, typische Beispiele für deren Auftreten und einige Kennzahlen kennen
- Zusammenhänge zwischen den bekannten Verteilungen kennen

3.1 Was sind Zufallsvariablen?

In Zufallsexperimenten interessieren wir uns häufig nicht für die Elementarereignisse selbst, sondern für Funktionen davon. In Beispiel 2.28 wollten wir beispielsweise wissen, mit welcher Wahrscheinlichkeit die Augensumme 5 vor der Augensumme 7 beobachtet wird, wenn beliebig oft zwei faire Würfel geworfen werden. Ist $\omega := (\omega_1, \omega_2) \in \{1, \dots, 6\}^2$ ein Elementarereignis, dann interessiert uns in diesem Fall nur $X(\omega) := \omega_1 + \omega_2$.

Während der Grundraum Ω beliebige Elemente enthalten kann (also nicht nur reelle Zahlen, sondern auch Vektoren, Buchstaben u.s.w.), sollen Zufallsvariablen ausschließlich reelle Zahlen als Werte annehmen können. Definieren wir formal, was wir unter einer Zufallsvariablen verstehen. Erinnern wir dafür nochmal an unsere Festlegung aus dem vorherigen Kapitel:

Festlegung: In allen folgenden Definitionen und Sätzen sei, sofern nichts anderes gesagt wird, $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. 

Definition 3.1

Eine *Zufallsvariable* (ZV) oder *Zufallsgröße* auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ ist eine (messbare) Abbildung $X: \Omega \rightarrow \mathbb{R}$.

Bemerkung 3.2 (Messbarkeit). *Was bedeutet Messbarkeit einer Abbildung? Nehmen wir die Zufallsgröße Augensumme zweier Würfel, d.h. $X(\omega) := \omega_1 + \omega_2$. In Beispiel 2.28 haben wir die Wahrscheinlichkeiten $\mathbb{P}(X = 5) = \mathbb{P}(X^{-1}(5))$, $\mathbb{P}(X = 7) = \mathbb{P}(X^{-1}(7))$ und $\mathbb{P}(X \notin \{5, 7\}) =$*

$\mathbb{P}(X^{-1}(\{5, 7\}^c))$ berechnet. Im Allgemeinen wollen wir für Ereignisse $B \in \mathcal{B}(\mathbb{R})$ (oder ggf. einer anderen σ -Algebra auf der Zielmenge) die Wahrscheinlichkeiten der Ereignisse $\{X \in B\} = X^{-1}(B)$ bestimmen können, d.h. für unserem Wahrscheinlichkeitsmaß \mathbb{P} sollen diese Ereignisse messbar sein. Ist Ω diskret, so ist diese Bedingung immer erfüllt, da die σ -Algebra auf Ω in diesem Fall immer $\mathcal{P}(\Omega)$ ist. Nur wenn Ω überabzählbar ist, ist es möglich, Mengen in $\mathcal{P}(\Omega) \setminus \mathcal{A}$ zu treffen – das soll mit der Messbarkeit ausgeschlossen werden.

Notation

Es ist üblich, die Mengenklammern von $\{X = k\}$ oder $\{X \in B\}$ innerhalb von Wahrscheinlichkeiten wegzulassen, d.h. wir schreiben $\mathbb{P}(X \in B)$ statt $\mathbb{P}(\{X \in B\})$. Die ausführliche Schreibweise für dieses Ereignis ist im Übrigen

$$\mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\}).$$



Beispiel 3.3. Drei faire Münzen werden geworfen. Sei X die Anzahl der Zahl-Würfe. Mit welcher Wahrscheinlichkeit nimmt X die Werte 0, 1, 2 und 3 an?

Antwort: Wir setzen $\Omega = \{K, Z\}^3$, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \mathcal{A}$, d.h. es handelt sich um ein Laplace-Experiment. Überlegen wir zunächst, welche Werte von X durch welche Elementarereignisse erreicht werden.

ω	(K, K, K)	(K, K, Z)	(K, Z, K)	(Z, K, K)	(K, Z, Z)	(Z, K, Z)	(Z, Z, K)	(Z, Z, Z)
$X(\omega)$								

Jedes dieser Elementarereignisse hat die Wahrscheinlichkeit

$$\mathbb{P}(\omega) =$$

Damit gilt

$$\mathbb{P}(X = 0) = \quad \mathbb{P}(X = 1) = \quad \mathbb{P}(X = 2) = \quad \mathbb{P}(X = 3) =$$



Beispiel 3.4. Aus einer Urne mit 20 Kugeln, nummeriert von 1 bis 20, werden drei Kugeln rein zufällig gezogen (ohne Zurücklegen). Wir wetten, dass mindestens eine der drei Kugeln eine Nummer größer oder gleich 17 hat. Mit welcher Wahrscheinlichkeit gewinnen wir die Wette?

Antwort: Die Zufallsgröße sei

X : die maximale Nummer der drei gezogenen Kugeln.

X kann Werte in der Menge

$$K := \{ \quad \quad \quad \}$$

annehmen. Für jedes $k \in K$ gilt dann

$$\mathbb{P}(X = k) = \text{---}$$

Die Wahrscheinlichkeit $\mathbb{P}(X \geq 17)$ berechnen wir als Summe der folgenden Einzelwahrscheinlichkeiten:

$$\mathbb{P}(X = 20) = \text{---} = \text{---}$$

$$\mathbb{P}(X = 19) = \text{---} = \text{---}$$

$$\mathbb{P}(X = 18) = \text{---} = \text{---}$$

$$\mathbb{P}(X = 17) = \text{---} = \text{---}$$

Damit ist

$$\mathbb{P}(X \geq 17) \approx 0,15 + 0,134 + 0,119 + 0,105 = 0,508.$$

Es gibt eine Vielzahl von Kenngrößen, die Zufallsgrößen charakterisieren. Damit ist es möglich, von dem konkreten Problem zu abstrahieren und Lösungsmethoden für ganze Klassen von Problemen zu bieten. Aus Sicht der Mathematik ist es am Ende egal, ob wir eine Münze werfen und auf das erste Auftreten von *Zahl* warten oder ob wir einen Würfel werfen und auf das erste Auftreten einer geraden Zahl warten.

3.2 Die Massenfunktion diskret verteilter Zufallsvariablen

Definition 3.5

Eine Zufallsvariable X heißt *diskret (verteilt)*, falls sie höchstens abzählbar viele Werte $(x_k)_{k \in \mathbb{N}}$ mit positiver Wahrscheinlichkeit annimmt. Die Funktion

$$p_X(x_k) := \mathbb{P}(X = x_k), \quad k \in \mathbb{N},$$

heißt *Massenfunktion* von X .

Notation

Zufallsgrößen werden in der Regel mit Großbuchstaben bezeichnet; Variablen, die für die Werte stehen, die die Zufallsgrößen annehmen können, hingegen mit Kleinbuchstaben.

Beispiel 3.6. Betrachten wir das Zufallsexperiment des zweifachen Münzwurfes, d.h. $\Omega = \{K, Z\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(\omega) = \frac{1}{4}$ für alle $\omega \in \Omega$. Sei

X : Anzahl der Münzen, die Zahl zeigen.

Dann ist $X(\Omega) = \{0, 1, 2\}$ und die Massenfunktion von X ist gegeben durch

$$p_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(\{(K, K)\}) = \frac{1}{4},$$

$$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(\{(K, Z), (Z, K)\}) = \frac{1}{2},$$

$$p_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(\{(Z, Z)\}) = \frac{1}{4}.$$



Beispiel 3.7. In jeder Packung Kekse ist eine von N Sammelfiguren enthalten. In einer beliebigen Packung kann jede der Sammelfiguren mit gleicher Wahrscheinlichkeit enthalten sein und dies unabhängig von bisher bereits gefundenen Figuren. Sei Z die Anzahl von Kekspackungen, die man kaufen muss, bis man alle N Figuren besitzt. Wie ist Z verteilt?

Antwort: Wir nummerieren die Sammelfiguren von 1 bis N . Sei

$$A_j^n : \text{„Figur } j \text{ ist in den ersten } n \text{ Packungen nicht enthalten.“}$$

Für jedes $j \in \{1, \dots, N\}$ gilt dann

$$\mathbb{P}(A_j^n) =$$

Wir suchen die Wahrscheinlichkeit $\mathbb{P}(Z = n)$. Diese berechnen wir folgendermaßen:

$$\mathbb{P}(Z = n) = \mathbb{P}(Z > n - 1) - \mathbb{P}(Z > n).$$

Die Wahrscheinlichkeit, durch den Kauf von n Packungen noch nicht alle Figuren zu haben, lässt sich mit Hilfe der definierten Ereignisse ausdrücken:

$$\mathbb{P}(Z > n) = \mathbb{P}\left(\bigcup_{j=1}^N A_j^n\right).$$

Für die Siebformel benötigen wir die Wahrscheinlichkeiten der Schnitte der Ereignisse:

$$\mathbb{P}(A_{j_1}^n \cap A_{j_2}^n) =$$

$$\mathbb{P}(A_{j_1}^n \cap A_{j_2}^n \cap A_{j_3}^n) =$$

$$\vdots$$

$$\mathbb{P}(A_{j_1}^n \cap \dots \cap A_{j_k}^n) =$$

Somit ist

$$\begin{aligned} & \mathbb{P}(Z > n) \\ &= \mathbb{P}\left(\bigcup_{j=1}^N A_j^n\right) \\ &= \sum_{j=1}^N \mathbb{P}(A_j^n) - \sum_{j_1 < j_2} \mathbb{P}(A_{j_1}^n \cap A_{j_2}^n) + \sum_{j_1 < j_2 < j_3} \mathbb{P}(A_{j_1}^n \cap A_{j_2}^n \cap A_{j_3}^n) + \dots + (-1)^{N+1} \mathbb{P}\left(\bigcap_{j=1}^N A_j^n\right) \\ &= \\ &= \sum_{j=1}^{N-1} \end{aligned}$$

Für die Massenfunktion von Z erhalten wir durch geschicktes Umformen die Formel

$$\mathbb{P}(Z = n) = \mathbb{P}(Z > n - 1) - \mathbb{P}(Z > n) = \sum_{j=1}^{N-1} (-1)^{j+1} \binom{N}{j} \left(\frac{N-j}{N}\right)^{n-1} \frac{j}{N}.$$

Da diskrete Zufallsvariablen in vielen Beispielen nur endlich viele Werte annehmen, ist eine Massenfunktion dadurch eindeutig bestimmt, dass die Funktionswerte dieser Werte – aufgelistet oder in Tabellenform – angegeben werden.



Notation

Ist $A \in \mathcal{A}$ ein Ereignis, so können wir damit eine wichtige Zufallsvariable definieren, die *Indikatorfunktion* von A :

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

Beispiel 3.8. Sei $A \in \mathcal{A}$ ein Ereignis auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und sei $X := \mathbb{1}_A$. Die Massenfunktion von X ist dann gegeben durch



$$p_X(\cdot) = \mathbb{P}(X = \cdot) = \mathbb{P}(\mathbb{1}_A = \cdot)$$

Die Indikatorfunktion $\mathbb{1}_A$ eines Ereignisses gibt an, ob ein Ereignis eingetreten ist oder nicht. Dies ermöglicht es uns, die Anzahl eingetretener Ereignisse als Zufallsvariable zu formulieren:

Sind $A_1, \dots, A_n \in \mathcal{A}$ Ereignisse, so gibt die Zufallsvariable

$$X := \sum_{k=1}^n \mathbb{1}_{A_k}$$

an, wie viele der Ereignisse A_1, \dots, A_n eingetreten sind. Man nennt X daher auch *Zählvariable*.

Lemma 3.9

Seien $A, B \in \mathcal{A}$ Ereignisse. Für die zugehörigen Indikatorfunktionen gelten folgende Rechenregeln:

1. $\mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$;
2. $\mathbb{1}_{A^c} = \mathbb{1}_\Omega - \mathbb{1}_A$;
3. $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B}$;
4. $A \subseteq B$ gilt genau dann, wenn $\mathbb{1}_A \leq \mathbb{1}_B$.

Beweis.

1. Falls $\omega \in A \cap B$ ist, so ist auch $\omega \in A$ und $\omega \in B$ und es gilt in der Tat

$$\mathbb{1}_{A \cap B}(\omega) = 1 = 1 \cdot 1 = \mathbb{1}_A(\omega) \cdot \mathbb{1}_B(\omega).$$

Ist $\omega \notin A \cap B$, so ist entweder $\omega \notin A$ oder $\omega \notin B$ oder beides. O.B.d.A. sei $\omega \in A^c$. Folglich ist

$$\mathbb{1}_{A \cap B}(\omega) = 0 = 0 \cdot \mathbb{1}_B(\omega) = \mathbb{1}_A(\omega) \cdot \mathbb{1}_B(\omega).$$

2. Falls $\omega \in A$ ist, so gilt

$$\mathbb{1}_{A^c}(\omega) = 0 = 1 - 1 = \mathbb{1}_\Omega(\omega) - \mathbb{1}_A(\omega).$$

Ist hingegen $\omega \in A^c$, so gilt

$$\mathbb{1}_{A^c}(\omega) = 1 = 1 - 0 = \mathbb{1}_\Omega(\omega) - \mathbb{1}_A(\omega).$$

Aussagen 3 und 4 können ebenfalls mit Fallunterscheidung bewiesen werden und sind Übungsaufgaben. \square

Satz 3.10

Sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in (M, \mathcal{B}) , wobei $M \subseteq \mathbb{R}$ und $\mathcal{B} \subseteq \mathcal{P}(M)$ eine σ -Algebra ist. Dann wird durch

$$\mathbb{P}^X : \mathcal{B} \rightarrow [0, 1], \quad \mathbb{P}^X(B) := \mathbb{P}(X \in B)$$

ein Wahrscheinlichkeitsmaß definiert. Man bezeichnet \mathbb{P}^X auch als *Verteilung von X* .

Beweis. Übungsaufgabe. \square

Bemerkung 3.11. Ist X eine diskrete Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$, so ist $X(\Omega)$ höchstens abzählbar und wir können $\mathcal{B} := \mathcal{P}(M)$ setzen. Nach Satz 3.10 ist $(X(\Omega), \mathcal{P}(X(\Omega)), \mathbb{P}^X)$ ein Wahrscheinlichkeitsraum.

Beispiel 3.12. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $A \in \mathcal{A}$ ein Ereignis. Ist $X = \mathbb{1}_A$ die Indikatorfunktion von A , so nennt man \mathbb{P}^X Bernoulli-Verteilung auf $(\{0, 1\}, \mathcal{P}(\{0, 1\}))$ mit Erfolgswahrscheinlichkeit $p = \mathbb{P}^X(\{1\}) = \mathbb{P}(A)$. Die Verteilung ist durch Angabe der Massenfunktion eindeutig spezifiziert.

Im nächsten Abschnitt werden wir einige wichtige diskrete Verteilungen kennen lernen.

3.3 Die wichtigsten diskreten Verteilungen

3.3.1 Bernoulli- und Binomialverteilung

In Beispiel 2.27 haben wir eine Folge von Versuchen mit zwei möglichen Ausgängen – Erfolg und Misserfolg – untersucht. Einen einzelnen Versuch dieser Art nennt man *Bernoulli-Experiment* und die n -fache Wiederholung dieses Versuchs nennt man *Bernoulli-Kette* der Länge n .

Ist $p \in (0, 1)$ die Wahrscheinlichkeit für einen Erfolg in einem Versuch, so ist die Wahrscheinlichkeit, in n Versuchen genau k Erfolge zu erzielen gegeben durch

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

Definition 3.13

Eine Zufallsvariable X , deren Massenfunktion durch

$$p_X(1) = p, \quad p_X(0) = 1 - p,$$

für ein $p \in [0, 1]$ gegeben ist, heißt *Bernoulli-verteilt* mit Parameter p .

Kurzschreibweise:

$$X \sim \text{Ber}(p).$$

Analog führen wir die Verteilung ein, welche die Anzahl der Erfolge bei einer Bernoulli-Kette charakterisiert.

Definition 3.14

Eine Zufallsvariable X besitzt eine *Binomialverteilung* mit den Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$, falls ihre Massenfunktion gegeben ist durch

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

Kurzschreibweise:

$$X \sim \text{Bin}(n, p).$$

Für die Extremfälle $p \in \{0, 1\}$ gilt wegen $0^0 = 1$:

$p = 0$: $\mathbb{P}(X = 0) = 1$ und $\mathbb{P}(X = k) = 0$ für alle $k \in \{1, \dots, n\}$;

$p = 1$: $\mathbb{P}(X = n) = 1$ und $\mathbb{P}(X = k) = 0$ für alle $k \in \{0, \dots, n-1\}$.

Bemerkung 3.15. Wir haben in keiner der Definitionen den Wertebereich der Zufallsvariablen genannt, d.h. der zu Grunde liegende Wahrscheinlichkeitsraum wird (abgesehen von \mathbb{P} , welches ohne explizit genannt werden zu müssen immer für ein Wahrscheinlichkeitsmaß steht) nicht mehr erwähnt. Einerseits liegt das daran, dass die Kenntnis von Ω hier überflüssig ist. Andererseits ist es in der Praxis sogar so, dass die Verteilung von Merkmalsausprägungen empirisch beobachtet werden kann und daraus eine Verteilung abgeleitet wird ohne dass man die „Ursache des Zufalls“ genauer untersuchen würde. Wir wissen also häufig gar nicht, wie die Verteilung entsteht – einen Grundraum müssten wir uns also künstlich ausdenken, was nicht sinnvoll ist.

Beispiel 3.16. Eine Firma produziert Schrauben, die in 10er-Packungen verkauft werden. Eine Schraube weist mit Wahrscheinlichkeit 0,01 einen Produktionsfehler auf. Die Firma überlegt, ob sie es sich leisten kann, eine Umtauschgarantie anzubieten – falls mindestens eine Schraube in der Packung einen Fehler aufweist, wird die Packung kostenfrei umgetauscht. Welchen Anteil an verkauften Packungen muss die Firma dann im Rahmen der Garantie umtauschen? 

Antwort: Bezeichne X die Anzahl fehlerhafter Schrauben in einer Packung. Wenn wir davon ausgehen, dass jede Schraube unabhängig von der nächsten einen Fehler aufweist, so kann

$$X \sim \text{Bin}\left(10, \frac{1}{100}\right)$$

angenommen werden. Damit gilt

$$\mathbb{P}(X \geq 1) =$$

Es müssten also (wenn alle Kunden von der Umtauschgarantie Gebrauch machen) knapp 10% der verkauften Packungen umgetauscht werden.

Bemerkung 3.17. Die Werte der Binomialverteilung werden in einigen Lehrbüchern tabellarisch angegeben. Hat man keine solche Tabelle zur Hand und möchte die Werte der Massenfunktion schnell berechnen, so sind diese beispielsweise in Excel¹ bereits implementiert:

$$=\text{BINOM. VERT}(k; n; p; \text{FALSCH})$$

gibt $\mathbb{P}(X = k)$ für $X \sim \text{Bin}(n, p)$ wider. Ersetzt man *FALSCH* durch *WAHR*, so berechnet Excel nicht die Werte der Massenfunktion, sondern die der Verteilungsfunktion F_X , welche wir noch kennen lernen werden.

¹Die Befehle können von Version zu Version und je nach Systemsprache variieren.

Satz 3.18

Für feste Parameter $n \in \mathbb{N}$ und $p \in [0, 1]$ sei $X \sim \text{Bin}(n, p)$. Dann ist \mathbb{P}^X ein Wahrscheinlichkeitsmaß auf $(\{0, 1, \dots, n\}, \mathcal{P}(\{0, 1, \dots, n\}))$.

Beweis. Diese Aussage folgt direkt aus Satz 3.10 sobald wir wissen, dass die angegebene Massenfunktion tatsächlich eine ist. Um dies zu sehen, überprüfen wir, ob \mathbb{P}^X ausschließlich nicht-negative Werte annehmen kann und ob $\mathbb{P}^X(\{0, 1, \dots, n\}) = 1$ ist.

- Da alle Summanden nichtnegativ sind, gilt $\mathbb{P}^X(k) = \mathbb{P}(X = k) \geq 0$ für alle $k \in \{0, \dots, n\}$.
- Der binomische Lehrsatz (Satz A.4) liefert für $x = p$ und $y = 1 - p$ sofort die gewünschte Aussage:

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

□

Die Berechnung einzelner Werte der Massenfunktion oder das Summieren vieler Werte derselben ist rechenintensiv und numerisch schwierig für große Werte für n und kleine Wahrscheinlichkeiten p . Wir werden daher mehrere Möglichkeiten kennen lernen, die Binomialverteilung zu approximieren durch Verteilungen, die leichter auszuwerten sind.

3.3.2 Poisson-Verteilung**Satz 3.19**

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge binomialverteilter Zufallsvariablen, wobei $X_n \sim \text{Bin}(n, p_n)$ sei für $n \in \mathbb{N}$ und $p_n \in (0, 1)$. Gilt für $\lambda_n := n \cdot p_n$

$$\lim_{n \rightarrow \infty} \lambda_n = \lambda > 0,$$

so konvergieren die Massenfunktionen p_{X_n} punktweise gegen eine neue Massenfunktion, welche gegeben ist durch

$$p(k) := \lim_{n \rightarrow \infty} p_{X_n}(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0.$$

Beweis. Sei $k \in \mathbb{N}_0$ fest gewählt. Dann gilt

$$\begin{aligned} p_{X_n}(k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \underbrace{\frac{\lambda_n^k}{k!}}_{(I)} \cdot \underbrace{\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}}_{(II)} \cdot \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^n}_{(III)} \cdot \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^{-k}}_{(IV)} \\ &\xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda}. \end{aligned}$$

Um die Konvergenz zu sehen, benötigt man zunächst, dass aus $\lim_{n \rightarrow \infty} a_n = a$ und $\lim_{n \rightarrow \infty} b_n = b$ sofort $\lim_{n \rightarrow \infty} (a_n b_n) = ab$ folgt, d.h. wir können alle Faktoren getrennt untersuchen.

- (I) $\lim_{n \rightarrow \infty} \frac{\lambda_n^k}{k!} = \frac{\lambda^k}{k!}$
- (II) Jeder der Faktoren konvergiert gegen 1, so dass auch das Produkt der k Faktoren gegen 1 konvergiert.
- (III) Da $\lambda_n \xrightarrow{n \rightarrow \infty} \lambda$ und $(1 + \frac{x}{n})^n \xrightarrow{n \rightarrow \infty} e^x$ gehen, konvergiert dieser Ausdruck insgesamt gegen $e^{-\lambda}$. Die Details findet man im *Fundamentallemma* in Abschnitt 8.1 in [Kö99].
- (IV) Wegen $\lim_{n \rightarrow \infty} \frac{\lambda_n}{n} = 0$ konvergiert der gesamte Term gegen 1.

□

Die als Limes der Binomialverteilung auftretende Verteilung ist benannt nach dem französischen Physiker und Mathematiker Siméon Denis Poisson (1781-1840).

Definition 3.20

Eine diskret verteilte Zufallsvariable X mit Werten in \mathbb{N}_0 heißt Poisson-verteilt mit Parameter $\lambda > 0$, falls ihre Massenfunktion durch $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ ($k \in \mathbb{N}_0$) gegeben ist.

Kurzschreibweise:

$$X \sim \text{Poisson}(\lambda).$$

Dieses Grenzverhalten kann man sich anhand ausgewählter Massenfunktionen verdeutlichen.

$\lambda = 2$	Bin(n,p)						Poisson(λ)
$p \rightarrow$	0,5	0,2	0,1	0,05	0,01	0,001	0
$k \downarrow n \rightarrow$	4	10	20	40	200	2000	∞
0	0,062500	0,107374	0,121577	0,128512	0,133980	0,135200	0,135335
1	0,250000	0,268435	0,270170	0,270552	0,270666	0,270671	0,270671
2	0,375000	0,301990	0,285180	0,277672	0,272033	0,270806	0,270671
3	0,250000	0,201327	0,190120	0,185114	0,181355	0,180537	0,180447
4	0,062500	0,088080	0,089779	0,090122	0,090220	0,090223	0,090224
5		0,026424	0,031921	0,034151	0,035723	0,036053	0,036089
6		0,005505	0,008867	0,010485	0,011727	0,012000	0,012030
7		0,000786	0,001970	0,002680	0,003283	0,003422	0,003437
8		0,000074	0,000356	0,000582	0,000800	0,000853	0,000859
9		0,000004	0,000053	0,000109	0,000172	0,000189	0,000191
10		0,000000	0,000006	0,000018	0,000033	0,000038	0,000038

Bemerkung 3.21 (Faustregel). Die Poisson-Verteilung gilt als gute Approximation der Binomialverteilung für $n \geq 30$ und $p \leq 0,5$ – oft sind die Werte jedoch schon vorher gut genug als Approximation.



Bemerkung 3.22 (Interpretation). Die Poisson-Verteilung wird verwendet um die Anzahl von Ereignissen zu modellieren, die unabhängig voneinander und mit einer konstanten Rate innerhalb eines vorgegebenen Zeitraumes eintreten. Man spricht auch von der Verteilung seltener Ereignisse. Beispiele hierfür sind die Anzahl Druckfehler auf einer Seite; die Anzahl radioaktiver Zerfälle einer Substanz in einem Zeitintervall, welches klein ist im Verhältnis zur Halbwertszeit; die Anzahl Blitzeinschläge pro Hektar Land in einem Jahr oder die Anzahl von Impfschäden pro Jahr in einem Staat. Auch in der Warteschlangentheorie werden Ankünfte von Kunden in einem gegebenen Zeitraum bisweilen als Poisson-verteilt angenommen.

3.3.3 Geometrische Verteilung

Während die Anzahl der Erfolge in einem wiederholten Bernoulli-Experiment durch die *Binomialverteilung* modelliert wird, führt das Warten auf den ersten Erfolg bei einem beliebig oft wiederholten Bernoulli-Experiment zu der *geometrischen Verteilung*.

Beispiel 3.23. Wie in Beispiel 2.27 wird eine unendliche Folge von Versuchen mit zwei möglichen Ausgängen – Erfolg und Misserfolg – durchgeführt. Sei $p \in (0, 1)$ die Wahrscheinlichkeit für einen Erfolg in einem Versuch. Wie groß ist die Wahrscheinlichkeit,

1. dass im k -ten Versuch der Erste Erfolg auftritt?
2. dass bis zum ersten Erfolg k Misserfolge aufgetreten sind?

Antwort: Wir definieren folgende Ereignisse:

E_k : „Erfolg im k -ten Versuch“

A_k : „Erster Erfolg im k -ten Versuch“

B_k : „ k Misserfolge vor dem ersten Erfolg“

1. Gesucht wird $\mathbb{P}(A_k)$.

$k = 1$: In diesem Fall ist $\mathbb{P}(A_1) = \mathbb{P}(E_1) = p$.

$k = 2$: In diesem Fall ist $\mathbb{P}(A_2) = \mathbb{P}(E_1^c \cap E_2) = p \cdot (1 - p)$.

$k \geq 3$: In diesem Fall ist

$$\mathbb{P}(A_k) = \mathbb{P}(E_1^c \cap \dots \cap E_{k-1}^c \cap E_k) = p \cdot (1 - p)^{k-1}.$$

2. Gesucht wird $\mathbb{P}(B_k) = \mathbb{P}(A_{k+1})$. Nach den Überlegungen aus 1. ist demnach

$$\mathbb{P}(B_k) = p \cdot (1 - p)^k, \quad k \in \mathbb{N}_0.$$

Definition 3.24

Eine diskret verteilte Zufallsvariable X heißt *geometrisch verteilt* mit Parameter $p \in (0, 1)$, falls ihre Massenfunktion gegeben ist durch $p_X(k) = p(1 - p)^{k-1}$ für $k \in \mathbb{N}$.

Kurzschreibweise: $X \sim \text{Geo}(p)$.

Bemerkung 3.25.

1. Wenn eine Zufallsvariable Y die Massenfunktion $p_Y(k) = p(1 - p)^k$ für $k \in \mathbb{N}_0$ besitzt, so wird diese Zufallsvariable auch als *geometrisch verteilt* bezeichnet. Das Beispiel hat gezeigt, dass in der Tat beide zu Grunde liegenden Fragestellungen mit den gleichen Berechnungen beantwortet werden können. Um in dieser Vorlesung jedoch eine eindeutige Bezeichnung zu haben, werden wir ausschließlich mit der in der Definition gegebenen Massenfunktion arbeiten.
2. Die geometrische Verteilung wird gerne für die Modellierung von Wartezeiten verwendet – wir werden gleich sehen, warum sie dafür prädestiniert ist.

Satz 3.26: Gedächtnislosigkeit der geometrischen Verteilung

Sei X eine geometrisch verteilte Zufallsvariable zum Parameter $p \in (0, 1)$. Dann gilt

$$\mathbb{P}(X = n + k | X > n) = \mathbb{P}(X = k), \quad \forall k, n \in \mathbb{N}.$$

Beweis. Übungsaufgabe. □

3.3.4 Hypergeometrische Verteilung

Beispiel 3.27 (Verallgemeinerte Lotto-Ziehung). Eine Urne enthält N Kugeln. Unter den Kugeln befinden sich K weiße und $N - K$ schwarze Kugeln. Die Kugeln werden gut gemischt und es werden n Kugeln ohne Zurücklegen gezogen. Wie wahrscheinlich ist es, dass unter den n Kugeln genau k weiße Kugeln sind?

Vorüberlegung: Es gibt folgende Einschränkungen für die Parameter in diesem Beispiel. Warum? 

- $N, n, K, k \in \mathbb{N}_0$
- $n \leq N$
- $K \leq N$
- $k \leq \min\{n, K\}$
- $k \geq \max\{0, n + K - N\}$

Antwort auf die Frage: Wir nehmen an, dass alle Anforderungen an die Parameter erfüllt sind.

Wir wollen das Experiment als Laplace-Experiment auffassen, daher nummerieren wir die Kugeln von 1 bis N , wobei Kugeln 1 bis K weiß seien und der Rest schwarz. Der Wahrscheinlichkeitsraum ist damit

$$\begin{aligned} \Omega &= \{\omega \in \{1, \dots, N\}^n \mid \omega_k < \omega_{k+1}, k \in \{1, \dots, n-1\}\} \\ \mathcal{A} &= \mathcal{P}(\Omega) \\ \mathbb{P}(\omega) &= \frac{1}{|\Omega|}, \quad \forall \omega \in \Omega. \end{aligned}$$

Es gibt $\binom{N}{n}$ Möglichkeiten, n Kugeln unter den N vorhandenen Kugeln auszuwählen, d.h. $|\Omega| = \binom{N}{n}$. Unter Verwendung dieser Notation suchen wir die Wahrscheinlichkeit des Ereignisses

$$A_k = \{\omega \in \Omega \mid \omega_k \leq K, \omega_{k+1} > K\}.$$

Um genau k weiße Kugeln zu ziehen, müssen wir zunächst k Kugeln von den K weißen Kugeln auswählen – dafür gibt es $\binom{K}{k}$ Möglichkeiten – und im zweiten Schritt $n - k$ schwarze Kugeln unter den $N - K$ verfügbaren schwarzen Kugeln auswählen – dafür gibt es $\binom{N-K}{n-k}$ Möglichkeiten. Mit dem Zählprinzip gilt also

$$\mathbb{P}(A_k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}.$$

Definieren wir die Zufallsvariable X , welche die Anzahl weißer Kugeln in einer Stichprobe von n Kugeln angibt, so ist die Massenfunktion gegeben durch

$$p_X(k) = \mathbb{P}(X = k) = \mathbb{P}(A_k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \in \{\max\{0, n + K - N\}, \dots, \min\{K, n\}\}.$$

Notation

Für zwei Zahlen $x, y \in \mathbb{R}$ verwenden wir der Kürze halber für das Minimum und Maximum beider Zahlen folgende Notation:

$$x \vee y := \max\{x, y\}, \quad x \wedge y := \min\{x, y\}.$$

Bemerkung 3.28. Die Symbole werden in der Logik für oder und und zwischen mathematischen Aussagen verwendet. Man kann sich ihre Bedeutung als Maximum und Minimum wie folgt merken:

$$\max\{x, y\} = x \vee y \geq 0 \iff (x \geq 0) \vee (y \geq 0)$$

$$\min\{x, y\} = x \wedge y \geq 0 \iff (x \geq 0) \wedge (y \geq 0)$$

Definition 3.29

Eine diskret verteilte Zufallsvariable heißt *hypergeometrisch verteilt* mit Parametern (N, K, n) , falls ihre Massenfunktion gegeben ist durch

$$p_X(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \in \{0 \vee (n + K - N), \dots, K \wedge n\}.$$

Kurzschreibweise: $X \sim \text{Hyp}(N, K, n)$.



Bemerkung 3.30. Indem wir $\binom{n}{k} := 0$ setzen für $k > n$, können wir den Wertebereich von X vereinfachen zu $\{0, \dots, n\}$.

Da das Berechnen der vielen Binomialkoeffizienten nicht einfach ist, hilft folgender Approximationssatz, die hypergeometrische Verteilung für große N (relativ zu n) durch die Binomialverteilung zu ersetzen.

Satz 3.31

Sei $n \in \mathbb{N}_0$ fest gewählt und sei (K_N) eine Folge derart, dass $\lim_{N \rightarrow \infty} \frac{K_N}{N} = p \in (0, 1)$. Dann gilt

$$\lim_{N \rightarrow \infty} \frac{\binom{K_N}{k} \binom{N-K_N}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, \dots, n\},$$

d.h. die Massenfunktion der $\text{Hyp}(N, K_N, n)$ -Verteilung konvergiert gegen die Massenfunktion der $\text{Bin}(n, p)$ -Verteilung.

Bemerkung 3.32. Wenn für feste Werte $N, K = K_N \in \mathbb{N}$ das Verhältnis $\frac{K}{N} = p$ gilt, dann ist möglicherweise $K_{N+1} := p \cdot (N+1) \notin \mathbb{N}$. Daher verlangen wir nicht $\frac{K_N}{N} = p$ für alle $N \in \mathbb{N}$. Allerdings gilt für jedes $j \in \mathbb{N}$, dass $\frac{jK}{jN} = \frac{K}{N} = p$ und $aK \in \mathbb{N}$ ist. Somit erfüllt die Teilfolge

$$K_{jN} := j \cdot K, \quad j \in \mathbb{N}$$

die gewünschten Bedingungen und die Folgenglieder dazwischen können geeignet gewählt werden, dass der Grenzwert erhalten bleibt.

Beweis. Die Massenfunktionen seien gegeben durch

$$h_{N,K,n}(k) := \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{und} \quad b_{n,p}(k) := \binom{n}{k} p^k (1-p)^{n-k}$$

für $k \in \{0, \dots, n\}$. Dann gilt

$$\begin{aligned} h_{N,K,n}(k) &= \frac{K \cdot (K-1) \cdot \dots \cdot (K-k+1)}{k!} \cdot \frac{(N-K) \cdot (N-K-1) \cdot \dots \cdot (N-K-(n-k)+1)}{(n-k)!} \\ &= \binom{n}{k} \cdot \frac{\frac{K \cdot (K-1) \cdot \dots \cdot (K-k+1)}{N^k} \cdot \frac{(N-K) \cdot (N-K-1) \cdot \dots \cdot (N-K-(n-k)+1)}{N^{n-k}}}{\frac{N \cdot (N-1) \cdot \dots \cdot (N-n+1)}{N^n}} \\ &= \binom{n}{k} \frac{\left(\frac{K}{N} \cdot \frac{K-1}{N} \cdot \dots \cdot \frac{K-k+1}{N}\right) \cdot \left(\frac{N-K}{N} \cdot \frac{N-K-1}{N} \cdot \dots \cdot \frac{N-K-n+k+1}{N}\right)}{\frac{N}{N} \cdot \frac{N-1}{N} \cdot \dots \cdot \frac{N-n+1}{N}} \\ &\stackrel{*}{\rightarrow} \binom{n}{k} \frac{(p \cdot p \cdot \dots \cdot p) \cdot ((1-p) \cdot (1-p) \cdot \dots \cdot (1-p))}{1 \cdot 1 \cdot \dots \cdot 1} \\ &= \binom{n}{k} p^k (1-p)^{n-k} = b_{n,p}(k), \end{aligned}$$

wobei in * gleichzeitig $N \rightarrow \infty$ und $\frac{K}{N} \rightarrow p$ konvergieren. □

Für $n = 10$ und $p = \frac{1}{4}$ zeigt folgende Tabelle die Werte von $h_{N,K,n}$ und $b_{n,p}$ – man sieht in der Tat, dass für große Werte von N (und zugehöriges $K = p \cdot N$) die Binomialverteilung eine gute Approximation der hypergeometrischen Verteilung liefert.

k→	0	1	2	3	4	5	6	7	8	9	10
$h(100,25,10)$	0,04788656	0,18138847	0,29238739	0,26372196	0,14714921	0,05297371	0,01243514	0,00187514	0,00017339	0,00000885	0,00000019
$h(200,50,10)$	0,05209363	0,18472918	0,28685060	0,25676137	0,14665710	0,05583084	0,01434011	0,00245273	0,00026723	0,00001674	0,00000046
$h(400,100,10)$	0,05420250	0,18626289	0,28417849	0,25346523	0,14634599	0,05714935	0,01528488	0,00276436	0,00032351	0,00002212	0,00000067
$h(1000,250,10)$	0,05546892	0,18714209	0,28260473	0,25154230	0,14614134	0,05790728	0,01584817	0,00295808	0,00036037	0,00002587	0,00000083
$b(10,1/4)$	0,05631351	0,18771172	0,28156757	0,25028229	0,14599800	0,05839920	0,01622200	0,00308990	0,00038624	0,00002861	0,00000095

3.3.5 Diskrete Gleichverteilung

Die diskrete Gleichverteilung wird auch als *Laplace-Verteilung* bezeichnet.

Definition 3.33

Nimmt eine Zufallsvariable X nur Werte in einer endlichen Menge M an, so heißt X *gleichverteilt* auf M , falls die Massenfunktion gegeben ist durch

$$p_X(m) = \frac{1}{|M|}, \quad \forall m \in M.$$

Kurzschreibweise: $X \sim U(M)$.

Bemerkung 3.34. Die Kurzschreibweise $X \sim U(M)$ stammt von der englischen Bezeichnung der Gleichverteilung – *uniform distribution*.

Beispiel 3.35. Die Zufallsvariable X beschreibe die Augenzahl, die ein fairer Würfel zeigt. Dann ist $X \sim U(\{1, 2, 3, 4, 5, 6\})$.

Bemerkung 3.36. Kann man eine Gleichverteilung auf den natürlichen Zahlen definieren?

Um zu sehen, warum dies nicht geht, nehmen wir an, p_X sei die Massenfunktion einer gleichverteilten Zufallsvariablen X auf $M = \mathbb{N}$. Dann wäre $p_X(n) = c$ für alle $n \in \mathbb{N}$ und einen konstanten Wert $c \in [0, 1]$. Damit \mathbb{P}^X tatsächlich ein Wahrscheinlichkeitsmaß definiert, muss

$$\sum_{k \in \mathbb{N}} p_X(k) = \sum_{k \in \mathbb{N}} c \stackrel{!}{=} 1$$

gelten. Diese Gleichung wird durch keine Zahl $c \in [0, 1]$ erfüllt, d.h. es gibt kein Wahrscheinlichkeitsmaß auf der Menge der natürlichen Zahlen, welches allen natürlichen Zahlen den gleichen Wert zuordnet, also keine Gleichverteilung auf \mathbb{N} .

3.4 Erwartungswert und Varianz

Definition 3.37

Sei X eine diskret verteilte Zufallsvariable mit Massenfunktion p_X . Gilt

$$\sum_{x: p_X(x) > 0} |x| \cdot p_X(x) < \infty,$$

so ist der Erwartungswert von X definiert durch

$$\mathbb{E}[X] := \sum_{x: p_X(x) > 0} x \cdot p_X(x). \quad (3.4.1)$$

Der Erwartungswert drückt das durch die Massenfunktion gewichtete Mittel der von X angenommenen Werte aus.

Bemerkung 3.38. Wozu die Bedingung $\sum_{x: p_X(x) > 0} |x| \cdot p_X(x) < \infty$? Wann immer eine Zufallsvariable nur endlich viele endliche Werte annehmen kann, ist diese Bedingung grundsätzlich erfüllt. Da nur diskret verteilte Zufallsvariablen betrachtet werden, geht es bei der Bedingung also um Zufallsvariablen, die abzählbar viele Werte, bezeichnen wir sie mit x_1, x_2, x_3, \dots , annehmen können. Ist die Reihe $\sum_{k \in \mathbb{N}} x_k p_X(x_k)$ absolut konvergent, so ist der Erwartungswert definiert als Wert dieser Reihe, welcher dank der absoluten Konvergenz nicht von der Reihenfolge der Summation abhängig ist (vgl. Satz A.14). Ist $\sum_{x: p_X(x) > 0} |x| \cdot p_X(x) = \infty$, so sagen wir, dass $\mathbb{E}[X]$ nicht definiert ist bzw. nicht existiert.

Bemerkung 3.39. Alternativ kann der Erwartungswert einer diskret verteilten Zufallsvariablen X auch wie folgt definiert werden:

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}). \quad (3.4.2)$$

Während in (3.4.1) über die angenommenen Werte aus dem Wertebereich von X summiert wird, wird in (3.4.2) über die Werte aus dem Definitionsbereich von X summiert. Da für jedes $x \in X(\Omega)$

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\}) = \mathbb{P}(X^{-1}(x)) = \sum_{\omega \in X^{-1}(x)} \mathbb{P}(\{\omega\})$$

ist, sind (3.4.1) und (3.4.2) inhaltlich gleichwertig. Der Vorteil von (3.4.1) ist jedoch, dass man den Grundraum $(\Omega, \mathcal{A}, \mathbb{P})$ nicht benötigt, sofern man bereits die Verteilung der Zufallsvariablen (beispielsweise in Form der Massenfunktion) gegeben hat.

Beispiel 3.40. Sehen wir uns die Zufallsvariable aus Beispiel 3.6 (zweifacher Münzwurf) an. Die Massenfunktion ist gegeben durch

$$p_X(0) = \frac{1}{4}, \quad p_X(1) = \frac{1}{2}, \quad p_X(2) = \frac{1}{4}.$$

Folglich ist

$$\mathbb{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

Beispiel 3.41. Betrachten wir die Indikatorfunktion $\mathbb{1}_A$ zu einem Ereignis $A \in \mathcal{A}$. Was ist der Erwartungswert dieser Zufallsvariablen? 

Antwort: $\mathbb{1}_A$ kann nur zwei Werte annehmen, nämlich 1 und 0. Die Wahrscheinlichkeiten dafür sind

$$\begin{aligned} p_{\mathbb{1}_A}(1) &= & \text{und} \\ p_{\mathbb{1}_A}(0) &= \end{aligned}$$

Somit berechnen wir den Erwartungswert als

$$\mathbb{E}[\mathbb{1}_A] =$$

Beispiel 3.42. Sei $X \sim \text{Geo}(p)$. Welchen Erwartungswert hat X ?

Antwort: Nach Definition des Erwartungswertes gilt zunächst

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k p (1-p)^{k-1} = p \sum_{k=1}^{\infty} \left[(1-p)^{k-1} \sum_{j=1}^k 1 \right] = p \sum_{k=1}^{\infty} \left[\sum_{j=1}^k (1-p)^{k-1} \right].$$

Da alle Summanden nichtnegativ sind, können wir die Reihenfolge ändern. Es ist

$$\{(j, k) \in \mathbb{N}^2 \mid j \leq k\} = \{(j, k) \in \mathbb{N}^2 \mid k \geq j\}$$

und somit

$$\mathbb{E}[X] = p \sum_{j=1}^{\infty} \left[\sum_{k=j}^{\infty} (1-p)^{k-1} \right] = p \sum_{j=1}^{\infty} \left[(1-p)^{j-1} \sum_{k=j}^{\infty} (1-p)^{k-1} (1-p)^{1-j} \right].$$

Nach dem Kürzen der Exponenten machen wir eine Indexverschiebung an der inneren Summe und erhalten

$$\mathbb{E}[X] = p \sum_{j=1}^{\infty} \left[(1-p)^{j-1} \sum_{k=j}^{\infty} (1-p)^{k-j} \right] = p \sum_{j=1}^{\infty} \left[(1-p)^{j-1} \sum_{k=0}^{\infty} (1-p)^k \right].$$

Indem wir den Wert der geometrischen Reihe

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

einsetzen, erhalten wir schlussendlich

$$\mathbb{E}[X] = p \sum_{j=1}^{\infty} \left[(1-p)^{j-1} \frac{1}{p} \right] = \sum_{j=1}^{\infty} (1-p)^{j-1} = \sum_{j=0}^{\infty} (1-p)^j = \frac{1}{p}.$$



Beispiel 3.43. Sei $X \sim \text{Hyp}(N, K, n)$. Welchen Erwartungswert hat X ?

Antwort: Der Kürze halber führen wir zwei Indexmengen ein:

$$I_1 := \{0 \vee n + K - N, \dots, K \wedge n\},$$

$$I_2 := \{0 \vee n + K - N - 1, \dots, K \wedge n - 1\}$$

Damit ist

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k \in I_1} k \cdot p_X(k) \\ &= \sum_{k \in I_1} k \cdot \frac{\binom{K-n}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \\ &= \sum_{k \in I_1} \frac{K \cdot \binom{K-n-1}{k-1} \binom{N-K}{n-k}}{\binom{N}{n}} \\ &\stackrel{j=k-1}{=} \frac{nK}{N} \cdot \sum_{j \in I_2} \frac{\binom{K-n-1}{j} \binom{N-K}{n-j}}{\binom{N-1}{n-1}} \\ &= \frac{nK}{N} \sum_{j \in I_2} \mathbb{P}(Y = j) \quad \text{mit } Y \sim \text{Hyp}(N-1, K-1, n-1). \end{aligned}$$

Folglich ist

$$\mathbb{E}[X] = \frac{nK}{N} \cdot \frac{N-1}{N} = \frac{nK}{N} \cdot \frac{N-1}{N}.$$

Beispiel 3.44. Erinnern wir uns nochmal an die Zufallsvariable X aus Beispiel 3.6 (Anzahl Kopf beim zweifachen Münzwurf). Die Massenfunktion von X ist gegeben durch

$$p_X(0) = \frac{1}{4}, \quad p_X(1) = \frac{1}{2}, \quad p_X(2) = \frac{1}{4}$$

und den Erwartungswert hatten wir in Beispiel 3.40 bereits berechnet als

$$\mathbb{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

Sei nun $Y := X^2$. Dies ist ebenfalls eine Zufallsvariable und ihre Massenfunktion ist wie folgt gegeben:

$$p_Y(0) = \frac{1}{4}, \quad p_Y(1) = \frac{1}{2}, \quad p_Y(4) = \frac{1}{4}.$$

Ihr Erwartungswert ist demnach

$$\mathbb{E}[Y] = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} = \frac{3}{2}.$$

Insbesondere gilt $\mathbb{E}[X^2] \neq (\mathbb{E}[X])^2$.

Sehen wir uns im Folgenden an, wie der Erwartungswert einer Zufallsvariablen berechnet wird, die als Funktion einer anderen Zufallsvariablen gegeben ist.

Satz 3.45

Sei X eine diskret verteilte Zufallsvariable mit Massenfunktion p_X . Sei zudem $g: \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion, welche die Bedingung

$$\sum_{x: p_X(x) > 0} |g(x)| \cdot p_X(x) < \infty$$

erfüllt. Dann gilt

$$\mathbb{E}[g(X)] = \sum_{x: p_X(x) > 0} g(x) \cdot p_X(x).$$

Beweis. Sei $Y := g(X)$. Ist p_Y die Massenfunktion von Y , so gilt

$$\mathbb{E}[g(X)] = \sum_{y: p_Y(y) > 0} y \cdot p_Y(y) = \sum_{y: p_Y(y) > 0} \left(y \sum_{x \in g^{-1}(y): p_X(x) > 0} p_X(x) \right),$$

denn

$$\mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \mathbb{P}(X \in g^{-1}(y)).$$

Ersetzen wir nun y durch $g(x)$, so erhalten wir

$$\mathbb{E}[g(X)] = \sum_{y: p_Y(y) > 0} \left(\sum_{x: g(x)=y, p_X(x) > 0} g(x) \cdot p_X(x) \right) = \sum_{x: p_X(x) > 0} g(x) \cdot p_X(x).$$

□

Korollar 3.46: Linearität des Erwartungswertes

Seien $a, b \in \mathbb{R}$ und sei X eine diskrete Zufallsvariable mit Massenfunktion p_X , deren Erwartungswert existiert. Dann gilt

$$\mathbb{E}[aX + b] = a \cdot \mathbb{E}[X] + b.$$

Beweis. Wir wählen $g(x) = ax + b$ in Satz 3.45. Es ist

$$\begin{aligned} \sum_{x: p_X(x) > 0} |g(x)| p_X(x) &= \sum_{x: p_X(x) > 0} |ax + b| p_X(x) \\ &\leq |a| \sum_{x: p_X(x) > 0} |x| p_X(x) + |b| \sum_{x: p_X(x) > 0} p_X(x) \\ &= |a| \sum_{x: p_X(x) > 0} |x| p_X(x) + |b| < \infty. \end{aligned}$$

Damit ist die Voraussetzung erfüllt und wir können den Erwartungswert von $aX + b$ berechnen:

$$\begin{aligned} \mathbb{E}[aX + b] &= \sum_{x: p_X(x) > 0} (ax + b) \cdot p_X(x) = a \cdot \underbrace{\sum_{x: p_X(x) > 0} x p_X(x)}_{=\mathbb{E}[X]} + b \cdot \underbrace{\sum_{x: p_X(x) > 0} p_X(x)}_{=1} \\ &= a \cdot \mathbb{E}[X] + b. \end{aligned}$$

□



Beispiel 3.47. Eine faire Münze wird dreimal nacheinander geworfen. Liegt nach einem Wurf Zahl oben, so erhält man 2 Euro, liegt Kopf oben, so bekommt man für diesen Wurf nichts.

- Sei X die Auszahlung des Spiels. Wie ist X verteilt?
- Wie hoch ist die erwartete Auszahlung des Spiels?
- Wie hoch muss der Spieleinsatz sein, damit das Spiel fair ist, d.h. dass der erwartete Gewinn 0 Euro beträgt?

Antwort:

- In Beispiel 3.3 hatten wir bereits die Massenfunktion der Anzahl Zahl-Würfe bestimmt. Dies ist genau die Hälfte der Auszahlung, d.h.

$$\mathbb{P}(X = 0) = \frac{1}{8}, \quad \mathbb{P}(X = 2) = \frac{3}{8}, \quad \mathbb{P}(X = 4) = \frac{3}{8}, \quad \mathbb{P}(X = 6) = \frac{1}{8}.$$

- Die erwartete Auszahlung ist

$$\begin{aligned} \mathbb{E}[X] &= \\ &= \end{aligned}$$

- Ist $a \in \mathbb{R}$ der Spieleinsatz und bezeichnen wir mit G den Gewinn des Spiels, so ist $G = X - a$. Damit ist

$$\mathbb{E}[G] = \quad = 0 \iff a =$$

Der Erwartungswert einer Zufallsvariablen beschreibt den Schwerpunkt der Verteilung von X ; er sagt jedoch nichts darüber aus, wie stark die angenommenen Werte um diesen Schwerpunkt herum gestreut sind.

Beispiel 3.48. Seien X , Y und Z Zufallsvariablen mit

$$p_X(0) = 1, \quad p_Y(-1) = \frac{1}{2} = p_Y(1), \quad p_Z(-100) = \frac{1}{2} = p_Z(100).$$

Dann gilt $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[Z] = 0$.

Um die Abweichung vom Erwartungswert zu messen, führen wir die *Varianz* einer Zufallsvariablen ein.

Definition 3.49

Sei X eine diskret verteilte Zufallsvariable mit Massenfunktion p_X . Gilt

$$\sum_{x: p_X(x) > 0} x^2 \cdot p_X(x) < \infty,$$

so ist die *Varianz* von X definiert durch

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

Bemerkung 3.50.

1. Gilt $\sum_{x: p_X(x) > 0} x^2 \cdot p_X(x) < \infty$, so existiert der Erwartungswert und auch die Varianz ist wohldefiniert, d.h. die zu Grunde liegende Reihe konvergiert (absolut).
2. Die Varianz beschreibt die mittlere quadratische Abweichung einer Zufallsvariablen von deren Erwartungswert.
3. In Beispiel 3.44 haben wir gesehen, dass im Allgemeinen $\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$ ist, d.h. im Allgemeinen ist die Varianz nicht Null. In der Tat ist sie nur dann Null, wenn die Zufallsvariable fast sicher konstant ist, d.h. wenn $\mathbb{P}(X = c) = 1$ für ein $c \in \mathbb{R}$ ist (vgl. Korollar 3.53).

Mit folgender Formel lässt sich die Varianz in der Praxis recht gut berechnen.

Lemma 3.51

Ist X eine **diskret verteilte** Zufallsvariable mit existierender Varianz, so gilt $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Beweis. Die diskret verteilte Zufallsvariable X habe die Massenfunktion p_X . Dann gilt

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \stackrel{\text{Satz 3.45}}{=} \sum_{x: p_X(x) > 0} (x - \mathbb{E}[X])^2 \cdot p_X(x) \\ &= \sum_{x: p_X(x) > 0} (x^2 - 2x\mathbb{E}[X] + \mathbb{E}[X]^2) \cdot p_X(x) \\ &= \sum_{x: p_X(x) > 0} x^2 p_X(x) - 2\mathbb{E}[X] \cdot \sum_{x: p_X(x) > 0} x p_X(x) + \mathbb{E}[X]^2 \cdot \sum_{x: p_X(x) > 0} p_X(x) \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X] \cdot \mathbb{E}[X] + \mathbb{E}[X]^2 \cdot 1 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2. \end{aligned}$$

□

Beispiel 3.52. Sei X die geworfene Augenzahl eines fairen Würfels. Wie groß ist ihre Varianz?
Antwort: Die Massenfunktion von X ist gegeben durch



$$p_X(k) = \frac{1}{6}, \quad \forall k \in \{1, 2, 3, 4, 5, 6\}.$$

Damit ist

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = \frac{1}{6} \sum_{k=1}^6 k = \\ \mathbb{E}[X^2] &= \frac{1}{6} \cdot (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{1}{6} \sum_{k=1}^6 k^2 = \end{aligned}$$

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

Alternativ könnte man berechnen:

$$\begin{aligned} \text{Var}(X) &= \sum_{k=1}^6 (k - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = k) = \sum_{k=1}^6 \left(k - \frac{7}{2}\right)^2 \cdot \frac{1}{6} \\ &= \frac{1}{6} \cdot ((-2, 5)^2 + (-1, 5)^2 + (-0, 5)^2 + 0, 5^2 + 1, 5^2 + 2, 5^2). \end{aligned}$$

Lemma 3.51 ist nicht nur für die Berechnung der Varianz nützlich, sondern wir können damit auch folgende Aussage beweisen:

Korollar 3.53

Ist X eine diskrete Zufallsvariable mit existierender Varianz, so gilt $\text{Var}(X) \geq 0$ und somit $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$. Ist X nicht konstant, so gilt sogar $\mathbb{E}[X^2] > \mathbb{E}[X]^2$.

Beweis. Übungsaufgabe. □

Lemma 3.54

Sei X eine diskret verteilte Zufallsvariable mit existierender Varianz und seien $a, b \in \mathbb{R}$ Konstanten. Dann gilt

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

Insbesondere folgt aus diesem Lemma, dass stets $\text{Var}(-X) = \text{Var}(X)$ ist.

Beweis.

$$\begin{aligned} \text{Var}(aX + b) &= \mathbb{E} \left[(aX + b - \mathbb{E}[aX + b])^2 \right] \\ &= \mathbb{E} \left[(aX + b - a\mathbb{E}[X] - b)^2 \right] \\ &= \mathbb{E} \left[a^2 \cdot (X - \mathbb{E}[X])^2 \right] \\ &= a^2 \cdot \text{Var}(X). \end{aligned}$$

□

Sobald wir mit physikalischen Größen oder generell mit Einheiten rechnen, haben wir das Problem, dass die Varianz durch das Quadrat schwer zu interpretieren ist. Daher ist es nützlich, dass wir neben der Varianz als Maß für die Streuung auch folgenden Begriff zur Verfügung haben:

Definition 3.55

Ist X eine diskrete Zufallsvariable mit existierender Varianz $\text{Var}(X)$, so heißt $\sigma(X) := \sqrt{\text{Var}(X)}$ Standardabweichung von X .

Sehen wir uns die Varianzen einer uns bekannter Verteilungen an. Dabei werden einige schon gesehene „Tricks“ wieder auftreten – diese können auch bei den hier nicht betrachteten Verteilungen angewendet werden.

Beispiel 3.56. Sei $X \sim \text{Ber}(p)$, d.h. X ist Bernoulli-verteilt mit Parameter p . Dann gilt für die Massenfunktion $\mathbb{P}(X = 1) = p$ und $\mathbb{P}(X = 0) = 1 - p$. Folglich ist

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = (1^2 \cdot p + 0^2 \cdot (1 - p)) - (1 \cdot p + 0 \cdot (1 - p))^2 = p - p^2 = p(1 - p).$$

Beispiel 3.57. Sei $X \sim \text{Geo}(p)$. Welche Varianz hat X ?

Antwort: In Beispiel 3.42 haben wir bereits den Erwartungswert als $\mathbb{E}[X] = \frac{1}{p}$ berechnet. Für die Varianz benötigen wir $\mathbb{E}[X^2]$.

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{k=1}^{\infty} k^2 p_X(k) = \sum_{k=1}^{\infty} k(k+1)p_X(k) - \sum_{k=1}^{\infty} kp_X(k) \\
 &= \sum_{k=1}^{\infty} k(k+1)p(1-p)^{k-1} - \frac{1}{p} \\
 &= 2p \sum_{k=1}^{\infty} \frac{k(k+1)}{2} \cdot (1-p)^{k-1} - \frac{1}{p} \\
 &= 2p \sum_{k=1}^{\infty} \left((1-p)^{k-1} \cdot \sum_{j=1}^k j \right) - \frac{1}{p} \\
 &= 2p \sum_{k=1}^{\infty} \sum_{j=1}^k \left((1-p)^{k-1} \cdot j \right) - \frac{1}{p} && \text{analog zu Beispiel 3.42} \\
 &= 2p \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \left((1-p)^{k-1} \cdot j \right) - \frac{1}{p} \\
 &= 2p \sum_{j=1}^{\infty} \left(j \cdot \sum_{k=j}^{\infty} \left((1-p)^{k-1} \cdot \frac{(1-p)^{j-1}}{(1-p)^{j-1}} \right) \right) - \frac{1}{p} && \text{Indexverschiebung} \\
 &= 2p \sum_{j=1}^{\infty} \left(j \cdot (1-p)^{j-1} \cdot \sum_{l=0}^{\infty} (1-p)^l \right) - \frac{1}{p} && \text{Formel der geom. Reihe} \\
 &= 2p \sum_{j=1}^{\infty} \left(j \cdot (1-p)^{j-1} \cdot \frac{1}{p} \right) - \frac{1}{p} \\
 &= \frac{2}{p} \sum_{j=1}^{\infty} (jp(1-p)^{j-1}) - \frac{1}{p} && \text{Erwartungswert von } \text{Geo}(p) \\
 &= \frac{2}{p^2} - \frac{1}{p}.
 \end{aligned}$$

Zusammen mit dem bekannten Erwartungswert ist also

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \left(\frac{2}{p^2} - \frac{1}{p} \right) - \left(\frac{1}{p} \right)^2 = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}.$$

Beispiel 3.58. Sei $X \sim \text{Bin}(n, p)$ mit $n \in \mathbb{N}$ und $p \in (0, 1)$. Wie lauten Erwartungswert und Varianz von X ? 

Antwort: Berechnen wir zunächst den Erwartungswert.

$$\mathbb{E}[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{(n-1)-(k-1)},$$

denn

Nach einer Indexverschiebung erhalten wir

$$\mathbb{E}[X] = np \sum_{j=0}^{n-1}$$

In Satz 3.18 haben wir bereits gezeigt, dass für $Y \sim \text{Bin}(m, q)$

$$\sum_{i=0}^m \binom{m}{i} q^i (1-q)^{m-i} = 1$$

ist. Setzen wir $m =$ und $q =$, so ergibt sich

$$\mathbb{E}[X] = np.$$

Für die Varianz berechnen wir mit den gleichen Schritten $\mathbb{E}[X^2]$.

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^n k \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{j=0}^{n-1} (j+1) \binom{n-1}{j} p^j (1-p)^{n-1-j} = np \cdot (\mathbb{E}[Y] + 1) \end{aligned}$$

mit $Y \sim \text{Bin}(n-1, p)$

$$= np \cdot \left[\right].$$

Somit ist

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



Übersicht über die Kennzahlen der wichtigsten diskreten Verteilungen

	$k \in$	$p_X(k)$	$\mathbb{E}[X]$	$\text{Var}(X)$	$\sum p_X(k) = 1$
allg. Formel					Satz 3.10
$U(\{1, \dots, n\})$	$\{1, \dots, n\}$				
Ber(p)		$(1-p) \cdot \mathbb{1}_{\{0\}}(k) + p \cdot \mathbb{1}_{\{1\}}(k)$			
Bin(n, p)					
Geo(p)					
Poisson(λ)					
Hyp(N, K, n)					

3.5 Verteilungsfunktionen von Zufallsvariablen

Während sich bei den meisten Kenngrößen die Formeln je nach Wertebereich der Zufallsvariablen unterscheiden, ist dies bei der Verteilungsfunktion nicht der Fall – die folgende Formel gilt ausdrücklich nicht nur für diskret verteilte Zufallsvariablen.

Definition 3.59

Für eine Zufallsvariable X heißt

$$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq x\}), \quad x \in \mathbb{R},$$

Verteilungsfunktion von X .

Bemerkung 3.60. Es gibt in der Literatur auch \mathbb{R}^n -wertige Zufallsvariablen – diese klammern wir vorerst ausdrücklich aus, da Verteilungsfunktionen so wie wir sie definiert haben sonst keinen Sinn ergeben. Sollten wir doch einer Abbildung $X: \Omega \rightarrow \mathbb{R}^n$ begegnen, so sprechen wir in Abgrenzung zu Zufallsvariablen mit Werten in \mathbb{R} in diesem Fall von Zufallsvektoren.

Bevor wir sehen, wie wir mit Verteilungsfunktionen rechnen, sammeln wir einige Eigenschaften von Verteilungsfunktionen.

Satz 3.61

Sei X eine reellwertige Zufallsvariable. Dann hat ihre Verteilungsfunktion F_X folgende Eigenschaften:

1. $0 \leq F_X(x) \leq 1$ für alle $x \in \mathbb{R}$;
2. F_X ist monoton wachsend, d.h. aus $x \leq y$ folgt stets $F_X(x) \leq F_X(y)$;
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$;
4. $\lim_{x \rightarrow \infty} F_X(x) = 1$;
5. F_X ist rechts-stetig, d.h. $\lim_{y \downarrow x} F_X(y) = F_X(x)$.

Beweis.

1. Da \mathbb{P} ein Wahrscheinlichkeitsmaß ist, gilt $\mathbb{P}(A) \in [0, 1]$ für alle Ereignisse $A \in \mathcal{A}$. Insbesondere gilt also

$$0 \leq \mathbb{P}(X \leq x) = F_X(x) \leq 1, \quad \forall x \in \mathbb{R}.$$

2. Ist $x \leq y$, so ist $\{\omega \in \Omega \mid X(\omega) \leq x\} \subseteq \{\omega \in \Omega \mid X(\omega) \leq y\}$.

Folglich ist $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$.

3. Die Folge von Ereignissen $(A_n)_{n \in \mathbb{N}}$, welche gegeben ist durch $A_n := \{\omega \in \Omega \mid X(\omega) \leq -n\}$, ist eine absteigende Folge, d.h. $A_n \supseteq A_{n+1}$ für alle $n \in \mathbb{N}$. Aus der Stetigkeit des Wahrscheinlichkeitsmaßes von oben (siehe Satz 2.29) folgt also

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow \infty} F_X(-n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \mathbb{P}(\emptyset) = 0.$$

4. Für die aufsteigende Folge $(B_n)_{n \in \mathbb{N}}$, welche gegeben ist durch $B_n := \{\omega \in \Omega \mid X(\omega) \leq n\}$, gilt wegen der Stetigkeit des Wahrscheinlichkeitsmaßes von unten

$$\lim_{x \rightarrow \infty} F_X(x) = \lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}(\Omega) = 1.$$

5. Sei $(x_n)_{n \in \mathbb{N}}$ eine monoton fallende und gegen $x \in \mathbb{R}$ konvergierende Folge, d.h. $x_1 \geq x_2 \geq x_3 \geq \dots$ und $\lim_{n \rightarrow \infty} x_n = x$. Wir wollen zeigen, dass

$$\lim_{n \rightarrow \infty} F_X(x_n) = F_X(x)$$

ist. Da die Ereignisse $C_n := \{\omega \in \Omega \mid X \leq x_n\}$ eine absteigende Folge bilden, gilt analog zu den vorherigen Aussagen

$$\lim_{y \downarrow x} F_X(y) = \lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} \mathbb{P}(C_n) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} \{X \leq x_n\}\right) = \mathbb{P}(X \leq x) = F_X(x).$$

□

Aus der Massenfunktion kann man einfach die Verteilungsfunktion einer diskret verteilten Zufallsvariablen bestimmen.

Beispiel 3.62. Sei $X \sim U(\{1, 2, 3, 4, 5, 6\})$ das Ergebnis eines Würfelwurfes. Dann gilt

$$p_X(k) = \frac{1}{6}, \quad \forall k \in \{1, 2, 3, 4, 5, 6\}.$$

Folglich ist für $x \in \mathbb{R}$

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{k=1}^6 \frac{1}{6} \mathbb{1}_{\{x \geq k\}} = \begin{cases} 0 & , x < 1 \\ \frac{1}{6} & , 1 \leq x < 2 \\ \frac{2}{6} & , 2 \leq x < 3 \\ \frac{3}{6} & , 3 \leq x < 4 \\ \frac{4}{6} & , 4 \leq x < 5 \\ \frac{5}{6} & , 5 \leq x < 6 \\ 1 & , 6 \leq x. \end{cases}$$

Kennen wir die Verteilungsfunktion einer Zufallsvariablen, so können wir daraus andersherum auch die Massenfunktion und somit alle möglichen für uns interessanten Wahrscheinlichkeiten berechnen. Wie das geht, illustrieren wir anhand eines weiteren Beispiels.

Beispiel 3.63. Eine Zufallsvariable habe folgende Verteilungsfunktion:

$$F_X(x) = \begin{cases} 0 & , x < 0 \\ \frac{1}{2} & , 0 \leq x < 1 \\ \frac{2}{3} & , 1 \leq x < 2 \\ \frac{11}{12} & , 2 \leq x < 3 \\ 1 & , 3 \leq x. \end{cases}$$

Berechnen Sie folgende Wahrscheinlichkeiten

$$\mathbb{P}(X = 1), \quad \mathbb{P}(X < 3), \quad \mathbb{P}\left(X > \frac{1}{2}\right), \quad \mathbb{P}(2 < X \leq 4).$$

Antwort: Es ist

$$\mathbb{P}(X = 1) = \mathbb{P}(X \leq 1) - \mathbb{P}(X < 1) \stackrel{(*)}{=} F_X(1) - \lim_{n \rightarrow \infty} F_X\left(1 - \frac{1}{n}\right) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6},$$

wobei wir in (*) die Stetigkeit des Wahrscheinlichkeitsmaßes verwendet haben. Es ist nämlich

$$(-\infty, 1) = \bigcup_{n \in \mathbb{N}} \left(-\infty, 1 - \frac{1}{n}\right]$$

eine aufsteigende Folge von Ereignissen, daher gilt

$$\mathbb{P}^X((-\infty, 1)) = \lim_{n \rightarrow \infty} \mathbb{P}^X\left(\left(-\infty, 1 - \frac{1}{n}\right]\right) = \lim_{n \rightarrow \infty} F_X\left(1 - \frac{1}{n}\right).$$

Mit der gleichen Idee und den bekannten Rechenregeln berechnen wir die übrigen Wahrscheinlichkeiten: 

$$\begin{aligned} \mathbb{P}(X < 3) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(X \leq \quad\quad\quad\right) = \lim_{n \rightarrow \infty} F_X(\quad\quad\quad) = \\ \mathbb{P}\left(X > \frac{1}{2}\right) &= \\ \mathbb{P}(2 < X \leq 4) &= \end{aligned}$$

Bemerkung 3.64. Die Verteilung einer diskreten Zufallsvariablen X kann durch das zugehörige Wahrscheinlichkeitsmaß \mathbb{P}^X , durch die Massenfunktion p_X oder durch die Verteilungsfunktion F_X eindeutig beschrieben werden.

Sofern nichts anderes gesagt wird, soll bei der Frage nach der Verteilung einer diskreten Zufallsvariablen deren Massenfunktion angegeben werden. Die Verteilungsfunktion ist hingegen gegeben durch $F_X(x) = \mathbb{P}(X \leq x)$. Wenn nach der Art der Verteilung gefragt wird, könnte die Antwort beispielsweise $X \sim \text{Bin}(n, p)$ lauten, d.h. im Falle eines bekannten Verteilungstyps ist es nicht nötig, die Massenfunktion oder Verteilungsfunktion anzugeben. 

Kapitel 4

Absolutstetige Zufallsvariablen und Verteilungen

Ziele

- Erwartungswerte, Varianzen und andere Verteilungsfunktionen absolutstetig verteilter ZVen berechnen können
- die wichtigsten absolutstetigen Verteilungen kennen, typische Beispiele für deren Auftreten und einige Kennzahlen kennen
- Zusammenhänge zu den bisher bekannten Verteilungen kennen

4.1 Wann heißt eine Verteilung absolutstetig?

Definition 4.1

Eine integrierbare Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ heißt *Dichte*, falls gilt:

1. $f(x) \geq 0$ für alle $x \in \mathbb{R}$,
2. $\int_{\mathbb{R}} f(x) dx = 1$.

Notation

Ist $B \subseteq \mathbb{R}$, so ist die Notation $\int_B f(x) dx$ wie folgt zu verstehen:

$$\int_B f(x) dx = \int_{-\infty}^{\infty} \mathbb{1}_B(x) f(x) dx.$$

Beispiel 4.2. Sei $f: \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch

$$f(x) = \begin{cases} c \cdot (4x - 2x^2), & 0 < x < 2 \\ 0, & \text{sonst.} \end{cases}$$

Kann man $c \in \mathbb{R}$ derart wählen, dass f eine Dichte ist?

Antwort: Überprüfen wir zunächst die Nichtnegativität: Ist $x \notin (0, 2)$, so ist dies offensichtlich erfüllt. Sei daher $x \in (0, 2)$. Dann gilt

$$f(x) \geq 0 \iff c \cdot (4x - 2x^2) \geq 0 \iff c \cdot 2x \cdot (2 - x) \geq 0.$$

Da für $x \in (0, 2)$ sowohl $2x > 0$ als auch $2 - x > 0$ ist, ist die Nichtnegativität genau dann erfüllt, wenn $c \geq 0$ ist.

Kommen wir nun zur Normierung (und gleichzeitig zur Integrierbarkeit):

$$1 \stackrel{!}{=} \int_{\mathbb{R}} f(x) dx = \int_{-\infty}^{\infty} c(4x - 2x^2) \mathbb{1}_{(0,2)}(x) dx = c \int_0^2 (4x - 2x^2) dx. \quad (4.1.1)$$



Berechnen wir das Integral:

$$\int_0^2 (4x - 2x^2) dx =$$

Setzen wir diesen Wert für das Integral in (4.1.1) ein und stellen die Gleichung nach c um, so erhalten wir $c = \frac{3}{8}$. Die Integrierbarkeit von f ist für diesen Parameter damit offensichtlich gegeben und die Nichtnegativität ist für $c = \frac{3}{8} > 0$ erfüllt. Folglich ist f für $c = \frac{3}{8}$ eine Dichte.

Hat man eine Dichte, so kann man sich mit deren Hilfe ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ konstruieren. Die Idee hinter der Konstruktion ist die, dass man die Wahrscheinlichkeiten aller Mengen eines Erzeugers der Borel- σ -Algebra festlegt und alle anderen Wahrscheinlichkeiten daraus hergeleitet werden können. Ein möglicher Erzeuger sind die offenen Mengen, ein anderer die linksoffenen Intervalle.

Satz 4.3

Sei $f: \mathbb{R} \rightarrow \mathbb{R}$ eine Dichte. Dann existiert genau ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, welches

$$\mathbb{P}((-\infty, x]) = \int_{-\infty}^x f(y) dy$$

für alle $x \in \mathbb{R}$ erfüllt.

Nun kann man sich fragen, ob jedes Wahrscheinlichkeitsmaß eine Dichte besitzt. Dies ist nicht der Fall – Gegenbeispiele sind u.a. sämtliche diskreten Verteilungen, die wir im vorherigen Kapitel kennen gelernt haben. Mit Hilfe von Dichten können wir definieren, wann eine Zufallsvariable bzw. deren Verteilung *absolutstetig* genannt wird.

Definition 4.4

Eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt (*absolut*) *stetig* (*verteilt*), falls eine Dichte f existiert, so dass für alle (messbaren) Mengen $B \in \mathcal{B}(\mathbb{R})$ gilt:

$$\mathbb{P}^X(B) = \mathbb{P}(X \in B) = \int_B f(x) dx.$$

In diesem Fall nennen wir auch das Wahrscheinlichkeitsmaß \mathbb{P}^X *absolutstetig*.

Bemerkung 4.5. Sei f_X die Dichte und F_X die Verteilungsfunktion von X . Setzt man $B = (-\infty, x]$, so ist

$$F_X(x) = \mathbb{P}^X((-\infty, x]) = \int_{-\infty}^x f_X(y) dy$$

die Verteilungsfunktion von X . Insbesondere gilt somit auch $f_X(x) = F'_X(x)$.

Bemerkung 4.6 (Notation). Möchte man betonen, dass f die Dichte und F die Verteilungsfunktion einer Zufallsvariablen X ist, so notiert man beide mit Index, d.h. f_X bzw. F_X . Ist dies nicht nötig, da keine Verwechslungsgefahr besteht, so können die Indizes auch weggelassen werden.

Bemerkung 4.7 (Begriffsherkunft). Der Begriff eines absolutstetigen (Wahrscheinlichkeits-)Maßes wird in der Maßtheorie in Relation zu einem zweiten Maß genannt. In unserem Setting wäre das zweite Maß Lebesgue-Maß λ und \mathbb{P} heißt absolutstetig bezüglich λ , wenn aus $\lambda(B) = 0$ immer $\mathbb{P}(B) = 0$ für alle messbaren Mengen B folgt. In diesem Fall besagt der Satz von Radon-Nikodym, dass eine Dichte f existiert, so dass $\mathbb{P}(B) = \int_B f d\lambda$ ist. In der uns bekannten Schreibweise für das Riemann-Integral entspricht $d\lambda$ gerade dx und wir erhalten die Formel der Definition für das Wahrscheinlichkeitsmaß \mathbb{P}^X .

Andererseits heißt eine Funktion $F: \mathbb{R} \rightarrow \mathbb{R}$ absolut stetig, falls für alle $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass für jede Folge paarweise disjunkter Intervalle $(x_k, y_k) \subseteq I$ mit $\sum_{k=1}^n (y_k - x_k) < \delta$ stets $\sum_{k=1}^n |F(y_k) - F(x_k)| < \varepsilon$ ist. Ist F monoton wachsend (wie jede Verteilungsfunktion), so ist $F: [a, b] \rightarrow \mathbb{R}$ genau dann absolut stetig, wenn F fast überall differenzierbar ist mit integrierbarer Ableitung f und $F(x) - F(a) = \int_a^x f(y) dy$ für alle $x \in [a, b]$ gilt.

Beispiel 4.8. Für jedes Intervall $[a, b] \subseteq [0, 1]$ sei $\mathbb{P}([a, b]) = b - a$ und für ein Intervall $I \not\subseteq [0, 1]$ gelte $\mathbb{P}(I) := \mathbb{P}(I \cap [0, 1])$. Ist dieses Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ absolutstetig?

Antwort: Zunächst einmal ist

$$\mathbb{P}(\mathbb{R}) = \mathbb{P}(\mathbb{R} \cap [0, 1]) = \mathbb{P}([0, 1]) = 1 - 0 = 1.$$

Da \mathbb{P} keine negativen Werte annehmen kann, gilt somit $\mathbb{P}(A) \in [0, 1]$ für alle $A \in \mathcal{B}(\mathbb{R})$. Die σ -Additivität gilt ebenfalls.¹ Zudem gilt für jedes $x \in \mathbb{R}$

$$\begin{aligned} \mathbb{P}((-\infty, x]) &= \begin{cases} \mathbb{P}(\emptyset) = 0, & x < 0, \\ \mathbb{P}([0, x]) = x - 0 = x, & x \in [0, 1], \\ \mathbb{P}([0, 1]) = 1 - 0 = 1, & x > 1 \end{cases} \\ &= \int_{-\infty}^x \mathbb{1}_{[0,1]}(y) dy \end{aligned}$$

Die Dichte dieser Verteilung, der Gleichverteilung auf $[0, 1]$, ist also gegeben durch $f(x) = \mathbb{1}_{[0,1]}(x)$. Folglich handelt es sich um eine absolutstetige Verteilung.

Beispiel 4.9. Die Dauer (in Stunden), die ein Drucker ohne Wartung am Stück funktioniert, sei eine Zufallsvariable mit folgender Dichte:

$$f(x) = \lambda \cdot e^{-x/100} \mathbb{1}_{[0,\infty)}(x) = \begin{cases} \lambda e^{-x/100}, & x \geq 0 \\ 0, & \text{sonst.} \end{cases}$$

- Wie muss $\lambda \in \mathbb{R}$ gewählt werden, damit es sich bei f um eine Dichte handelt?
- Wie lautet die Verteilungsfunktion F ?

¹Das definierte Wahrscheinlichkeitsmaß erbt seine σ -Additivität vom Lebesgue-Maß.

- c) Wie hoch ist die Wahrscheinlichkeit, dass der Drucker zwischen 50 und 100 Stunden funktioniert bevor er ausfällt?
- d) Wie hoch ist die Wahrscheinlichkeit, dass der Drucker (strikt) weniger als 100 Stunden funktioniert?

Antwort:

- a) Damit die Dichte nichtnegativ ist, muss $\lambda \geq 0$ sein. Zudem soll gelten

$$1 \stackrel{!}{=} \int_{\mathbb{R}} f(x) dx =$$

Folglich ist $\lambda =$.

- b) Wir müssen zwischen zwei Fällen unterscheiden.

$x < 0$: In diesem Fall ist

$$F_X(x) = \int_{-\infty}^x f(y) dy =$$

$x \geq 0$: In diesem Fall ist

$$F_X(x) = \int_{-\infty}^x f(y) dy =$$

- c) Mit der schon bekannten Stammfunktion ist

$$\begin{aligned} \mathbb{P}(50 < X < 100) &= \int_{50}^{100} f(x) dx = \\ &= \end{aligned}$$

- d) Hier sehen wir deutlich den Unterschied zu den diskreten Verteilungen:

$$\mathbb{P}(X < 100) = \int_{-\infty}^{100} f(x) dx = F_X(100) =$$

Hat eine Zufallsvariable X eine bekannte Dichte f_X und ist $Y = g(X)$, so ist Y , sofern g bestimmte Voraussetzungen erfüllt, ebenfalls absolutstetig verteilt und wir können die zugehörige Dichte f_Y wie folgt bestimmen:

Satz 4.10: Dichtetransformationssatz

Sei X absolutstetig verteilt mit Dichte f_X und sei $g: \mathbb{R} \rightarrow \mathbb{R}$ strikt monoton und differenzierbar. Dann hat $Y := g(X)$ die Dichte

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)|, & y \in g(X(\Omega)), \\ 0, & \text{sonst,} \end{cases}$$

wobei g^{-1} die Umkehrfunktion von g bezeichnet.

Beweis. Sei g strikt monoton wachsend. Dann ist g^{-1} ebenfalls strikt monoton wachsend und für $y \in g(\mathbb{R})$ gilt

$$F_Y(y) = \mathbb{P}(g(X) \leq y) \stackrel{(1)}{=} \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Durch Ableiten erhalten wir die Dichte von Y :

$$f_Y(y) = F'_Y(y) = (F_X(g^{-1}(y)))' = f_X(g^{-1}(y)) \cdot (g^{-1})'(y) = f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)|.$$

Ist andererseits g strikt monoton fallend, so dreht sich das Vorzeichen in (1) um, d.h.

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) \stackrel{(2)}{=} 1 - F_X(g^{-1}(y)),$$

wobei (2) auf Grund der Absolutstetigkeit gilt, d.h. für beliebige x^* gilt

$$\mathbb{P}(X \geq x^*) = 1 - \mathbb{P}(X < x^*) = 1 - \mathbb{P}(X \leq x^*) = 1 - F_X(x^*).$$

Durch Ableiten der Verteilungsfunktion erhalten wir wieder die Dichte als

$$f_Y(y) = F'_Y(y) = -(F_X(g^{-1}(y)))' = -f_X(g^{-1}(y)) \cdot (g^{-1})'(y) = f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)|.$$

□

Die Formel der Dichtetransformation muss man sich nicht merken um die Verteilung von $g(X)$ zu bestimmen, wenn die Dichte f_X bekannt ist. Die Methode aus dem Beweis kann auch direkt angewendet werden und sie ermöglicht es sogar, die Anforderungen an g ein wenig abzuschwächen, wie wir im anschließenden Beispiel sehen werden.

Methode zur Bestimmung der Dichte von $g(X)$

Sei f_X die Dichte von X und $Y = g(X)$. Gesucht ist f_Y .

1. Aus f_X mittels Integration F_X bestimmen:

$$F_X(x) = \int_{-\infty}^x f(y) dy$$

2. $g(X) \leq y$ nach X umstellen:

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \in B_y)$$

Ist g (ggf. lokal) strikt monoton und invertierbar, so ist B_y von der Form $(-\infty, g^{-1}(y)]$ bzw. $[g^{-1}(y), \infty)$.

3. Die gefundene Wahrscheinlichkeit mit Hilfe von F_X ausdrücken und anschließend den gefundenen Ausdruck nach y ableiten.

Beispiel 4.11. Sei X absolutstetig mit Dichte f_X und sei $Y := X^2$. Für $y \leq 0$ ist dann $F_Y(y) \equiv 0$ und somit $f_Y(y) \equiv 0$. Für $y > 0$ berechnen wir die Dichte wie beschrieben (Schritte 2 und 3):

$$F_Y(y) =$$

Ableiten nach y ergibt

$$f_Y(y) =$$



4.2 Erwartungswert und Varianz

Definition 4.12

Sei X eine absolutstetige Zufallsvariable mit Dichte f . Gilt

$$\int_{\mathbb{R}} |x| f(x) dx < \infty,$$

so ist der *Erwartungswert* von X definiert als

$$\mathbb{E}[X] := \int_{\mathbb{R}} x f(x) dx.$$

Beispiel 4.13. Sei $f(x) = \mathbb{1}_{[0,1]}(x)$. Dies ist offensichtlich eine Dichte, denn

$$\int_{\mathbb{R}} f(x) dx = \int_0^1 dx = [x]_0^1 = 1.$$

Ist X eine Zufallsvariable, deren Dichte durch f gegeben ist, so ist ihr Erwartungswert

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx = \int_0^1 x dx = \left[\frac{1}{2} x^2 \right]_0^1 = \frac{1}{2}.$$

Bevor wir weitere Beispiele berechnen, wollen wir eine Formel für den Erwartungswert der Funktion einer Zufallsvariablen, d.h. von $\mathbb{E}[g(X)]$, angeben und beweisen.

Satz 4.14

Sei X eine absolutstetig verteilte Zufallsvariable mit Dichte f_X und sei $Y = g(X)$ für eine Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$, welche die Bedingung

$$\int_{\mathbb{R}} |g(x)| f_X(x) dx < \infty$$

erfüllt. Dann gilt für den Erwartungswert von Y

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

Zum Beweis des Satzes benötigen wir folgendes Hilfsresultat:

Lemma 4.15

Sei X eine nichtnegative

- a) diskrete Zufallsvariable mit Massenfunktion p_X ;
- b) absolutstetige Zufallsvariable mit Dichte f_X .

Dann gilt

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > x) dx.$$

Beweis.

a) Es gilt

$$\begin{aligned} \int_0^\infty \mathbb{P}(X > x) dx &= \int_0^\infty \sum_{y>x: p_X(y)>0} \mathbb{P}(X = y) dx \\ &= \sum_{y: p_X(y)>0} p_X(y) \int_0^y dx \\ &= \sum_{y: p_X(y)>0} p_X(y) \cdot y \\ &= \mathbb{E}[X]. \end{aligned}$$

b) Der Beweis funktioniert analog, indem die Summen durch geeignete Integrale ersetzt werden. (Übungsaufgabe)

□

Beweis von Satz 4.14. Sei $g(x) \geq 0$ für alle $x \in \mathbb{R}$. Dann gilt für jede messbare Menge $B \subseteq \mathbb{R}$:

$$\mathbb{P}(X \in B) = \int_B f_X(x) dx.$$

Für jedes $y \in \mathbb{R}_+$ definieren wir die Menge¹

$$A_y := \{x \in \mathbb{R} \mid g(x) > y\}.$$

Dann gilt

$$\mathbb{P}(Y > y) = \mathbb{P}(g(X) > y) = \mathbb{P}(X \in A_y) = \int_{A_y} f_X(x) dx = \int_{\mathbb{R}} \mathbb{1}_{A_y}(x) f_X(x) dx.$$

Wenden wir Lemma 4.15 an, so erhalten wir wieder durch Vertauschen der Integrale

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^\infty \mathbb{P}(Y > y) dy \\ &= \int_0^\infty \int_{\mathbb{R}} \mathbb{1}_{A_y}(x) f_X(x) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{B_x}(y) f_X(x) dy dx \\ &= \int_{\mathbb{R}} f_X(x) \cdot \left(\int_{\mathbb{R}} \mathbb{1}_{B_x}(y) dy \right) dx \\ &= \int_{\mathbb{R}} f_X(x) \cdot \left(\int_0^{g(x)} dy \right) dx \\ &= \int_{\mathbb{R}} f_X(x) g(x) dx, \end{aligned}$$

wobei die Menge B_x für alle $x \in \mathbb{R}$ gegeben ist durch

$$B_x := \{y \geq 0 \mid g(x) > y\}.$$

¹Streng genommen muss g bestimmte Bedingungen erfüllen, damit eine solche Menge in der Borel- σ -Algebra $\mathcal{B}(\mathbb{R})$ liegt. Details dazu gehen über diese Vorlesung hinaus.

Nimmt g anders als bisher angenommen auch negative Werte an, so können wir die Funktion in Positiv- und Negativteil zerlegen, d.h. $g = g^+ - g^-$ mit $g^+, g^- : \mathbb{R} \rightarrow [0, \infty)$. Auf beide Teile können wir das Gezeigte anwenden und im Anschluss die Resultate wieder zusammensetzen. \square



Beispiel 4.16. Sei X die Zufallsvariable mit der Dichte aus Beispiel 4.13, d.h. $f_X(x) = \mathbb{1}_{[0,1]}(x)$. Sei $Y := e^X$ eine weitere Zufallsvariable. Wie lautet deren Erwartungswert $\mathbb{E}[Y]$?

Antwort: Wenden wir Satz 4.14 an auf die gegebene Dichte und $g(x) = e^x$, so erhalten wir sofort den Erwartungswert

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{\mathbb{R}} e^x \cdot \mathbb{1}_{[0,1]}(x) dx = \int_0^1 e^x dx = e - 1.$$

Um zu sehen, wie viel Rechensparnis uns der Satz gebracht hat, berechnen wir den Erwartungswert noch einmal indem wir zunächst die Dichte f_Y bestimmen und dann den Erwartungswert mit der Formel $\mathbb{E}[Y] = \int_{\mathbb{R}} y f_Y(y) dy$ berechnen.

Berechnung der Verteilungsfunktion von Y : Da die Dichte von X nur auf $[0, 1]$ positiv ist, nimmt $Y = e^X$ nur die Werte in $[e^0, e^1] = [1, e]$ an. Folglich gilt $F_Y(y) = 0$ für $y < 1$, $F_Y(y) = 1$ für $y > e$ und für $y \in [1, e]$ ist

$$F_Y(y) =$$

Berechnung der Dichte von Y : Für $y \notin (1, e)$ ist $f_Y(y) \equiv 0$. Im Innern des Intervalls leiten wir die Verteilungsfunktion ab und erhalten

$$f_Y(y) = F'_Y(y) =$$

$$\text{Zusammengefasst ist also } f_Y(y) = \quad .$$

Alternative Berechnung der Dichte mit dem Transformationssatz: $g(x) = e^x$ ist auf ganz \mathbb{R} strikt monoton und differenzierbar und X ist absolutstetig verteilt. Damit ist

$$\begin{aligned} f_Y(y) &= \begin{cases} f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)|, & y \in g(X(\Omega)), \\ 0, & \text{sonst} \end{cases} \\ &= \begin{cases} & y \in [1, e], \\ 0, & \text{sonst} \end{cases} \\ &= \begin{cases} & y \in [1, e], \\ 0, & \text{sonst.} \end{cases} \end{aligned}$$

Berechnung des Erwartungswertes von Y : Nach der bekannten Formel ist

$$\mathbb{E}[Y] = \int_{\mathbb{R}} y f_Y(y) dy =$$

Das folgende Resultat ist das Analogon zu Satz 3.46 für absolutstetige Verteilungen.

Korollar 4.17

Sei X absolutstetig mit Dichte f und existierendem Erwartungswert $\mathbb{E}[X]$. Dann gilt für alle $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

Beweis. Sei $g(x) = ax + b$. Dann ist

$$\mathbb{E}[aX + b] =$$

□

Definition 4.18

Sei X eine absolutstetige Zufallsvariable mit Dichte f , für die $\int_{\mathbb{R}} x^2 f(x) dx < \infty$ erfüllt ist. Dann ist die *Varianz* von X definiert als

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

Bemerkung 4.19 (Zur Existenz der Varianz). Die Bedingung $\int_{\mathbb{R}} x^2 f(x) dx < \infty$ bedeutet, dass $\mathbb{E}[X^2]$ existiert. Wegen $|x| \leq 1 + x^2$ impliziert dies, dass auch $\mathbb{E}[|X|]$, also auch $\mathbb{E}[X]$ existiert. Schlussendlich folgt aus $(x - \mathbb{E}[X])^2 \leq x^2 + 2|x|\mathbb{E}[X] + (\mathbb{E}[X])^2$ die Existenz der Varianz.

Die Aussage von Lemma 3.51 können wir auch für absolutstetig verteilte Zufallsvariablen beweisen.

Lemma 4.20

Ist X eine absolutstetig verteilte Zufallsvariable mit existierender Varianz, so gilt $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Beweis. Ist f die Dichte von X , so gilt

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x) dx \\ &= \int_{\mathbb{R}} x^2 f(x) dx - 2\mathbb{E}[X] \cdot \int_{\mathbb{R}} x f(x) dx + (\mathbb{E}[X])^2 \cdot \int_{\mathbb{R}} f(x) dx \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X] \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2 \cdot 1 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \end{aligned}$$

□

Bemerkung 4.21. Aus dieser Aussage folgen automatisch die Aussagen von Korollar 3.53 und Lemma 3.54 für absolutstetig verteilte Zufallsvariablen – es gilt also insbesondere

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

für $a, b \in \mathbb{R}$.



Beispiel 4.22. Sei $f(x) = \mathbb{1}_{[0,1]}(x)$ die Dichte einer Zufallsvariablen X . Dann ist deren Varianz gegeben durch

$$\begin{aligned}\operatorname{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \int_{\mathbb{R}} x^2 f(x) dx - \left(\int_{\mathbb{R}} x f(x) dx \right)^2 \\ &= \int_0^1 x^2 dx - \left(\int_0^1 x dx \right)^2 \\ &= \frac{1}{3} - \left(\frac{1}{2} \right)^2 = \frac{1}{12}.\end{aligned}$$



Beispiel 4.23. Sei $f(x) = 2x \mathbb{1}_{[0,1]}(x)$ die Dichte einer Zufallsvariablen X . Wie groß sind ihr Erwartungswert und ihre Varianz?

Antwort: Überprüfen wir zunächst, da wir diese Dichte noch nicht gesehen haben, dass es sich tatsächlich um eine Dichte handelt:

- $f(x) \geq 0$ für alle $x \in \mathbb{R}$
- Normierung:

$$\int_{\mathbb{R}} f(x) dx =$$

Als nächstes berechnen wir die Erwartungswerte von X und X^2 :

$$\mathbb{E}[X] =$$

$$\mathbb{E}[X^2] =$$

Damit ist die Varianz

$$\operatorname{Var}(X) =$$

4.3 Die wichtigsten absolutstetigen Verteilungen

4.3.1 Gleichverteilung

In den Beispielen 4.8 und 4.13 haben wir bereits für einen Spezialfall die stetige Gleichverteilung kennen gelernt. In diesem Abschnitt sehen wir uns die Verallgemeinerung des Beispiels an.

Definition 4.24

Eine Zufallsvariable X heißt *gleichverteilt* auf $[a, b] \subset \mathbb{R}$ mit $a < b$, falls X absolutstetig verteilt ist mit Dichte

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{sonst.} \end{cases}$$

Kurzschreibweise: $X \sim U([a, b])$.

Bemerkung 4.25. Die Kurzschreibweise ist in Bezug auf die Klammern nicht notwendigerweise eindeutig. Da bei der Integration die Grenzen des Intervalls keinen Einfluss auf den Wert des Integrals haben, könnten wir die Gleichverteilung analog auch für das offene Intervall (a, b) definieren oder für ein halboffenes Intervall der Form $(a, b]$ oder $[a, b)$. In [FLP15] wird die Gleichverteilung auf abgeschlossenen, in [Hen08] auf offenen Intervallen definiert. Demnach werden die inneren Klammern häufig weggelassen, d.h. man schreibt kurz $X \sim U(a, b)$, da keine Verwechslungsgefahr mit der diskreten Gleichverteilung auf $\{a, b\}$ besteht, welche man mit geschweiften Klammern als $X \sim U(\{a, b\})$ notieren würde.

Satz 4.26: Eigenschaften der Gleichverteilung

Sei $X \sim U(a, b)$. Dann hat X die Verteilungsfunktion

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b. \end{cases}$$

Zudem ist der Erwartungswert $\mathbb{E}[X] = \frac{a+b}{2}$ und die Varianz $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Beweis. Zunächst einmal ist definitionsgemäß $F_X(x) = \mathbb{P}(X \leq x)$ und für $x < a$ ist diese Wahrscheinlichkeit Null. Andererseits ist 

$$F_X(b) =$$

und auf Grund der Monotonie der Verteilungsfunktion ist $F_X(x) = 1$ für alle $x > b$. Ist $x \in [a, b]$, so ist

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

wie behauptet. Berechnen wir nun die Erwartungswerte von X und X^2 :

$$\mathbb{E}[X] =$$

$$\mathbb{E}[X^2] =$$

Folglich ist die Varianz gegeben durch

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

□

Beispiel 4.27. An einer Bushaltestelle kommt alle 15 Minuten ein Bus – der erste um 7:00 Uhr. Ein Fahrgast kommt mit der Bahn zu einer Zeit an die Haltestelle, welche gleichverteilt ist zwischen 7:00 Uhr und 7:30 Uhr. Wie groß ist die Wahrscheinlichkeit, dass der Fahrgast 

- weniger als 5 Minuten auf den Bus warten muss?
- mehr als 10 Minuten auf den Bus warten muss?

Antwort: Wir suchen die Wahrscheinlichkeiten der Ereignisse

A : „Fahrgast wartet weniger als 5 Minuten.“ und

B : „Fahrgast wartet mehr als 10 Minuten.“

Sei X die Zufallsvariable, welche die Ankunftszeit nach 7:00 Uhr (in Minuten, nicht gerundet) beschreibt. Laut Text ist

$$X \sim \text{Exp}(\lambda).$$

Damit gilt

$$\mathbb{P}(A) = \mathbb{P}(X \in (10, 15] \cup (25, 30]) =$$

$$\mathbb{P}(B) =$$

4.3.2 Exponentialverteilung

Diese Verteilungsart haben wir bereits in Beispiel 4.9 kennen gelernt als die Verteilung, die beschreibt, wie lange ein Drucker ohne Wartung am Stück funktioniert.

Definition 4.28

Sei $\lambda > 0$. Eine absolutstetig verteilte Zufallsvariable X mit Dichte

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0, \infty)}(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

heißt *exponentialverteilt* zum Parameter λ .

Kurzschreibweise: $X \sim \text{Exp}(\lambda)$.

Satz 4.29: Eigenschaften der Exponentialverteilung

Sei $X \sim \text{Exp}(\lambda)$. Dann hat X die Verteilungsfunktion

$$F_X(x) = (1 - e^{-\lambda x}) \mathbb{1}_{[0, \infty)}(x),$$

den Erwartungswert $\mathbb{E}[X] = \frac{1}{\lambda}$ und die Varianz $\text{Var}(X) = \frac{1}{\lambda^2}$.



Beweis. Die Verteilungsfunktion berechnen wir wieder als $F_X(x) = 0$, falls $x < 0$ ist, da in diesem Bereich die Dichte konstant Null ist. Für $x \geq 0$ gilt hingegen

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

Insbesondere folgt hiermit

$$\int_{\mathbb{R}} f(x) dx = \lim_{x \rightarrow \infty} F_X(x) = \lim_{x \rightarrow \infty} (1 - e^{-\lambda x}) = 1.$$

Berechnen wir nun die Erwartungswerte von X und von X^2 (jeweils mit Hilfe der Produktregel) und damit im Anschluss die Varianz.

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx =$$

=

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x) dx =$$

=

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

□

Das Warten auf den ersten Erfolg in einer Bernoulli-Kette haben wir als geometrisch verteilte Zufallsvariable modelliert; das Warten auf ein Ereignis mit stetigen Ergebnissen modellieren wir häufig als exponentialverteilte Zufallsvariable. Analog zur geometrischen ist auch die Exponentialverteilung gedächtnislos in folgendem Sinne (Vgl. Satz 3.26):

Satz 4.30: Gedächtnislosigkeit der Exponentialverteilung

Sei $X \sim \text{Exp}(\lambda)$ für $\lambda > 0$. Dann gilt für alle $s, t > 0$:

$$\mathbb{P}(X > s + t | X > t) = \mathbb{P}(X > s).$$

Beweis. Zunächst einmal ist für $s > 0$

$$\mathbb{P}(X > s) = 1 - \mathbb{P}(X \leq s) = 1 - F_X(s) = 1 - (1 - e^{-\lambda s}) \mathbb{1}_{[0, \infty)}(s) = e^{-\lambda s}.$$

Nach Definition der bedingten Wahrscheinlichkeit ist zudem mit der gleichen Rechnung für $s, t > 0$

$$\mathbb{P}(X > s + t | X > t) = \frac{\mathbb{P}(X > s + t, X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s}.$$

Damit wurde die Behauptung bereits gezeigt. □

Beispiel 4.31. Peter teilt sich ein Büro mit Uwe. Im Schnitt dauern Uwes Gespräche 15 Minuten, daher geht Peter davon aus, dass die Gesprächslänge (in Minuten) exponentialverteilt ist zum Parameter $\lambda = \frac{1}{15}$. 

- Uwe beginnt gerade ein Telefonat. Wie wahrscheinlich ist es, dass das Telefonat weniger als 10 bzw. zwischen 10 und 20 Minuten dauern wird?
- Uwe telefoniert bereits seit 10 Minuten. Wie wahrscheinlich ist es, dass das Telefonat innerhalb der nächsten 10 Minuten beendet sein wird?

Antwort:

a) Die gesuchten Wahrscheinlichkeiten sind

$$\begin{aligned}\mathbb{P}(X < 10) &= \\ \mathbb{P}(10 < X < 20) &= \end{aligned}$$

b) Wegen der Gedächtnislosigkeit gilt sofort

$$\mathbb{P}(X \leq 20 | X > 10) =$$

Wir haben bereits erwähnt, dass sowohl die geometrische als auch die Exponential-Verteilung für die Modellierung von Wartezeiten (diskret bzw. stetig) verwendet werden und wir haben gesehen, dass beide Verteilungen gedächtnislos sind. Einen direkten Zusammenhang gibt es, wenn wir die eigentlich stetige Exponentialverteilung aufrunden und somit diskretisieren.

Satz 4.32

Für $\lambda > 0$ sei $X \sim \text{Exp}(\lambda)$. Sei zudem

$$Y := \lceil X \rceil := \min \{ n \in \mathbb{N} \mid n \geq X \}.$$

Dann ist $Y \sim \text{Geo}(p)$ mit $p = (1 - e^{-\lambda})$.

Beweis. Übungsaufgabe. □

4.3.3 Normalverteilung

Wir führen die Normalverteilung zunächst über ihre Dichte ein und sehen uns einige Eigenschaften an. Erst im nächsten Abschnitt werden wir sehen, welchen Zusammenhang diese Verteilung zu der uns bereits bekannten Binomialverteilung besitzt.

Definition 4.33

Seien $\mu \in \mathbb{R}$ und $\sigma > 0$. Eine absolutstetige Zufallsvariable X , deren Dichte gegeben ist durch

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right), \quad x \in \mathbb{R},$$

heißt *normalverteilt* mit Parametern $\mathbb{E}[X] = \mu$ und $\text{Var}(X) = \sigma^2$.

Kurzschreibweise: $X \sim \mathcal{N}(\mu, \sigma^2)$.

In der Definition wurden Erwartungswert und Varianz bereits genannt, da dies gerade die Parameter der Verteilung sind. Wir müssen jedoch noch überprüfen, dass beide Kenngrößen tatsächlich diese Werte haben und dass f tatsächlich eine Dichte ist. Vorher führen wir jedoch noch einen Spezialfall ein.

Notation

Im Spezialfall $X \sim \mathcal{N}(0, 1)$ nennt man die Verteilung von X die *Standardnormalverteilung*. In diesem Fall benutzt man für die Dichte anders als bei allen anderen Verteilungen die Bezeichnung

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

und die zugehörige Verteilungsfunktion bezeichnet man mit

$$\Phi(x) := \int_{-\infty}^x \varphi(y) dy, \quad x \in \mathbb{R}.$$

Für beliebige $a, b \in \mathbb{R}$ mit $a < b$ gilt dann

$$\mathbb{P}(X \in [a, b]) = \int_a^b \varphi(x) dx = \Phi(b) - \Phi(a).$$

Um nachzuweisen, dass f eine Dichte ist, benötigen wir folgendes Resultat:

Lemma 4.34

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

Beweis. Wir berechnen das Quadrat des gewünschten Integrals – die Aussage folgt im Anschluss sofort.

$$\begin{aligned} \left(\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx\right)^2 &= \left(\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx\right) \cdot \left(\int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy\right) \\ &\stackrel{(*)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy \\ &= \int_0^{\infty} \int_0^{2\pi} \exp\left(-\frac{(r \cos \varphi)^2 + (r \sin \varphi)^2}{2}\right) r d\varphi dr \\ &= \int_0^{\infty} \int_0^{2\pi} \exp\left(-\frac{r^2}{2}\right) r d\varphi dr \\ &= 2\pi \left[-\exp\left(-\frac{r^2}{2}\right)\right]_0^{\infty} \\ &= 2\pi. \end{aligned}$$

In $(*)$ wird zu Polarkoordinaten übergegangen, d.h. $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix}$ mit $r \geq 0$ und $\varphi \in [0, 2\pi)$.

Die Determinante der Jacobi-Matrix der Abbildung $(r, \varphi) \mapsto \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix}$ ist

$$\begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix} = r(\cos \varphi)^2 + r(\sin \varphi)^2 = r,$$

daher taucht gemäß dem Transformationssatz (Satz A.29) – dem mehrdimensionalen Analogon zur Substitutionsregel – der Faktor r im Doppelintegral auf. \square

Korollar 4.35

Die Abbildung $f: \mathbb{R} \rightarrow \mathbb{R}$, welche gegeben ist durch

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad x \in \mathbb{R},$$

ist eine Dichte.

Beweis. Es gilt offensichtlich $f(x) \geq 0$ für alle $x \in \mathbb{R}$. Zudem gilt mit der Substitution $y = \frac{x-\mu}{\sigma}$, also $dx = \sigma dy$, und vorherigem Lemma

$$\int_{\mathbb{R}} f(x) dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2}\right) \sigma dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{y^2}{2}\right) dy = 1.$$

□

Sehen wir uns nun Erwartungswert und Varianz der Normalverteilung an:

Satz 4.36

Für Parameter $\mu \in \mathbb{R}$ und $\sigma > 0$ sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann hat X den Erwartungswert $\mathbb{E}[X] = \mu$ und die Varianz $\text{Var}(X) = \sigma^2$.

Beweis. Mit der schon bekannten Substitution $y = \frac{x-\mu}{\sigma}$ berechnen wir:

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\ &= \int_{\mathbb{R}} (x-\mu) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx + \int_{\mathbb{R}} \mu \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\ &= \int_{\mathbb{R}} \frac{\sigma y}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) dy + \mu \\ &= \frac{\sigma}{\sqrt{2\pi}} \cdot \left[-\exp\left(-\frac{1}{2}y^2\right)\right]_{-\infty}^{\infty} + \mu \\ &= \mu \\ \text{Var}(X) &= \int_{\mathbb{R}} \frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} y^2 \exp\left(-\frac{y^2}{2}\right) dy \\ &\stackrel{(*)}{=} \frac{\sigma^2}{\sqrt{2\pi}} \cdot \left\{ \left[-y \exp\left(-\frac{y^2}{2}\right)\right]_{-\infty}^{\infty} + \int_{\mathbb{R}} \exp\left(-\frac{y^2}{2}\right) dy \right\} \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \cdot (0 + \sqrt{2\pi}) \\ &= \sigma^2, \end{aligned}$$

wobei in (*) partielle Integration mit $g(y) = y$ und $f'(y) = ye^{-y^2/2}$ durchgeführt wurde. □

Kommen wir nun zur Verteilungsfunktion der Normalverteilung, d.h. $x \mapsto F_X(x)$ für $X \sim \mathcal{N}(\mu, \sigma^2)$. Ist f_X die zugehörige Dichte, so ist formal

$$F_X(x) = \int_{-\infty}^x f_X(y) dy.$$

Dieses Integral hat *keine geschlossene Form*, d.h. obwohl wir $\lim_{x \rightarrow \infty} F_X(x) = 1$ gezeigt haben, können wir keine Stammfunktion von f_X angeben und somit auch nicht F_X . Für die Standardnormalverteilung gibt es jedoch Tabellen, in denen die Werte der Verteilungsfunktion Φ abgelesen werden können.

Ist $Z \sim \mathcal{N}(0, 1)$, so kann man die Werte von $\Phi(z) = \mathbb{P}(Z \leq z)$ in folgender Tabelle ablesen. Die Zeile entspricht der Zahl einschließlich der ersten Nachkommastelle und die Spalte entspricht ausschließlich der zweiten Nachkommastelle.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997

Die Dichte φ existiert auch für negative Werte, diese werden jedoch grundsätzlich in keiner Tabelle auftauchen. Das ist jedoch kein Problem dank des folgenden Ergebnisses.

Lemma 4.37

Sei f die Dichte der $\mathcal{N}(\mu, \sigma^2)$ -Verteilung, d.h.

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

f ist symmetrisch zur Achse $x = \mu$ und ihr Maximum ist im Punkt $\left(\mu, \frac{1}{\sqrt{2\pi\sigma^2}}\right)$.

Beweis. Übungsaufgabe. □



Bemerkung 4.38. Für negative Argumente $-z < 0$ gilt, da die Dichte φ symmetrisch ist bezüglich der y -Achse,

$$\Phi(-z) = \mathbb{P}(Z \leq -z) = \mathbb{P}(Z \geq z) = 1 - \mathbb{P}(Z < z) = 1 - \Phi(z).$$



Folgende Werte kann man mit Hilfe der Tabelle ablesen:

$$\Phi(1, 11) \approx 0, 86650$$

$$\Phi(1, 96) \approx$$

$$\Phi(-0, 5) = 1 - \Phi(0, 5) \approx 1 - 0, 69146$$

$$\Phi(-2, 06) \approx$$

$$\Phi^{-1}(0, 881) \approx 1, 18$$

$$\Phi^{-1}(0, 7054) \approx$$

Wie berechnet man nun die Verteilungsfunktion einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariablen, wenn $\mu \neq 0$ und/oder $\sigma \neq 1$ ist?

Satz 4.39

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ und seien $a > 0$ und $b \in \mathbb{R}$. Dann ist $aX + b$ ebenfalls normalverteilt, und zwar ist $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Beweis. Wenden wir den Dichtetransformationssatz 4.10 auf $g(x) = ax + b$ an, so erhalten wir für $Y := aX + b$ die Dichte

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)| \\ &= f_X\left(\frac{y-b}{a}\right) \left|\frac{1}{a}\right| \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2\right) \cdot \frac{1}{a} \\ &= \frac{1}{\sqrt{2\pi\sigma^2 a^2}} \exp\left(-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2\right), \quad y \in \mathbb{R}. \end{aligned}$$

Das ist gerade die Dichte der $\mathcal{N}(a\mu + b, a^2\sigma^2)$ -Verteilung. □

Aus diesem Satz folgt unmittelbar folgendes Resultat, welches für das Rechnen mit Normalverteilungen extrem nützlich ist.

Korollar 4.40

Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $Z := \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$, d.h. Z ist standardnormalverteilt. Insbesondere gilt für jedes $x \in \mathbb{R}$

$$F_X(x) = \mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Beispiel 4.41. Sei X eine normalverteilte Zufallsvariable mit Erwartungswert 3 und Varianz 9. Wie groß sind die Wahrscheinlichkeiten

$$\mathbb{P}(X > 0), \quad \mathbb{P}(2 < X < 5) \quad \text{und} \quad \mathbb{P}(|X - 3| > 6)?$$

Antwort:



$$\mathbb{P}(X > 0) =$$

$$\approx$$

$$\mathbb{P}(2 < X < 5) =$$

$$=$$

$$\approx$$

$$\mathbb{P}(|X - 3| > 6) =$$

$$=$$

$$\approx$$

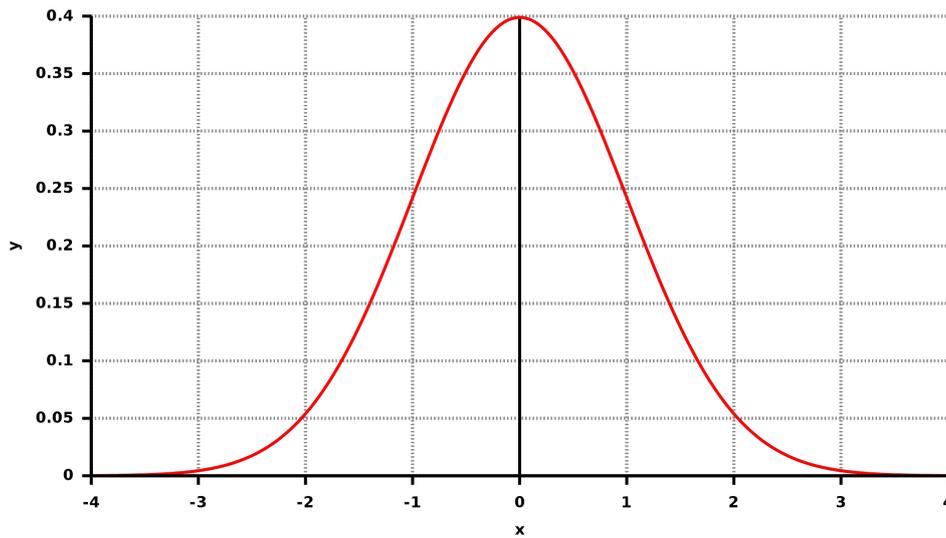
Die Normalverteilung wird in vielen verschiedenen Zusammenhängen verwendet. Beispielsweise werden folgende Größen häufig als normalverteilt angenommen:

- zufällige (d.h. nicht systematische) Fehler bei Messwerten
- zufällige Abweichungen vom Sollmaß bei der Herstellung von Objekten oder allgemein bei der Produktion von Gütern

Bemerkung 4.42 (Historischer Hintergrund). Die Normalverteilung wird auch als Gauß-Verteilung bezeichnet und auf Grund der glockenartigen Form wird der Graph der Dichte der Standardnormalverteilung auch als Glockenkurve bezeichnet.

Im nächsten Abschnitt werden wir uns ansehen, wie die Binomialverteilung durch die Normalverteilung approximiert werden kann – dieses Resultat geht auf Abraham de Moivre (1667-1754) zurück. Die Dichte der Normalverteilung wird zu Ehren von Carl Friedrich Gauß (1777-1855) auch als Gauß'sche Glockenkurve bezeichnet und den 10-DM-Schein zierte einst nicht nur ein Bild des Mathematikers Gauß, sondern auch die Glockenkurve. Gauß selbst hat Beiträge zur numerischen Integration geliefert, welche die näherungsweise Berechnung der Fläche unter der Glockenkurve ermöglicht und somit die Berechnung der Verteilungsfunktion der Normalverteilung. Auf der Rückseite des besagten 10-DM-Scheins ist übrigens ein Gerät zur Landvermessung, das Vize-Heliotrop, abgebildet, da Gauß in diesem Bereich einen sehr praktischen Beitrag geleistet hat.

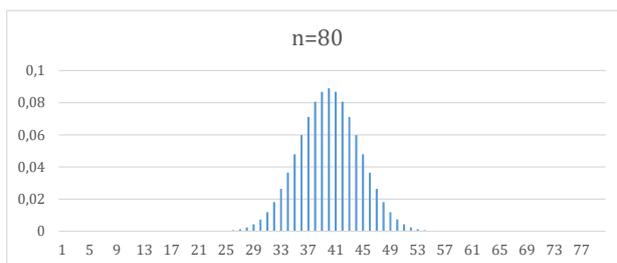
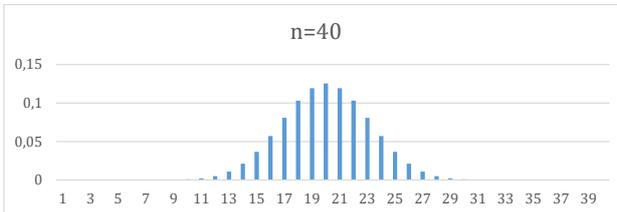
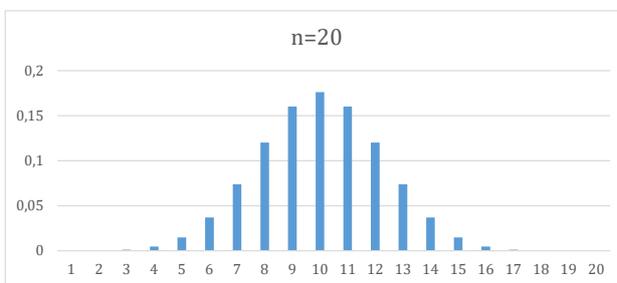
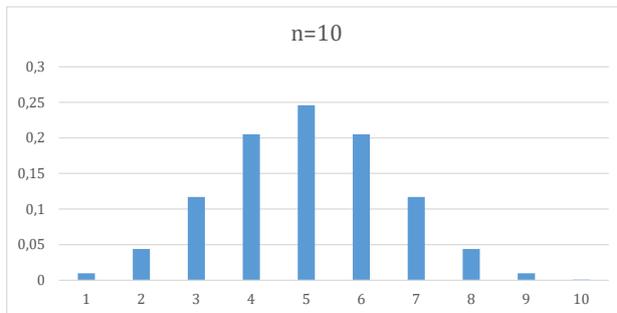
Anschauungsmaterial:

Abbildung 4.1: Quelle: <https://commons.wikimedia.org/w/index.php?curid=31516368>Abbildung 4.2: Quelle: <https://commons.wikimedia.org/w/index.php?curid=3813487>

4.4 Der Satz von de Moivre-Laplace

Erinnern wir uns an die Massenfunktion einer $\text{Bin}(n, p)$ -verteilten Zufallsvariable X :

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, \dots, n\}.$$



Das Aussehen dieser Massenfunktion lässt sich für den Spezialfall $p = \frac{1}{2}$ mit Hilfe eines *Galton-Bretts* nachvollziehen.

Die Binomialverteilung $\text{Bin}(n, p)$ entsteht, wenn ein Experiment mit Erfolgswahrscheinlichkeit p unabhängig voneinander n -mal durchgeführt wird und die Anzahl der Erfolge gezählt wird.

Im Falle des Galton-Bretts besteht ein Experiment darin, ob die Kugel an einer Stelle den Weg nach links oder rechts wählt und die Position der Kugel am Ende (d.h. die Spalte) entspricht der Anzahl der Erfolge, d.h. der Rechts-Bewegungen der Kugel.

Für die Werte $n \in \{10, 20, 40, 80\}$ sind die Massenfunktionen der $\text{Bin}(n, \frac{1}{2})$ -Verteilung (der Sichtbarkeit halber als Säulendiagramme) abgebildet.

Der Satz von Moivre-Laplace beschreibt die Beobachtung, dass die Massenfunktion, wenn wir sie geeignet skalieren und verstetigen, das Aussehen der Normalverteilungsdichte annimmt.

Zunächst sehen wir uns eine Aussage an, was der Satz von de Moivre-Laplace praktisch bedeutet bevor wir die Mathematik im Hintergrund genauer untersuchen.

Approximationsaussage von de Moivre-Laplace

Sei $X \sim \text{Bin}(n, p)$ für $np(1-p) \geq 9$. Dann gilt für $k, k_1, k_2 \in \mathbb{R}$ mit $k_1 < k_2$

$$\mathbb{P}(X \leq k) \approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right),$$

$$\mathbb{P}(k_1 \leq X \leq k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{np(1-p)}}\right).$$

Auch wenn dieser Aussage eine mathematisch saubere Konvergenzaussage zu Grunde liegt, müssen wir diese Approximation mit Vorsicht verwenden. Insbesondere ist zwar $\lim_{x \rightarrow -\infty} \Phi(x) = 0$, aber für jeden endlichen Wert $x \in \mathbb{R}$ ist $\Phi(x) > 0$. Insbesondere ist damit auch

$$\Phi\left(\frac{-1 - np}{\sqrt{np(1-p)}}\right) > 0 \quad \text{und} \quad \Phi\left(\frac{n - np}{\sqrt{np(1-p)}}\right) < 1,$$

obwohl $\mathbb{P}(X \leq -1) = 0$ und $\mathbb{P}(X \leq n) = 1$ ist für $X \sim \text{Bin}(n, p)$.



Beispiel 4.43. Eine faire Münze wird 50 Mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 23 und höchstens 27 Mal Zahl geworfen wird?

Antwort: Der wiederholte Münzwurf ist ein Bernoulli-Experiment. Interpretieren wir einen Münzwurf, der Zahl ergibt, als Erfolg, so ist die Anzahl der Erfolge binomialverteilt. Sei daher

X : Anzahl der Zahl-Würfe unter 50 Münzwürfen.

Es gilt $X \sim \text{Bin}(50, \frac{1}{2})$. Gesucht ist $\mathbb{P}(23 \leq X \leq 27)$. Die exakte Wahrscheinlichkeit ist

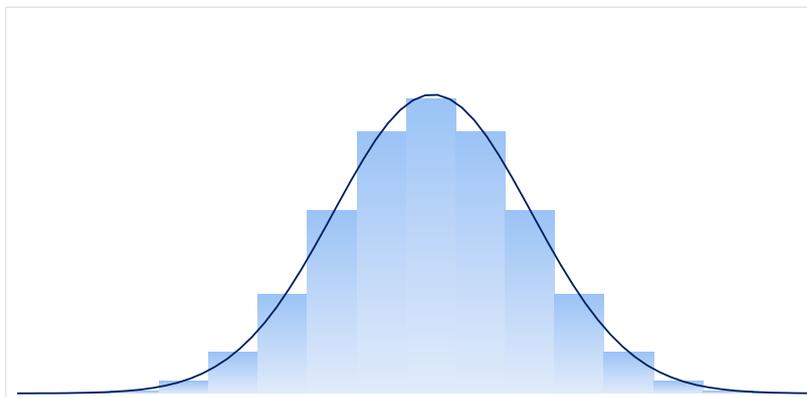
$$\begin{aligned} \mathbb{P}(23 \leq X \leq 27) &= \sum_{k=23}^{27} \binom{50}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{50-k} = \left(\frac{1}{2}\right)^{50} \cdot \left[2 \cdot \binom{50}{23} + 2 \cdot \binom{50}{24} + \binom{50}{25}\right] \\ &\approx 0,5201. \end{aligned}$$

In diesem Fall ist $np(1-p) = \frac{50}{4} = 12,5 > 9$, daher sehen wir uns das Ergebnis der Approximation durch die Normalverteilung an:

$$\begin{aligned} \mathbb{P}(23 \leq X \leq 27) &\approx \\ &\approx \Phi(0,566) - \Phi(-0,566) \\ &= \\ &\approx \end{aligned}$$

Obwohl in dem Beispiel $np(1-p) > 9$ ist, ist die Approximation nicht gut. Sehen wir uns an, wie wir sie noch verbessern können:

Sei $X \sim \text{Bin}(n, p)$ Für $k, l \in \mathbb{N}_0$ mit $k < l \leq n$ gilt $\mathbb{P}(X = k) > 0$, aber $\mathbb{P}\left(Z = \frac{x - np}{\sqrt{np(1-p)}}\right) = 0$ für jedes $x \in \mathbb{R}$ und $Z \sim \mathcal{N}(0, 1)$, d.h. die Verteilungsfunktion der Normalverteilung hat – im Gegensatz zu der Verteilungsfunktion der Binomialverteilung – keine Sprünge. Folgende Skizze illustriert, warum die vorgestellte Approximation durch die Normalverteilung die gesuchte Wahrscheinlichkeit tendenziell unterschätzt:



Um diesen systematischen Fehler zu beheben, wenden wir folgende *Stetigkeitskorrektur* an:

$$\mathbb{P}(k_1 \leq X \leq k_2) = \mathbb{P}\left(k_1 - \frac{1}{2} \leq X \leq k_2 + \frac{1}{2}\right) \approx \Phi\left(\frac{k_2 + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Fassen wir diese Methode zusammen:

Stetigkeitskorrektur

Sei $X \sim \text{Bin}(n, p)$ für $np(1-p) \geq 9$. Dann gilt für $k_1, k_2 \in \mathbb{R}$ mit $k_1 < k_2$

$$\mathbb{P}(k_1 \leq X \leq k_2) \approx \Phi\left(\frac{k_2 + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

Beispiel 4.44 (Fortsetzung von Beispiel 4.43). Wenden wir nun die Stetigkeitskorrektur auf Beispiel 4.43 an: 

$$\begin{aligned} \mathbb{P}(23 \leq X \leq 27) &\approx \\ &\approx \Phi(0,7071) - \Phi(-0,7071) = \end{aligned}$$

Diese Approximation ist in der Tat sehr viel besser als die durch die Normalverteilung ohne Stetigkeitskorrektur.

Sehen wir uns nun eine mathematisch sauber formulierte Aussage an, mit deren Hilfe die Approximationsformel für den Spezialfall $p = \frac{1}{2}$ hergeleitet werden kann. Sie wurde von Abraham de Moivre 1733 gezeigt und von Pierre-Simon Laplace 1812 verallgemeinert.

Satz 4.45: Satz von de Moivre-Laplace

Sei $c > 0$ und seien

$$x_n(k) := \frac{k - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \quad \text{und} \quad p_n(k) = \binom{n}{k} \left(\frac{1}{2}\right)^n$$

für $k \in \{0, \dots, n\}$ und $n \in \mathbb{N}$. Dann gilt

$$\lim_{n \rightarrow \infty} \max_{k: |x_n(k)| \leq c} \left| \frac{p_n(k)}{\varphi(x_n(k)) \sqrt{\frac{n}{4}}} - 1 \right| = 0.$$

Bemerkung 4.46. In dem Satz ist $p_n(\cdot)$ die Massenfunktion der $\text{Bin}(n, \frac{1}{2})$ -Verteilung. Die Aussage impliziert folgende lokale Approximationseigenschaft: Mit der Notation $\mu = np = \frac{n}{2}$ und $\sigma = \sqrt{np(1-p)} = \sqrt{\frac{n}{4}}$ gilt

$$p_n(k) \approx \frac{1}{\sigma} \varphi\left(\frac{k - \mu}{\sigma}\right).$$

Der Beweis benötigt folgende Formel aus der Analysis, deren Beweis beispielsweise unter <https://www.math.uni-augsburg.de/htdocs/emeriti/pukelsheim/2002f.pdf> zu finden ist. Sie ist benannt nach dem schottischen Mathematiker James Stirling (1692-1770), der diese Formel 1730 in seinem Werk *Methodus Differentialis: sive tractatus de summatione et interpolatione serierum infinitarum* veröffentlichte.

Lemma 4.47: Stirling'sche Formel

Für jede natürliche Zahl $n \in \mathbb{N}$ gilt

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad \text{d.h.} \quad \lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n} = 1.$$

An dieser Stelle sei an die Notation für das asymptotische Verhalten von Folgen (und Funktionen) erinnert:

Definition 4.48: Asymptotisches Verhalten von Folgen

Zwei Folgen $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ heißen *asymptotisch äquivalent*, in Zeichen $a_n \sim b_n$, wenn

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1 \quad \text{ist.}$$

Ist $(b_n)_{n \in \mathbb{N}}$ eine asymptotische obere Schranke für $(a_n)_{n \in \mathbb{N}}$, d.h. gilt

$$\limsup_{n \rightarrow \infty} \left| \frac{a_n}{b_n} \right| < \infty,$$

so schreibt man mit Hilfe des Landau-Symbols $a_n \in \mathcal{O}(b_n)$. Ist $(a_n)_{n \in \mathbb{N}}$ gegenüber $(b_n)_{n \in \mathbb{N}}$ asymptotisch vernachlässigbar, d.h. gilt

$$\lim_{n \rightarrow \infty} \left| \frac{a_n}{b_n} \right| = 0,$$

so schreibt man $a_n \in \mathcal{o}(b_n)$.

Beweis von Satz 4.45. Sei $X_n^* := \frac{X_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$. Dann gilt für $k \in \{0, \dots, n\}$

$$\mathbb{P}(X_n^* = x_n(k)) = \mathbb{P}(X_n = k) = p_n(k) \quad \text{und} \quad \frac{k}{n} = \frac{1}{2} \left(1 + \frac{x_n(k)}{\sqrt{n}} \right). \quad (4.4.1)$$

Mit der Stirling-Formel gilt dann

$$p_n(k) = \binom{n}{k} \left(\frac{1}{2}\right)^n = \frac{n!}{k!(n-k)!} \left(\frac{1}{2}\right)^n \sim \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sqrt{k(n-k)}} \cdot \frac{n^n}{k^k (n-k)^{n-k}} \left(\frac{1}{2}\right)^n$$

Fassen wir die Terme geeignet zusammen, so gilt also

$$p_n(k) \sim \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} \left(\frac{1}{2}\right)^n. \quad (4.4.2)$$

Unter Verwendung von (4.4.1) können wir den Term unter der Wurzel wie folgt umformen:

$$\frac{k(n-k)}{n} = \frac{n}{4} \cdot \frac{2k}{n} \cdot \frac{2(n-k)}{n} = \frac{n}{4} \cdot \left(1 + \frac{x_n(k)}{\sqrt{n}} \right) \cdot \left(1 - \frac{x_n(k)}{\sqrt{n}} \right) \sim \frac{n}{4},$$

Für alle $k \in \{0, \dots, n\}$, für die $|x_n(k)| \leq c$ ist für eine feste Schranke $c > 0$, gilt $\lim_{n \rightarrow \infty} \frac{x_n(k)}{\sqrt{n}} = 0$ und folglich ist

$$\frac{k(n-k)}{n} \sim \frac{n}{4}$$

gleichmäßig in k mit $|x_n(k)| \leq c$. Dadurch wird aus (4.4.2)

$$p_n(k) \cdot \sqrt{\frac{n}{4}} \sim \frac{1}{\sqrt{2\pi}} \binom{n}{k}^k \left(\frac{n}{n-k}\right)^{n-k} \left(\frac{1}{2}\right)^n. \quad (4.4.3)$$

Um das (asymptotische) Grenzverhalten von $\binom{n}{k}^k \left(\frac{n}{n-k}\right)^{n-k} \left(\frac{1}{2}\right)^n$ für $n \rightarrow \infty$ zu untersuchen, wenden wir den natürlichen Logarithmus auf diesen Term an.

$$\begin{aligned} & -\ln \left[\binom{n}{k}^k \left(\frac{n}{n-k}\right)^{n-k} \left(\frac{1}{2}\right)^n \right] \\ &= \ln \left[\left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} 2^n \right] \\ &= k \ln \left(\frac{2k}{n}\right) + (n-k) \ln \left(\frac{2(n-k)}{n}\right) \\ &= \frac{n}{2} \cdot \frac{2k}{n} \cdot \ln \left(\frac{2k}{n}\right) + \frac{n}{2} \cdot \frac{2(n-k)}{n} \cdot \ln \left(\frac{2(n-k)}{n}\right) \\ &= \frac{n}{2} \left(1 + \frac{x_n(k)}{\sqrt{n}}\right) \ln \left(1 + \frac{x_n(k)}{\sqrt{n}}\right) + \frac{n}{2} \left(1 - \frac{x_n(k)}{\sqrt{n}}\right) \ln \left(1 - \frac{x_n(k)}{\sqrt{n}}\right) \\ &\stackrel{(*)}{=} \frac{n}{2} \left[\left(1 + \frac{x_n(k)}{\sqrt{n}}\right) \left(\frac{x_n(k)}{\sqrt{n}} - \frac{x_n(k)^2}{2n} + \mathcal{O}(n^{-3/2})\right) + \left(1 - \frac{x_n(k)}{\sqrt{n}}\right) \left(-\frac{x_n(k)}{\sqrt{n}} - \frac{x_n(k)^2}{2n} + \mathcal{O}(n^{-3/2})\right) \right] \\ &= \frac{n}{2} \left[2\frac{x_n(k)^2}{n} - 2\frac{x_n(k)^2}{2n} + \mathcal{O}(n^{-3/2}) \right] \\ &= \frac{1}{2} x_n(k)^2 + \mathcal{O}(n^{-1/2}), \end{aligned}$$

wobei wir in (*) die Taylor-Entwicklung des Logarithmus verwendet haben:

$$\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k} = x - \frac{1}{2}x^2 + \mathcal{O}(x^3).$$

Aus (4.4.3) wird mit Hilfe dieser Beobachtung

$$p_n(k) \cdot \sqrt{\frac{n}{4}} \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_n(k)^2}{2}\right) \quad \text{für } k \text{ mit } |x_n(k)| \leq c.$$

Es folgt also, indem wir die Definition der asymptotischen Äquivalenz anwenden und die Dichte der Standardnormalverteilung als solche schreiben,

$$\lim_{n \rightarrow \infty} \frac{p_n(k)}{\varphi(x_n(k))} \cdot \sqrt{\frac{n}{4}} = 1, \quad \text{für alle } k \text{ mit } |x_n(k)| \leq c.$$

Da die Aussage für alle solche Argumente k gilt, können wir (vor dem Anwenden des Grenzwertes) das Maximum über alle zulässigen k bilden und 1 subtrahieren und erhalten die Behauptung des Satzes. \square

Aus dem Satz von de Moivre-Laplace folgt die Aussage, auf der die Approximationsaussage vom Beginn dieses Abschnittes basiert:

Korollar 4.49

Sei $X_n \sim \text{Bin}(n, p)$ für $n \in \mathbb{N}$ und $p \in (0, 1)$. Dann gilt für $a, b \in \mathbb{R}$, $a < b$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq b \right) = \int_a^b \varphi(x) dx.$$

Beweis. Wir zeigen die Aussage nur für den Spezialfall $p = \frac{1}{2}$. Sei $X_n^* := \frac{X_n - np}{\sqrt{np(1-p)}}$.

Mit den Bezeichnungen aus Satz 4.45 gilt

$$\lim_{n \rightarrow \infty} \max_{k: a \leq x_n(k) \leq b} \left| \frac{p_n(k)}{\varphi(x_n(k))} \sqrt{\frac{n}{4}} - 1 \right| = 0.$$

Zu jedem beliebigen $\varepsilon > 0$ existiert folglich $n_0 \in \mathbb{N}$ derart, dass für alle $n \geq n_0$

$$\max_{k: a \leq x_n(k) \leq b} \left| \frac{p_n(k)}{\varphi(x_n(k))} \sqrt{\frac{n}{4}} - 1 \right| < \varepsilon$$

erfüllt ist. Damit gilt die Abschätzung

$$\begin{aligned} & \left| \mathbb{P}(a \leq X_n^* \leq b) - \sum_{k: a \leq x_n(k) \leq b} \frac{\varphi(x_n(k))}{\sqrt{\frac{n}{4}}} \right| \\ &= \left| \sum_{k: a \leq x_n(k) \leq b} \mathbb{P}(X_n^* = x_n(k)) - \sum_{k: a \leq x_n(k) \leq b} \frac{\varphi(x_n(k))}{\sqrt{\frac{n}{4}}} \right| \\ &\leq \sum_{k: a \leq x_n(k) \leq b} \left| p_n(k) - \frac{\varphi(x_n(k))}{\sqrt{\frac{n}{4}}} \right| \\ &\leq \varepsilon \sum_{k: a \leq x_n(k) \leq b} p_n(k) \\ &\leq \varepsilon \end{aligned}$$

für alle $n \geq n_0$. Da die Riemann-Summe $\sum_{k: a \leq x_n(k) \leq b} \frac{\varphi(x_n(k))}{\sqrt{\frac{n}{4}}}$ das Integral $\int_a^b \varphi(x) dx$ approximiert, folgt hieraus die Behauptung. \square

Abschließend soll ein Beispiel zeigen, dass der Satz von de Moivre-Laplace durch doppelte Approximation sogar bei der hypergeometrischen Verteilung nützlich sein kann.

Beispiel 4.50. Eine Bürgerinitiative meint, dass 52% der Berliner ihr Anliegen unterstützt und will dies mit einer Umfrage untermauern. Wie viele Berliner müssen befragt werden, damit mit einer Wahrscheinlichkeit von mindestens 95% mindestens die Hälfte der Befragten die Bürgerinitiative befürwortet?

Antwort: Seien N die Anzahl der Berliner insgesamt und $K = 0,52N$ die Anzahl der Befürworter der Initiative. Ist n die Anzahl der Befragten und S_n die Anzahl der Befürworter unter den Befragten, so ist $S_n \sim \text{Hyp}(N, K, n)$, d.h.

$$\mathbb{P}(S_n = k) = \frac{\binom{0,52N}{k} \binom{0,48N}{n-k}}{\binom{N}{n}}, \quad k \in \{0, \dots, n\}.$$

Nach Satz 3.31 gilt, da N und $K = 0,52N$ groß sind,

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

für $p = \frac{K}{N} = 0,52$. Wir wollen $n \in \mathbb{N}$ derart bestimmen, dass $\mathbb{P}(S_n \geq \frac{n}{2}) \geq 0,95$ ist. Mit der Moivre-Laplace (ohne Stetigkeitskorrektur) gilt dann näherungsweise

$$\begin{aligned} \mathbb{P}\left(S_n \geq \frac{n}{2}\right) &= \mathbb{P}\left(\frac{S_n - 0,52 \cdot n}{\sqrt{0,52 \cdot 0,48 \cdot n}} \geq \frac{\frac{n}{2} - 0,52 \cdot n}{\sqrt{0,52 \cdot 0,48 \cdot n}}\right) \\ &\approx 1 - \Phi\left(\frac{-0,02n}{\sqrt{0,2496n}}\right) \\ &= \Phi\left(\frac{0,02n}{\sqrt{0,2496n}}\right) \\ &\approx \Phi(0,04\sqrt{n}) \\ &\stackrel{!}{\geq} 0,95. \end{aligned}$$

Umstellen nach n liefert

$$\Phi(0,04\sqrt{n}) \geq 0,95 \iff 0,04\sqrt{n} \geq \Phi^{-1}(0,95) \approx 1,645 \iff n \geq \left(\frac{1,645}{0,04}\right)^2 \approx 1691,3.$$

Da $n \in \mathbb{N}$ sein soll, müssen mindestens $n = 1692$ Berliner befragt werden.

Bemerkung 4.51. Mit Stetigkeitskorrektur hätten wir im Beispiel die Bedingung

$$\begin{aligned} 1 - \Phi\left(\frac{\frac{n}{2} - \frac{1}{2} - 0,52 \cdot n}{\sqrt{0,52 \cdot 0,48 \cdot n}}\right) \geq 0,95 &\iff \Phi\left(\frac{0,02n + 0,5}{\sqrt{0,2496n}}\right) \geq 0,95 \\ &\iff \frac{0,02}{\sqrt{0,2496}}\sqrt{n} + \frac{0,5}{\sqrt{0,2496n}} \geq \Phi^{-1}(0,95) \\ &\stackrel{approx.}{\iff} 0,04n + 1 \geq 1,645\sqrt{n}. \end{aligned}$$

Da beide Seiten positiv sind, können wir die Terme quadrieren und erhalten die Bedingung

$$\begin{aligned} (0,04n + 1)^2 &\geq 1,645^2 n \\ \iff 0,0016n^2 + 0,08n + 1 &\geq 2,706n \\ \iff 0,0016n^2 - 2,626n + 1 &\geq 0 \\ \iff \left(n \leq \frac{1,313}{0,0016} - \sqrt{\frac{1,313^2}{0,0016^2} - \frac{1}{0,0016}}\right) &\text{ oder } \left(n \geq \frac{1,313}{0,0016} + \sqrt{\frac{1,313^2}{0,0016^2} - \frac{1}{0,0016}}\right) \\ \iff (n \leq 0,38) \text{ oder } (n \geq 1640,869). \end{aligned}$$

Damit ergibt sich die etwas schwächere Bedingung $n \geq 1641$.

Kapitel 5

Gemeinsame Verteilungen von Zufallsvariablen

Ziele

- aus der gemeinsamen Massenfunktion $p_{X,Y}$ oder Dichte $f_{X,Y}$ zweier ZVen die Verteilungen von X und Y bestimmen, d.h. p_X und p_Y bzw. f_X und f_Y
- den Begriff der stochastischen Unabhängigkeit von ZVen kennen
- basierend auf der gemeinsamen Verteilung zweier ZVen entscheiden, ob diese stochastisch unabhängig sind
- Kovarianz und Korrelation zweier ZVen bestimmen
- die Massenfunktion, Dichte oder Verteilungsfunktion der Summe zweier ZVen bestimmen
- die Verteilung der Summen unabhängiger ZVen für bestimmte Spezialfälle kennen

5.1 Gemeinsame Verteilung

In diesem Kapitel geht es darum, dass wir die Verteilungen von zwei oder mehr Zufallsvariablen gleichzeitig betrachten, d.h. gerade im Fall zweier Zufallsvariablen X und Y müssen wir auch die Abhängigkeiten zwischen X und Y beachten.

Definition 5.1

Seien X und Y zwei auf $(\Omega, \mathcal{A}, \mathbb{P})$ diskret verteilte Zufallsvariablen. In diesem Fall heißt das Paar (X, Y) *gemeinsam diskret verteilt* und die Massenfunktion von (X, Y) ist gegeben durch

$$(x, y) \mapsto p(x, y) := \mathbb{P}(X = x, Y = y).$$

Bemerkung 5.2. Analog zur Massenfunktion einer einzelnen Zufallsvariablen gilt auch hier, dass die Summe über alle möglichen Werte 1 ergibt. Bei einem Paar von Zufallsvariablen (X, Y) heißt das:

$$\sum_{(x,y): p(x,y)>0} p(x, y) = 1.$$

Aus der gemeinsamen Massenfunktion von (X, Y) kann man die Massenfunktionen p_X von X und p_Y von Y bestimmen. Man nennt die durch p_X und p_Y charakterisierte Verteilung auch die *Randverteilung*. Es gilt:

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}\left(\bigcup_{y: p(x,y)>0} \{X = x, Y = y\}\right) = \sum_{y: p(x,y)>0} p(x, y).$$

und analog

$$p_Y(y) = \sum_{x: p(x,y)>0} p(x, y).$$



Beispiel 5.3. In einer Gemeinde haben 15% der Familien keine Kinder, 20% ein Kind, 35% zwei Kinder, und 30% drei Kinder. In jeder Familie ist jedes Kind mit gleicher Wahrscheinlichkeit ein Junge oder ein Mädchen. Eine Familie wird zufällig ausgewählt und seien J und M die zufälligen Zahlen von Jungen und Mädchen; K sei die Anzahl der Kinder. Berechnen Sie die gemeinsame Massenfunktion von (J, M) , sowie die Randverteilungen p_J und p_M .

Antwort: Sei $p(j, m) = \mathbb{P}(J = j, M = m)$.

Im Text gegeben sind folgende totale Wahrscheinlichkeiten:

$$\mathbb{P}(K = 0) = 0,15 \quad \mathbb{P}(K = 1) = 0,2 \quad \mathbb{P}(K = 2) = 0,35 \quad \mathbb{P}(K = 3) = 0,3.$$

Zudem kann man aus dem Text bedingte Wahrscheinlichkeiten herauslesen, beispielsweise

$$\mathbb{P}(J = 0, M = 1 | K = 1) = \frac{1}{2} \quad \text{und} \quad \mathbb{P}(J = 1, M = 1 | K = 2) = \frac{1}{2},$$

wobei Letzteres gilt, da unter der Voraussetzung, dass eine Familie zwei Kinder hat, die Wahrscheinlichkeit zweier Kinder gleichen Geschlechts jeweils die Wahrscheinlichkeit $\frac{1}{4}$ hat.

Berechnen wir alle Werte der Massenfunktion p :

$$p(0, 0) = \mathbb{P}(K = 0) = 0,15$$

$$p(0, 1) = \mathbb{P}(J = 0, M = 1) = \mathbb{P}(J = 0, M = 1 | K = 1)\mathbb{P}(K = 1) = \frac{1}{2} \cdot 0,2 = 0,1$$

$$p(1, 0) =$$

$$p(1, 1) =$$

$$p(0, 2) =$$

$$p(2, 0) =$$

$$p(1, 2) =$$

$$p(2, 1) =$$

$$p(0, 3) =$$

$$p(3, 0) =$$

Diese Werte können wir in eine Tabelle eintragen und damit die Randverteilung leicht berechnen:

$j \setminus m$	0	1	2	3	$p_J(j)$
0	0,15	0,1			
1					
2					
3					
$p_M(m)$					

Kommen wir nun zum Analogon für zwei absolutstetig verteilte Zufallsvariablen.

Definition 5.4

Seien X und Y zwei Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$. Das Paar (X, Y) heißt *gemeinsam absolutstetig verteilt*, falls eine integrierbare Funktion $f: \mathbb{R}^2 \rightarrow [0, \infty)$ existiert mit

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) dx dy = 1,$$

so dass für jede messbare Menge $A \subseteq \mathbb{R}^2$ gilt:

$$\mathbb{P}((X, Y) \in A) = \int_A f(x, y) d(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_A(x, y) f(x, y) dx dy.$$

Die Funktion f heißt *gemeinsame Dichte* von (X, Y) .

Aus der gemeinsamen Dichte kann man die Randverteilung berechnen. Wenn zwei Zufallsvariablen (X, Y) gemeinsam absolutstetig verteilt sind, so sind auch X und Y für sich genommen absolutstetig verteilt und ihre Dichten, genannt *Randdichten* oder *marginale Dichten*, sind gegeben durch

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy \quad \text{bzw.} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) dx. \quad (5.1.1)$$

Um dieses Resultat jedoch formal herleiten zu können, benötigen wir die Verteilungsfunktion, welche wieder für beliebige (d.h. diskret verteilte, absolutstetig verteilte und sonstige) Paare von Zufallsvariablen definiert wird.

Definition 5.5

Seien X und Y Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Die *gemeinsame Verteilungsfunktion* von (X, Y) ist dann gegeben durch

$$F(x, y) := \mathbb{P}(X \leq x, Y \leq y) := \mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}), \quad x, y \in \mathbb{R}.$$

Aus der gemeinsamen Verteilungsfunktion von (X, Y) kann man die Verteilungsfunktion von X wie folgt berechnen:

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y < \infty) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \{X \leq x, Y \leq n\}\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(X \leq x, Y \leq n) = \lim_{n \rightarrow \infty} F(x, n) \end{aligned}$$

für beliebige $x \in \mathbb{R}$. Analog gilt für beliebige $y \in \mathbb{R}$

$$F_Y(y) = \lim_{n \rightarrow \infty} F(n, y).$$

Die Randdichten erhält man nun durch Ableiten der Verteilungsfunktionen. Ist (X, Y) gemeinsam absolutstetig verteilt mit gemeinsamer Dichte f , so gilt

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds.$$

Somit ist insbesondere

$$f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y), \quad x, y \in \mathbb{R}.$$

Die Verteilungsfunktion von X ist nun gegeben durch

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

und die Dichte ist (wie in (5.1.1) vorhergesagt)

$$f_X(x) = F'_X(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$



Beispiel 5.6. Die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ sei gegeben durch

$$f(x, y) = 2e^{-x}e^{-2y}\mathbb{1}_{(0, \infty)}(x)\mathbb{1}_{(0, \infty)}(y).$$

- Überprüfen Sie, ob es sich bei f um eine Dichte zweier Zufallsvariablen (X, Y) handelt.
- Berechnen Sie $\mathbb{P}(X > 1, Y < 1)$, $\mathbb{P}(X < Y)$ und $\mathbb{P}(X < a)$ für ein festes $a \in \mathbb{R}$.

Antwort:

- Es ist offensichtlich $f(x, y) \geq 0$ für alle $(x, y) \in \mathbb{R}^2$. Zudem ist f integrierbar und normiert, wie folgende Rechnung zeigt:

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) dx dy &= \int_0^{\infty} \int_0^{\infty} 2e^{-x}e^{-2y} dx dy = \left(\int_0^{\infty} e^{-x} dx \right) \cdot \left(\int_0^{\infty} 2e^{-2y} dy \right) \\ &= \\ &= 1. \end{aligned}$$

- Da es sich bei f um eine Dichte handelt, können wir die erste gesuchte Wahrscheinlichkeiten analog berechnen.

$$\begin{aligned} \mathbb{P}(X > 1, Y < 1) &= \int_{-\infty}^1 \left(\int_1^{\infty} f(x, y) dx \right) dy = \left(\int_1^{\infty} e^{-x} dx \right) \cdot \left(\int_0^1 2e^{-2y} dy \right) \\ &= \\ &= \end{aligned}$$

Für das zweite Integral ist es sinnvoll, zunächst eine Vorüberlegung anzustellen:

$$\{(x, y) \in (0, \infty)^2 \mid x < y\} = \{(x, y) \in \mathbb{R}^2 \mid \quad \quad \quad \}.$$

Damit ist

$$\begin{aligned} \mathbb{P}(X < Y) &= \int_{\mathbb{R}^2} \mathbb{1}_{(-\infty, y)}(x) f(x, y) d(x, y) = \int_0^\infty \left(\int_0^y 2e^{-x} e^{-2y} dx \right) dy \\ &= \\ &= \\ &= \\ &= \\ &= \end{aligned}$$

Kommen wir nun zur letzten gesuchten Wahrscheinlichkeit:

$$\begin{aligned} \mathbb{P}(X < a) &= F_X(a) = \int_{-\infty}^\infty \int_{-\infty}^a f(x, y) dx dy \\ &= \\ &= \\ &= \end{aligned}$$

Beispiel 5.7. Ein Punkt in einem Kreis von Radius $R > 0$ wird zufällig gewählt, so dass die Wahrscheinlichkeit, ihn in einem Gebiet zu finden, proportional zu dessen Fläche sei. Der Mittelpunkt des Kreises sei $(0, 0)$ und (X, Y) seien die kartesischen Koordinaten des Punktes. Ist $A = [x_1, x_2] \times [y_1, y_2]$ ein Rechteck innerhalb des Kreises, dann gilt insbesondere 

$$\mathbb{P}((X, Y) \in A) = c \cdot |A| = c(x_2 - x_1)(y_2 - y_1) = c \int_{x_1}^{x_2} \int_{y_1}^{y_2} dy dx,$$

d.h. (X, Y) ist absolutstetig verteilt mit Dichte

$$f(x, y) = c \mathbb{1}_{\{x^2 + y^2 \leq R^2\}}(x, y) = \begin{cases} c, & x^2 + y^2 \leq R^2 \\ 0, & \text{sonst.} \end{cases}$$

Sei zudem D der Abstand eines Punktes (X, Y) vom Koordinatenursprung $(0, 0)$.

- Wie muss der Parameter c gewählt werden, damit f eine Dichte ist?
- Wie lauten die Randdichten f_X und f_Y ?
- Wie groß ist die Wahrscheinlichkeit $\mathbb{P}(D \leq a)$ für ein festes $a \geq 0$?
- Welchen Erwartungswert hat die Zufallsvariable D ?

Antwort:

- Die Nichtnegativität ist für jeden Parameter $c \geq 0$ erfüllt. Für die Normierung nutzen wir wieder den Transformationssatz (Satz A.31) um das Integral in Polarkoordinaten zu berechnen.

$$\int_{\mathbb{R}^2} f(x, y) d(x, y) = \int_0^R \int_0^{2\pi} c r d\varphi dr = 2\pi c \left[\frac{r^2}{2} \right]_0^R = c \cdot \pi R^2 \stackrel{!}{=} 1.$$

Folglich ist $c =$

- b) Ist $|x| > R$, so ist $x^2 + y^2 > R^2$ und folglich $f(x, y) = 0$. Damit ist auch $f_X(x) = 0$ für $|x| > R$. Für $|x| \leq R$ gilt innerhalb des Kreises $|y| \leq \sqrt{R^2 - x^2}$, also ist

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_{-\sqrt{R^2 - x^2}}^{\sqrt{R^2 - x^2}} \frac{1}{\pi R^2} dy =$$

Da die Dichte $f(x, y)$ bezüglich x und y symmetrisch ist, können wir f_Y ohne Rechnung angeben als

$$f_Y(y) =$$

- c) Es ist $D = \sqrt{X^2 + Y^2}$. Sei zudem $K(0, r) := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < r^2\}$ die offene Kreisscheibe um den Koordinatenursprung mit Radius $r > 0$. Dann ist für festes $a \leq R$

$$\mathbb{P}(D \leq a) = \mathbb{P}(X^2 + Y^2 \leq a^2) = \frac{|K(0, a)|}{|K(0, R)|} = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}$$

Für $a > R$ ist $\mathbb{P}(D \leq a) =$

- d) In c) haben wir die Verteilungsfunktion von D berechnet. Indem wir diese ableiten, erhalten wir die Dichte. Für $a \notin [0, R]$ ist $f_D(a) = 0$ und für $a \in [0, R]$ ist

$$f_D(a) = \frac{d}{da} \left(\frac{a^2}{R^2} \right) = \frac{2a}{R^2}.$$

Damit ist der Erwartungswert

$$\mathbb{E}[D] =$$

Übersicht über Formeln der Integration



Was wird mit folgenden Formeln ausgerechnet, wenn f_X die Dichte einer absolutstetig verteilten Zufallsvariablen X und f die gemeinsame Dichte von (X, Y) ist?

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= \\ \int_{-\infty}^a f_X(x) dx &= \\ \int_{-\infty}^{\infty} x f_X(x) dx &= \\ \int_{-\infty}^{\infty} f(x, y) dy &= \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy &= \\ \int_{-\infty}^{\infty} \int_{-\infty}^a f(x, y) dx dy &= \end{aligned}$$

5.2 Stochastische Unabhängigkeit von Zufallsvariablen

In den Definitionen 2.19 und 2.24 haben wir die stochastische Unabhängigkeit von zwei oder mehr Ereignissen einer σ -Algebra $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ definiert.

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ also wie üblich ein Wahrscheinlichkeitsraum und seien $A, B \in \mathcal{A}$. Die Ereignisse sind genau dann stochastisch unabhängig, wenn $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ gilt. Sind \mathcal{A}_1 und \mathcal{A}_2 Teil- σ -Algebren von \mathcal{A} , d.h. $\mathcal{A}_1 \subseteq \mathcal{A}$ und $\mathcal{A}_2 \subseteq \mathcal{A}$, so nennen wir die Teil- σ -Algebren stochastisch unabhängig, wenn je zwei Elemente daraus stochastisch unabhängig sind, d.h. falls gilt:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2), \quad \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2. \quad (5.2.1)$$

Um die stochastische Unabhängigkeit zweier Zufallsvariablen X und Y auf $(\Omega, \mathcal{A}, \mathbb{P})$ zu definieren, ziehen wir uns auf die von ihnen erzeugten σ -Algebren zurück, d.h. X und Y (je mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$) heißen stochastisch unabhängig, wenn $\sigma(X)$ und $\sigma(Y)$ stochastisch unabhängig sind, wobei $\sigma(X)$ über die Urbilder definiert ist:

$$\sigma(X) := \sigma(\{X^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\}).$$

Dabei ist

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{A}.$$

Mit der Definition der Unabhängigkeit zweier σ -Algebren aus (5.2.1) heißt das Folgendes:

Definition 5.8

Seien X und Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Die Zufallsvariablen heißen genau dann *stochastisch unabhängig voneinander*, wenn

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$$

für alle $A, B \in \mathcal{B}(\mathbb{R})$ gilt.

Diese Definition ist leider sehr unpraktisch, da die Borel- σ -Algebra $\mathcal{B}(\mathbb{R})$ zu groß ist um alle Mengen, die darin enthalten sind, überprüfen zu können. Glücklicherweise sind zwei σ -Algebren bereits dann voneinander stochastisch unabhängig, wenn ihre Erzeuger voneinander unabhängig sind. In Formeln heißt das: Ist $\mathcal{A}_1 = \sigma(M_1)$ und $\mathcal{A}_2 = \sigma(M_2)$, so sind \mathcal{A}_1 und \mathcal{A}_2 genau dann stochastisch unabhängig, wenn

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2), \quad \forall A_1 \in M_1, A_2 \in M_2$$

erfüllt ist. Es ist $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, x] \mid x \in \mathbb{R}\})$, d.h. die Borel- σ -Algebra wird erzeugt durch die halboffenen Intervalle. Daher haben wir folgende Definition, mit der wir wesentlich besser arbeiten können:

Satz 5.9

Seien X und Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Die Zufallsvariablen sind genau dann *stochastisch unabhängig voneinander*, wenn

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y), \quad \forall x, y \in \mathbb{R}$$

gilt.

Mit anderen Worten, zwei Zufallsvariablen sind genau dann stochastisch unabhängig voneinander, wenn ihre gemeinsame Verteilungsfunktion das Produkt der beiden Randverteilungsfunktionen ist.

Für die zwei Spezialfälle dieser Vorlesung – diskrete und absolutstetig verteilte Zufallsvariablen – können wir noch weitere Kriterien herleiten.

Diskret verteilte Zufallsvariablen: Ein weiterer Erzeuger der Borel- σ -Algebra sind die einelementigen Mengen, d.h.

$$\mathcal{B}(\mathbb{R}) = \sigma(\{\{x\} \mid x \in \mathbb{R}\}).$$

Da $\mathbb{P}(X \in \{x\}) = \mathbb{P}(X = x)$ ist, folgt hieraus, dass zwei diskret verteilte Zufallsvariablen genau dann stochastisch unabhängig sind, wenn

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y), \quad \forall x, y \in \mathbb{R}$$

erfüllt ist, d.h. wenn die gemeinsame Massenfunktion das Produkt der marginalen Massenfunktionen ist.

Absolutstetig verteilte Zufallsvariablen: Leiten wir die Formel aus Satz 5.9 nach x und y partiell ab, so erhalten wir die Bedingung

$$\frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} (F_X(x) \cdot F_Y(y)) \iff f(x, y) = f_X(x) \cdot f_Y(y),$$

welche für alle $x, y \in \mathbb{R}$ gelten soll. Zwei gemeinsam absolutstetig verteilte Zufallsvariablen sind also genau dann stochastisch unabhängig voneinander, wenn ihre gemeinsame Dichte das Produkt der Randdichten ist.



Beispiel 5.10. Zwei Zufallsvariablen seien gemeinsam stetig verteilt mit der Dichte aus Beispiel 5.6, d.h.

$$f(x, y) = 2e^{-x}e^{-2y}\mathbb{1}_{(0,\infty)}(x)\mathbb{1}_{(0,\infty)}(y).$$

Sind X und Y stochastisch unabhängig voneinander?

Antwort: Um diese Frage zu beantworten, müssen wir beide Randdichten bestimmen.

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy =$$

=

=

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) dx =$$

=

=

Folglich ist in der Tat $f(x, y) = f_X(x)f_Y(y)$ für alle $(x, y) \in \mathbb{R}^2$, d.h. X und Y sind stochastisch unabhängig.



Beispiel 5.11. Ein fairer Würfel wird $n + m$ -mal hintereinander geworfen. Als Erfolg werten wir alle Würfe, bei denen eine Eins oben liegt. Wir betrachten folgende Zufallsvariablen:

X : Anzahl Einsen in den ersten n Würfeln,

Y : Anzahl Einsen in den letzten m Würfeln,

Z : Anzahl Einsen in den gesamten $n + m$ Würfeln.

Sind X und Y bzw. X und Z stochastisch unabhängig?

Antwort: Davon ausgehend, dass die Würfelergebnisse unabhängig voneinander sind, sind die Zufallsvariablen wie folgt verteilt:

$$X \sim \text{Bin}(n, p), \quad Y \sim \text{Bin}(m, p), \quad Z \sim \text{Bin}(n + m, p),$$

wobei $p = \frac{1}{6}$ ist. Einerseits gilt für beliebige $k \in \{0, \dots, n\}$ und $l \in \{0, \dots, m\}$

$$\mathbb{P}(X = k, Y = l) =$$

d.h. X und Y sind voneinander stochastisch unabhängig. Andererseits gilt

$$\mathbb{P}(X = n) \cdot \mathbb{P}(Z = n - 1) =$$

aber

$$\mathbb{P}(X = n, Z = n - 1) =$$

d.h. X und Z sind nicht stochastisch unabhängig.

Beispiel 5.12. Die Zahl der Kunden, die ein Postamt an einem Tag besuchen, sei Poisson-verteilt zum Parameter $\lambda > 0$. Jeder Kunde sei mit Wahrscheinlichkeit $p \in (0, 1)$ weiblich und mit Wahrscheinlichkeit $1 - p$ männlich. Sei X die Anzahl der weiblichen und Y die Anzahl männlicher Kunden. Wie sind X und Y gemeinsam verteilt? Sind sie stochastisch unabhängig voneinander? 

Antwort: Die gemeinsame Verteilung können wir mit Hilfe der bedingten Wahrscheinlichkeit herleiten. Für beliebige $k, l \in \mathbb{N}_0$ ist

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k, Y = l | X + Y = k + l) \cdot \mathbb{P}(X + Y = k + l)$$

=

=

Folglich gilt für die Massenfunktion von X

$$\mathbb{P}(X = k) = \sum_{l \in \mathbb{N}_0} \mathbb{P}(X = k, Y = l) =$$

Folglich ist $X \sim \text{Poisson}(\lambda p)$. Analog rechnet man nach, dass $Y \sim \text{Poisson}(\lambda(1 - p))$ gilt. Es ist insgesamt

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k)\mathbb{P}(Y = l), \quad \forall k, l \in \mathbb{N}_0,$$

d.h. X und Y sind stochastisch unabhängig voneinander.

5.3 Summen unabhängiger Zufallsvariablen

5.3.1 Diskrete Zufallsvariablen

Sehen wir uns zunächst ein Beispiel an, welches wir auch ohne die Theorie zu kennen lösen können.



Beispiel 5.13. Sie möchten beim Bäcker 2 Croissants kaufen. Vor Ihnen stehen noch zwei Kunden und im Regal sind noch 5 Croissants vorrätig. Wie wahrscheinlich ist es, dass Sie die gewünschten 2 Croissants kaufen können, wenn Kunden unabhängig voneinander Croissants mit folgenden Wahrscheinlichkeiten kaufen?

Anzahl	0	1	2	3	4
Wahrscheinlichkeit	0,3	0,2	0,3	0,1	0,1

Antwort: Seien X und Y die Anzahl gekaufter Croissants der vor Ihnen stehenden Kunden. Wir suchen die Wahrscheinlichkeit $\mathbb{P}(X + Y \leq 3)$. Dafür bestimmen wir die ersten Werte der Massenfunktion von $X + Y$:

$$\mathbb{P}(X + Y = 0) = \mathbb{P}(X = 0, Y = 0) \stackrel{\text{Unabh.}}{=} \mathbb{P}(X = 0)\mathbb{P}(Y = 0) = 0,3 \cdot 0,3 = 0,09$$

$$\mathbb{P}(X + Y = 1) =$$

=

$$\mathbb{P}(X + Y = 2) =$$

=

$$\mathbb{P}(X + Y = 3) =$$

=

Damit ist insgesamt

$$\mathbb{P}(X + Y \leq 3) = \sum_{k=0}^3 \mathbb{P}(X = k) = 0,61,$$

d.h. mit einer Wahrscheinlichkeit von 61% können Sie die gewünschten 2 Croissants kaufen.

Wir können die entscheidende Beobachtung wie folgt zusammenfassen: Sind X und Y stochastisch unabhängige diskret verteilte Zufallsvariablen, die nur Werte in \mathbb{N}_0 (oder einer Teilmenge davon) annehmen, so gilt

$$\mathbb{P}(X + Y = n) = \sum_{k \leq n} \mathbb{P}(X = k, Y = n - k) \stackrel{\text{Unabh.}}{=} \sum_{k \leq n} \mathbb{P}(X = k)\mathbb{P}(Y = n - k).$$

Damit können wir auch folgende Sätze beweisen.

Satz 5.14

Sind X_1 und X_2 stochastisch unabhängige Zufallsvariablen mit $X_1 \sim \text{Bin}(n_1, p)$ und $X_2 \sim \text{Bin}(n_2, p)$ für $n_1, n_2 \in \mathbb{N}$ und $p \in (0, 1)$, so gilt

$$X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p).$$

Beweis. Für $k \in \{0, \dots, n_1 + n_2\}$ gilt stets

$$\begin{aligned}
 \mathbb{P}(X_1 + X_2 = k) &= \sum_{l=0}^k \mathbb{P}(X_1 = l, X_2 = k - l) \\
 &= \sum_{l=0}^k \mathbb{P}(X_1 = l) \mathbb{P}(X_2 = k - l) \\
 &= \sum_{l=0 \vee (k-n_2)}^{k \wedge n_1} \binom{n_1}{l} p^l (1-p)^{n_1-l} \binom{n_2}{k-l} p^{k-l} (1-p)^{n_2-k+l} \\
 &= p^k (1-p)^{n_1+n_2-k} \sum_{l=0 \vee (k-n_2)}^{k \wedge n_1} \binom{n_1}{l} \binom{n_2}{k-l} \\
 &= p^k (1-p)^{n_1+n_2-k} \binom{n_1+n_2}{k},
 \end{aligned}$$

wobei die letzte Gleichung wie folgt begründet werden kann: Die Anzahl an Möglichkeiten, k Elemente aus $\{1, \dots, n_1 + n_2\}$ auszuwählen, ist gleich der Anzahl, zunächst l Elemente aus $\{1, \dots, n_1\}$ und anschließend $k - l$ Elemente aus $\{n_1 + 1, \dots, n_1 + n_2\}$ auszuwählen und anschließend diese Anzahl über alle möglichen Werte für l zu summieren. \square

Bemerkung 5.15. Die Verteilungen $\text{Ber}(p)$ hat die gleiche Massenfunktion wie $\text{Bin}(1, p)$. Ist $X \sim \text{Bin}(1, p)$, so gilt

$$\mathbb{P}(X = 1) = \binom{1}{1} p^1 (1-p)^{1-1} = p \quad \text{und} \quad \mathbb{P}(X = 0) = \binom{1}{0} p^0 (1-p)^{1-0} = 1-p$$

Verallgemeinern wir obigen Satz auf $n \in \mathbb{N}$ Bernoulli-verteilte Zufallsvariablen X_1, \dots, X_n (zum gemeinsamen Parameter $p \in (0, 1)$), welche voneinander stochastisch unabhängig seien, d.h. es gelte

$$\mathbb{P}\left(\bigcap_{i \in I} \{X_i = k_i\}\right) = \prod_{i \in I} \mathbb{P}(X_i = k_i)$$

für $k_i \in \{0, 1\}$ für alle $i \in I$ und für alle Indextmengen $I \subseteq \{1, \dots, n\}$. Dann ist

$$\sum_{i=1}^n X_i \sim \text{Bin}(n, p),$$

d.h. die Summe von n $\text{Ber}(p)$ -verteilter Zufallsvariablen ist in der Tat $\text{Bin}(n, p)$ -verteilt.

Satz 5.16

Sind X_1 und X_2 stochastisch unabhängige Zufallsvariablen mit $X_1 \sim \text{Poisson}(\lambda_1)$ und $X_2 \sim \text{Poisson}(\lambda_2)$ für $\lambda_1, \lambda_2 > 0$, so gilt

$$X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

Beweis. Übungsaufgabe. \square

5.3.2 Absolutstetige Zufallsvariablen

Seien X und Y stochastisch unabhängige Zufallsvariablen mit Dichten f_X und f_Y . Was kann man dann über die Verteilung von $X + Y$ sagen? Sehen wir uns dazu die Verteilungsfunktion von $X + Y$ an. Für alle $z \in \mathbb{R}$ gilt

$$\begin{aligned} F_{X+Y}(z) &= \mathbb{P}(X + Y \leq z) = \mathbb{P}((X, Y) \in \{(x, y) \mid x + y \leq z\}) \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y \leq z\}}(x, y) f(x, y) d(x, y). \end{aligned}$$

Es ist

$$\{(x, y) \in \mathbb{R}^2 \mid x + y \leq z\} = \{(x, y) \in \mathbb{R}^2 \mid x \leq z - y\}$$

und damit, da $f(x, y) = f_X(x) \cdot f_Y(y)$ ist,

$$\begin{aligned} F_{X+Y}(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^z f_X(\tilde{x} - y) f_Y(y) d\tilde{x} dy \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} f_X(\tilde{x} - y) f_Y(y) dy d\tilde{x}. \end{aligned}$$

Dabei haben wir eine Substitution verwendet und die Integrale vertauscht (gemäß dem Satz von Fubini). Durch Ableiten erhalten wir (gemäß dem Fundamentalsatz der Analysis) die Dichte:

$$f_{X+Y}(z) = F'_{X+Y}(z) = \int_{\mathbb{R}} f_X(z - y) f_Y(y) dy, \quad z \in \mathbb{R}.$$

Notation

Sind $f, g: \mathbb{R} \rightarrow \mathbb{R}$ Funktionen, für die die folgenden Integrale existieren, so ist die Faltung $f * g: \mathbb{R} \rightarrow \mathbb{R}$ definiert als

$$(f * g)(x) := \int_{\mathbb{R}} f(x - y) g(y) dy.$$

Halten wir die hergeleitete Formel einmal in Form eines Satzes fest.

Satz 5.17

Sind X und Y stochastisch unabhängige Zufallsvariablen mit Dichten f_X und f_Y , so ist $X + Y$ ebenfalls absolutstetig verteilt mit Dichte

$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(z - y) f_Y(y) dy, \quad z \in \mathbb{R}.$$

Beispiel 5.18. Seien X und Y stochastisch unabhängige auf $[0, 1]$ gleichverteilte Zufallsvariablen. Welche Dichte hat $X + Y$?

Antwort: Da X und Y jeweils nur Werte aus $[0, 1]$ mit positiver Wahrscheinlichkeit annehmen, kann $X + Y$ nur die Werte aus $[0, 2]$ mit positiver Wahrscheinlichkeit annehmen – für alle übrigen Werte sollte die Dichte Null sein.

Die Dichte von X können wir mit Hilfe der Indikatorfunktion kompakt schreiben als $f_X(x) = \mathbb{1}_{[0,1]}(x)$. Setzen wir diese Dichte in die Faltungsformel ein, so erhalten wir

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(z-y)f_Y(y)dy = \int_{\mathbb{R}} \mathbb{1}_{[0,1]}(z-y)\mathbb{1}_{[0,1]}(y)dy = \int_{\mathbb{R}} \mathbb{1}_{[0,1] \cap [z-1,z]}(y)dy.$$

Darauf kommt man mit folgender Überlegung:

$$0 \leq z-y \leq 1 \iff -z \leq -y \leq 1-z \iff z \geq y \geq z-1.$$

Der Schnitt der Intervalle der Indikatorfunktion ist

$$[0,1] \cap [z-1,z] = [0 \vee z-1, 1 \wedge z] = \begin{cases} [0, z], & z \in [0, 1] \\ [z-1, 1], & z \in (1, 2]. \end{cases}$$

Die Berechnung der Dichte von f_{X+Y} erfordert also eine Fallunterscheidung. Ist $z \in [0, 1]$, so ist

$$f_{X+Y}(z) = \int_0^z dy = z.$$

Ist hingegen $z \in (1, 2]$, so ist

$$f_{X+Y}(z) = \int_{z-1}^1 dy = 2-z.$$

Insgesamt ist die Dichte von $X+Y$ also gegeben durch

$$f_{X+Y}(z) = \begin{cases} z, & z \in [0, 1] \\ 2-z, & z \in (1, 2] \\ 0, & \text{sonst.} \end{cases}$$

Beispiel 5.19. Für ein $\lambda > 0$ seien $X \sim \text{Exp}(\lambda)$ und $Y \sim \text{Exp}(\lambda)$ zwei stochastisch unabhängige Zufallsvariablen. Welche Dichte hat $X+Y$?

Antwort: Da X und Y nur nichtnegative Werte annehmen, trifft dies auch auf $X+Y$ zu, d.h. $f_{X+Y}(z) = 0$ für alle $z < 0$. Für $z \geq 0$ gilt mit der Faltungsformel

$$\begin{aligned} f_{X+Y}(z) &= \int_{\mathbb{R}} f_X(z-y)f_Y(y)dy \\ &= \int_{\mathbb{R}} \lambda e^{-\lambda(z-y)} \mathbb{1}_{[0,\infty)}(z-y) \cdot \lambda e^{-\lambda y} \mathbb{1}_{[0,\infty)}(y)dy \\ &\stackrel{(*)}{=} \int_0^z \lambda^2 e^{-\lambda z} dy \\ &= \lambda^2 e^{-\lambda z} \int_0^z dy \\ &= \lambda^2 z e^{-\lambda z}, \end{aligned}$$

wobei wir in (*) die Bedingung $z-y \geq 0$ wieder nach y umformen und $y \leq z$ erhalten. Zusammen mit der Bedingung $y \geq 0$ bedeutet das, dass $0 \leq y \leq z$ sein muss, was wie vorhergesagt nur für $z \geq 0$ möglich ist.

Das Beispiel können wir mit einer Zwischenüberlegung auch auf drei oder mehr Zufallsvariablen verallgemeinern.

Lemma 5.20

Seien X, Y und Z stochastisch unabhängige gemeinsam absolutstetig verteilte Zufallsvariablen. Dann sind auch $X + Y$ und Z stochastisch unabhängig voneinander.

Beweis. Für alle $a, b \in \mathbb{R}$ ist

$$\begin{aligned} \mathbb{P}(X + Y \leq a, Z \leq b) &= \mathbb{P}((X, Y, Z) \in \{(x, y, z) \mid x + y \leq a, z \leq b\}) \\ &= \int_{\mathbb{R}^3} \mathbb{1}_{\{x \leq a, y \leq a-x, z \leq b\}}(x, y, z) f(x, y, z) d(x, y, z) \\ &= \int_{-\infty}^a \int_{-\infty}^{a-x} \int_{-\infty}^b f_X(x) f_Y(y) f_Z(z) dz dy dx \quad \text{wegen Unabh.} \\ &= \int_{-\infty}^a \int_{-\infty}^{a-x} f_X(x) f_Y(y) dy dx \cdot \int_{-\infty}^b f_Z(z) dz \\ &= \mathbb{P}(X + Y \leq a) \cdot \mathbb{P}(Z \leq b), \end{aligned}$$

d.h. $F_{(X+Y), Z} = F_{X+Y} \cdot F_Z$. □



Beispiel 5.21. Für einen Parameter $\lambda > 0$ seien $X, Y, Z \sim \text{Exp}(\lambda)$ stochastisch unabhängige Zufallsvariablen. Welche Dichte hat $X + Y + Z$?

Antwort: Aus Beispiel 5.19 kennen wir bereits die Dichte von $X + Y$. Da $X + Y$ von Z unabhängig ist, können wir das gleiche Verfahren nochmals anwenden und erhalten die Dichte

$$\begin{aligned} f_{X+Y+Z}(a) &= \int_{\mathbb{R}} f_{X+Y}(a-z) f_Z(z) dz \\ &= \\ &= \\ &= \end{aligned}$$

Summiert man mehrere unabhängige $\text{Exp}(\lambda)$ -verteilte Zufallsvariablen, so erhält man die *Gamma-Verteilung*, welche wir im Folgenden zunächst allgemein einführen.

Definition 5.22

Eine absolutstetig verteilte Zufallsvariable X heißt *Gamma-verteilt* zu den Parametern $s, \lambda > 0$, falls ihre Dichte durch

$$f(x) = \frac{\lambda}{\Gamma(s)} e^{-\lambda x} (\lambda x)^{s-1} \mathbb{1}_{[0, \infty)}(x), \quad x \in \mathbb{R},$$

gegeben ist, wobei die Γ -Funktion wie folgt definiert ist:

$$\Gamma(s) := \int_0^{\infty} e^{-y} y^{s-1} dy, \quad s > 0.$$

Kurzschreibweise: $X \sim \Gamma(s, \lambda)$.

Folgendes Lemma impliziert, dass $\Gamma(n) = (n - 1)!$ für $n \in \mathbb{N}$ ist und somit in der Tat in obigem Beispiel die Dichte der Gamma-Verteilung $\Gamma(3, \lambda)$ berechnet wurde.

Lemma 5.23

Die Gamma-Funktion

$$\Gamma(s) := \int_0^{\infty} e^{-y} y^{s-1} dy, \quad s > 0,$$

hat folgende Eigenschaften:

- i) $\Gamma(1) = 1$,
- ii) $\Gamma(x + 1) = x \cdot \Gamma(x)$ für alle $x > 0$,
- iii) $\Gamma(n) = (n - 1)!$ für alle $n \in \mathbb{N}$,
- iv) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Beweis.



i) $\Gamma(1) =$

ii) Für alle $x > 0$ berechnen wir mit partieller Integration

$$\begin{aligned} \Gamma(x + 1) &= \int_0^{\infty} e^{-y} y^x dy = \\ &= \end{aligned}$$

iii) Diese Aussage zeigt man mittels vollständiger Induktion, wobei der Induktionsanfang in i) gemacht wurde und ii) ist der Induktionsschritt, wenn man ihn auf natürliche Zahlen einschränkt: Nehmen wir als Induktionsvoraussetzung an, dass $\Gamma(n) = (n - 1)!$ ist für ein $n \in \mathbb{N}$, so folgt mit ii)

$$\Gamma(n + 1) = n \cdot \Gamma(n) \stackrel{\text{IV}}{=} n \cdot (n - 1)! = n!.$$

iv) Mit Substitution $y = \frac{1}{2}x^2$ (und $dy = x dx$) ist

$$\begin{aligned} \Gamma\left(\frac{1}{2}\right) &= \int_0^{\infty} e^{-y} y^{-1/2} dy \stackrel{\text{Subst.}}{=} \\ &= \end{aligned}$$

□

Satz 5.24

Seien X_1, \dots, X_n stochastisch unabhängige $\text{Exp}(\lambda)$ -verteilte Zufallsvariablen für einen Parameter $\lambda > 0$ und für $n \in \mathbb{N}$. Dann ist $\sum_{k=1}^n X_k \sim \Gamma(n, \lambda)$.

Beweis. Übungsaufgabe.

□

Bemerkung 5.25 (Spezialfälle der Gamma-Verteilung). Ist $X \sim \Gamma(1, \lambda)$, so ist $X \sim \text{Exp}(\lambda)$.

Ist $X \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ für $n \in \mathbb{N}$, so sagt man, X ist χ^2 -verteilt mit n Freiheitsgraden und man schreibt $X \sim \chi_n^2$ (Sprechweise: Chi-Quadrat-Verteilung). Diese Verteilung wird in diesem Kapitel nochmals auftauchen und sie spielt eine wichtige Rolle bei Hypothesentests.

Satz 5.26

Seien X_1, \dots, X_n stochastisch unabhängige Zufallsvariablen mit $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ für alle $k \in \{1, \dots, n\}$. Dann ist

$$\sum_{k=1}^n X_k \sim \mathcal{N}\left(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2\right).$$

Beweis. Wir zeigen die Aussage zunächst für $n = 2$ Zufallsvariablen mit $\mu_1 = \mu_2 = 0$. Setzen wir die Normalverteilungsdichte in die Faltungsformel ein, so erhalten wir

$$f_{X_1+X_2}(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-y)^2}{2\sigma_1^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{y^2}{2\sigma_2^2}\right) dy.$$

Der (negative) Exponent im Integranden ist

$$\begin{aligned} \frac{(x-y)^2}{2\sigma_1^2} + \frac{y^2}{2\sigma_2^2} &= \frac{1}{2} \left[\frac{1}{\sigma_1^2} x^2 - \frac{2xy}{\sigma_1^2} + \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) y^2 \right] \\ &= \frac{1}{2} \left[\left(\frac{x}{\sigma_1^2 \sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}} - \sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \cdot y \right)^2 + x^2 \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_1^4 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)} \right) \right] \\ &= \frac{1}{2} \left[(a-z)^2 + x^2 \left(\frac{1}{\sigma_1^2 + \sigma_2^2} \right) \right] \end{aligned}$$

mit

$$a = \frac{x}{\sigma_1^2 \sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}} \quad \text{und} \quad z = \sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \cdot y.$$

Mit der Substitution $z = g(y) = \sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \cdot y$ (a betrifft nicht die Integrationsvariable, dies ist also keine Subst.) erhalten wir die Dichte von $X_1 + X_2$:

$$\begin{aligned} f_{X_1+X_2}(x) &= \exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}(a-z)^2\right) \frac{1}{\sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}} dz \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2\sigma_2^2}} \cdot \frac{1}{\sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}} \cdot \exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(a-z)^2\right) dz \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma_1^2 + \sigma_2^2}\right) \end{aligned}$$

als die Dichte der $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$ -Verteilung.

Sind nun $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, so folgt aus Satz 4.39, dass $X_1 - \mu_1 \sim \mathcal{N}(0, \sigma_1^2)$ und $X_2 - \mu_2 \sim \mathcal{N}(0, \sigma_2^2)$ sind. Aus der gerade gezeigten Eigenschaft folgt sofort

$$(X_1 - \mu_1) + (X_2 - \mu_2) \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2).$$

Aus Satz 4.39 folgt nun wiederum

$$X_1 + X_2 = [(X_1 - \mu_1) + (X_2 - \mu_2)] + (\mu_1 + \mu_2) \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Die Verallgemeinerung auf $n > 2$ zeigt man nun induktiv, wobei der Induktionsanfang bereits gemacht ist. Die Induktionsvoraussetzung besagt, dass

$$X := X_1 + \dots + X_{n-1} \sim \mathcal{N}\left(\sum_{k=1}^{n-1} \mu_k, \sum_{k=1}^{n-1} \sigma_k^2\right)$$

ist. Im Induktionsschritt zeigt man analog zu obiger Rechnung, dass

$$X + X_n \sim \mathcal{N}\left(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2\right)$$

gilt. □

Aus Satz 4.39 und dem gerade bewiesenen Satz 5.26 folgt sofort folgende Aussage:

Korollar 5.27

Seien X_1, \dots, X_n stochastisch unabhängige Zufallsvariablen mit $X_k \sim \mathcal{N}(\mu, \sigma^2)$ für alle $k \in \{1, \dots, n\}$. Dann ist

$$\frac{1}{n} \sum_{k=1}^n X_k \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Beweis. Übungsaufgabe. □

5.4 Erwartungswert von Funktionen von Zufallsvariablen

Analog zu Satz 3.45 und Satz 4.14 definieren wir nun den Erwartungswert von $g(X, Y)$ unter der Voraussetzung, dass die Zufallsvariablen X und Y entweder gemeinsam absolutstetig oder gemeinsam diskret verteilt sind und dass der Erwartungswert existiert. Insbesondere darf g nur Werte in \mathbb{R} annehmen.

Satz 5.28

1. Seien X, Y diskret verteilt mit gemeinsamer Massenfunktion p und sei $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ derart, dass

$$\sum_{(x,y): p(x,y)>0} |g(x,y)| p(x,y) < \infty.$$

Dann ist der Erwartungswert der Zufallsvariablen $g(X, Y)$ gegeben durch

$$\mathbb{E}[g(X, Y)] = \sum_{(x,y): p(x,y)>0} g(x,y)p(x,y).$$

2. Seien X, Y gemeinsam absolutstetig verteilt mit gemeinsamer Dichte f und sei $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ messbar und erfülle

$$\int_{\mathbb{R}^2} |g(x,y)| f(x,y) d(x,y) < \infty.$$

Dann ist

$$\mathbb{E}[g(X, Y)] = \int_{\mathbb{R}^2} g(x,y) f(x,y) d(x,y).$$

Der Beweis ist analog zu den genannten Sätzen, daher überspringen wir ihn an dieser Stelle.



Beispiel 5.29. Peter und Tony verabreden sich für ein Treffen am Alexanderplatz zwischen 15:00 und 15:15 Uhr. Beide kommen unabhängig voneinander in diesem Zeitraum an und ihre Ankunftszeiten können als gleichverteilt angenommen werden. Wer zuerst ankommt, muss auf den anderen warten. Wie lang ist die zu erwartende Wartezeit?

Antwort: Seien $X, Y \sim U(0, T)$ die Ankunftszeiten nach 15:00 Uhr in Minuten, d.h. $T = 15$ (Minuten) und X und Y sind stochastisch unabhängig. Gesucht ist $\mathbb{E}[|X - Y|]$. Wegen der Unabhängigkeit ist die gemeinsame Dichte von X und Y gegeben durch

$$f(x, y) = f_X(x) \cdot f_Y(y) = \frac{1}{T^2} \mathbb{1}_{[0,T]}(x) \mathbb{1}_{[0,T]}(y).$$

Damit ist

$$\begin{aligned}\mathbb{E}[|X - Y|] &= \int_{\mathbb{R}^2} |x - y| f(x, y) d(x, y) \\ &= \int_0^T \int_0^T |x - y| \frac{1}{T^2} dx dy \\ &= \end{aligned}$$

Korollar 5.30

Seien X und Y gemeinsam diskret (absolutstetig) verteilte Zufallsvariablen mit gemeinsamer Massenfunktion p (Dichte f). Dann gilt $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Beweis. Wir zeigen die Aussage für gemeinsam diskret verteilte Zufallsvariablen. Für gemeinsam absolutstetig verteilte Zufallsvariablen ist der Beweis Inhalt einer Übungsaufgabe.

$$\begin{aligned}\mathbb{E}[X + Y] &= \sum_{(x,y): p(x,y)>0} (x + y)p(x, y) \\ &= \sum_{x: p_X(x)>0} \sum_{y: p(x,y)>0} xp(x, y) + \sum_{y: p_Y(y)>0} \sum_{x: p(x,y)>0} yp(x, y) \\ &= \sum_{x: p_X(x)>0} x \left(\sum_{y: p(x,y)>0} p(x, y) \right) + \sum_{y: p_Y(y)>0} y \left(\sum_{x: p(x,y)>0} p(x, y) \right) \\ &= \sum_{x: p_X(x)>0} xp_X(x) + \sum_{y: p_Y(y)>0} yp_Y(y) \\ &= \mathbb{E}[X] + \mathbb{E}[Y].\end{aligned}$$

□

Beispiel 5.31. In Satz 5.14 und Bemerkung 5.15 haben wir festgestellt, dass die Summe von n Bernoulli-verteilten Zufallsvariablen zum Parameter $p \in (0, 1)$ selbst $\text{Bin}(n, p)$ -verteilt ist. Damit können wir den Erwartungswert der Binomialverteilung sehr einfach herleiten.

Seien $X_1, \dots, X_n \sim \text{Ber}(p)$ stochastisch unabhängige Zufallsvariablen, die beispielsweise ein wiederholtes Bernoulli-Experiment repräsentieren. Nennen wir das Ereignis $X_k = 1$ einen Erfolg, so repräsentiert $X := \sum_{k=1}^n X_k$ die Anzahl der Erfolge bei n Versuchen. Nach Satz 5.14 bzw. Bemerkung 5.15 ist $X \sim \text{Bin}(n, p)$. Es gilt

$$\mathbb{E}[X_k] = 1 \cdot \mathbb{P}(X_k = 1) + 0 \cdot \mathbb{P}(X_k = 0) = \mathbb{P}(X_k = 1) = p$$

für $k \in \{1, \dots, n\}$ und somit

$$\mathbb{E}[X] = \sum_{k=1}^n \mathbb{E}[X_k] = np.$$

Korollar 5.32

Seien X und Y gemeinsam diskret oder absolutstetig verteilte Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Erwartungswerten $\mathbb{E}[X]$ und $\mathbb{E}[Y]$. Gilt $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$, so folgt daraus

$$\mathbb{E}[X] \leq \mathbb{E}[Y].$$



Beweis. Nach dem vorherigen Korollar ist

$$\mathbb{E}[Y] \geq \mathbb{E}[X] \iff \mathbb{E}[Y - X] \geq 0.$$

Dieser Erwartungswert ist in der Tat nichtnegativ, denn für gemeinsam diskret verteilte Zufallsvariablen ist

$$\mathbb{E}[Y - X] =$$

und für gemeinsam stetig verteilte Zufallsvariablen ist

$$\mathbb{E}[Y - X] =$$

□



Beispiel 5.33 (Hutproblem – Fortsetzung). Die Hüte von n Personen werden gemischt und nacheinander nimmt jede Person zufällig einen Hut. Wie groß ist die erwartete Anzahl der Personen, die ihren eigenen Hut ziehen?

Antwort: In Beispiel 1.46 haben wir bereits die Wahrscheinlichkeit des Ereignisses

$$E_i: \text{„Person } i \text{ erhält ihren eigenen Hut.“}$$

als

$$\mathbb{P}(E_i) = \frac{1}{n}, \quad i \in \{1, \dots, n\}$$

berechnet. Ist X die Anzahl der Personen, die ihren eigenen Hut bekommen, so ist

$$X = \sum_{i=1}^n \mathbb{1}_{E_i}$$

und somit (vgl. Beispiel 3.41 für den Erwartungswert der Indikatorfunktion)

$$\mathbb{E}[X] =$$

5.4.1 Kovarianz und Korrelation

In Korollar 5.30 haben wir gesehen, dass der Erwartungswert additiv ist, d.h. es gilt immer $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ – ganz gleich, ob die Zufallsvariablen stochastisch unabhängig sind oder nicht. Um zu sehen, ob dies auch auf Varianzen zutrifft, müssen wir untersuchen, ob auch der Erwartungswert vom Produkt zweier Zufallsvariablen gleich dem Produkt von deren Erwartungswerten ist.

Satz 5.34

Seien X und Y zwei stochastisch unabhängige Zufallsvariablen und seien $g, h: \mathbb{R} \rightarrow \mathbb{R}$ messbare Funktionen, so dass die Erwartungswerte von $g(X)$ und $h(Y)$ existieren. Dann gilt

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)].$$

Beweis von Satz 5.34. Wir verlangen hier nicht, dass X und Y gemeinsam absolutstetig oder diskret verteilt sind. Die Idee ist Folgende: Man kann jede Zufallsvariable in einen absolutstetigen und einen diskreten Anteil zerlegen wie wir uns an einer Skizze verdeutlichen. Damit genügt es dann, den Beweis für zwei diskrete, zwei absolutstetige oder eine diskrete und eine absolutstetige ZV zu führen. Details zu dem Übergang zu beliebigen ZVen geht über das Ziel dieser Vorlesung hinaus.

X, Y **diskret:** Auch $g(X)$ und $h(Y)$ sind diskret verteilt, so dass wir den Erwartungswert wie folgt berechnen:

$$\begin{aligned} \mathbb{E}[g(X)h(Y)] &= \sum_{(x,y): p(x,y)>0} g(x)h(y)p(x,y) \\ &= \left(\sum_{x: p_X(x)>0} g(x)p_X(x) \right) \cdot \left(\sum_{y: p_Y(y)>0} h(y)p_Y(y) \right) \\ &= \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)]. \end{aligned}$$

X, Y **absolutstetig:** Wegen der Unabhängigkeit beider ZVen gilt $f(x, y) = f_X(x)f_Y(y)$ und somit

$$\begin{aligned} \mathbb{E}[g(X)h(Y)] &= \int_{\mathbb{R}^2} g(x)h(y)f(x, y)d(x, y) \\ &= \left(\int_{\mathbb{R}} g(x)f_X(x)dx \right) \cdot \left(\int_{\mathbb{R}} h(y)f_Y(y)dy \right) \\ &= \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)]. \end{aligned}$$

X **diskret**, Y **absolutstetig:** Analog zum Beweis von Satz 4.14 nehmen wir an, dass $g, h \geq 0$ (d.h. nichtnegative Funktionen) sind. Dann gilt mit Lemma 4.15

$$\begin{aligned}
\mathbb{E}[g(X)h(Y)] &= \int_0^\infty \mathbb{P}(g(X)h(Y) > z) dz \\
&= \int_0^\infty \sum_{x: p_X(x) > 0} \mathbb{P}(g(X)h(Y) > z | X = x) \mathbb{P}(X = x) dz \\
&= \int_0^\infty \sum_{x: p_X(x) > 0} \mathbb{P}(g(x)h(Y) > z | X = x) \mathbb{P}(X = x) dz \\
&= \int_0^\infty \sum_{x: p_X(x) > 0} \mathbb{P}(g(x)h(Y) > z) \mathbb{P}(X = x) dz \\
&= \sum_{x: p_X(x) > 0} \left(\mathbb{P}(X = x) \int_0^\infty \mathbb{P}(g(x)h(Y) > z) dz \right) \\
&= \sum_{x: p_X(x) > 0} \mathbb{P}(X = x) \mathbb{E}[g(x)h(Y)] \\
&= \mathbb{E}[h(Y)] \sum_{x: p_X(x) > 0} g(x) \mathbb{P}(X = x) \\
&= \mathbb{E}[h(Y)] \cdot \mathbb{E}[g(X)].
\end{aligned}$$

□

Kommen wir nun zur Varianz der Summe zweier Zufallsvariablen.

Satz 5.35

Seien X und Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, deren Erwartungswerte und Varianzen existieren. Falls zudem

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

existiert, so gilt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \quad (5.4.1)$$

Beweis. Es gilt

$$\begin{aligned}
\text{Var}(X + Y) &= \mathbb{E} \left[((X + Y) - \mathbb{E}[X + Y])^2 \right] \\
&= \mathbb{E} \left[(X + Y)^2 + (\mathbb{E}[X + Y])^2 - 2(X + Y)\mathbb{E}[X + Y] \right] \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] + (\mathbb{E}[X] + \mathbb{E}[Y])^2 - 2(\mathbb{E}[X] + \mathbb{E}[Y])(\mathbb{E}[X] + \mathbb{E}[Y]) \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 + 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] \\
&= \text{Var}(X) + \text{Var}(Y) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].
\end{aligned}$$

□

Bemerkung 5.36. Falls X und Y stochastisch unabhängig sind, so zeigt Satz 5.34 (in größerer Allgemeinheit), dass $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ ist. In diesem Fall wird aus (5.4.1)

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

d.h. die Varianz der Summe zweier stochastisch unabhängiger Zufallsvariablen ist gleich der Summe der Varianzen beider Zufallsvariablen. Der im allgemeinen zusätzlich auftretende Term entsteht durch Abhängigkeiten zwischen beiden Zufallsvariablen (er hängt von der gemeinsamen Verteilung ab und nicht nur von den Randverteilungen!) und wir führen für ihn einen neuen Begriff ein.

Definition 5.37

Seien X und Y Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit existierenden Varianzen. Dann existiert der Ausdruck

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

und wir bezeichnen ihn als Kovarianz von X und Y .

Bemerkung 5.38. Falls $\text{Var}(X)$ und $\text{Var}(Y)$ existierten, so existiert auch die Kovarianz $\text{Cov}(X, Y)$, denn dann existieren die Erwartungswerte $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\mathbb{E}[Y]$ und $\mathbb{E}[Y^2]$ und aus

$$|XY| \leq \frac{X^2 + Y^2}{2}$$

folgt

$$\mathbb{E}[|XY|] \leq \frac{1}{2} (\mathbb{E}[X^2] + \mathbb{E}[Y^2]) < \infty.$$

Bemerkung 5.39. Die Kovarianz (englisch covariance) wird in der Literatur bisweilen auch durch $C(X, Y)$ abgekürzt. Mit Hilfe der Kovarianz wird aus (5.4.1)

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

Satz 5.40: Eigenschaften der Kovarianz

Seien X, X_1, \dots, X_n und Y, Y_1, \dots, Y_m Zufallsvariablen, deren Erwartungswerte, Varianzen und Kovarianz existieren. Dann gelten folgende Eigenschaften.

1. $\text{Cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$;
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
3. $\text{Cov}(X, X) = \text{Var}(X)$;
4. $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$ für beliebige $a, b, c, d \in \mathbb{R}$;
5. $\text{Cov}\left(\sum_{k=1}^n X_k, \sum_{j=1}^m Y_j\right) = \sum_{k=1}^n \sum_{j=1}^m \text{Cov}(X_k, Y_j)$.

Beweis.



1. Rechnen wir diesen Schritt aus dem Beweis von Satz 5.35 nochmal genau nach:

$$\begin{aligned} \text{Cov}(X, Y) &= \\ &= \\ &= \end{aligned}$$

2. Die Aussage folgt sofort aus 1., da die Formel symmetrisch ist bezüglich X und Y .

3. Mit 1. und der Zerlegungsformel der Varianz aus Lemma 3.51 bzw. Lemma 4.20 gilt

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

4. Mit Hilfe von 1. und den Rechenregeln für den Erwartungswert erhalten wir

$$\begin{aligned} & \text{Cov}(aX + b, cY + d) \\ &= \mathbb{E}[(aX + b) \cdot (cY + d)] - \mathbb{E}[aX + b] \cdot \mathbb{E}[cY + d] \\ &= \end{aligned}$$

5. Wir verwenden wieder die Darstellung aus 1.

$$\begin{aligned} \text{Cov}\left(\sum_{k=1}^n X_k, \sum_{j=1}^m Y_j\right) &= \mathbb{E}\left[\sum_{k=1}^n X_k \cdot \sum_{j=1}^m Y_j\right] - \mathbb{E}\left[\sum_{k=1}^n X_k\right] \cdot \mathbb{E}\left[\sum_{j=1}^m Y_j\right] \\ &= \mathbb{E}\left[\sum_{k=1}^n \sum_{j=1}^m X_k Y_j\right] - \left(\sum_{k=1}^n \mathbb{E}[X_k]\right) \cdot \left(\sum_{j=1}^m \mathbb{E}[Y_j]\right) \\ &= \sum_{k=1}^n \sum_{j=1}^m (\mathbb{E}[X_k Y_j] - \mathbb{E}[X_k] \mathbb{E}[Y_j]) \\ &= \sum_{k=1}^n \sum_{j=1}^m \text{Cov}(X_k, Y_j). \end{aligned}$$

□

Korollar 5.41

Sind X und Y stochastisch unabhängige Zufallsvariablen mit existierenden Erwartungswerten, so gilt $\text{Cov}(X, Y) = 0$.

Beweis. Dies folgt sofort aus der Formel aus dem vorherigen Satz und aus Satz 5.34 wegen der stoch. Unabhängigkeit:

$$\text{Cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] \stackrel{\text{Satz 5.34}}{=} \mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0.$$

□

Die Umkehrung des Korollars gilt nicht, was folgendes Beispiel zeigt.



Beispiel 5.42. Die Zufallsvariable X sei gleichverteilt auf $\{-1, 0, 1\}$ und $Y := \mathbb{1}_{\{X=0\}}$. Sind X und Y stochastisch unabhängig? Wie groß ist ihre Kovarianz?

Antwort: X hat die Massenfunktion

und Y hat die Massenfunktion

Damit ist

$$\mathbb{P}(X = 0, Y = 0) =$$

d.h. X und Y sind nicht stochastisch unabhängig. Die Erwartungswerte, die wir für die Kovarianz benötigen, sind

$$\mathbb{E}[X] =$$

$$\mathbb{E}[Y] =$$

$$\mathbb{E}[XY] =$$

Damit ist

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$$

Mit Hilfe von 5. aus Satz 5.40 kann man die Varianz der Summe von n Zufallsvariablen berechnen und erhält als Verallgemeinerung von Bemerkung 5.36

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \text{Cov}\left(\sum_{k=1}^n X_k, \sum_{j=1}^n X_j\right) = \sum_{k=1}^n \text{Var}(X_k) + \sum_{k \neq j} \text{Cov}(X_k, X_j). \quad (5.4.2)$$

Beispiel 5.43. Seien X_1, \dots, X_n stochastisch unabhängige Bernoulli-verteilte Zufallsvariablen zum Parameter $p \in (0, 1)$. Welche Varianz besitzt $X := \sum_{k=1}^n X_k$? 

Antwort: Einerseits wissen wir bereits, dass $X \sim \text{Bin}(n, p)$ ist und haben die Varianz der Binomialverteilung berechnet. Andererseits können wir die Varianz mit den neuen Methoden bestimmen:

- Für jedes $k \in \{1, \dots, n\}$ ist

$$\text{Var}(X_k) =$$

- Ist $k \neq j$, so gilt wegen der stochastischen Unabhängigkeit mit Satz 5.34

$$\text{Cov}(X_k, X_j) =$$

- Aus Formel (5.4.2) folgt damit

$$\text{Var}(X) =$$

Beispiel 5.44 (Stichprobenmittel und -varianz). Seien X_1, \dots, X_n stochastisch unabhängige und identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_k] =: \mu$ und $\text{Var}(X_k) =: \sigma^2$. Welchen Erwartungswert haben

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

und welche Varianz hat \bar{X} ?

Antwort: Betrachten wir zunächst das Stichprobenmittel \bar{X} .

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

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Beide Eigenschaften zusammen zeigen, warum das arithmetische Mittel einer Stichprobe ein guter Schätzer für den Erwartungswert der zugrunde liegenden Zufallsvariablen ist. Sehen wir uns nun die Stichprobenvarianz an:

$$\begin{aligned} \mathbb{E}[S] &= \frac{1}{n-1} \sum_{k=1}^n \mathbb{E}[(X_k - \bar{X})^2] \\ &= \frac{1}{n-1} \sum_{k=1}^n \mathbb{E}[(X_k - \mu + \mu - \bar{X})^2] \\ &= \frac{1}{n-1} \sum_{k=1}^n \left(\mathbb{E}[(X_k - \mu)^2] + 2\mathbb{E}[(X_k - \mu)(\mu - \bar{X})] + \mathbb{E}[(\bar{X} - \mu)^2] \right) \\ &= \frac{1}{n-1} \sum_{k=1}^n \left(\text{Var}(X_k) + 2(\mu\mathbb{E}[X_k] - \mu^2 + \mu\mathbb{E}[\bar{X}] - \mathbb{E}[X_k \cdot \bar{X}]) + \text{Var}(\bar{X}) \right) \\ &= \frac{1}{n-1} \sum_{k=1}^n \left(\sigma^2 + 2 \left(\mu^2 - \frac{1}{n} \left(\mathbb{E}[X_k^2] + \sum_{k \neq j} \mathbb{E}[X_k X_j] \right) \right) + \frac{\sigma^2}{n} \right) \\ &= \frac{1}{n-1} \sum_{k=1}^n \left(\sigma^2 + 2 \left(\mu^2 - \frac{1}{n} (\sigma^2 + \mu^2 + (n-1)\mu^2) \right) + \frac{\sigma^2}{n} \right) \\ &= \frac{n}{n-1} \cdot \left(\frac{n-2+1}{n} \sigma^2 + 0\mu \right) \\ &= \sigma^2. \end{aligned}$$

Man sagt daher, S ist ein erwartungstreuer Schätzer für die Varianz σ^2 .

Eigenschaft 4 aus Satz 5.40 sagt uns, dass die Kovarianz zweier Zufallsvariablen mit den Zufallsvariablen wächst – damit ist sie nicht geeignet, um den Grad der (linearen) Abhängigkeit zwischen Zufallsvariablen zu messen. Dazu führen wir eine neue geeignet skalierte Größe ein.

Definition 5.45

Seien X und Y zwei nicht konstante Zufallsvariablen mit existierenden Varianzen. Dann heißt

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

Korrelationskoeffizient von X und Y . Ist $\varrho(X, Y) = 0$, so heißen X und Y unkorreliert.

Bemerkung 5.46. Sind X und Y nicht konstant, so sind ihre Varianzen strikt positiv und damit ist der Korrelationskoeffizient wohldefiniert. Damit gilt auch

$$\varrho(X, Y) = 0 \iff \text{Cov}(X, Y) = 0,$$

d.h. zwei nicht konstante Zufallsvariablen sind genau dann unkorreliert, wenn ihre Kovarianz Null ist.

Mit Hilfe dieser Bezeichnung können wir uns merken:

Stochastisch unabhängige Zufallsvariablen sind stets unkorreliert, aber unkorrelierte Zufallsvariablen müssen nicht stochastisch unabhängig sein. 

Bevor wir ein Beispiel ansehen, halten wir fest, dass der Korrelationskoeffizient gegenüber der Kovarianz tatsächlich beschränkt ist.

Lemma 5.47

Seien X und Y nicht-konstante Zufallsvariablen mit existierenden Varianzen. Dann gilt

$$|\rho(X, Y)| \leq 1.$$

Beweis. Seien $\sigma_X^2 := \text{Var}(X)$ und $\sigma_Y^2 := \text{Var}(Y)$. Dann ist

$$\begin{aligned} 0 &\leq \text{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) \\ &= \text{Var}\left(\frac{X}{\sigma_X}\right) + \text{Var}\left(\frac{Y}{\sigma_Y}\right) \pm 2 \text{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\ &= \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} \pm 2 \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\ &= 2 \pm 2\rho(X, Y) \end{aligned}$$

Umstellen liefert $-1 \leq \rho(X, Y) \leq 1$. □

Beispiel 5.48. Zu einer Zufallsvariablen X mit $\mathbb{E}[X] = \mu$ und $\text{Var}(X) = \sigma^2$ sei $Y := aX + b$ für $a, b \in \mathbb{R}$. Wie groß ist $\rho(X, Y)$? 

Antwort: Es ist

$$\begin{aligned} \mathbb{E}[Y] &= \\ \text{Var}(Y) &= \\ \text{Cov}(X, Y) &= \end{aligned}$$

$$\rho(X, Y) =$$

5.4.2 Anwendung: Lineare Regression

Das Ziel besteht darin, basierend auf der Realisierung einer Zufallsvariablen X die Realisierung einer Zufallsvariablen Y mit Hilfe eines linearen Modells vorherzusagen.

Mit anderen Worten: Wir kennen $X(\omega)$ und wollen $a, b \in \mathbb{R}$ derart wählen, dass $\hat{Y}(\omega) := aX(\omega) + b$ ein guter Schätzer für $Y(\omega)$ ist. Die Güte wird durch die mittlere quadratische Abweichung gemessen, d.h. wir wollen

$$\mathbb{E}[(Y - aX - b)^2]$$

über $a, b \in \mathbb{R}$ minimieren.

Wäre X konstant, d.h. $\mathbb{P}(X = x) = 1$ für ein $x \in \mathbb{R}$, so wären X und Y stochastisch unabhängig und da die Varianz die quadratische Abweichung minimiert (Vgl. Ergänzungsfrage 3 Kapitel 4 siehe B.4), wird $\mathbb{E}[(Y - aX - b)^2]$ für all die $a, b \in \mathbb{R}$ minimal, welche

$$ax + b = \mathbb{E}[Y]$$

erfüllen, also insb. für $a = 0$ und $b = \mathbb{E}[Y]$.

Nehmen wir ab jetzt an, dass X nicht konstant ist und somit $\text{Var}(X) > 0$ ist. Wir wollen die Funktion

$$g(a, b) := \mathbb{E}[(Y - aX - b)^2] = \mathbb{E}[Y^2] + a^2\mathbb{E}[X^2] + b^2 - 2a\mathbb{E}[XY] + 2ab\mathbb{E}[X] - 2b\mathbb{E}[Y]$$



minimieren. Dazu leiten wir sie zunächst nach a und b partiell ab und setzen diese Null:

$$\begin{aligned} \frac{\partial}{\partial a} g(a, b) &= \\ \frac{\partial}{\partial b} g(a, b) &= \end{aligned}$$

Aus der zweiten Gleichung erhalten wir

$$b =$$

und Einsetzen in die erste Gleichung und Umstellen liefert

$$\iff a =$$

Das hinreichende Optimalitätskriterium kann durch die Definitheit der Hessematrix

$$Hg(a, b) =$$

überprüft werden. Diese ist positiv definit, denn ihr charakteristisches Polynom

$$p_H(t) =$$

hat die Nullstellen

$$t_{1/2} =$$

Einerseits ist offensichtlich

$$t_1 = 1 + \mathbb{E}[X^2] + \sqrt{(\mathbb{E}[X^2] - 1)^2 + 4\mathbb{E}[X]^2} > 0.$$

Andererseits ist wegen $\text{Var}(X) > 0$

$$\mathbb{E}[X^2]^2 - 2\mathbb{E}[X^2] + 1 + 4\mathbb{E}[X]^2 < \mathbb{E}[X^2]^2 - 2\mathbb{E}[X^2] + 1 + 4\mathbb{E}[X]^2$$

und somit auch

$$t_2 = 1 + \mathbb{E}[X^2] - \sqrt{\mathbb{E}[X^2]^2 - 2\mathbb{E}[X^2] + 1 + 4\mathbb{E}[X]^2} > 0.$$

Folglich ist $Hg(a, b)$ in der Tat positiv definit, d.h.

$$(a^*, b^*) = \left(\frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \mathbb{E}[Y] - a^*\mathbb{E}[X] \right)$$

ist ein Minimierer von g . Da es keine weiteren kritischen Punkte gibt, hat g in (a^*, b^*) sogar sein globales Minimum.

Bemerkung 5.49. Alternativ zur Untersuchung der Vorzeichen der Eigenwerte der Hesse-Matrix gibt es das *Minoren-Kriterium*: Sind beide Hauptminoren von $Hg(a, b)$ positiv, so ist die Matrix positiv definit. In unserem Fall sind $2\mathbb{E}[X^2] > 0$ und $\det(Hg(a, b)) = 4 \operatorname{Var}(X) > 0$.

Fassen wir das gefundene Ergebnis als Satz zusammen.

Satz 5.50: Lineare Regression

Ist $\operatorname{Var}(X) > 0$, so besitzt das Optimierungsproblem

$$\min_{a, b \in \mathbb{R}} \mathbb{E} [(Y - aX - b)^2]$$

die eindeutige Lösung

$$a^* = \frac{\operatorname{Cov}(X, Y)}{\operatorname{Var}(X)} \quad \text{und} \quad b^* = \mathbb{E}[Y] - a^* \mathbb{E}[X].$$

Wie groß ist der Vorhersagefehler F ?



$$\begin{aligned} F &= \mathbb{E} [(Y - a^*X - b^*)^2] \\ &= \\ &= \\ &= \\ &= \\ &= \end{aligned}$$

Der Korrelationskoeffizient bestimmt also die Güte der Approximation bei der linearen Regression – der Vorhersagefehler wird minimal für $\varrho(X, Y) \in \{\pm 1\}$ und maximal für $\varrho(X, Y) = 0$.

- Sind X und Y unkorreliert, so ist die beste lineare Vorhersage $\hat{Y} = \mathbb{E}[Y]$ unabhängig von X und der Vorhersagefehler ist $\operatorname{Var}(Y)$.
- Ist $|\varrho(X, Y)| = 1$, so ist der Vorhersagefehler $F = 0$, d.h. es gibt $a, b \in \mathbb{R}$ derart, dass $\mathbb{P}(Y = aX + b) = 1$ ist. Andersherum folgt aus $Y = aX + b$ sofort

$$\varrho(X, Y) = \frac{\operatorname{Cov}(X, aX + b)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(aX + b)}} = \frac{a \operatorname{Cov}(X, X)}{\sqrt{a^2 \operatorname{Var}(X)^2}} = \frac{a}{|a|} = \operatorname{sgn}(a) \in \{\pm 1\}.$$

Kapitel 6

Grenzwertsätze

Ziele

- Ungleichungen von Markow und Tschebyschow kennen
- Schwaches Gesetz der großen Zahlen kennen und beweisen können
- die ZGWS-Eigenschaft kennen und dafür hinreichende Voraussetzungen kennen
- ZGWS anwenden können (ggf. mit Stetigkeitskorrektur)

6.1 Zwei wichtige Ungleichungen

Wenn wir eine faire Münze 1000-mal werfen, so erwarten wir, ungefähr 500-mal *Zahl* zu sehen. Es könnten jedoch auch 487 oder 504 *Zahl*-Würfe sein – die tatsächliche Anzahl sollte jedoch „in der Nähe“ des Wertes 500 liegen.

In diesem Abschnitt wollen wir die Wahrscheinlichkeit, dass eine Zufallsvariable um mehr als einen vorgegebenen Wert von ihrem Erwartungswert abweicht, nach oben abschätzen. Dafür benötigen wir zunächst ein Hilfsmittel.

Nach dem russischen Mathematiker Andrei Andrejewitsch Markow (1856-1922) – im englischsprachigen Raum auch Markov geschrieben – ist folgende nützliche Ungleichung benannt:

Satz 6.1: Markow-Ungleichung (allgemein)

Sei $h: \mathbb{R} \rightarrow \mathbb{R}_+$ eine monoton wachsende Funktion und sei X eine Zufallsvariable mit existierendem Erwartungswert $\mathbb{E}[h(X)]$. Dann gilt für alle $a \in \mathbb{R}$

$$h(a)\mathbb{P}(X \geq a) \leq \mathbb{E}[h(X)].$$

Beweis. Es gilt dank der Monotonie des Erwartungswertes aus Korollar 5.32

$$\begin{aligned} h(a)\mathbb{P}(X \geq a) &= h(a)\mathbb{E}[\mathbb{1}_{\{X \geq a\}}] \\ &= \mathbb{E}[h(a)\mathbb{1}_{\{X \geq a\}}] \\ &\leq \mathbb{E}[h(X)\mathbb{1}_{\{X \geq a\}}] \\ &\leq \mathbb{E}[h(X)]. \end{aligned}$$

□

Setzen wir $h(x) = x$, so erhalten wir unmittelbar folgendes teilweise auch als Markow-Ungleich bezeichnetes Resultat:

Korollar 6.2: Markow-Ungleichung (speziell)

Sei $X \geq 0$ eine Zufallsvariable mit existierendem Erwartungswert. Dann gilt für alle $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Mit Hilfe dieser Ungleichung können wir das eigentliche Ziel dieses Abschnitts – die nach dem ebenfalls russischen Mathematiker Pafnuti Lwowitsch Tschebyschow (1821-1894) benannte Ungleichung – beweisen. Alternative Schreibweisen dieses Namens sind *Tschebyscheff* und (im englischsprachigen Raum) *Chebyshev*.

Satz 6.3: Tschebyschow-Ungleichung

Sei X eine Zufallsvariable mit existierender Varianz σ^2 und Erwartungswert μ . Dann gilt für jedes $k > 0$

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

Beweis. Wenden wir die spezielle Markow'sche Ungleichung auf $Z := (X - \mu)^2$ an, so erhalten wir direkt

$$\mathbb{P}(|X - \mu| \geq k) = \mathbb{P}((X - \mu)^2 \geq k^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

□



Beispiel 6.4. Die Zahl der Produkte, die eine Firma pro Tag herstellt, sei eine Zufallsvariable X mit Erwartungswert $\mathbb{E}[X] = 50$ und Varianz $\text{Var}(X) = 25$.

- Wie wahrscheinlich ist es, dass an einem Tag mindestens 75 Produkte hergestellt werden?
- Mit welcher Wahrscheinlichkeit werden zwischen 40 und 60 Produkte an einem Tag hergestellt?

Antwort:

- Mit der Markow'schen Ungleichung berechnen wir

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}[X]}{75} =$$

- Wir wenden die Ungleichung von Tschebyschow an und erhalten

$$\begin{aligned} \mathbb{P}(40 < X < 60) &= \\ &= \end{aligned}$$

6.2 Gesetze der großen Zahlen

Die Aussage um die es in diesem Abschnitt geht kann man wie folgt zusammenfassen: Das arithmetische Mittel einer Folge von zentrierten Zufallsvariablen (d.h. ihr Erwartungswert ist Null) konvergiert gegen Null.

Wir müssen dazu jedoch festhalten was wir unter Konvergenz einer Folge von Zufallsvariablen verstehen und unter welchen Bedingungen an die Zufallsvariablen diese Konvergenzaussage gesichert ist.

Da wir von *Gesetzen der großen Zahlen* im Plural sprechen, halten wir zunächst fest, welche Fassungen der Konvergenzaussage es gibt:

Definition 6.5

1. Die Zufallsvariablen $(X_k)_{k \in \mathbb{N}}$ genügen dem *schwachen Gesetz der großen Zahlen*, falls

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{1}{n} \sum_{k=1}^n (X_k - \mathbb{E}[X_k]) \right| > \varepsilon \right) = 0, \quad \forall \varepsilon > 0.$$

2. Die Zufallsvariablen $(X_k)_{k \in \mathbb{N}}$ genügen dem *starken Gesetz der großen Zahlen*, falls

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{k=1}^n (X_k - \mathbb{E}[X_k]) \right| = 0 \right) = 1.$$

Wie der Name bereits suggeriert ist das starke G.d.g.Z. stärker als das schwache G.d.g.Z. und es erfordert folglich stärkere Voraussetzungen. Für diese Vorlesung soll es uns genügen, hinreichende Bedingungen für das schwache G.d.g.Z. anzugeben und die Gültigkeit für diesen einen Fall zu beweisen.

6.2.1 Schwaches Gesetz der großen Zahlen

Satz 6.6: Schwaches Gesetz der großen Zahlen

Sei $(X_k)_{k \in \mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $\mathbb{E}[X_k] = \mu$ und $\text{Var}(X_k) = \sigma^2$ für $k \in \mathbb{N}$. Dann gilt für alle $\varepsilon > 0$

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

Bemerkung 6.7. Die angegebene Formel ist tatsächlich die des schwachen G.d.g.Z., denn wenn $\mathbb{E}[X_k] = \mu$ ist für alle $k \in \mathbb{N}$, dann ist

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mathbb{E}[X_k]) = \frac{1}{n} \sum_{k=1}^n (X_k - \mu) = \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n \mu = \frac{1}{n} \sum_{k=1}^n X_k - \mu.$$

und zudem ist für beliebige Zufallsvariablen Z_n und beliebige Konstanten $\varepsilon > 0$

$$\mathbb{P}(Z_n > \varepsilon) \leq \mathbb{P}(Z_n \geq \varepsilon),$$

was insbesondere für $Z_n = \frac{1}{n} \sum_{k=1}^n (X_k - \mathbb{E}[X_k])$ gilt.

Beweis von Satz 6.6. Für $n \in \mathbb{N}$ sei

$$Z_n := \frac{1}{n} \sum_{k=1}^n X_k.$$

Erwartungswert und Varianz von Z_n berechnen wir mit den bekannten Eigenschaften:

$$\begin{aligned} \mathbb{E}[Z_n] &= \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n X_k \right] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] = \frac{1}{n} \sum_{k=1}^n \mu = \mu \\ \text{Var}(Z_n) &= \text{Var} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) \stackrel{(*)}{=} \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{\sigma^2}{n}, \end{aligned}$$

wobei (*) dank der stochastischen Unabhängigkeit der Zufallsvariablen gilt. Durch Anwendung der Tschebyschow'schen Ungleichung (Satz 6.3) auf Z_n erhalten wir

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| \geq \varepsilon \right) = \mathbb{P} (|Z_n - \mathbb{E}[Z_n]| \geq \varepsilon) \leq \frac{\text{Var}(Z_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

□

Bemerkung 6.8. Aus den genannten Voraussetzungen folgt sogar die Gültigkeit des starken Gesetzes der großen Zahlen – dessen Beweis geht jedoch über die zur Verfügung stehenden Mittel hinaus, daher begnügen wir uns mit dieser Aussage.

Aus dem schwachen Gesetz der großen Zahlen folgt, dass bei einem wiederholten Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p die relativen Häufigkeiten gegen die tatsächliche Wahrscheinlichkeit p konvergieren.

Korollar 6.9

Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge stochastisch unabhängiger Ereignisse, die mit gleicher Wahrscheinlichkeit $\mathbb{P}(A_n) = p \in (0, 1)$ eintreten. Dann gilt für alle $\varepsilon > 0$

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{A_k} - p \right| \geq \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0.$$



Beweis. Sind die Ereignisse A_1, A_2, \dots stochastisch unabhängig, so sind die zugehörigen Indikatorfunktionen stochastisch unabhängige Zufallsvariablen mit

$$\mathbb{E}[\mathbb{1}_{A_k}] = p \quad \text{und} \quad \text{Var}(\mathbb{1}_{A_k}) = p(1-p)$$

für alle $k \in \mathbb{N}$. Die Behauptung folgt damit unmittelbar aus dem schwachen Gesetz der großen Zahlen in der Form von Satz 6.6. □

6.2.2 Exkurs: Konvergenz von Folgen von Zufallsvariablen

Sehen wir uns einmal die Konvergenz einer Folge $(X_k)_{k \in \mathbb{N}}$ von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ genauer an. Jede einzelne Zufallsvariable der Folge ist eine Funktion $X_k: \Omega \rightarrow \mathbb{R}$, d.h. es handelt sich insgesamt um eine *Funktionenfolge*. Bereits aus Analysis ist bekannt, dass es verschiedene Konvergenzbegriffe für Funktionenfolgen gibt, nämlich u.a. die punktweise und die gleichmäßige Konvergenz (siehe Definition A.16 im Anhang). Für Folgen von Zufallsvariablen haben wir u.a. folgende Konvergenzbegriffe:

Definition 6.10

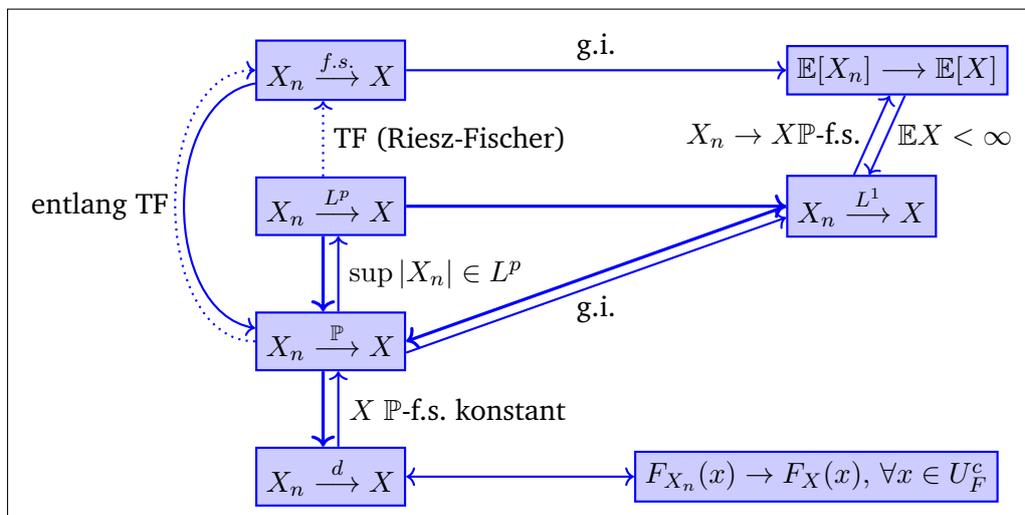
Eine Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ konvergiert gegen X

1. \mathbb{P} -fast sicher (Notation $X_n \xrightarrow{f.s.} X$), falls $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$;
2. nach Maß / stochastisch (Notation $X_n \xrightarrow{\mathbb{P}} X$), falls $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$ für alle $\varepsilon > 0$;
3. in L^p (Notation $X_n \xrightarrow{L^p} X$), falls $\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0$;
4. in Verteilung (Notation $X_n \xrightarrow{d} X$), falls \mathbb{P}_{X_n} schwach gegen \mathbb{P}_X konvergiert (Notation $\mathbb{P}_{X_n} \xrightarrow{w} \mathbb{P}_X$).

Es gilt $\mathbb{P}_{X_n} \xrightarrow{w} \mathbb{P}_X$ genau dann, wenn $F_X(x) = \lim_{n \rightarrow \infty} F_{X_n}(x)$ in allen Stetigkeitspunkten x von F .

Das *schwache Gesetz der großen Zahlen* besagt also, dass das arithmetische Mittel einer zentrierten Folge von Zufallsvariablen $\frac{1}{n} \sum_{k=1}^n (X_k - \mathbb{E}[X_k])$ *stochastisch* gegen $X \equiv 0$ konvergiert, wohingegen das am Anfang des Abschnittes genannte *starke Gesetz der großen Zahlen* die \mathbb{P} -fast sichere Konvergenz vorhersagt.

Die Zusammenhänge zwischen den Konvergenzarten können wir anhand eines Schemas verdeutlichen – alles Weitere geht über diese Vorlesung hinaus.



6.3 Zentraler Grenzwertsatz

Georg Pólya (1887-1985) stellte 1920 in seinem Aufsatz *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung* fest, dass beispielsweise bei oft wiederholten Versuchen mit sehr kleinen Fehlern immer wieder die Normalverteilungsdichte auftaucht und dass der zu Grunde liegende Satz "eine zentrale Rolle" in der Wahrscheinlichkeitsrechnung spiele. Somit setzte sich für die Sätze, die wir in diesem Abschnitt behandeln, die Bezeichnung "zentraler Grenzwertsatz" (bzw. "zentrale Grenzwertsätze") durch.

Auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ seien unabhängige Zufallsvariablen $(X_k)_{k \in \mathbb{N}}$ gegeben mit $\mu_k := \mathbb{E}[X_k]$ und $\sigma_k := \sqrt{\text{Var}(X_k)} > 0$, $k \in \mathbb{N}$. Zudem seien für $n \in \mathbb{N}$

$$S_n := \sum_{k=1}^n X_k, \quad s_n := \sqrt{\sum_{k=1}^n \sigma_k^2} \quad \text{und} \quad S_n^* := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - \mathbb{E}[S_n]}{s_n}.$$

Als *Zentraler Grenzwertsatz* (ZGWS) werden Aussagen bezeichnet, die die Frage beantworten, unter welchen Bedingungen

$$\mathbb{P}_{S_n^*} \xrightarrow{w} \mathcal{N}(0, 1) \quad (\text{ZGE})$$

gilt für $n \rightarrow \infty$; wir bezeichnen diese Eigenschaft als *zentrale Grenzwerteigenschaft*.

Für unsere Zwecke soll eine Version des ZGWS genügen.

Satz 6.11: Zentraler Grenzwertsatz

Sei $(X_k)_{k \in \mathbb{N}}$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit $\text{Var}(X_k) =: \sigma^2 \in (0, \infty)$. Seien zudem

$$S_n := \sum_{k=1}^n X_k \quad \text{und} \quad S_n^* := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n}\sigma}$$

für $n \in \mathbb{N}$. Dann gilt

$$\lim_{n \rightarrow \infty} F_{S_n^*}(x) = \Phi(x), \quad x \in \mathbb{R},$$

d.h. die Verteilungsfunktion von S_n^* konvergiert punktweise gegen die Verteilungsfunktion der Standardnormalverteilung.

Bemerkung 6.12. Unter den Voraussetzungen des Satzes gilt die Näherungsformel

$$\mathbb{P}(S_n^* \leq x) \approx \Phi(x), \quad \forall x \in \mathbb{R}.$$

Im Spezialfall $X_k \sim \text{Ber}(p)$ für $k \in \{1, \dots, n\}$ ist $S_n \sim \text{Bin}(n, p)$ und aus dem ZGWS folgt direkt der Satz von de Moivre-Laplace (Satz 4.45 bzw. Korollar 4.49).

Analog zu de Moivre-Laplace gilt auch beim ZGWS, dass im Falle diskret verteilter $(X_k)_{k \in \mathbb{N}}$ die Stetigkeitskorrektur eine bessere Approximation liefert.

Stetigkeitskorrektur

Sei $(X_k)_{k \in \mathbb{N}}$ eine Folge unabhängiger identisch verteilter \mathbb{Z} -wertiger Zufallsvariablen mit $\mathbb{E}[X_k] = \mu$ und $\text{Var}(X_k) = \sigma^2 \in (0, \infty)$ für $k \in \mathbb{N}$. Seien weiterhin $a, b \in \mathbb{Z}$ mit $a < b$ und $S_n := \sum_{k=1}^n X_k$. Ist n groß genug, so gilt näherungsweise

$$\mathbb{P}(a \leq S_n \leq b) \approx \mathbb{P}\left(a - \frac{1}{2} < Z_n < b + \frac{1}{2}\right) = \Phi\left(\frac{b + \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a - \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right)$$

mit $Z_n \sim \mathcal{N}(n\mu, n\sigma^2)$.

Beispiel 6.13. Die Bedingung, dass die Varianz nicht verschwindet, ist nicht harmlos. Seien $X_k \equiv a$ für $k \in \mathbb{N}$ und eine feste Konstante $a \in \mathbb{R}$. Dann ist offensichtlich $\mathbb{E}[X_k] = a$ und $\text{Var}(X_k) = 0$ für $k \in \mathbb{N}$. Weiterhin sind die Zufallsvariablen unabhängig (also iid). Allerdings gilt $S_n = \frac{1}{n}(X_1 + \dots + X_n) = c$ \mathbb{P} -fast sicher und $\text{Var}(S_n) = 0$, d.h. S_n^* ist gar nicht definiert. Insbesondere erfüllt $(X_k)_{k \in \mathbb{N}}$ nicht (ZGE).

Beispiel 6.14. 10 Würfel werden geworfen. Mit welcher Wahrscheinlichkeit liegt die Augensumme zwischen (einschließlich der Grenzen) 30 und 40? 

Antwort: Ist X_k die Augenzahl des k -ten Würfels, so ist für $k \in \{1, \dots, 10\}$

$$\mathbb{P}(X_k = j) = \frac{1}{6}, \quad j \in \{1, \dots, 6\}, \quad \mathbb{E}[X_k] = \frac{7}{2}, \quad \text{Var}(X_k) = \frac{35}{12}.$$

Somit gilt für die Augensumme $S_{10} := \sum_{k=1}^{10} X_k$

$$\mathbb{P}(30 \leq S_{10} \leq 40) \approx$$

Beispiel 6.15. Seien X_1, \dots, X_n stochastisch unabhängige Zufallsvariablen mit $X_k \sim U(0, 1)$ für alle $k \in \{1, \dots, 10\}$. Wie groß ist näherungsweise die Wahrscheinlichkeit 

$$\mathbb{P}\left(\sum_{k=1}^{10} X_k \geq 6\right)?$$

Antwort: Da die Zufallsvariablen bereits absolutstetig verteilt sind, ist keine Stetigkeitskorrektur nötig. Für $k \in \{1, \dots, 10\}$ gilt nach Satz 4.26

$$\mathbb{E}[X_k] = \frac{1}{2} \quad \text{und} \quad \text{Var}(X_k) = \frac{1}{12}.$$

Folglich gilt

$$\begin{aligned} \mathbb{P}\left(\sum_{k=1}^{10} X_k \geq 6\right) &= \\ &\approx 1 - \Phi\left(\quad\right) \\ &\approx 1 - \Phi(1,095) \\ &\approx \end{aligned}$$

Beispiel 6.16. Ein Astronom möchte die Entfernung eines weit entfernten Sterns bestimmen. Er weiß, dass seine Messungen keine exakten Werte liefern. Deshalb macht er eine Reihe von unabhängigen Messungen X_1, \dots, X_n und nimmt deren Mittelwert. Der Astronom nimmt an, dass X_1, \dots, X_n identisch verteilt sind mit Erwartungswert d und der Varianz 4 Lichtjahre. Wie viele Messungen muss er durchführen, um die Distanz d mit 95%-iger Wahrscheinlichkeit auf $\pm 0,5$ Lichtjahre sicher bestimmen zu können? 

Antwort: Laut Text ist für $k \in \{1, \dots, n\}$

$$\mathbb{E}[X_k] = d \quad \text{und} \quad \text{Var}(X_k) = 4.$$

Setzen wir $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$, so ist gemäß dem ZGWS

$$S_n^* := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{1}{2\sqrt{n}}(n\bar{X}_n - nd)$$

approximativ $\mathcal{N}(0, 1)$ -verteilt. Wir wollen $n \in \mathbb{N}$ derart bestimmen, dass

$$\mathbb{P}(|\bar{X}_n - d| \leq 0,5) \stackrel{!}{=} 0,95$$

gilt.

$$\begin{aligned} \mathbb{P}(|\bar{X}_n - d| \leq 0,5) &= \mathbb{P}\left(\left|\sum_{k=1}^n X_k - nd\right| \leq \frac{n}{2}\right) \\ &= \mathbb{P}\left(|S_n^*| \leq \right) \approx \end{aligned}$$

Damit muss gelten:

Der Astronom muss also mindestens 62 Messungen durchführen um die gewünschte Genauigkeit und Sicherheit zu haben (wobei wir die zusätzliche Ungenauigkeit durch die Approximation durch die Normalverteilung vernachlässigen).

Kapitel 7

Elemente der Statistik

Ziele

- Maximum-Likelihood-Prinzip verstehen
- Konfidenzintervalle (KI) bestimmen
- Hypothesentests durchführen und auswerten
- Fehler 1. und 2. Art bei Hypothesentests bestimmen
- die unterschiedlichen Herangehensweisen beim Hypothesentest und dem KI kennen
- empirische Kenngrößen (Stichprobenmittel, -varianz) kennen
- lineare Regression durchführen
- Daten durch Kennzahlen, Histogramme und Boxplots darstellen

Die Stochastik hat die Teilgebiete *Wahrscheinlichkeitstheorie* und *Statistik*. Bisher haben wir uns mit Themen der Wahrscheinlichkeitstheorie befasst, wobei wir basierend auf einem Modell gewisse Eigenschaften von Daten vorhersagen. Die Statistik hingegen beschreibt vorhandene Daten (*deskriptive Statistik*) und leitet aus den Daten geeignete Modelle und damit wiederum Schlussfolgerungen (wie Vorhersagen) ab (*induktive Statistik*).

7.1 Schätzen von unbekanntem Parametern

Bisher sind wir in der Regel davon ausgegangen, dass wir den Verteilungstyp einer Zufallsvariablen und die zugehörigen Parameter kennen. In diesem Abschnitt geht es darum, bei bekannter Verteilung einen unbekanntem Parameter auf Grund einer Stichprobe zu schätzen. Sehen wir uns zunächst ein Beispiel dazu an.

Beispiel 7.1. Es soll der Anteil p der Bürger geschätzt werden, die die Blutgruppe 0 besitzen. Dazu betrachten wir eine Stichprobe von n Personen, d.h. n voneinander unabhängig rein zufällig gewählte Personen. Wenn in der Stichprobe k Personen mit Blutgruppe 0 registriert werden, wie groß ist dann vermutlich der gesuchte Anteil p ? 

Antwort: Sei S_n die Anzahl der Personen in der Stichprobe, die Blutgruppe 0 besitzen, d.h. $S_n \sim \text{Bin}(n, p)$ mit dem unbekanntem Parameter $p \in [0, 1]$. Wenn p der wahre Parameter ist, so ist also

$$L_k(p) := \mathbb{P}_p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Wir suchen den Wert für $p \in [0, 1]$, der diese Wahrscheinlichkeit maximiert, d.h. wir suchen den Maximierer von $p \mapsto L_k(p)$ für ein festes k . Wir untersuchen zunächst die Randfälle:

- Ist $k = 0$, so ist $L_k(p) =$ maximal für $\hat{p} :=$
- Ist $k = n$, so ist $L_k(p) =$ maximal für $\hat{p} :=$

Sei nun $k \in \{1, \dots, n-1\}$. Wir leiten die Funktion L_k nach p ab und setzen sie gleich Null:

$$L'_k(p) =$$

$$\stackrel{!}{=} 0.$$

Da $\binom{n}{k} p^{k-1} (1-p)^{n-k-1} > 0$ ist für $p \in (0, 1)$ und $k \in \{1, \dots, n-1\}$, ist diese Bedingung genau dann erfüllt, wenn

$$\iff p = \frac{k}{n}$$

ist. Überprüfen wir noch die zweite Ableitung:

$$L''_k(p) =$$

$$= \binom{n}{k} \left(k(k-1)p^{k-2}(1-p)^{n-k} - 2k(n-k)p^{k-1}(1-p)^{n-k-1} \right. \\ \left. + (n-k)(n-k-1)p^k(1-p)^{n-k-2} \right).$$

Einsetzen des kritischen Wertes $\hat{p} := \frac{k}{n}$ liefert

$$L''_k(\hat{p}) = \binom{n}{k} \cdot \left(\frac{1}{n} \right)^{n-2} \cdot \left[k(k-1)k^{k-2}(n-k)^{n-k} - 2k(n-k)k^{k-1}(n-k)^{n-k-1} \right. \\ \left. + (n-k)(n-k-1)k^k(n-k)^{n-k-2} \right]$$

Folglich ist in allen Fällen $L_k(\hat{p}) = \max \{ L_k(p) \mid p \in [0, 1] \}$, d.h. für $\hat{p} = \frac{k}{n}$ ist die Wahrscheinlichkeit am größten, k Personen in der Stichprobe mit Blutgruppe 0 zu beobachten.

Bemerkung 7.2. Der Wert $\hat{p} = \frac{k}{n}$ ist eine Realisierung der Zufallsvariablen $R_n := \frac{S_n}{n}$, der relativen Häufigkeit der Personen der Blutgruppe 0 innerhalb der Stichprobe der Größe n . Es gilt

$$\mathbb{E}_p[R_n] = \frac{1}{n} \mathbb{E}_p[S_n] = \frac{1}{n} \cdot np = p.$$

Somit ist \hat{p} ein erwartungstreuer Schätzer für den unbekannt Parameter p . Zudem gilt

$$\text{Var}_p(R_n) = \frac{1}{n^2} \text{Var}(S_n) = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n} \xrightarrow{n \rightarrow \infty} 0,$$

d.h. durch Vergrößern der Stichprobe kann p durch \hat{p} beliebig genau geschätzt werden. Die Folge der Schätzer $(\hat{p}_n)_{n \in \mathbb{N}}$ mit $\hat{p}_n = R_n$ bezeichnet man in diesem Fall als konsistent.

Halten wir das in dem Beispiel verwendete Prinzip und die genannten Begriffe einmal fest.

Definition 7.3: Schätzer

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $X: (\Omega, \mathcal{A}) \rightarrow (X(\Omega), \mathcal{B})$ (mit $X(\Omega) \subseteq \mathbb{R}$) eine Zufallsvariable. Sei weiterhin $\Theta \subseteq \mathbb{R}$ eine Parametermenge und damit $(X(\Omega), \mathcal{B}, \mathbb{P}_\vartheta)$ für $\vartheta \in \Theta$ eine Familie von Wahrscheinlichkeitsräumen, genannt *statistisches Modell*. Dann gilt:

1. Der unbekannte Wert $\vartheta \in \Theta$ ist der zu schätzende Parameter.
2. Jede Funktion $T: X(\Omega) \rightarrow \Theta$ heißt *Schätzer* für $\vartheta \in \Theta$.
3. Für $x \in X(\Omega)$ ist $T(x) \in \Theta$ ein konkreter Schätzwert für ϑ .

Insbesondere ist ein Schätzer zunächst einmal eine Funktion von der Menge der möglichen Beobachtungen, d.h. dem *Stichprobenraum*, in die Parametermenge Θ . Im Beispiel ist

- $X(\Omega) = S_n(\Omega) = \{0, \dots, n\}$,
- $T(x) = \frac{x}{n}$ für alle $x \in X(\Omega)$,
- $\Theta = [0, 1]$ mit $\vartheta = p$.

Soll der Parameter $\vartheta \in \Theta$ auf Grund der Realisierung(en) einer Zufallsvariablen X mit Hilfe eines Schätzers T berechnet werden, so bezeichnet man auch die Zufallsvariable $\hat{\vartheta} := T(X)$ als *Schätzer* für ϑ . 

Da wir den wahren Parameter nicht kennen, müssen Wahrscheinlichkeiten, Erwartungswerte und Varianzen in Abhängigkeit des Parameters $\vartheta \in \Theta$ formulieren. Dazu verwenden wir folgende Notation:

Notation

Für $\vartheta \in \Theta$ notieren wir das Wahrscheinlichkeitsmaß, welches zum Parameter ϑ gehört, als $\mathbb{P}_\vartheta(\cdot)$ und der zugehörige Erwartungswert und die Varianz analog als $\mathbb{E}_\vartheta[\cdot]$ bzw. $\text{Var}_\vartheta(\cdot)$.

Idee des Maximum-Likelihood-Schätzers: Stehen verschiedene Modelle $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ zur Auswahl, so halten wir bei vorliegenden Daten dasjenige Modell für das glaubwürdigste, unter welchem die beobachteten Daten die größte Wahrscheinlichkeit des Auftretens besitzen.

Definition 7.4: Maximum-Likelihood-Schätzer

Sei $(X(\Omega), \mathcal{B}, \mathbb{P}_\vartheta)$ für $\vartheta \in \Theta$ ein statistisches Modell und sei

- $L(\vartheta, x) := p_\vartheta(X = x)$, falls X diskret verteilt ist mit Massenfunktion p_ϑ unter dem Parameter ϑ ;
- $L(\vartheta, x) := f_\vartheta(x)$, falls X absolutstetig verteilt ist mit Dichte f_ϑ unter ϑ .

Dann heißt $L: \Theta \times X(\Omega) \rightarrow \mathbb{R}_+$ *Likelihood-Funktion* und ein Schätzer $\hat{\vartheta}$ heißt *Maximum-Likelihood-Schätzer*, falls

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

für (fast) alle $x \in X(\Omega)$.

Im Beispiel ist (mit $\vartheta = p$)

- $L(p, x) = \mathbb{P}_p(S_n = x) = \binom{n}{x} p^x (1-p)^{n-x}$ für $x \in S_n(\Omega) = \{0, \dots, n\}$;
- $\hat{p} = \frac{S_n}{n}$ ist der Maximum-Likelihood-Schätzer für den Parameter p ;
- $\hat{p} = \frac{k}{n}$ ist der Wert des Maximum-Likelihood-Schätzers bei beobachteter Realisierung $S_n(\omega) = k$.

Zudem heißt ein Schätzer *erwartungstreu*, falls der Erwartungswert des Schätzers angewandt auf die Zufallsvariable berechnet unter der Annahme, dass ϑ der wahre Parameter ist, stets gleich ϑ ist. Man spricht auch von einem *unverzerrten* Schätzer.

Definition 7.5: erwartungstreuer Schätzer

Sei $(X(\Omega), \mathcal{B}, \mathbb{P}_\vartheta)$ für $\vartheta \in \Theta$ ein statistisches Modell. Ein Schätzer $\hat{\vartheta}$ heißt *erwartungstreu*, falls gilt:

$$\mathbb{E}_\vartheta[\hat{\vartheta}] = \vartheta, \quad \forall \vartheta \in \Theta.$$

Im Beispiel ist $\mathbb{E}_p[\hat{p}] = \mathbb{E}_p\left[\frac{S_n}{n}\right] = p$ für alle $p \in [0, 1]$, d.h. \hat{p} ist ein erwartungstreuer Schätzer.

Sehen wir uns nun ein weiteres Beispiel für einen Maximum-Likelihood-Schätzer an.



Beispiel 7.6. Um die Anzahl der Fische in einem Teich (N) zu schätzen, werden K Fische gefangen, markiert und wieder ausgesetzt. Anschließend wartet man bis eine gute Vermischung der markierten Fische angenommen werden kann und fängt dann n Fische. Falls sich unter diesen k markierte Fische befinden, wie schätzt man N mit Hilfe der Maximum-Likelihood-Methode?

Antwort: Sei X die Anzahl der markierten Fische unter den n gefangenen. Dann gilt

$$X \sim$$

wobei N unbekannt ist, d.h. der Parameterraum ist

$$\Theta =$$

Die zu maximierende Likelihood-Funktion ist

$$L(N, k) = \mathbb{P}_N(X = k) =$$

Da N nur Werte in (einer Teilmenge von) \mathbb{N} annehmen kann, finden wir das Optimum indem wir vorgehen wie bereits in Aufgabe 2 von Übungsblatt 6 – wir betrachten die Quotienten zweier aufeinanderfolgender Wert der Likelihood-Funktion.

$$\frac{L(N, k)}{L(N-1, k)} =$$

Die Likelihood-Funktion ist genau dann wachsend zwischen $N-1$ und N , wenn dieser Quotient größer ist als 1. Das ist genau dann der Fall, wenn gilt:

$$\iff \frac{Kn}{k} > N.$$

Die Likelihood-Funktion wächst also bis zum Wert $\hat{N} := \lfloor \frac{Kn}{k} \rfloor$ und es gilt insbesondere $L(\hat{N}, k) > L(\hat{N}-1, k)$. Für alle $N > \hat{N}$ ist die Funktion fallend, d.h. es gilt insbesondere $L(\hat{N}+1, k) < L(\hat{N}, k)$. Damit haben wir insgesamt

$$\hat{N} = \left\lfloor \frac{Kn}{k} \right\rfloor = \arg \max_{N \in \Theta} L(N, k).$$

Bemerkung 7.7. Alternativ zur Betrachtung des Quotienten $\frac{L(N, k)}{L(N-1, k)} \stackrel{\geq}{\leq} 1$ hätte man berechnen können, für welche $N \in \Theta$ die Differenz positiv/negativ ist, d.h. ob $L(N, k) - L(N-1, k) \stackrel{\geq}{\leq} 0$ ist.

Halten wir zuletzt noch einmal das Ergebnis aus Beispiel 5.44 in Hinblick auf die neuen Begriffe fest.

Beispiel 7.8 (Schätzer für Erwartungswert und Varianz). Seien X_1, \dots, X_n stochastisch unabhängige und identisch verteilte (d.h. i.i.d.) Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Seien $\mu := \mathbb{E}[X_1]$ der Erwartungswert und $\sigma^2 := \text{Var}(X_1)$ die Varianz dieser Zufallsvariablen. Dann sind erwartungstreue Schätzer von μ und σ^2 gegeben durch das arithmetische Mittel

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

und die Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

7.2 Konfidenzintervalle

Im vorherigen Abschnitt haben wir Schätzer für unbekannte Modellparameter kennen gelernt. Allerdings ist der konkrete Schätzwert von der Realisierung der Stichprobe ab.

Beispiel 7.9. Der Erwartungswert einer Zufallsvariablen X soll durch die Realisierung des arithmetischen Mittels geschätzt werden. Eine erste Stichprobe liefert die Werte

$$41, 38, 40, 44, 38, 39, 37, 40, 43, 39.$$

In diesem Fall ist das arithmetische Mittel und damit der Schätzer für den Erwartungswert

$$\hat{\mu}_1 = \frac{1}{10} \sum_{k=1}^{10} x_k = 39,9.$$

In einer anderen Stichprobe erhält man die Werte

$$37, 41, 42, 40, 44, 41, 38, 42, 44, 41.$$

In diesem Fall ist das arithmetische Mittel und damit der Schätzer für den Erwartungswert

$$\hat{\mu}_2 = \sum_{k=1}^{10} x_k = 41,0.$$

Der wahre Parameter wird vermutlich irgendwo in der Nähe des Schätzwertes liegen. Die Idee hinter dem Konfidenzintervall ist es, ein Intervall um den Schätzwert herum anzugeben, in dem der wahre Parameter mit einer vorgegebenen Wahrscheinlichkeit liegt.



Beispiel 7.10. Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\text{Var}(X) = \sigma^2 = \frac{1}{4}$ aber unbekanntem Parameter $\mu \in \mathbb{R}$. Eine Stichprobe vom Umfang $n = 50$ ergibt den Mittelwert $\bar{x} = 3,6$. In welchem Intervall liegt der wahre Parameter μ mit Wahrscheinlichkeit 90%?

Antwort: Der Einfachheit wählen wir das Intervall symmetrisch um das Stichprobenmittel. Dann muss folgende Bedingung erfüllt sein:

$$\mathbb{P}_\mu(|\bar{X}_{50} - \mu| \leq \gamma) = 0,9.$$

Wir suchen das kleinstmögliche $\gamma \in \mathbb{R}$, welches diese Bedingung erfüllt. Da die Zufallsvariablen X_1, \dots, X_n alle stochastisch unabhängig und $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind, gilt

$$\bar{X}_{50} = \frac{1}{50} \sum_{k=1}^{50} X_k \sim$$

Insbesondere ist damit

$$\frac{\bar{X}_n - \mu}{\sigma_{\bar{X}_n}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = 2\sqrt{50} \cdot (\bar{X}_{50} - \mu) \sim \mathcal{N}(0, 1).$$

Um das gesuchte Intervall zu bestimmen, berechnen wir zunächst die angegebene Wahrscheinlichkeit in Abhängigkeit von γ und setzen diese dann gleich 0,9.

$$\mathbb{P}_\mu(|\bar{X}_{50} - \mu| \leq \gamma) =$$

Diese Gleichung ist genau dann erfüllt, wenn Folgendes gilt:

$$\Phi\left(2\sqrt{50}\gamma\right) = 0,95 \quad \iff$$

Damit lautet das Konfidenzintervall für den Parameter μ bei den gemachten Beobachtungen zur Wahrscheinlichkeit 90%

$$\begin{aligned} KI &= [\bar{x} - \gamma, \bar{x} + \gamma] \\ &= [3,6 - 0,12; 3,6 + 0,12] \\ &= [3,48; 3,72]. \end{aligned}$$

Fassen wir die Ergebnisse dieses Beispiels für ein etwas allgemeineres Setting zusammen.

Satz 7.11

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit unbekanntem Erwartungswert μ und bekannter Varianz σ^2 . Seien X_1, \dots, X_n stochastisch unabhängige und identisch wie X verteilte Zufallsvariablen. Das Konfidenzintervall für den Parameter μ zum Konfidenzniveau (Vertrauensniveau) $1 - \alpha$ ist gegeben durch

$$KI = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{X}_n + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right].$$

Bemerkung 7.12. KI ist das gesuchte Konfidenzintervall, falls

$$\mathbb{P}_\mu(\mu \in KI) = 1 - \alpha \quad (7.2.1)$$

ist. Hierbei ist μ zwar unbekannt, aber keine Zufallsvariable. Wieso sollte das Ereignis $\{\mu \in KI\}$ also vom Zufall abhängen?

Antwort: Das Konfidenzintervall KI selbst hängt vom Zufall ab, denn die beiden Grenzen des Intervalls sind Funktionen der Zufallsvariablen \bar{X}_n . Es handelt sich also um ein Intervall fester Länge, welches um die Zufallsgröße \bar{X}_n herum zentriert ist.

Beweis. Mit $\mathbb{E}_\mu[\bar{X}_n] = \mu$ und $\text{Var}_\mu(\bar{X}_n) = \frac{\sigma^2}{n}$ gilt

$$\begin{aligned} \mathbb{P}_\mu(\mu \in KI) &= \mathbb{P}_\mu\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \\ &= \mathbb{P}_\mu\left(-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \frac{\mu - \bar{X}_n}{\frac{\sigma}{\sqrt{n}}} \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \\ &= \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) - \Phi\left(-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \\ &= \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} \\ &= 1 - \alpha. \end{aligned}$$

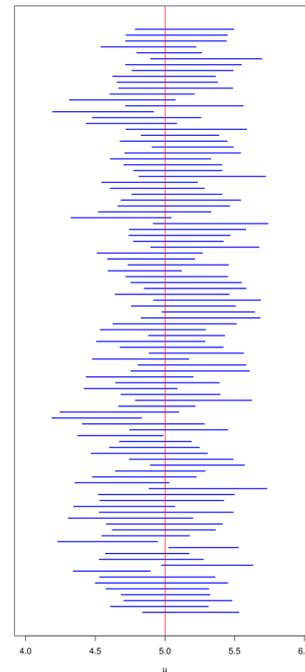
□

Analog kann man auch Konfidenzintervalle für nicht normalverteilte Stichprobenmittel oder andere Parameter ermitteln. Wichtig ist dabei, dass man zunächst die Verteilung des Schätzers kennt und die analoge Bedingung zu (7.2.1) aufstellt. Kann man diese nach dem unbekanntem Parameter umstellen, so erhält man das gewünschte Konfidenzintervall (oder allgemeiner die Konfidenzmenge).

Auf der rechts abgebildeten Grafik sind die Konfidenzintervalle zum Niveau 95% für 100 Stichproben vom Umfang 30 aus einer normalverteilten Grundgesamtheit abgebildet. Davon überdecken 94 Intervalle den exakten Erwartungswert $\mu = 5$; die übrigen 6 tun das nicht.

Quelle:

Von HilberTraum – Eigenes Werk, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=31793077>



7.3 Hypothesentests

Statistische Tests basieren auf einem zunächst festgelegten statistischen Modell $(X(\Omega), \mathcal{B}, \mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ (vgl. Definition 7.3) und einer Partition des Parameterraumes in zwei disjunkte Teilmengen:

$$\Theta = \Theta_0 \dot{\cup} \Theta_1.$$

Im Gegensatz zu der Situation bei der Berechnung des Konfidenzintervalls gibt es bereits eine Hypothese, deren Gültigkeit mit einer gewissen Sicherheit überprüft werden soll.

Definition 7.13

Ein *Signifikanztest* ist eine Entscheidungsregel, die für jede mögliche Realisierung einer Stichprobe X_1, \dots, X_n festlegt, ob man die *Nullhypothese*

$$H_0: \vartheta \in \Theta_0$$

beibehält oder zu Gunsten der *Alternative*

$$H_1: \vartheta \in \Theta_1$$

ablehnt.

Je nach Testergebnis wird die Nullhypothese also beibehalten oder abgelehnt, was je nachdem, in welchem Teilintervall der wahre Parameter liegt, eine richtige oder falsche Entscheidung ist. Die Möglichkeiten kann man tabellarisch verdeutlichen:

	$\vartheta \in \Theta_0$	$\vartheta \in \Theta_1$
H_0 beibehalten	richtige Entscheidung	Fehler 2. Art
H_0 ablehnen	Fehler 1. Art	richtige Entscheidung

Man kann in der Praxis nicht sowohl den Fehler 1. Art als auch den Fehler 2. Art kontrollieren. Daher legt man Nullhypothese und Alternative so fest, dass der Fehler 1. Art der „gravierendere“ Fehler ist; dessen Wahrscheinlichkeit wird im Anschluss beim Testdesign nach oben beschränkt.



Beispiel 7.14. Ein Pilzsammler findet einen Pilz, der giftig sein könnte. Welchen Fehler kann der Pilzsammler bei der Überprüfung der Hypothese



„Der Pilz ist giftig.“

begehen? Welchen Fehler wird er versuchen zu vermeiden? Wie müssen demnach H_0 und H_1 gewählt werden?

Antwort: Die möglichen Fehler sind:

(A)

(B)

Fehler sollte minimiert werden, d.h. das wird der Fehler 1. Art. Folglich ist

H_0 : „Der Pilz ist “

H_1 : „Der Pilz ist “

Die Daten, auf deren Grundlage wir unsere Entscheidung fällen wollen, sind in der Regel gegeben durch eine Stichprobe, d.h. gegeben wird uns eine Realisierung von n stochastisch unabhängigen und identisch verteilten (d.h. i.i.d.) Zufallsvariablen X_1, \dots, X_n . Von diesen kennen wir in unserem Setting den Verteilungstyp, jedoch nicht den bzw. nicht alle Parameter der Verteilung.

Um zu einer Entscheidungsregel gelangen zu können, müssen wir eine Testgröße als Funktion der Stichprobe definieren, deren Verteilung wir kennen. Diese Testgröße notieren wir als $T(X_1, \dots, X_n)$. Wir zerlegen dann den Wertebereich dieser Testfunktion in zwei disjunkte Teilmengen:

- den kritischen Bereich \mathcal{K}_1 ,
- den Annahmehereich $\mathcal{K}_0 = (\mathcal{K}_1)^c$.

Die Entscheidungsregel wird dann wie folgt definiert:

Entscheidungsregel

Ist (x_1, \dots, x_n) eine Realisierung der Stichprobe (X_1, \dots, X_n) , so entscheide bei gegebenem kritischen Bereich \mathcal{K}_1 wie folgt:

- Ist $T(x_1, \dots, x_n) \in \mathcal{K}_1$, so lehne H_0 zu Gunsten von H_1 ab;
- ist $T(x_1, \dots, x_n) \in \mathcal{K}_0$, so behalte H_0 bei.

Man kann keinen der möglichen Fehler komplett ausschließen, aber in dem gegebenen Modell kann man den Test so kalibrieren, dass die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, begrenzt ist. Die Obergrenze für diese Wahrscheinlichkeit, welche (willkürlich) festgesetzt wird, nennt man *Signifikanzniveau* und bezeichnet es mit $\alpha \in (0, 1)$.

Signifikanztest

Ein Hypothesentest

$$H_0: \vartheta \in \Theta_0 \quad \text{gegen} \quad H_1: \vartheta \in \Theta_1$$

ist ein Test zum *Signifikanzniveau* $\alpha \in (0, 1)$, falls der kritische Bereich \mathcal{K}_1 derart gewählt ist, dass gilt:

$$\mathbb{P}_\vartheta (T(X_1, \dots, X_n) \in \mathcal{K}_1) \leq \alpha, \quad \forall \vartheta \in \Theta_0. \quad (7.3.1)$$

Ein Test zum Signifikanzniveau α kontrolliert nur den Fehler 1. Art, d.h. bei wiederholter Durchführung des Tests und bei Gültigkeit der Nullhypothese (das wissen wir nicht!) würden wir uns nur in $\alpha \cdot 100\%$ der Fälle gegen die Nullhypothese entscheiden. Den Fehler 2. Art können wir nicht kontrollieren – er hängt von dem unbekanntem Parameter ϑ ab. Wird H_0 nicht verworfen, so bedeutet das lediglich, dass die vorliegende Beobachtung nicht im Widerspruch zur Nullhypothese steht. Die Gültigkeit der Nullhypothese wurde keinesfalls bewiesen!

Bemerkung 7.15. *In der Literatur wird teilweise nicht davon gesprochen, dass H_0 angenommen wird, sondern dass H_0 auf Grund der Daten nicht abgelehnt werden kann.*

Ist weder der Fehler 1. Art noch der Fehler 2. Art „gravierender“, so begründet die gerade gemachte Überlegung eine weitere Herangehensweise, was als Nullhypothese und was als Alternative festgelegt wird:



Will man eine Behauptung mit einer gewissen Sicherheit treffen können, so wird die gegenteilige Behauptung zur Nullhypothese mit dem Ziel, sie mit einer vorgegebenen maximalen Irrtumswahrscheinlichkeit ablehnen zu können.



Beispiel 7.16. *Firma A behauptet, dass ihr Produkt länger hält als das Konkurrenzprodukt von Firma B. Wie sollte sie Nullhypothese und Alternative bei einem Hypothesentest aufstellen, um mit einer Fehlerwahrscheinlichkeit von 5% verkünden zu können, dass das eigene Produkt besser ist?*

Antwort: Ziel ist es, die Aussage

ablehnen zu können. Folglich ist das die Nullhypothese und die Alternative ist

Bevor wir einen konkreten Hypothesentest durchführen, fassen wir das Vorgehen zusammen:

Schema eines Hypothesentests

1. Verteilungsannahme an die Stichprobe formulieren, Parameterraum festlegen;
2. Hypothesen formulieren und Signifikanzniveau (=Irrtumswahrscheinlichkeit) festlegen bzw. aus Vorgaben herauslesen;
3. Testgröße $T(X_1, \dots, X_n)$, deren Verteilung unter H_0 bekannt ist, konstruieren;
4. (größtmöglichen) kritischen Bereiches \mathcal{K}_1 festlegen, so dass (7.3.1) erfüllt ist;
5. Entscheidungsregel formulieren;
6. Realisierung der Testgröße $T(x_1, \dots, x_n)$ aus der Stichprobe berechnen;
7. Entscheidung fällen.

Beispiel 7.17. Eine Abfüllmaschine für Wasserflaschen ist so konstruiert, dass die Abfüllmenge als annähernd $\mathcal{N}(\mu, \sigma^2)$ -verteilt angenommen werden kann mit $\sigma^2 = 4$. Eine Verbraucherorganisation möchte anhand einer Stichprobe von n Flaschen nachweisen, dass die vom Hersteller angegebene mittlere Füllmenge von 1000 ml signifikant unterschritten wird. Bei einer Stichprobe aus $n = 50$ Flaschen beträgt der Mittelwert der Füllmengen $\bar{x} = 999,5$ ml. Unterstützt das Testergebnis die Behauptung der Verbraucherorganisation zur Sicherheit von 99%?

Antwort: Eine Stichprobe besteht aus stochastisch unabhängigen und identisch $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariablen X_1, \dots, X_{50} . Der interessierende Parameter ist $\vartheta = \mu$ und der Parameterraum¹ ist $\Theta = \mathbb{R}$.

Sollte fälschlicherweise die Füllmenge als ausreichend bezeichnet werden, wäre das nicht so gravierend wie eine öffentliche Anklage des Herstellers, die sich später als unbegründet herausstellt. Wir wählen also

$$H_0: \mu \geq \mu_0 = 1000, \quad H_1: \mu < \mu_0.$$

Somit ist $\Theta_1 = (-\infty, 1000)$ und $\Theta_0 = [1000, \infty)$ und $\alpha = 0,01$ ist laut Text die Irrtumsw'keit.

Der erste Kandidat für unsere Testgröße ist in unserem Fall $\tilde{T}(X_1, \dots, X_n) := \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, da das arithmetische Mittel der ML-Schätzer für den Erwartungswert μ ist. Allerdings ist dann $\tilde{T}(X_1, \dots, X_n) \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, d.h. wir können die Werte der Verteilungsfunktion nicht direkt berechnen. Wir müssen diesen Ausdruck also noch normieren:

$$T(X_1, \dots, X_n) = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}\left(\sqrt{n} \frac{\mu - \mu_0}{\sigma}, 1\right).$$

Unter der Annahme $\mu = \mu_0$ ist also $T(X_1, \dots, X_n) \sim \mathcal{N}(0, 1)$.

Desto kleiner μ ist, desto kleiner sollte auch $T(X_1, \dots, X_n)$ sein. Folglich müssen wir $\mathcal{K}_1 = (-\infty, c)$ und $\mathcal{K}_0 = [c, \infty)$ setzen mit einem noch zu bestimmenden (kritischen) Punkt c .

Wir wollen die W'keit eines Fehlers 1. Art durch die W'keit $\alpha = 1\%$ beschränken. Wir nehmen also an, dass H_0 gilt (d.h. $\mu \geq \mu_0$) und beschränken die W'keit, dass H_0 abgelehnt wird, da $T(X_1, \dots, X_n) \in \mathcal{K}_1$ ist, also kleiner als c . Fassen wir das in Formeln zusammen:

$$\mathbb{P}_\mu(T(X_1, \dots, X_n) < c) \stackrel{!}{\leq} \alpha, \quad \forall \mu \geq \mu_0.$$

Für $\mu \neq \mu_0$ können wir diese Wahrscheinlichkeit, zwar nicht berechnen, aber es gilt die Abschätzung

$$\mathbb{P}_\mu(T(X_1, \dots, X_n) < c) \leq \mathbb{P}_{\mu_0}(T(X_1, \dots, X_n) < c) = \Phi(c), \quad \forall \mu \geq \mu_0.$$

Schätzen wir diese Wahrscheinlichkeit durch $\alpha = 0,01$ ab, so erhalten wir durch Umstellen den gewünschten kritischen Wert:

$$\Phi(c) \leq \alpha \iff c \leq \Phi^{-1}(\alpha) = -\Phi^{-1}(1 - \alpha) = -\Phi^{-1}(0,99) \approx -2,33.$$

Der kritische Bereich ist damit $\mathcal{K}_1 = (-\infty, -2,33)$, d.h. die Entscheidungsregel lautet:

- ist $T(x_1, \dots, x_n) < -2,33$, so verwirfe H_0 ;
- ist $T(x_1, \dots, x_n) \geq -2,33$, so behalte H_0 bei.

Nun werten wir die Stichprobe aus: $\bar{x} = 999,5$ bei einer Stichprobengröße von $n = 50$, d.h.

$$T(x_1, \dots, x_n) = \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma} = \sqrt{50} \cdot \frac{999,5 - 1000}{2} \approx -1,77 > -2,33.$$

Da dieser Wert nicht im kritischen Bereich (=Ablehnbereich) liegt, behalten wir H_0 bei. Das Testergebnis unterstützt also nicht die Behauptung der Verbraucherorganisation.

¹Inhaltlich kann es keine negativen Füllmengen geben, d.h. $\mu < 0$ ist nicht sinnvoll. Allerdings wurde die Füllmenge als normalverteilt angenommen, so dass dies rein mathematisch möglich ist. Daher ist $\Theta = \mathbb{R}$ statt $\Theta = \mathbb{R}_+$.

7.3.1 Typen von Tests

Die Parameter der $\mathcal{N}(\mu, \sigma^2)$ -Verteilung repräsentieren gerade Erwartungswert und Varianz, so dass sich die Techniken von Hypothesentests daran gut illustrieren lassen. In den nächsten drei Abschnitten geht es also jeweils um die Parameter μ und σ^2 , wobei jedoch wahlweise einer von beiden oder gar beide unbekannt sind.

Gaußtest

Die Stichprobe X_1, \dots, X_n bestehe aus stochastisch unabhängigen und identisch verteilten Zufallsvariablen mit $X_k \sim \mathcal{N}(\mu, \sigma^2)$ für $k \in \{1, \dots, n\}$, wobei μ unbekannt und $\sigma > 0$ bekannt (und unabhängig von μ) sei. Damit ist der Parameterraum im Folgenden $\Theta = \mathbb{R}$.

Für die Hypothesen gibt es verschiedene Möglichkeiten, die wir jetzt systematisch durchgehen werden, nämlich

- **zweiseitiger Test:** $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$
- **rechtsseitiger Test:** $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$
- **linksseitiger Test:** $H_0: \mu \geq \mu_0$ vs. $H_1: \mu < \mu_0$

In allen drei Fällen ist das Stichprobenmittel $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ ein erwartungstreu-er Schätzer für den unbekannt Parameter μ . Als Testgröße verwenden wir die unter $\mu = \mu_0$ standardnormalverteilte Zufallsgröße

$$T(X_1, \dots, X_n) := \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}\left(\sqrt{n} \frac{\mu - \mu_0}{\sigma}, 1\right).$$

Den kritischen Bereich \mathcal{K}_1 müssen wir so festlegen, dass folgende Bedingung erfüllt ist:

$$\mathbb{P}_{\mu_0}(T(X_1, \dots, X_n) \in \mathcal{K}_1) = \alpha.$$

Beim zweiseitigen Test ist dies genau die Wahrscheinlichkeit eines Fehlers 1. Art und beim einseitigen Test ist es eine obere Schranke für den Fehler 1. Art, so dass in allen Fällen ein Signifikanztest zum Signifikanzniveau α konstruiert wird. Wir werden diese Eigenschaft für alle drei Fälle nachweisen und den kritischen Bereich konstruieren.

Zweiseitiger Test: Es ist definitionsgemäß

$$\mathbb{P}_{H_0}(T(X_1, \dots, X_n) \in \mathcal{K}_1) = \mathbb{P}_{\mu_0}(T(X_1, \dots, X_n) \in \mathcal{K}_1) = \alpha$$

genau die Wahrscheinlichkeit eines Fehlers 1. Art. Unter H_0 (d.h. angenommen $\mu = \mu_0$ ist der wahre Parameter) ist

$$T(X_1, \dots, X_n) \sim \mathcal{N}(0, 1),$$

d.h. die Realisierung der Testgröße sollte in der Nähe des Wertes Null liegen. Als Annahemebereich wählen wir ein symmetrisches Intervall um Null herum und erhalten somit den kritischen Bereich

$$\mathcal{K}_1 = \mathbb{R} \setminus [-k, k] = (-\infty, -k) \cup (k, \infty),$$

wobei k derart zu bestimmen ist, dass

$$\begin{aligned} \alpha &\stackrel{!}{=} \mathbb{P}_{\mu_0}(T(X_1, \dots, X_n) \in \mathcal{K}_1) \\ &= \mathbb{P}_{\mu_0}(|T(X_1, \dots, X_n)| > k) \\ &= \mathbb{P}_{\mu_0}(T(X_1, \dots, X_n) < -k) + \mathbb{P}_{\mu_0}(T(X_1, \dots, X_n) > k) \\ &= \Phi(-k) + (1 - \Phi(k)) \\ &= 2 - 2\Phi(k). \end{aligned}$$

Umstellen nach k liefert

$$\Phi(k) = \frac{2 - \alpha}{2} = 1 - \frac{\alpha}{2} \iff k = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) =: z_{1-\frac{\alpha}{2}},$$

wobei

$$z_a := \Phi^{-1}(a) = \min\{x \in \mathbb{R} \mid \Phi(x) \geq a\}$$

als das a -Quantil¹ der Standardnormalverteilung bezeichnet wird. Der kritische Bereich ist somit

$$\mathcal{K}_1 = \left(-\infty, -z_{1-\frac{\alpha}{2}}\right) \cup \left(z_{1-\frac{\alpha}{2}}, \infty\right).$$

Rechtsseitiger Test: Die Testgröße ist eine monoton wachsende Funktion des Stichprobenmittels und das Stichprobenmittel ist ein Schätzer für den Erwartungswert. Folglich sollten wir H_0 genau dann ablehnen, wenn das Stichprobenmittel deutlich größer ist als μ_0 , d.h. wenn die Testgröße einen kritischen Wert $k > 0$ überschreitet. Bestimmen wir diesen kritischen Wert: Die Beschränkung des Fehlers 1. Art bedeutet jetzt

$$\mathbb{P}_\mu(T(X_1, \dots, X_n) \in \mathcal{K}_1) \leq \alpha, \forall \mu \leq \mu_0.$$

Für alle $\mu \in H_0$, also $\mu \leq \mu_0$, gilt

$$\mathbb{P}_\mu(T(X_1, \dots, X_n) > k) \leq \mathbb{P}_{\mu_0}(T(X_1, \dots, X_n) > k) = 1 - \Phi(k).$$

Setzen wir diesen Wert gleich dem Signifikanzniveau und stellen nach α um, so erhalten wir

$$1 - \Phi(k) = \alpha \iff k = \Phi^{-1}(1 - \alpha) = z_{1-\alpha},$$

d.h. der kritische Bereich ist

$$\mathcal{K}_1 = (k, \infty) = (z_{1-\alpha}, \infty).$$

Linksseitiger Test: Analog zum rechtsseitigen Test sollten wir H_0 nun ablehnen, wenn die Testgröße einen kritischen Wert $-k < 0$ unterschreitet. Für alle $\mu \in H_0$, also $\mu \geq \mu_0$, gilt



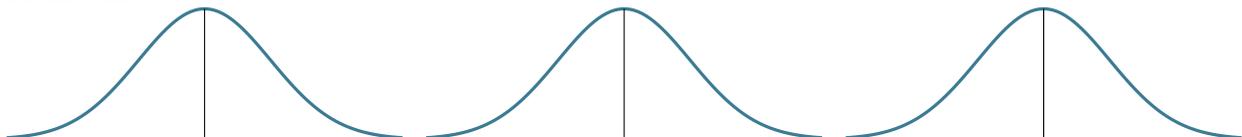
Setzen wir diesen Wert gleich dem Signifikanzniveau und stellen nach α um, so erhalten wir

Der kritische Bereich ist in diesem Fall

$$\mathcal{K}_1 =$$

Das Aufstellen der Entscheidungsregel folgt in allen drei Fällen dem gleichen Schema: Fällt die Realisierung der Testgröße in den kritischen Bereich, so lehnen wir H_0 ab; andernfalls wird H_0 beibehalten.

Sehen wir uns einmal die Interpretation der gefundenen kritischen Bereiche an Hand der $\mathcal{N}(\mu_0, 1)$ -Dichte an:



Sehen wir uns analog zum linksseitigen Test in Beispiel 7.17 einen zweiseitigen Test an.

¹Da die Dichte der Normalverteilung überall strikt positiv ist, ist die zugehörige Verteilungsfunktion streng monoton wachsend und besitzt somit eine Umkehrfunktion. Bei diskreten Verteilungen ist dies nicht der Fall, so dass in diesem Fall mit der angegebenen Definition des Quantils gearbeitet werden muss.



Beispiel 7.18. Eine Abfüllmaschine für Wasserflaschen ist so konstruiert, dass die Abfüllmenge als annähernd $\mathcal{N}(\mu, \sigma^2)$ -verteilt angenommen werden kann mit $\sigma^2 = 4$. Nach einer Produktionspause soll die Maschine wieder in Betrieb genommen werden, jedoch soll vorher überprüft werden, ob die mittlere Abfüllmenge korrekt auf 1000 ml justiert ist. Bei 100 Probe-Abfüllungen werden die Füllmengen notiert und das Stichprobenmittel berechnet. Konstruieren Sie einen Signifikanztest zu den Niveaus $\alpha = 10\%$, $\alpha = 5\%$ und $\alpha = 1\%$, der entscheidet, ob die Maschine vor Aufnahme der Produktion neu justiert werden muss, wobei unnötige Neujustieren möglichst vermieden werden soll. Vergleichen Sie die Ergebnisse.

Antwort: Es soll überprüft werden, ob $\mu = \mu_0$ oder $\mu \neq \mu_0$ ist. Wir setzen also

und

Mit $n = 100$, $\mu_0 = 1000$ und $\sigma = 2$ ergibt sich die Testgröße

$$T(X_1, \dots, X_{100}) =$$

Der kritische Bereich ist für $\alpha = 10\%$ gegeben durch

$$\mathcal{K}_1 = (-\infty, -z_{0,95}) \cup (z_{0,95}, \infty) =$$

und im Fall $\alpha = 5\%$

$$\mathcal{K}_1 =$$

und im Fall $\alpha = 1\%$

$$\mathcal{K}_1 =$$

Der Anschauung halber formulieren wir die Entscheidungsregeln für das Stichprobenmittel

$$\bar{X}_{100} = \frac{T(X_1, \dots, X_{100})}{5} + 1000.$$

Für $\alpha = 10\%$ haben wir die Entscheidungsregel:

- Ist $|\bar{X}_{100}| > \frac{1,645}{5} + 1000 = 1000,329$, so
- ist $|\bar{X}_{100}| \leq 1000,329$, so

Für $\alpha = 5\%$ haben wir die etwas weniger strenge Entscheidungsregel:

- Ist $|\bar{X}_{100}| >$
- ist $|\bar{X}_{100}| \leq$

Für $\alpha = 1\%$ haben wir die noch weniger ablehnende Entscheidungsregel:

- Ist $|\bar{X}_{100}| >$
- ist $|\bar{X}_{100}| \leq$

Möchte man die Wahrscheinlichkeit für eine Fehler 1. Art, d.h. das unnötige Neujustieren, verkleinern, so muss der Annahmehereich, also die Toleranz bei den beobachteten Füllmengen, vergrößert werden.

t-Test

Die Stichprobe X_1, \dots, X_n bestehe nach wie vor aus stochastisch unabhängigen und identisch verteilten Zufallsvariablen mit $X_k \sim \mathcal{N}(\mu, \sigma^2)$ für $k \in \{1, \dots, n\}$, wobei jedoch sowohl μ als auch σ unbekannt seien und μ sei der uns interessierende Parameter. Der Parameterraum für (μ, σ) ist im Folgenden $\Theta = \mathbb{R} \times \mathbb{R}_+$.

Die Unterscheidung in zwei-, rechts- und linksseitigen Test ist identisch zum Gaußtest:

- **zweiseitiger Test:** $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$
- **rechtsseitiger Test:** $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$
- **linksseitiger Test:** $H_0: \mu \geq \mu_0$ vs. $H_1: \mu < \mu_0$

Allerdings eignet sich $\sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma}$ nicht als Testgröße, da σ nicht bekannt ist. Die Lösung besteht darin, zunächst σ^2 durch die Stichprobenvarianz S_n^2 (vgl. Beispiele 5.44 und 7.8) zu schätzen, d.h. die Testgröße ist dann

$$T(X_1, \dots, X_n) := \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} \quad \text{mit} \quad S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Während die Testgröße beim Gaußtest normalverteilt war, ist die neue Testgröße nicht mehr normalverteilt.

Lemma 7.19

Seien X_1, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen mit $X_k \sim \mathcal{N}(\mu, \sigma^2)$ für alle $k \in \{1, \dots, n\}$. Dann ist

$$T_n := \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} \quad \text{mit} \quad S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

t -verteilt mit $n - 1$ Freiheitsgraden, d.h. T_n besitzt die Dichte

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

Hierbei bezeichnet Γ die Gamma-Funktion, welche wir in Definition 5.22 kennen gelernt und deren Eigenschaften wir in Lemma 5.23 bewiesen haben.

Ebenso wie die Verteilungsfunktion der Normalverteilung besitzt auch die der t -Verteilung zu $n - 1$ Freiheitsgraden ($n \geq 2$) i.d.R. keine explizite Darstellung – wir werden daher an keiner Stelle mit der Dichte rechnen. Die gefragten Werte der Verteilungsfunktion finden wir wieder in einer Tabelle.

Notation

Das $(1 - \alpha)$ -Quantil der t -Verteilung mit $n - 1$ Freiheitsgraden bezeichnen wir mit

$$t_{n-1; 1-\alpha} := F_{T_n}^{-1}(1 - \alpha) = \min \{x \in \mathbb{R} \mid \mathbb{P}(T_n \leq x) \geq 1 - \alpha\}$$

für $\alpha \in (0, 1)$.

Da wir uns nicht für die Werte der Verteilungsfunktion, sondern für die Quantile interessieren, gibt es diese direkt in einer Tabelle:

Quantile der t -Verteilung

n	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

Üblicherweise benötigt man nur die Quantile für kleine α , also große $1 - \alpha$. Sollten dennoch auch andere Quantile benötigt werden, so kann man die Symmetrie der t -Verteilung nutzen:

Ist f_n die Dichte der t -Verteilung mit n Freiheitsgraden, so gilt $f_n(x) = f_n(-x)$ für alle $x \in \mathbb{R}$. Folglich ist

$$t_{n;\alpha} = -t_{n;1-\alpha}.$$

Bemerkung 7.20. Wie sind die Werte für „ $n = \infty$ “ in der Tabelle der Quantile entstanden?

Antwort: Für $n \rightarrow \infty$ konvergiert die t -Verteilung mit n Freiheitsgraden gegen die Standardnormalverteilung.

Die kritischen Bereiche finden wir analog zu denen im Gauß-Test, wenn wir das z -Quantil der Standardnormalverteilung gegen das entsprechende t -Quantil der t -Verteilung mit $n - 1$ Freiheitsgraden ersetzen:

- **zweiseitiger Test:** $\mathcal{K}_1 = \left(-\infty, -t_{n-1;1-\frac{\alpha}{2}}\right) \cup \left(t_{n-1;1-\frac{\alpha}{2}}, \infty\right)$;
- **rechtsseitiger Test:** $\mathcal{K}_1 = (t_{n-1;1-\alpha}, \infty)$;
- **linksseitiger Test:** $\mathcal{K}_1 = (-\infty, -t_{n-1;1-\alpha})$.

Damit können wir ein Beispiel für einen t -Test ansehen.

Beispiel 7.21. Nach Angaben eines Autohändlers soll der Benzinverbrauch eines Wagens bei einer Geschwindigkeit von 120 km/h unter 8,8 l je 100 km liegen. Bei 36 Testfahrten erhält man insgesamt die folgenden (gerundeten) Werte mit den angegebenen Häufigkeiten:

8,5	8,6	8,7	8,8	8,9	9,0	9,1	9,2	9,3	9,4
1x	4x	4x	8x	7x	5x	1x	2x	1x	3x

Der Benzinverbrauch soll als annähernd normalverteilt angenommen werden. Lässt sich die Behauptung des Händlers zu einem Signifikanzniveau von $\alpha = 0,05$ aufrechterhalten?

Antwort: Gehen wir systematisch die Schritte durch

1. Sei X_k der Benzinverbrauch bei der k -ten Fahrt (in Liter je 100 km). Diese Zufallsvariablen sind i.i.d. und es gilt $X_k \sim \mathcal{N}(\mu, \sigma^2)$ mit unbekanntem Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$. Der Parameterraum ist

$$\Theta =$$

2. Nullhypothese und Alternative sind

$$H_0: \mu = \mu_0 = 8,8 \quad \text{und} \quad H_1: \mu \neq \mu_0.$$

Das Signifikanzniveau ist im Text vorgegeben als $\alpha =$

3. Die Testgröße ist

$$T(X_1, \dots, X_n) := \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} \quad \text{mit} \quad S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Mit den konkreten Werten $\mu_0 =$ und $n =$ ist somit unter der Annahme $\mu = \mu_0$

$$T(X_1, \dots, X_n) :=$$

4. Der kritische Bereich kann mit den hergeleiteten Formeln sofort notiert werden:

$$\mathcal{K}_1 =$$

5. Die Entscheidungsregel lautet:

- ist $T(x_1, \dots, x_{36})$, so lehne H_0 zu Gunsten von H_1 ab;
- ist $T(x_1, \dots, x_{36})$, so wird H_0 beibehalten.

6. Aus den uns gegebenen Werten berechnen wir zunächst die Realisierungen von X_{36} und S_{36} :

$$x_{36} =$$

$$s_{36} =$$

Damit ist

$$T(x_1, \dots, x_{36}) =$$

7. Entscheidung:

Bemerkung 7.22. Eine kleine Anekdote zum Schluss: Die t -Verteilung wird auch als Student- t -Verteilung bezeichnet, da William Sealy Gosset, der sie 1908 in einer Arbeit entwickelte, selbige Arbeit unter dem Pseudonym Student veröffentlichte.

χ^2 -Test

Die Stichprobe X_1, \dots, X_n bestehe nach wie vor aus stochastisch unabhängigen und identisch verteilten Zufallsvariablen mit $X_k \sim \mathcal{N}(\mu, \sigma^2)$ für $k \in \{1, \dots, n\}$, wobei σ^2 der uns interessierende Parameter ist. Beide Parameter, μ und $\sigma > 0$, sind unbekannt, daher ist der Parameterraum wieder $\Theta = \mathbb{R} \times \mathbb{R}_+$.

Mit der Festlegung $\sigma > 0$ ist es im Folgenden egal, ob wir Hypothesen für die Varianz σ^2 oder für die Standardabweichung σ formulieren. Allerdings ist der Schätzer für die Varianz bereits bekannt und wir werden sehen, dass wir auch die Verteilung des Schätzers schon kennen gelernt haben.

Die möglichen Hypothesen beim χ^2 -Test sind:

- **zweiseitiger Test:** $H_0: \sigma^2 = \sigma_0^2$ vs. $H_1: \sigma^2 \neq \sigma_0^2$
- **rechtsseitiger Test:** $H_0: \sigma^2 \leq \sigma_0^2$ vs. $H_1: \sigma^2 > \sigma_0^2$
- **linksseitiger Test:** $H_0: \sigma^2 \geq \sigma_0^2$ vs. $H_1: \sigma^2 < \sigma_0^2$

Als erwartungstreuen Schätzer für die Varianz haben wir in Beispielen 5.44 und 7.8 die Stichprobenvarianz kennen gelernt:

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

Bei der Verteilung unserer Testgröße stoßen wir auf die in Bemerkung 5.25 erwähnte *Chi-Quadrat-Verteilung* mit $n-1$ Freiheitsgraden, abgekürzt mit χ_{n-1}^2 .

Die Werte der χ^2 -Verteilungsfunktion sind ebenso wie die der Standardnormal- und t -Verteilung tabelliert:

Quantile der χ_n^2 -Verteilung

n	0.1%	0.5%	1.0%	2.5%	5.0%	10.0%	12.5%	20.0%	25.0%	33.3%	50.0%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

Quantile der χ_n^2 -Verteilung

n	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

Lemma 7.23

Seien X_1, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen mit $X_k \sim \mathcal{N}(\mu, \sigma^2)$ für alle $k \in \{1, \dots, n\}$. Dann ist

$$\frac{n-1}{\sigma^2} S_n^2 = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \sim \chi_{n-1}^2.$$

Wir wollen diese Aussage nicht in voller Allgemeinheit zeigen. Um jedoch die Ursache für den Verteilungstyp zu erkennen, zeigen wir folgenden Satz über die Verteilung der Summe der Quadrate von standardnormalverteilten Zufallsvariablen.

Satz 7.24

Seien X_1, \dots, X_n stochastisch unabhängige standardnormalverteilte Zufallsvariablen. Dann ist

$$\sum_{k=1}^n X_k^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) = \chi_n^2.$$

Beweis. Wir zeigen den Satz mit Hilfe vollständiger Induktion über $n \in \mathbb{N}$.

Induktionsanfang: Der Fall $n = 1$ ist eine Übungsaufgabe.

Induktionsannahme: $Z := \sum_{k=1}^{n-1} X_k^2 \sim \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right) = \chi_{n-1}^2$.

Induktionsschritt: Sind X_1, \dots, X_n stochastisch unabhängig, so sind insbesondere Z und X_n^2 stochastisch unabhängig. Die Verteilung von deren Summe berechnen wir mit Hilfe der Faltungformel und der Dichte der Gamma-Verteilung aus Definition 5.22:

$$\begin{aligned} f_{Z+X_n^2}(y) &= \int_{\mathbb{R}} f_Z(y-x) f_{X_n^2}(x) dx \\ &= \int_{\mathbb{R}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{y-x}{2}} \left(\frac{y-x}{2}\right)^{\frac{n-1}{2}-1} \mathbb{1}_{[0,\infty)}(y-x) \cdot \frac{1}{\Gamma\left(\frac{1}{2}\right)} e^{-\frac{x}{2}} \left(\frac{x}{2}\right)^{\frac{1}{2}-1} \mathbb{1}_{[0,\infty)}(x) dx \\ &= \frac{1}{4\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} e^{-\frac{y}{2}} \mathbb{1}_{[0,\infty)}(y) \int_0^y \left(\frac{y-x}{2}\right)^{\frac{n-3}{2}} \left(\frac{x}{2}\right)^{-\frac{1}{2}} dx \\ &\stackrel{(*)}{=} \frac{1}{4\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} e^{-\frac{y}{2}} \mathbb{1}_{[0,\infty)}(y) \int_0^1 \left(\frac{y-ty}{2}\right)^{\frac{n-3}{2}} \left(\frac{ty}{2}\right)^{-\frac{1}{2}} y dt \\ &= \frac{1}{2\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} e^{-\frac{y}{2}} \left(\frac{y}{2}\right)^{\frac{n-1}{2}-1} \mathbb{1}_{[0,\infty)}(y) \int_0^1 (1-t)^{\frac{n-3}{2}} t^{-\frac{1}{2}} dt \\ &= \underbrace{\frac{1}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{y}{2}} \left(\frac{y}{2}\right)^{\frac{n}{2}-1} \mathbb{1}_{[0,\infty)}(y)}_{\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)\text{-Dichte}} \cdot \underbrace{\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^1 (1-t)^{\frac{n-3}{2}} t^{-\frac{1}{2}} dt}_{=:c} \end{aligned}$$

wobei wir in (*) die Substitution $t = \frac{x}{y}$, also $x = ty$ und $dt = \frac{1}{y} dx$, durchgeführt haben. Es bleibt noch zu begründen, dass $c = 1$ ist. Wir wissen, dass $f_{Z+X_n^2}$ eine Dichte ist, ebenso wie die $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ -Dichte. Da c nur eine Konstante ist, muss diese Konstante den Wert 1 haben.

□

Sehen wir uns mit diesem Wissen nochmal die Verteilung in Lemma 7.23 an.

Hätten wir μ statt \bar{X}_n , so wäre die Zufallsgröße $\left(\frac{X_k - \mu}{\sigma}\right)^2 \sim \chi_1^2$ und die Summe von n solchen stochastisch unabhängigen Zufallsvariablen wäre χ_n^2 -verteilt. Indem wir den unbekanntem Erwartungswert durch das Stichprobenmittel schätzen, verlieren wir einen Freiheitsgrad, daher ist

$$\frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \sim \chi_{n-1}^2.$$

Damit haben wir unsere Testgröße gefunden: Um die Verteilung unter $\sigma = \sigma_0$ zu kennen, setzen wir

$$T_n := T(X_1, \dots, X_n) := \frac{1}{\sigma_0^2} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \stackrel{\sigma = \sigma_0}{\sim} \chi_{n-1}^2.$$

Führen wir zunächst Bezeichnungen für die Quantile der χ^2 -Verteilung ein:

Notation

Das α -Quantil der χ_{n-1}^2 -Verteilung mit $n - 1$ Freiheitsgraden bezeichnen wir mit

$$c_{n-1;\alpha} := F_{T_n}^{-1}(\alpha) = \min \{x \in \mathbb{R} \mid \mathbb{P}(T_n \leq x) \geq \alpha\}$$

für $\alpha \in (0, 1)$.

Insbesondere gilt unter $\sigma^2 = \sigma_0^2$:

$$\mathbb{P}_{\mu, \sigma_0^2}(T(X_1, \dots, X_n) \leq c_{n-1;\alpha}) = \alpha. \quad (7.3.2)$$

Leiten wir nun die Formeln für die Ablehnbereiche her, wobei wir beachten müssen, dass die kritischen Bereiche Teilmengen von $(0, \infty)$ sind, da der Parameter σ^2 positiv sein soll.

Zweiseitiger Test: Der kritische Bereich soll von der Form

$$\mathcal{K}_1 = (0, k_1) \cup (k_2, \infty)$$

sein für geeignete Grenzen $k_1, k_2 > 0$. Wir wollen die Grenzen so wählen, dass beide Intervalle die gleiche Wahrscheinlichkeit des Auftretens haben, d.h.

$$\mathbb{P}_{H_0}(T(X_1, \dots, X_n) < k_1) = \mathbb{P}_{H_0}(T(X_1, \dots, X_n) > k_2) = \frac{\alpha}{2}.$$

Da es sich bei der χ^2 -Verteilung um eine absolutstetige Verteilung handelt, folgt aus (7.3.2) sofort $k_1 = c_{n-1; \frac{\alpha}{2}}$. Zudem gilt

$$\frac{\alpha}{2} \stackrel{!}{=} \mathbb{P}_{H_0}(T(X_1, \dots, X_n) > k_2) = 1 - \mathbb{P}_{H_0}(T(X_1, \dots, X_n) \leq k_2),$$

also ist $k_2 = c_{n-1; 1 - \frac{\alpha}{2}}$.

Rechtsseitiger Test: Ist $\sigma^2 \leq \sigma_0^2$, so erwarten wir eher kleine Werte der Testgröße, wohingegen große Werte der Testgröße zu einer Ablehnung führen sollten. Damit ist der kritische Bereich von der Form $\mathcal{K}_1 = (k, \infty)$ für ein geeignetes $k > 0$.

Für alle $\sigma^2 \in \Theta_0$, also $\sigma^2 \leq \sigma_0^2$, gilt

$$\mathbb{P}_{\mu, \sigma^2}(T(X_1, \dots, X_n) > k) = \mathbb{P}_{\mu, \sigma^2}\left(\frac{(n-1)S_n^2}{\sigma^2} > k\right) \leq \mathbb{P}_{\mu, \sigma_0^2}\left(\frac{(n-1)S_n^2}{\sigma_0^2} > k\right) \stackrel{!}{=} \alpha.$$

Durch Übergang zum Gegenereignis erhalten wir die Bedingung

$$\mathbb{P}_{\mu, \sigma_0^2}(T(X_1, \dots, X_n) \leq k) = 1 - \alpha \iff k = c_{n-1; 1-\alpha}.$$

Linksseitiger Test: Ist $H_0: \sigma^2 \geq \sigma_0^2$, so sprechen kleine Werte der Testgröße gegen die Nullhypothese, d.h. der kritische Bereich ist von der Form $\mathcal{K}_1 = (0, k)$ für ein geeignetes $k > 0$. Für alle $\sigma^2 \in \Theta_0$, also $\sigma^2 \geq \sigma_0^2$, gilt

$$\mathbb{P}_{\mu, \sigma^2}(T(X_1, \dots, X_n) < k) \stackrel{\text{abs.-stetig}}{=} \mathbb{P}_{\mu, \sigma^2} \left(\frac{(n-1)S_n^2}{\sigma_0^2} \leq k \right) \leq \mathbb{P}_{\mu, \sigma_0^2} \left(\frac{(n-1)S_n^2}{\sigma_0^2} \leq k \right) \stackrel{!}{=} \alpha.$$

Das führt direkt auf den kritischen Wert $k = c_{n-1; \alpha}$.

Fassen wir die Beobachtungen zusammen:

- **zweiseitiger Test:** $\mathcal{K}_1 = \left(0, c_{n-1; \frac{\alpha}{2}}\right) \cup \left(c_{n-1; 1-\frac{\alpha}{2}}, \infty\right)$;
- **rechtsseitiger Test:** $\mathcal{K}_1 = (c_{n-1; 1-\alpha}, \infty)$;
- **linksseitiger Test:** $\mathcal{K}_1 = (0, c_{n-1; \alpha})$.

Bemerkung 7.25. Von einem rechtsseitigen Test sprechen wir, wenn große Werte der Testfunktion gegen die Nullhypothese sprechen und somit zur Ablehnung führen.

Sehen wir uns nun ein Beispiel für einen χ^2 -Test an:

Beispiel 7.26. In einer Bäckerei wird der Brotteig für Brotlaibe maschinell portioniert auf eine vorgegebene Masse in kg. Die bisher verwendete Maschine hatte eine Standardabweichung von 0,1 kg. Eine neu angeschaffte Maschine soll angeblich eine geringere Standardabweichung und somit auch eine kleinere Varianz aufweisen. 

Wir nehmen an, dass die Portionierungsgröße annähernd normalverteilt ist. Bei einer Stichprobe werden folgende 20 Messwerte erhoben (in kg):

1,966	1,962	2,012	2,114	2,059	1,965	2,019	1,815	2,085	1,980
1,997	1,924	1,877	2,026	2,051	2,074	1,960	1,962	1,980	1,945

Konstruieren Sie einen Signifikanztest zum Signifikanzniveau $\alpha = 0,05$.

Antwort:

1. Sei X_k die Masse des k -ten Brotröhlings. Diese ZVen sind stochastisch unabhängig und $\mathcal{N}(\mu, \sigma^2)$ verteilt mit unbekanntem $\mu \in \mathbb{R}$ und $\sigma > 0$. Der Parameterraum ist wieder $\Theta = \mathbb{R} \times \mathbb{R}_+$.
2. Nullhypothese und Alternative sind

$$H_0: \quad \quad \quad \text{vs.} \quad \quad H_1:$$

Das Signifikanzniveau ist im Text angegeben als $\alpha =$.

3. Die Testgröße ist

$$T(X_1, \dots, X_n) :=$$

4. Der kritische Bereich kann mit den hergeleiteten Formeln für den linksseitigen Test sofort notiert werden:

$$\mathcal{K}_1 =$$

5. Die Entscheidungsregel lautet:

- ist $T(x_1, \dots, x_{20})$, so lehne H_0 zu Gunsten von H_1 ab;
- ist $T(x_1, \dots, x_{20})$, so wird H_0 beibehalten.

6. Aus den uns gegebenen Werten berechnen wir zunächst die Realisierungen von \bar{X}_{20} und S_{20}^2 :

$$\bar{x}_{20} = \frac{1}{20} \sum_{k=1}^{20} x_k = 1,98865$$

$$S_{20}^2 = \frac{1}{19} \sum_{k=1}^{20} (x_k - \bar{x}_{20})^2 = 5,064 \cdot 10^{-3}.$$

Damit ist

$$T(x_1, \dots, x_{20}) =$$

7. Entscheidung:

Der Fehler 2. Art und die Gütefunktion

Bevor wir uns einem Test zuwenden, der nicht auf normalverteilten Stichproben basiert, wollen wir einmal den Fehler 2. Art ansehen. Betrachten wir konkret die Situation aus Beispiel 7.21:

Beispiel 7.27 (Fortsetzung von Beispiel 7.21). In diesem Beispiel wurde ein t -Test für die Hypothesen

$$H_0: \mu \leq \mu_0 = 8,8 \quad \text{vs.} \quad H_1: \mu > \mu_0$$

durchgeführt. Die Wahrscheinlichkeit für einen Fehler 1. Art wurde durch $\alpha = 0,05$ beschränkt, d.h. für die Testgröße $T(X_1, \dots, X_n) = \sqrt{n} \frac{\bar{X} - \mu_0}{S_n}$ gilt mit geeignetem gewähltem Ablehnbereich $\mathcal{K}_1 = (1,69; \infty)$:

$$\mathbb{P}_{H_0}(T(X_1, \dots, X_n) \in \mathcal{K}_1) \leq \alpha.$$

Der Trick um das für $\Theta_0 = (-\infty, \mu_0]$ zu erreichen bestand darin, dass für alle Parameter in Θ_0 , also $\mu \leq \mu_0$, stets

$$\mathbb{P}_{\mu, \sigma}(T(X_1, \dots, X_n) > 1,69) \leq \mathbb{P}_{\mu_0, \sigma}(T(X_1, \dots, X_n) > 1,69) = \alpha$$

gilt. Für den Fehler 2. Art müssen wir jedoch Parameter aus $\Theta_1 = (\mu_0, \infty)$, also $\mu > \mu_0$ wählen. Der Fehler 2. Art ist

$$\mathbb{P}_{H_1}(T(X_1, \dots, X_n) \leq 1,69).$$

Durch eine analoge Abschätzung bekommen wir also lediglich die wenig hilfreiche Abschätzung

$$\mathbb{P}_{\mu, \sigma}(T(X_1, \dots, X_n) \leq 1,69) \leq \mathbb{P}_{\mu_0, \sigma}(T(X_1, \dots, X_n) \leq 1,69) = 1 - \alpha = 0,95$$

für alle $\mu > \mu_0$. Der Fehler 2. Art kann also nach oben nur durch die Wahrscheinlichkeit 95% beschränkt werden.

Es gibt zwei Möglichkeiten, den Fehler 2. Art dennoch untersuchen zu können.

1. Um eine sinnvolle Schranke für den Fehler 2. Art angeben zu können, werden zwei konkurrierende Werte für den interessierenden Parameter ϑ angegeben, d.h. entgegen der bisherigen Aufspaltung $\Theta = \Theta_0 \dot{\cup} \Theta_1$ betrachten wir Hypothesen der Form

$$H_0: \vartheta = \vartheta_0 \quad \text{vs.} \quad H_1: \vartheta = \vartheta_1.$$

2. Die Wahrscheinlichkeiten für Fehler 1. und 2. Art kann man in Abhängigkeit des wahren Parameters auch graphisch darstellen. Wir betrachten dazu die Gütefunktion

$$g_n: \Theta \rightarrow [0, 1], \quad \vartheta \mapsto \mathbb{P}_\vartheta(T(X_1, \dots, X_n) \in \mathcal{K}_1).$$

Diese gibt an, mit welcher Wahrscheinlichkeit H_0 unter dem wahren Parameter ϑ abgelehnt wird.

Für jeden Parameter $\vartheta \in \Theta_0$ liefert $g_n(\vartheta)$ die Wahrscheinlichkeit für einen Fehler 1. Art.

Für jeden Parameter $\vartheta \in \Theta_1$ liefert $1 - g_n(\vartheta)$ die W'keit für einen Fehler 2. Art.

Bemerkung 7.28. *Achtung – Verwechslungsgefahr!*

Die Wahrscheinlichkeit für den Fehler 2. Art wird in der Literatur manchmal mit β und manchmal mit $1 - \beta$ bezeichnet. Zudem bezeichnet $OC = 1 - g_n$ die Operationscharakteristik oder Annahmekennlinie.

Wir werden im Folgenden die Wahrscheinlichkeit für den Fehler 2. Art mit β bezeichnen und wenn wir den Zusammenhang zwischen einem Parameter und der Fehlerwahrscheinlichkeit untersuchen, beschränken wir uns auf die eingeführte Gütefunktion.

Im Folgenden sehen wir uns ein Beispiel für einen Binomialtest an, d.h. anders als bisher ist die interessierende Zufallsvariable $\text{Bin}(n, p)$ -verteilt und der Parameter p ist unbekannt. Für dieses Beispiel können wir besonders gut sowohl den Fehler 2. Art als auch die Gütefunktion betrachten.

Beispiel 7.29 (Zwei konkurrierende Werte für einen Parameter – Binomialtest). *Der Bekanntheitsgrad einer Marke unter Jugendlichen beträgt nach Einschätzung der Firmenleitung 50%, nach Einschätzung der Werbeabteilung jedoch nur 30%. Durch die Befragung von 100 zufällig ausgewählten Jugendlichen soll entschieden werden, wer Recht hat, und damit, ob eine Werbekampagne nötig ist. Bestimmen Sie für beide möglichen Nullhypothesen eine Entscheidungsregel für $\alpha_1 = 10\%$ und $\alpha_2 = 1\%$ und bestimmen Sie jeweils auch die Wahrscheinlichkeit eines Fehlers 2. Art.*

Antwort: Sei X die Anzahl der Jugendlichen unter den Befragten, die die Marke kennen. Dann ist $X \sim \text{Bin}(n, p)$ mit $n = 100$ und unbekanntem Parameter p . X ist in diesem Fall auch unsere Testgröße, da wir unter H_0 die Verteilung von X genau kennen.

Variante 1: Wir wählen Nullhypothese und Alternative als

$$H_0: p = 0,5 \quad \text{und} \quad H_1: p = 0,3.$$

Wir sollten H_0 ablehnen, falls $X \leq k^*$ für einen noch zu bestimmenden Wert $k^* \in \{0, \dots, 100\}$ ist. Die Wahrscheinlichkeit für einen Fehler 1. Art ist dann

$$\mathbb{P}_{0,5}(X \leq k_j) \stackrel{!}{\leq} \alpha_j, \quad j \in \{1, 2\}.$$

Sehen wir uns einige Werte der Verteilungsfunktion der $\text{Bin}(100; 0,5)$ -Verteilung an:

k	36	37	38	39	40	41	42	43	44
$\mathbb{P}_{0,5}(X \leq k)$	0,003	0,006	0,010	0,018	0,028	0,044	0,067	0,097	0,136

Der größte Wert $k_1 \in \mathbb{N}$, für den $\mathbb{P}_{0,5}(X \leq k_1) \leq 0,1$ ist, ist $k_1^* = 43$; der größte Wert $k_2 \in \mathbb{N}$, für den $\mathbb{P}_{0,5}(X \leq k_2) \leq 0,01$ ist, ist $k_2^* = 38$ (bzw. ohne so starkes vorheriges Runden $k_2^* = 37$). Somit ist in beiden Fällen $\mathcal{K}_1 = \{0, \dots, k_j^*\}$ und die Entscheidungsregel lautet:

- Kennen höchstens 43 befragte Jugendliche die Marke, so lehne H_0 zum Signifikanzniveau $\alpha_1 = 0,1$ ab; kennen sogar höchstens 38 (bzw. 37) Jugendliche die Marke, so wird H_0 auch zum Signifikanzniveau $\alpha_2 = 0,01$ abgelehnt. In diesem Fall sollte eine Werbekampagne gestartet werden.
- Tritt obiger Fall nicht ein, so wird H_0 zum gewählten Signifikanzniveau nicht abgelehnt. Die Daten liefern also keinen Grund für eine Werbekampagne.

Der Fehler 2. Art hat die Wahrscheinlichkeit

$$\beta_1 = \mathbb{P}_{0,3}(X > 43) = \sum_{k=44}^{100} \binom{100}{k} 0,3^k \cdot 0,7^{100-k} \approx 0,002 \quad \text{bzw.}$$

$$\beta_2 = \mathbb{P}_{0,3}(X > 37) = \sum_{k=38}^{100} \binom{100}{k} 0,3^k \cdot 0,7^{100-k} \approx 0,05.$$

d.h. die Wahrscheinlichkeit für einen Fehler 2. Art beträgt ca 0,002 für $\alpha = 0,1$ und 0,05 für $\alpha = 0,01$. Eine geringere Wahrscheinlichkeit für einen Fehler 1. Art geht also mit einer höheren Wahrscheinlichkeit für einen Fehler 2. Art einher.



Variante 2: Wir wählen Nullhypothese und Alternative als

$$H_0: p = 0,3 \quad \text{und} \quad H_1: p = 0,5.$$

Wir sollten H_0 ablehnen, falls $X \geq k^*$ für einen noch zu bestimmenden Wert $k^* \in \{0, \dots, 100\}$ ist. Die Wahrscheinlichkeit für einen Fehler 1. Art ist dann

Sehen wir uns einige Werte der Verteilungsfunktion der Bin(100; 0,3)-Verteilung an:

k	34	35	36	37	38	39	40	41
$1 - \mathbb{P}_{0,3}(X \leq k)$	0,163	0,116	0,080	0,053	0,034	0,021	0,012	0,007

Bestimmen wir damit die kritischen Bereiche:

- Der kleinste Wert $k \in \mathbb{N}$, für den $1 - \mathbb{P}_{0,3}(X \leq k - 1) \leq 0,1$ ist, ist $k_1^* =$. Somit ist für $\alpha_1 = 0,1$ der kritische Bereich $\mathcal{K}_1 =$
- Der kleinste Wert $k \in \mathbb{N}$, für den $1 - \mathbb{P}_{0,3}(X \leq k - 1) \leq 0,01$ ist, ist $k_2^* =$. Somit ist für $\alpha_2 = 0,01$ der kritische Bereich $\mathcal{K}_1 =$

Damit haben wir folgende Entscheidungsregel:

- H_0 wird zum Niveau $\alpha_1 = 0,1$ verworfen, falls mindestens \quad der befragten Jugendlichen die Marke nicht kennen.
- H_0 wird zum Niveau $\alpha_2 = 0,01$ verworfen, falls mindestens \quad der befragten Jugendlichen die Marke nicht kennen.

In diesem Fall sollte keine Werbekampagne gestartet werden, da die von der Firmenleitung gewünschte Bekanntheit bereits da ist. Im anderen Fall sprechen die Daten nicht gegen den Start einer Werbekampagne.

Die Wahrscheinlichkeiten für den Fehler 2. Art können wir aus der oben vorhandenen Wertetabelle ablesen:

$$\beta_1 =$$

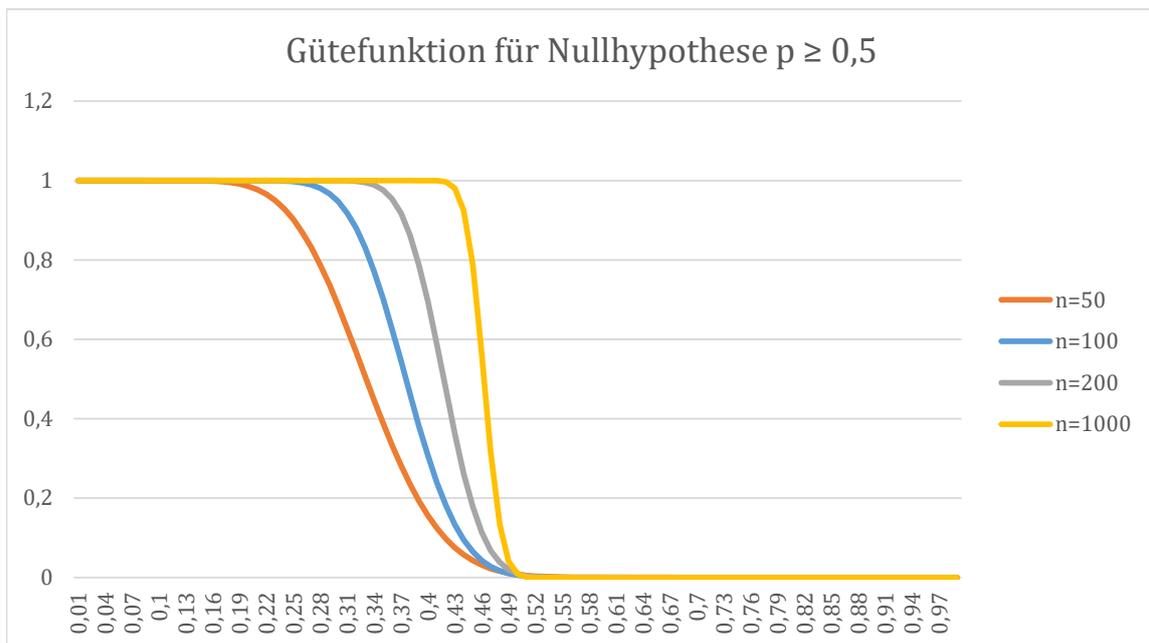
$$\beta_2 =$$

Kennen zwischen 37 und 43 Jugendliche unter den Befragten die Marke, hätten wir zum Niveau $\alpha = 0,1$ die Nullhypothese $p = 0,5$ abgelehnt, zugleich aber auch beim 2. Test dessen Nullhypothese $p = 0,3$ abgelehnt. Das liegt daran, dass bei zu großer Fehlerwahrscheinlichkeit 1. Art die kritischen Bereiche beider Tests eine nicht-leere Schnittmenge haben. Somit ist deutlich, warum das Testdesign einen wesentlichen Einfluss auf die Entscheidung hat – selbst bei identischen Daten.

Sehen wir uns nun für den Test $H_0: p \geq 0,5$ gegen $H_1: p < 0,5$ die Gütefunktion

$$g_n(p) := \mathbb{P}_p(X < k_n^*)$$

an, wobei $k_n^* \in \mathbb{N}$ die Grenze des kritischen Bereiches ist in Abhängigkeit vom Parameter $n \in \{50, 100, 200, 1000\}$.



Im Beispiel haben wir gesehen, dass für wachsendes n , also $n \rightarrow \infty$, die Wahrscheinlichkeit für einen Fehler 2. Art verringert werden kann. In Formeln ist das folgende Eigenschaft:

$$\lim_{n \rightarrow \infty} (1 - g_n(\vartheta)) = 0, \quad \forall \vartheta \in \Theta_1.$$

Das könnte dazu verleiten, eine größere Stichprobe grundsätzlich für besser zu halten, denn der Fehler 1. Art wird durch die Konstruktion des Tests bereits beschränkt und der Fehler 2. Art kann durch eine größere Stichprobe verringert werden. Allerdings muss man bei der Auswertung eines Tests auch bei großer Stichprobengröße vorsichtig sein.

Bemerkung 7.30 (Signifikanz vs. Relevanz). *Lehnt man die Nullhypothese bei einem Signifikanztest zum Niveau α nicht ab, so kann dies daran liegen, dass die beobachtete Abweichung von der Norm nicht inhaltlich relevant war, sondern zufällig aufgetreten ist, oder es kann daran liegen, dass der Stichprobenumfang zu klein war.*

Andersherum kann bei sehr großen Stichproben schon eine kleine Abweichung als signifikant gelten, obwohl rein inhaltlich keine relevante Abweichung von der Norm vorhanden ist.

Sehen wir uns zum Abschluss dieses Kapitels ein Beispiel dafür an, wie man eine Formel für die Gütefunktion angeben kann. Dazu kommen wir zurück zur Normalverteilung.



Beispiel 7.31. Seien x_1, \dots, x_n Realisierungen einer Stichprobe von $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariablen, wobei $\sigma > 0$ bekannt und $\mu \in \mathbb{R}$ unbekannt ist.

Wie berechnet man die Werte der Gütefunktion für den rechtsseitigen Gaußtest zu den Hypothesen

$$H_0: \mu \leq \mu_0 \quad \text{und} \quad H_1: \mu > \mu_0$$

zum Signifikanzniveau $\alpha \in (0, 1)$?

Antwort: Wir haben bereits die Entscheidungsregel für den rechtsseitigen Gaußtest zum Niveau α kennen gelernt: Die Nullhypothese H_0 wird genau dann abgelehnt, wenn die Realisierung der Testgröße im kritischen Bereich liegt, d.h. wenn

$$T(x_1, \dots, x_n) = \sqrt{n} \cdot \frac{\bar{x}_n - \mu_0}{\sigma} \in \mathcal{K}_1 =$$

gilt. Dabei begehen wir einen Fehler 1. Art, wenn wir H_0 ablehnen, obwohl $\mu \leq \mu_0$ ist. Die Wahrscheinlichkeit eines Fehlers 1. Art ist begrenzt durch α , d.h.

wobei die Testgröße unter dem wahren Parameter μ wie folgt verteilt ist:

$$T(X_1, \dots, X_n) :=$$

Berechnen wir mit diesem Wissen die Gütefunktion:

$$\begin{aligned} g_n(\mu) &= \mathbb{P}_\mu(T(X_1, \dots, X_n) \in \mathcal{K}_1) \\ &= \end{aligned}$$

Für jeden Wert $\mu \in \mathbb{R}$ liefert diese Funktion den Wert der Gütefunktion. Insbesondere können wir die Werte für $n \rightarrow \infty$ für $\mu > \mu_0$, $\mu = \mu_0$ und $\mu < \mu_0$ ablesen:

- Für $\mu = \mu_0$ ist $g_n(\mu) =$ für alle $n \in \mathbb{N}$.
- Für $\mu \geq \mu_0$, also $\mu \in \Theta_0$, ist

$$\lim_{n \rightarrow \infty} \left(\sqrt{n} \cdot \frac{\mu - \mu_0}{\sigma} \right) =$$

Folglich gilt für die Gütefunktion (dank der Stetigkeit von Φ und Φ^{-1})

$$\lim_{n \rightarrow \infty} g_n(\mu) =$$

- Für $\mu < \mu_0$, also $\mu \in \Theta_1$, ist

$$\lim_{n \rightarrow \infty} \left(\sqrt{n} \cdot \frac{\mu - \mu_0}{\sigma} \right) =$$

Folglich gilt für die Gütefunktion

$$\lim_{n \rightarrow \infty} g_n(\mu) =$$

Für $n \rightarrow \infty$ nähert sich die Gütefunktion folgender Treppenfunktion an:

$$g(\mu) :=$$

Es gibt noch diverse weitere Arten von Test – diese betreffen nicht nur Parameter einer Verteilung, sondern auch beispielsweise die Frage, ob Zufallsvariablen unkorreliert sind. Solche Tests gehen jedoch über das Ziel dieser Veranstaltung hinaus.

7.4 Beschreibende Statistik

Die bisherigen Abschnitte dieses Kapitels gehörten zum Bereich der *induktiven Statistik*, d.h. aus gegebenen Daten wollten wir Entscheidungen ableiten. In diesem Abschnitt soll es um die *deskriptive*, d.h. *beschreibende Statistik* gehen. Einige Begriffe sind bereits gefallen, aber der Vollständigkeit halber wiederholen wir sie an dieser Stelle.

7.4.1 Merkmale, Ausprägungen und Histogramme

Beschreiben wir zunächst abstrakt, was wir bei einer Erhebung von Daten tun.

In einer statistischen Erhebung werden Werte verschiedener *Merkmale* einer *Grundgesamtheit* beobachtet. Die *Grundgesamtheit* ist eine Menge G von Beobachtungseinheiten (z.B. Individuen), die verschiedene *Merkmale* (X, Y, \dots) aufweisen – das sind die zu untersuchenden Größen. Jedes *Merkmal* kann verschiedene *Ausprägungen* besitzen. Ist A die Menge der möglichen Ausprägungen des Merkmals X , so ist $X: G \rightarrow A$ eine Abbildung, die jedem Element der Grundgesamtheit eine Ausprägung zuordnet. Werden die Merkmalsausprägungen von allen Elementen der Grundgesamtheit erhoben, so handelt es sich um eine *Totalerhebung*; werden nur die Merkmalsausprägungen einer Teilmenge $M \subset G$ betrachtet, so bezeichnen wir M als *Teilerhebung* oder *Stichprobe*. Nummeriert man die Elemente der Stichprobe als $M = \{1, \dots, n\}$ und bezeichnet man die Merkmalsausprägungen als $x_k := X(k)$ für $k \in \{1, \dots, n\}$, so nennt man (x_1, \dots, x_n) eine *Messreihe*.

Anhand von Beispiel 7.29 wollen wir die eingeführten Begriffe illustrieren:

Beispiel 7.32. Es werden 100 Jugendliche befragt, ob sie eine Marke kennen und die Anzahl der bejahenden Antworten wird notiert in der Form $x_k = 1$, falls der k -te Jugendliche die Marke kennt, und $x_k = 0$, falls nicht.

- Als Grundgesamtheit gilt die Menge aller Jugendlichen der betrachteten Region, wobei wir Jugendliche beispielsweise als Personen im Alter von 14 bis 18 definieren können.
- Die Stichprobe besteht aus $n = 100$ rein zufällig ausgewählten Personen der Grundgesamtheit – nummerieren wir diese als $M = \{1, \dots, 100\}$.
- Das Merkmal X ist in diesem Fall die Eigenschaft, ob ein Merkmalsträger (d.h. jugendlicher) die vorgegebene Marke kennt. Die Merkmalsausprägung ist in einem ersten Schritt $A = \{\text{ja}, \text{nein}\}$. Dieses qualitative Merkmal wird in ein nominales Merkmal umgewandelt, indem „ja“ zu $X_k = 1$ und „nein“ zu $X_k = 0$ wird. Damit erhalten wir die Menge der Ausprägungen $A' = \{0, 1\}$.
- Eine Messreihe ist ein n -dimensionaler Vektor in $\{0, 1\}^n$.

Definition 7.33: Häufigkeiten

Ist eine Messreihe $(x_k)_{k \in M}$ mit $|M| = \#M =: n$ gegeben, so bezeichnet

$$h(X = a) := \#\{k \in M \mid x_k = a\} = |\{k \in M \mid x_k = a\}| = \sum_{k \in M} \mathbb{1}_{\{x_k = a\}}$$

die *absolute Häufigkeit* der Ausprägung $a \in A$ und

$$r(X = a) := \frac{h(X = a)}{n}$$

die *relative Häufigkeit* der Ausprägung $a \in A$.

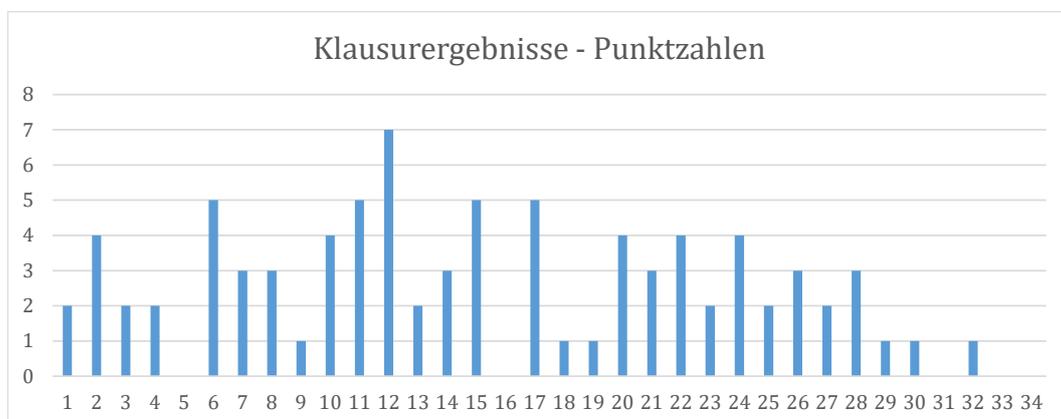
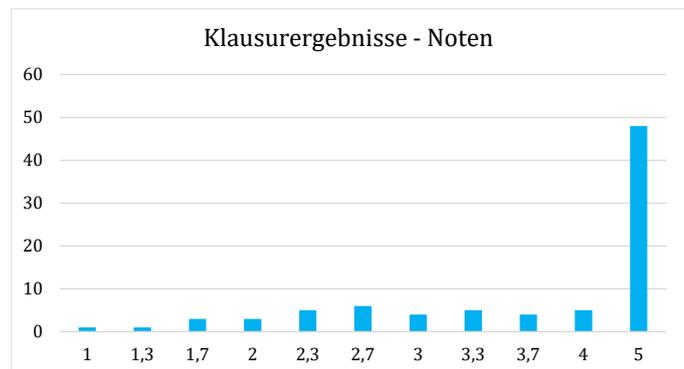
Beispiel 7.34 (Fortsetzung von Beispiel 7.32). Antworten 35 von 100 befragten Jugendlichen, dass sie die Marke kennen, so ist $h(X = 1) = 35$ und $r(X = 1) = \frac{35}{100} = 0,35$, sowie $h(X = 0) = 100 - 35 = 65$ und $r(X = 0) = \frac{65}{100} = 0,65$.

Während man bei wenigen Ausprägungen die absoluten und relativen Häufigkeiten aufzählen oder in einer Tabelle angeben kann, bietet sich bei mehr möglichen Ausprägungen eine graphische Darstellung an, beispielsweise in Form eines *Stab-* oder *Balkendiagramms*. Bei zu vielen möglichen Ausprägungen oder gar stetigen Merkmalsausprägungen fasst man die Ausprägungen zu Klassen zusammen. In diesem Fall sprechen wir von einem *Histogramm*.

Beispiel 7.35.

Die Ergebnisse einer Klausur sind in folgenden Balkendiagrammen dargestellt – einmal die Noten und einmal die erreichten Punkte (auf ganze Punkte gerundet).

Zur Note 5,0 gehören deutlich mehr Punktzahlen als zu den anderen Noten, was zu einem wesentlich höheren Balken führt.



Bei dem Balkendiagramm der Noten handelt es sich nicht um ein Histogramm, welches aus dem Balkendiagramm der Punktzahlen zusammengesetzt wurde. Wie ein solches aussehen müsste, sehen wir uns jetzt an.

Sortieren wir die auftretenden Merkmalsausprägungen der Größe nach, so können wir sie in K Klassen einteilen:

$$[a_1, a_2), [a_2, a_3), \dots, [a_{K-1}, a_K), [a_K, a_{K+1}] \quad \text{mit } a_k < a_{k+1} \text{ für } k \in \{1, \dots, K\}.$$

Die Klassenbreite ist definiert als $\Delta_k := a_{k+1} - a_k$ und die Häufigkeitsdichte ist

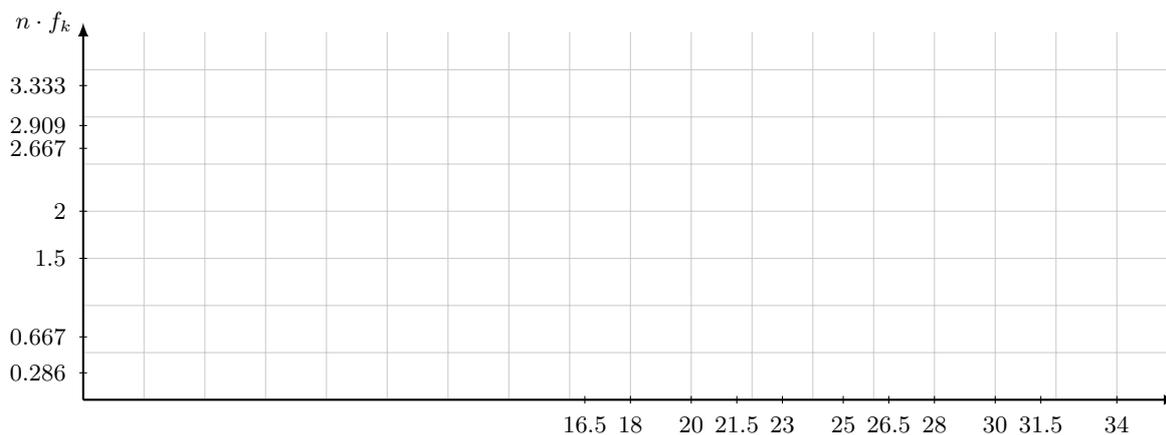
$$f_k := \frac{r(X \in [a_k, a_{k+1}))}{\Delta_k}, \quad k \in \{1, \dots, K\}.$$

Im Histogramm werden die Balken dann mit der Breite Δ_k und der Balkenhöhe f_k lückenlos nebeneinander dargestellt. Sehen wir uns ein solches Balkendiagramm für die Klausurergebnisse aus Beispiel 7.35 an.

Beispiel 7.36 (Fortsetzung von Beispiel 7.35). Die Klassen werden durch die Notengebung vorgegeben. In diesem Fall sei für die k -te Note $\Delta_k = a_{k+1} - a_k$ und h_k sei die absolute Häufigkeit dieser Noten. Mit $n = 85$ ergeben sich folgende Werte (mit $n \cdot f_k = \frac{h_k}{\Delta_k}$):

Note	1,0	1,3	1,7	2,0	2,3	2,7	3,0	3,3	3,7	4,0	5,0
Δ_k	3,5	1,5	2	1,5	1,5	2	1,5	1,5	2	1,5	16,5
h_k	1	1	3	3	5	6	4	5	4	5	48
$n \cdot f_k$	0,286	0,667	1,5	2	3,333	3	2,667	3,333	2	3,333	2,909

Skizzieren wir damit ein Histogramm:



Die relativen Häufigkeiten r_X eines Merkmals X entsprechen einer Massfunktion einer diskret verteilten Zufallsvariablen. Analog zur Verteilungsfunktion von Zufallsvariablen kann man auch die empirische Verteilungsfunktion eines Merkmals X betrachten. Dies ist jedoch nur dann sinnvoll, wenn das Merkmal quantitativ ist. Bei der Frage, ob Jugendliche eine Marke kennen, ist das nicht der Fall; auf die Klausurergebnisse (in Punkten oder Noten) trifft das hingegen zu.

Definition 7.37

Als empirische Verteilungsfunktion eines Merkmals X mit Werten in \mathbb{R} bezeichnet man die Abbildung

$$F_X: \mathbb{R} \rightarrow [0, 1], \quad F_X(x) := r(X \leq x).$$

7.4.2 Lage- und Streuungsparameter

Vergleichen wir die Dichten der $\mathcal{N}(-1, 1)$ - und der $\mathcal{N}(1, 1)$ -Verteilung, so ist die zweite um genau 2 Einheiten nach rechts verschoben. Das erkennen wir bereits daran, dass wir den Erwartungswert betrachten. Vergleichen wir die Dichten der $\mathcal{N}(0, 1)$ - und die der $\mathcal{N}(0, 25)$ -Verteilung, so

ist die Glockenkurve bei letzterer flacher und breiter, was wir an der größeren Varianz (oder Standardabweichung) erkennen.

Analog zu Erwartungswert und Varianz/Streuung von Zufallsvariablen gibt es diverse Parameter, die Aussagen über die Lage (wo ist das Zentrum der Verteilung?) oder die Streuung (wie stark sind die Werte vom Zentrum der Verteilung entfernt?) der empirischen Verteilung eines Merkmals beschreiben.

Definition 7.38: Lageparameter

Sei (x_1, \dots, x_n) eine Messreihe eines Merkmals X mit Werten in \mathbb{R} . Seien zudem $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ die der Größe nach geordneten Daten. Dann definieren wir

- das *arithmetische Mittel* als $\bar{x} := \frac{1}{n} \sum_{k=1}^n x_k$;
- den *Median* als

$$x_{0,5} := \begin{cases} x_{(\frac{n+1}{2})}, & \text{falls } n \text{ ungerade ist,} \\ \frac{1}{2} (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{falls } n \text{ gerade ist;} \end{cases}$$

- das *p-Quantil* (mit $p \in (0, 1)$) als

$$x_p := \begin{cases} x_{(\lfloor np \rfloor + 1)}, & \text{falls } np \text{ nicht ganzzahlig ist,} \\ \frac{1}{2} (x_{(np)} + x_{(np+1)}), & \text{falls } np \text{ ganzzahlig ist;} \end{cases}$$

- den *Modus* oder *Modalwert* x_M als die am häufigsten auftretende Merkmalsausprägung, d.h.

$$x_M \in \arg \max_a h(X = a).$$

Bemerkung 7.39.

1. Der Median ist gerade das 0,5-Quantil. Die Quantile $x_{0,25}$ und $x_{0,75}$ bezeichnet man auch als unteres bzw. oberes Quartil.
2. Handelt es sich bei X um ein quantitatives Merkmal (wie die Klausurergebnisse), so können alle genannten Parameter sinnvoll berechnet werden. Handelt es sich jedoch um ein Rangmerkmal, z.B. Platzierungen bei einem Wettkampf, so sind die Quantile und der Modalwert zwar sinnvoll, das arithmetische Mittel jedoch nicht, da die konkreten Werte des Merkmals willkürlich festgelegt wurden.
3. Das p -Quantil erfüllt die Bedingungen

$$r(X \leq x_p) \geq p \quad \text{und} \quad r(X \geq x_p) \geq 1 - p. \quad (7.4.1)$$

Dies ist allerdings nur eine notwendige, keine hinreichende Bedingung. Die Menge der Punkte x_p , die (7.4.1) erfüllen, ist nicht notwendigerweise einelementig, wohingegen obige Definition das p -Quantil eindeutig festlegt.

4. Ist $X: \{1, \dots, n\} \rightarrow \mathbb{R}$ ein Merkmal, so gilt

$$\bar{x} \in \arg \min_c \left\{ \sum_{k=1}^n (x_k - c)^2 \right\} \quad \text{und} \quad x_{0,5} \in \arg \min_c \left\{ \sum_{k=1}^n |x_k - c| \right\}.$$

Beispiel 7.40. Bei einer Familienfeier wird das Alter der anwesenden Personen notiert:



24 34 37 37 61 63 89.

Wie groß sind arithmetisches Mittel, Median und Modalwert?

Antwort:

- Das arithmetische Mittel dieser 7 Werte ist $\bar{x} =$
- Der Modalwert ist $x_M =$
- Der Median ist $x_{0,5} =$

Bemerkung 7.41. An den Werten im vorherigen Beispiel kann man gut sehen, dass Median und Modalwert tendenziell robuster gegenüber einzelnen falschen Werten sind als das arithmetische Mittel: Wir können leicht Werte verändern oder hinzunehmen ohne dass sich diese Kennzahlen ändern – wichtig ist nur, dass gleich viele Werte kleiner und größer als 37 sind und dass weiterhin 37 die am häufigsten auftretende Zahl ist. Ändert man hingegen einen einzigen Wert, so ändert sich grundsätzlich sofort auch das arithmetische Mittel.

Definition 7.42: Streuungsparameter

Sei (x_1, \dots, x_n) eine Messreihe eines Merkmals X mit Werten in \mathbb{R} . Dann definieren wir

- die empirische Varianz oder Stichprobenvarianz als

$$s_X^2 := \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2;$$

- die empirische Standardabweichung als

$$s_X := \sqrt{s_X^2}.$$

Bemerkung 7.43. Ein weiteres Maß für die Streuung ist die mittlere quadratische Abweichung

$$\sigma_X^2 := \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Wir haben allerdings bereits gesehen, dass dies keinen erwartungstreuen Schätzer für die Varianz einer Zufallsvariablen liefert – wir arbeiten daher in der Regel stattdessen mit der Stichprobenvarianz.

7.4.3 Schreibweise durch Datenvektoren und empirische lineare Regression

Beim Zusammenfassen der Daten als Vektor ist die Schreibweise mit Großbuchstaben üblich, d.h. wir schreiben, sofern keine Verwechslungsgefahr besteht,

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \quad \text{sowie} \quad \mathbf{1} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n \quad \text{und} \quad \delta_X := \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} = X - \bar{x} \cdot \mathbf{1}.$$

Gehen wir nun davon aus, dass wir zwei Datenvektoren $X, Y \in \mathbb{R}^n$ gegeben haben, zwischen denen möglicherweise ein Zusammenhang besteht.¹ Mit dem üblichen Skalarprodukt des \mathbb{R}^n , d.h. $\langle X, Y \rangle := \sum_{k=1}^n x_k y_k$ und der induzierten Norm $\|X\| := \sqrt{\langle X, X \rangle}$ gibt es für die Stichprobenvarianz dann die Kurzschreibweise

$$s_X^2 = \frac{1}{n-1} \|\delta_X\|^2.$$

Analog zu den Begriffen Kovarianz und Korrelationskoeffizient von Zufallsvariablen gibt es die *empirische Kovarianz* und den *empirischen Korrelationskoeffizienten* für Datenvektoren:

$$s_{X,Y} := \frac{1}{n-1} \langle \delta_X, \delta_Y \rangle \quad \text{und} \quad r_{X,Y} := \frac{s_{X,Y}}{s_X s_Y} = \frac{\langle \delta_X, \delta_Y \rangle}{\|\delta_X\| \|\delta_Y\|} = \frac{\langle X, Y \rangle - n\bar{x}\bar{y}}{\sqrt{(\|X\|^2 - n\bar{x}^2)(\|Y\|^2 - n\bar{y}^2)}}.$$

Bemerkung 7.44. Aus der Cauchy-Schwarz'schen Ungleichung folgt sofort $-1 \leq r_{X,Y} \leq +1$ und es gilt $|r_{X,Y}| = 1$ genau dann, wenn δ_X und δ_Y linear abhängig sind. Zudem gilt analog zu $\text{Cov}(X, X) = \text{Var}(X)$ auch hier $s_{X,X} = s_X^2$, d.h. die empirische Kovarianz eines Datenvektors mit sich selbst ist gleich dessen empirischer Varianz.

Bei Daten wird man i.d.R. schon wegen Messungenauigkeiten nicht $|r_{X,Y}| = 1$ oder $r_{X,Y} = 0$ erhalten, selbst wenn die Daten durch Zufallsvariablen erzeugt wurden, zwischen denen ein linearer Zusammenhang besteht oder die unkorreliert sind. Stattdessen ist es üblich, je nach Betrag des Korrelationskoeffizienten von *schwacher Korrelation* ($|r_{X,Y}| \leq 0,5$), *mittlerer Korrelation* ($0,5 < |r_{X,Y}| \leq 0,8$) oder *starker Korrelation* ($|r_{X,Y}| > 0,8$) zu sprechen.

Mit Hilfe dieser Größen können wir eine Formel für die *empirische Regressionsgerade* (oder *Trendgerade*) aufstellen, welche mit Hilfe der *Methode der kleinsten Quadrate* hergeleitet werden kann. Ziel ist es, falls ein linearer Zusammenhang zwischen Daten existiert, diesen durch eine Gerade zu beschreiben. Wenn es einen linearen Zusammenhang gäbe, wäre Y von der Form

$$Y = a \cdot X + b \cdot \mathbf{1}$$

mit Koeffizienten $a, b \in \mathbb{R}$. Die Koeffizienten der Trendgerade sollen so bestimmt werden, dass die Norm des Fehlervektors

$$\|v(a, b)\| = \|a \cdot X + b \cdot \mathbf{1} - Y\|$$

minimal wird.

Satz 7.45: Satz über die Trendgerade

Seien $X, Y \in \mathbb{R}^n$ zwei Datenvektoren mit $\delta_X \neq (0, \dots, 0)$. Dann ist der Fehler $\|v(a, b)\|$ mit $v(a, b) := a \cdot X + b \cdot \mathbf{1} - Y$ minimal für die Parameter

$$a^* := \frac{\langle \delta_X, \delta_Y \rangle}{\|\delta_X\|^2} = \frac{\langle X, Y \rangle - n\bar{x}\bar{y}}{\|X\|^2 - n\bar{x}^2} \quad \text{und} \quad b^* := \bar{y} - a^* \cdot \bar{x}.$$

Die Trendgerade ist gegeben durch

$$\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid y = a^* x + b \right\}.$$

¹Als Beispiele kann man sich u.a. folgende Paare von Merkmalen vorstellen: Größe und Gewicht von Personen; Gesamtpunktzahl in den Übungen und Punktzahl in der Klausur von Studierenden; Alter eines Autos und anfallende Kosten beim TÜV.

Da dieser Weg analog zu der bereits bekannten Herleitung der Regressionsgerade für Zufallsvariablen ist, verzichten wir auf den Beweis.

Um die Rechnungen möglichst einfach zu halten, betrachten wir im folgenden Beispiel eine sehr kleine Stichprobe von zwei Merkmalen, wobei die Ausprägungen jeweils zur gleichen Person oder Beobachtungseinheit gehören.

Beispiel 7.46. Gegeben sind folgende Daten:

k	1	2	3	4	5
x_k	1	1	4	3	6
y_k	3	2	1	2	0



Berechnen Sie, falls möglich, die zugehörige Trendgerade.

Antwort: Berechnen wir zunächst die Abweichungsvektoren δ_X und δ_Y , sowie die $\|\delta_X\|^2$, da dies darüber entscheidet, ob die Trendgerade von Y in Abhängigkeit von X existiert. Die arithmetischen Mittel sind

$$\bar{x} = \quad \text{und} \quad \bar{y} =$$

Damit sind die Abweichungsvektoren gegeben durch

$$\delta_X = \quad \text{und} \quad \delta_Y =$$

Es ist

$$\|\delta_X\|^2 =$$

d.h. die Trendgerade existiert und wir können die Koeffizienten a^* und b^* berechnen.

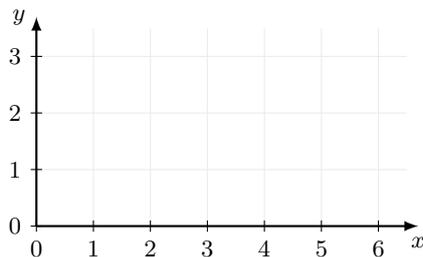
$$a^* =$$

$$b^* =$$

Somit ist die Regressionsgerade gegeben durch

$$\{(x, y)^T \in \mathbb{R}^2 \mid y = -0,5 \cdot x + 3,1\}.$$

Sehen wir uns das Ergebnis im Koordinatensystem an:



Bemerkung 7.47 (Korrelation vs. Kausalität). Besteht zwischen Merkmalen X und Y eine Kausalität, so gibt es einen ursächlichen Zusammenhang, der sich in den Messreihen widerspiegeln kann. In diesem Fall ist eine starke Korrelation zu erwarten, wobei wieder zu beachten ist, dass es auch nichtlineare Zusammenhänge gibt, die durch den Korrelationskoeffizienten nicht angezeigt werden. Andersherum kann jedoch von einer starken Korrelation grundsätzlich nicht auf eine Kausalität geschlossen werden. Eine ganze Sammlung von Merkmalen, die sich im Zeitverlauf auffallend ähnlich verhalten, bei denen man jedoch nicht von Kausalität ausgehen kann, sind auf der Seite <http://www.tylervigen.com/spurious-correlations> gesammelt.

7.4.4 Box-Plots

Zum Darstellen der Lage- und Streuungsparameter eignen sich sogenannte *Box-Plots*:

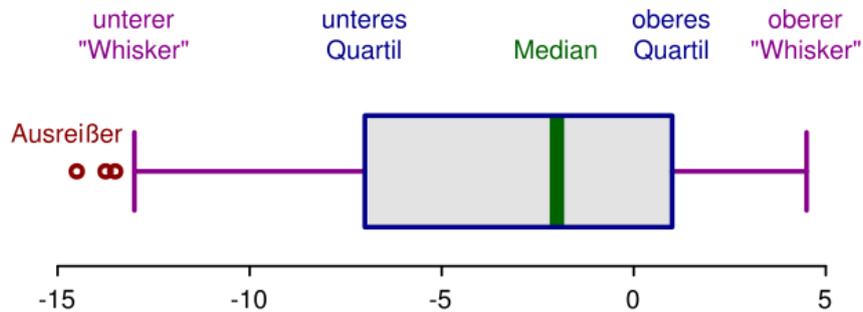


Abbildung 7.1: horizontaler Box-Plot — Quelle: Wikipedia

Wir kennen bereits die Quartile und den Median. Interessant ist die Frage, was als „Ausreißer“ gilt und was die „Whisker“ demnach markieren. Es gibt hier keine ganz einheitlich Vorgehensweise. Eine Möglichkeit besteht darin, *normale Beobachtungen* als diejenigen zu definieren, die innerhalb des Intervalls

$$[x_{0,25} - 1,5 \cdot (x_{0,75} - x_{0,25}), x_{0,75} + 1,5 \cdot (x_{0,75} - x_{0,25})] \quad (7.4.2)$$

liegen. Die kleinste und größte normale Beobachtung, d.h. die kleinste und größte Beobachtung, die noch im Intervall liegen, markieren wir dann durch die *Whisker*.

Beispiel 7.48. Wie sieht der Box-Plot zu folgenden Klausurergebnissen aus?

3 4 3.3 1 2 5 2.7 2.3 3

Antwort: Zunächst sortieren wir die $n = 9$ Werte:

Damit können wir die Quartile und den Median der Stichprobe berechnen:

$$x_{0,25} =$$

$$x_{0,5} =$$

$$x_{0,75} =$$

Berechnen wir damit das Intervall für die Whisker gemäß (7.4.2):

$$[x_{0,25} - 1,5 \cdot (x_{0,75} - x_{0,25}), x_{0,75} + 1,5 \cdot (x_{0,75} - x_{0,25})] =$$

Die Whisker haben damit die Werte \quad und \quad . Die einzige nicht normale Beobachtung ist \quad . Der Box-Plot sieht damit wie folgt aus:

Kapitel 8

Modellierung in der Stochastik – zwei Beispiele

Ziele

- Das Ziegenproblem kennen und sowohl die falsche Intuition als auch die korrekte Lösung verstehen
- Das Bertrand-Problem kennen und sich der Problematik beim Modellieren bewusst sein

8.1 Das Ziegenproblem

Das Problem kann man wie folgt beschreiben:

Am Ende einer Spielshow hat der siegreiche Kandidat die Chance, ein Auto zu gewinnen. Er steht vor drei Türen – hinter einer steht das Auto und hinter den anderen beiden sind Ziegen. Das Auto gehört ihm nur dann, wenn er die richtige Tür wählt.

Er entscheidet sich für Tür 1. Der Moderator (Monty Hall) öffnet Tür 3, hinter der eine Ziege steht. Dann bietet er dem Kandidaten an, seine Entscheidung noch einmal zu überdenken, also zu Tür 2 zu wechseln.

Sollte der Kandidat zu Tür 2 wechseln oder bei Tür 1 bleiben?

(Fehlgeleitete) Intuition: Zu Beginn hat der Kandidat drei Türen, hinter denen das Auto stehen kann – seine Gewinnwahrscheinlichkeit liegt also bei $\frac{1}{3}$. Nachdem der Moderator eine Tür geöffnet hat, sind nur noch zwei Türen da, hinter denen das Auto sein kann. Daher ist die Gewinnwahrscheinlichkeit bei beiden Türen nun $\frac{1}{2}$.

Problem: Der Moderator darf nicht die Tür mit dem Hauptgewinn öffnen. Falls der Kandidat bisher die Tür mit der Ziege gewählt haben sollte, kann der Moderator nur die andere Ziege zeigen – durch seine Reaktion verrät er also etwas, wodurch sich die Gewinnwahrscheinlichkeiten ändern.

Anschauliche Antwort: Sehen wir uns drei mögliche Fälle an, wenn wir davon ausgehen, dass der Kandidat ursprünglich Tür 1 gewählt hat.

Tür 1	Tür 2	Tür 3	Möglichk. des Moderators	Gewinn mit Wechsel	Gewinn ohne Wechsel
Auto	Ziege	Ziege	Tür 2 oder 3 öffnen	nein	ja
Ziege	Auto	Ziege	Tür 3 öffnen	ja	nein
Ziege	Ziege	Auto	Tür 2 öffnen	ja	nein

Wir sehen also, dass in 2 von 3 Fällen der Wechsel zum Gewinn führt.

Wahrscheinlichkeitstheoretische Beschreibung: Wir gehen im Folgenden immer davon aus, dass der Kandidat zunächst Tür 1 wählt und definieren folgende Ereignisse für $k \in \{1, \dots, 3\}$:

G_k : „Gewinn ist hinter Tür k “.

M_k : „Moderator öffnet Tür k “.

Wir suchen jeweils die Wahrscheinlichkeiten $\mathbb{P}(G_i|M_j)$ für $i \neq j$ und $j \in \{2, 3\}$, da der Moderator nicht die vom Kandidaten gewählte Tür öffnet und der Gewinn nie hinter der vom Moderator geöffneten Tür ist, d.h. es gilt $\mathbb{P}(G_k|M_k) = 0$ für alle $k \in \{1, 2, 3\}$. Somit ist, falls der Moderator Tür 3 öffnet,

$$\mathbb{P}(G_1|M_3) = \frac{\mathbb{P}(M_3|G_1)\mathbb{P}(G_1)}{\mathbb{P}(M_3|G_1)\mathbb{P}(G_1) + \mathbb{P}(M_3|G_2)\mathbb{P}(G_2) + \mathbb{P}(M_3|G_3)\mathbb{P}(G_3)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{1}{3}$$

$$\mathbb{P}(G_2|M_3) = \frac{\mathbb{P}(M_3|G_2)\mathbb{P}(G_2)}{\mathbb{P}(M_3|G_1)\mathbb{P}(G_1) + \mathbb{P}(M_3|G_2)\mathbb{P}(G_2) + \mathbb{P}(M_3|G_3)\mathbb{P}(G_3)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3}$$

$$\mathbb{P}(G_3|M_3) = 0,$$

d.h. durch das Wechseln erhöht sich die Gewinnwahrscheinlichkeit von $\frac{1}{3}$ auf $\frac{2}{3}$.

Woher kommt die falsche Intuition? Das Problem wird im ersten Moment in Gedanken so modelliert, als würde der Moderator rein zufällig eine der anderen Türen öffnen – in diesem Fall wäre es tatsächlich egal, ob der Kandidat seine Wahl ändert oder nicht. Die Bedingung für den Moderator, welche seine Entscheidung jedoch stark beeinflusst, wird dabei ignoriert. Somit geht die zusätzliche Information, die der Moderator liefert, verloren.

8.2 Das Bertrand-Paradoxon

Die Aussage, um die es bei dem Paradoxon geht, ist folgende:

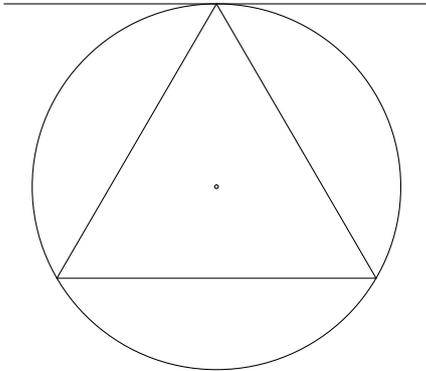
Wahrscheinlichkeiten müssen nicht wohldefiniert sein so lange der zu Grunde liegende Wahrscheinlichkeitsraum oder die Methode, die die Zufallsvariable produziert, nicht eindeutig definiert ist.

In *Calcul des probabilités* (1888) hat Joseph Louis François Bertrand (1822-1900) ein Problem formuliert und mehrere Lösungsansätze präsentiert, die zu verschiedenen Lösungen führen. Dieses wollen wir uns nun genauer ansehen.

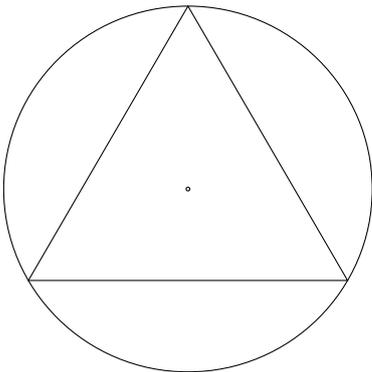
Fragestellung: Man zieht eine Sehne in einem Kreis *zufällig*. Was ist die Wahrscheinlichkeit dafür, dass sie kleiner als die Seite des einbeschriebenen gleichseitigen Dreiecks ist?

Problem: Was genau versteht man darunter, dass die Sehne *zufällig* einem Kreis einbeschrieben wird?

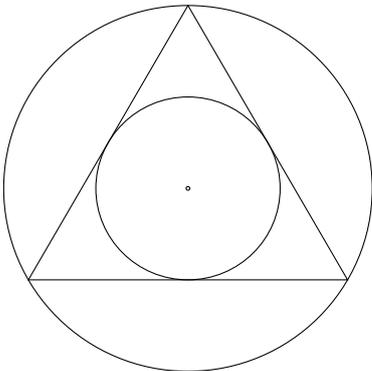
- 1. Lösungsansatz:** „Wenn einer der Endpunkte der Sehne bekannt ist, ändert diese Kenntnis die Wahrscheinlichkeit nicht.“ In diesem Fall kann die Richtung der Sehne frei gewählt werden, wobei wir darunter verstehen, dass wir den Winkel zwischen der Sehne und der Tangente an den Kreis im festgelegten Endpunkt der Sehne rein zufällig in $(0, \pi)$ wählen. Die Sehne ist genau dann länger als die Seite des Dreiecks, wenn der Winkel zwischen $\frac{\pi}{3}$ und $\frac{2\pi}{3}$ liegt.



- 2. Lösungsansatz:** „Wenn man die Richtung der Sehne kennt, ändert diese Kenntnis die Wahrscheinlichkeit nicht.“ In diesem Fall kann der Abstand R der Sehne zum Kreismittelpunkt rein zufällig in $[0, r)$ gewählt werden, wenn der Kreis Radius $r > 0$ besitzt. Die Sehne ist genau dann länger als die Seite des Dreiecks, wenn $R < \frac{r}{2}$ ist.



- 3. Lösungsansatz:** „Eine Sehne zufällig zu bestimmen bedeutet, ihren Mittelpunkt zufällig zu bestimmen.“ In diesem Fall wird der Mittelpunkt rein zufällig im Außenkreis des Dreiecks, also dem Kreis mit Radius r gewählt. Die Sehne ist genau dann länger als die Seite des Dreiecks, wenn der Mittelpunkt der Sehne innerhalb des Inkreises des Dreiecks liegt.



Anhang A

Relevante Inhalte der Analysis

Das Belegen von Analysis I und Analysis II wird für diesen Kurs vorausgesetzt. Ihre Lücken können Sie selbständig während des Semesters mit entsprechender Literatur, beispielsweise [Tim03] und [Tim02], schließen. Im Folgenden werden die wichtigsten Definitionen und Resultate zum Nachschlagen angegeben.

A.1 Fakultät und Binomialkoeffizient

Für das Kapitel 1.4 zur Kombinatorik benötigen wir die Begriffe *Fakultät* und *Binomialkoeffizient*.

Definition A.1

Die Fakultät einer natürlichen Zahl $n \in \mathbb{N}$ ist definiert als

$$n! := \prod_{k=1}^n k$$

und man setzt $0! := 1$.

Die Fakultät gibt uns die Anzahl möglicher Anordnungen der Elemente der Menge $\{1, \dots, n\}$.

Definition A.2

Seien $x \in \mathbb{R}$ und $k \in \mathbb{N}$. Dann ist der Binomialkoeffizient x über k definiert als

$$\binom{x}{k} := \frac{x(x-1) \cdot \dots \cdot (x-n+1)}{n!}.$$

Zudem setzen wir $\binom{x}{0} := 1$.

Der Binomialkoeffizient $\binom{n}{k}$ gibt uns für $n, k \in \mathbb{N}_0$ die Anzahl Möglichkeiten, k Elemente aus einer Menge von n Elementen auszuwählen ohne Beachtung der Reihenfolge und ohne Zurücklegen.

Folgende Eigenschaften sollten aus Analysis bekannt sein.

Satz A.3

1. Sind $n, k \in \mathbb{N}_0$, so gilt

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!}, & \text{falls } n \geq k, \\ 0, & \text{falls } n < k. \end{cases}$$

2. Für alle $x \in \mathbb{R}$ gilt $\binom{x}{1} = x$.

3. Erfüllen $n, k \in \mathbb{N}_0$ die Bedingung $n \geq k$, so ist $\binom{n}{k} = \binom{n}{n-k}$.

4. Es gilt

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

für alle $x \in \mathbb{R}$ und $k \in \mathbb{N}_0$.

Insbesondere gilt als Verallgemeinerung der binomischen Formeln der binomische Lehrsatz, welcher mit Hilfe vollständiger Induktion bewiesen wird.

Satz A.4: Binomischer Lehrsatz

Seien $x, y \in \mathbb{R}$ und $n \in \mathbb{N}$. Dann gilt

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

Zudem gilt folgendes Additionstheorem für Binomialkoeffizienten.

Satz A.5

Seien $x, y \in \mathbb{R}$ und $n \in \mathbb{N}_0$. Dann gilt:

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

A.2 Folgen, Reihen, Funktionenfolgen und Konvergenzbegriffe**Definition A.6**

Eine Folge $(a_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$ heißt konvergent, falls ein $a \in \mathbb{R}$ existiert, so dass gilt:

$$\forall \varepsilon > 0 \exists k_0 \in \mathbb{N} \forall k \geq k_0: |a_k - a| < \varepsilon.$$

Eine Folge, die gegen 0 konvergiert, heißt *Nullfolge*. Eine Folge in \mathbb{R} heißt *Cauchy-Folge*, falls folgende Bedingung erfüllt ist:

$$\forall \varepsilon > 0 \exists n \in \mathbb{N} \forall k, j \geq n: |a_k - a_j| < \varepsilon.$$

Wichtige Folgen

- $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$ für alle $x \in \mathbb{R}$;
- $\lim_{n \rightarrow \infty} \sqrt[n]{a} = 1$ für alle $a > 0$;
- $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$;
- $\lim_{n \rightarrow \infty} \sqrt[n]{n!} = \infty$.

Satz A.7: Konvergenzkriterien

1. Jede reelle Folge ist genau dann konvergent, wenn sie eine Cauchy-Folge ist.
2. Jede monoton wachsende (fallende), nach oben (unten) beschränkte Folge ist konvergent.
3. Wenn $\sum_{k \in \mathbb{N}} a_k$ konvergent ist, dann ist $(a_k)_{k \in \mathbb{N}}$ eine Nullfolge.

Mit Hilfe der Konvergenz von Folgen kann man Grenzwerte von Funktionen definieren.

Definition A.8

Seien $n \in \mathbb{N}$ und $M \subset \mathbb{R}^n$. Eine Funktion $f: M \rightarrow \mathbb{R}$ besitzt in $a \in \mathbb{R}^n$ den (uneigentlichen) Grenzwert $g \in \mathbb{R} \cup \{\pm\infty\}$, falls

1. eine Folge $(x_k)_{k \in \mathbb{N}} \subset M \setminus \{a\}$ existiert mit $\lim_{k \rightarrow \infty} x_k = a$ und
2. für jede Folge $(x_k)_{k \in \mathbb{N}} \subset M \setminus \{a\}$ mit $\lim_{k \rightarrow \infty} x_k = a$ gilt $\lim_{k \rightarrow \infty} f(x_k) = g$.

Notation: $\lim_{x \rightarrow a} f(x) = g$.

Der einseitige Grenzwert ist wie folgt definiert:

Definition A.9

Seien $f: M \rightarrow \mathbb{R}$ mit $M \subset \mathbb{R}$ und $a \in \mathbb{R}$.

Betrachtet man in Definition A.8 nur Folgen $(x_n)_{n \in \mathbb{N}}$ mit $x_n < a$ für alle $n \in \mathbb{N}$, so nennt man g *linksseitigen Grenzwert* von f in a und schreibt

$$f(a_-) = \lim_{x \rightarrow a_-} f(x) = \lim_{x \nearrow a} f(x) = \lim_{x \uparrow a} f(x) = g.$$

Betrachtet man in Definition A.8 nur Folgen $(x_n)_{n \in \mathbb{N}}$ mit $x_n > a$ für alle $n \in \mathbb{N}$, so nennt man g *rechtsseitigen Grenzwert* von f in a und schreibt

$$f(a_+) = \lim_{x \rightarrow a_+} f(x) = \lim_{x \searrow a} f(x) = \lim_{x \downarrow a} f(x) = g.$$

Für die Konvergenz von Reihen benötigt man noch den Begriff des Häufungspunktes einer Folge:

Definition A.10

Eine Zahl $a \in \mathbb{R}$ heißt Häufungspunkt (HP) einer Folge $(a_n)_{n \in \mathbb{N}}$, falls für ein beliebiges $\varepsilon > 0$ die Ungleichung $|a - a_n| < \varepsilon$ für unendlich viele $n \in \mathbb{N}$ erfüllt ist. Der größte HP einer Folge heißt Limes Superior, der kleinste HP heißt Limes Inferior und man schreibt $\limsup_{n \rightarrow \infty} a_n$ bzw. $\liminf_{n \rightarrow \infty} a_n$.

Bevor wir zu Reihen kommen, gibt es ein paar sehr nützliche Summenformeln:

Wichtige Summenformeln

$$\sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x} \quad \text{für alle } x \neq 1$$

$$\sum_{k=0}^n k = \frac{n(n+1)}{2} \quad \sum_{k=0}^n k^2 = \frac{n(n+1)(2n+1)}{6} \quad \sum_{k=0}^n k^3 = \frac{n^2(n+1)^2}{4}$$

Definition A.11

Sei $(a_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$ eine Folge. Dann heißt die Folge der Partialsummen $(s_n)_{n \in \mathbb{N}}$, welche gegeben ist durch

$$s_n := \sum_{k=1}^n a_k, \quad n \in \mathbb{N},$$

Reihe mit den Gliedern a_k , $k \in \mathbb{N}$. Die Reihe heißt *konvergent* (in \mathbb{R}), falls $s := \lim_{n \rightarrow \infty} s_n$ (in \mathbb{R}) existiert. Andernfalls heißt die Reihe *divergent*. Man schreibt für die Reihe auch kurz $\sum_{k=1}^{\infty} a_k$. Die Reihe heißt *absolut konvergent*, falls $\sum_{k=1}^{\infty} |a_k|$ konvergent ist.

Folgende Reihen muss man kennen.

Wichtige Reihen

- geometrische Reihe: $\sum_{k=0}^{\infty} ax^k = \frac{a}{1-x}$ für $|x| < 1$
- (verallgemeinerte) harmonische Reihe: $\sum_{k=1}^{\infty} \frac{1}{k^\alpha} = \infty$ für $\alpha \leq 1$
- Exponentialreihe: $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$ für alle $x \in \mathbb{R}$

Satz A.12: Konvergenzkriterien für Reihen

Leibniz-Kriterium: Ist $(a_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$ eine monoton fallende Nullfolge, so konvergiert die alternierende Reihe $\sum_{k \in \mathbb{N}} (-1)^k a_k$.

Quotientenkriterium: Sei $a_k \neq 0$ für $k \in \mathbb{N}$. Dann gilt:

Ist $\limsup_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} < 1$, dann konvergiert $\sum_{k=1}^{\infty} a_k$ absolut,

Ist $\liminf_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} > 1$, dann divergiert $\sum_{k=1}^{\infty} a_k$.

Wurzelkriterium:

Ist $\limsup_{k \rightarrow \infty} \sqrt[k]{|a_k|} < 1$, dann konvergiert $\sum_{k=1}^{\infty} a_k$ absolut,

Ist $\limsup_{k \rightarrow \infty} \sqrt[k]{|a_k|} > 1$, dann divergiert $\sum_{k=1}^{\infty} a_k$.

Wenn der Grenzwert $\lim_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} \in \mathbb{R}$ existiert, dann existiert auch der Grenzwert $\lim_{k \rightarrow \infty} \sqrt[k]{|a_k|} \in \mathbb{R}$ und beide Grenzwerte sind gleich. Die Umkehrung gilt nicht.

Anders als bei endlichen Summen dürfen bei Reihen die Summanden nicht einfach beliebig umgeordnet werden – es sei denn, die Reihe ist absolut konvergent.

Definition A.13

Ist $\sum_{k=1}^{\infty} a_k$ eine Reihe und $f: \mathbb{N} \rightarrow \mathbb{N}$ bijektiv, dann heißt die Reihe $\sum_{k=1}^{\infty} a_{f(k)}$ *Umordnung* von $\sum_{k=1}^{\infty} a_k$.

Satz A.14: Umordnungssatz

Konvergiert eine Reihe $\sum_{k=1}^{\infty} a_k$ absolut, so konvergiert auch jede Umordnung $\sum_{k=1}^{\infty} a_{f(k)}$ und die Werte stimmen überein: $\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} a_{f(k)}$.

Für Funktionenfolgen gibt es verschiedene Konvergenzbegriffe, deren Unterschied man sich an folgendem Beispiel überlegen kann.

Beispiel A.15. Für $n \in \mathbb{N}$ ist die Funktionenfolge

$$f_n(x) = \begin{cases} 0, & x \leq n \\ 1, & x > n \end{cases}$$

punktweise konvergent gegen die Funktion $f \equiv 0$, aber sie ist nicht gleichmäßig konvergent im Sinne der folgenden Definition.

Definition A.16

Für $n \in \mathbb{N}$ und $D \subseteq \mathbb{R}$ seien $f_n: D \rightarrow \mathbb{R}$ Funktionen.

1. Man sagt, f_n konvergiert *punktweise* gegen $f: D \rightarrow \mathbb{R}$, wenn

$$\forall \varepsilon > 0 \forall x \in D \exists N \in \mathbb{N} \forall n \geq N \quad |f_n(x) - f(x)| < \varepsilon.$$

2. Man sagt, f_n konvergiert *gleichmäßig* (in D) gegen $f: D \rightarrow \mathbb{R}$, wenn

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall x \in D \forall n \geq N \quad |f_n(x) - f(x)| < \varepsilon.$$

A.3 Stetigkeit, Differenzierbarkeit**Definition A.17: Stetigkeit 1**

Für $n \in \mathbb{N}$ sei $M \subset \mathbb{R}^n$. Weiterhin erfülle $a \in \mathbb{R}^n$ Bedingung 1 aus Definition A.8. Eine Funktion $f: M \rightarrow \mathbb{R}$ heißt *stetig im Punkt a* , falls $\lim_{x \rightarrow a} f(x) = f(a)$. Die Funktion heißt *stetig*, falls f stetig in jedem Punkt ihrer Definitionsmenge ist.

Eine Funktion heißt zudem stetig in jedem isolierten Punkt ihres Definitionsbereiches.

Definition A.18: Stetigkeit 2

$f: M \rightarrow \mathbb{R}$ stetig in $a \in M$ genau dann wenn

$$\forall \varepsilon > 0 \exists \delta > 0 : \quad x \in M, |x - a| < \delta \implies |f(x) - f(a)| < \varepsilon.$$

Es gibt auch stärkere Begriffe als Stetigkeit, beispielsweise gleichmäßige Stetigkeit, Hölder-Stetigkeit und Lipschitz-Stetigkeit. In der Vorlesung wird folgender Begriff auftauchen:

Definition A.19

Sei $I \subseteq \mathbb{R}$ ein Intervall. Eine Abbildung $f: I \rightarrow \mathbb{R}$ heißt *absolut stetig*, falls zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass zu endlich vielen disjunkten Intervallen $J_k = (a_k, b_k) \subseteq I$ mit Gesamtlänge $\sum_k |J_k| < \delta$ stets $\sum_k |f(b_k) - f(a_k)| < \varepsilon$ gilt.

Es gilt:

$$f \text{ Lipschitz-stetig} \implies f \text{ absolut stetig} \implies f \text{ gleichmäßig stetig} \implies f \text{ stetig.}$$

Absolut stetige Funktionen sind fast überall differenzierbar mit endlicher Ableitung. Damit gilt der Hauptsatz der Analysis in folgender Form:

Satz A.20: Hauptsatz der Analysis

Sei $F: [a, b] \rightarrow \mathbb{R}$ absolut stetig. Dann ist F fast überall differenzierbar, die Ableitung F' ist integrierbar und es gilt

$$\int_a^b F'(x) dx = F(b) - F(a).$$

Definition A.21

Seien $M \subset \mathbb{R}$, $f: M \rightarrow \mathbb{R}$ und $a \in M$. Falls der Grenzwert

$$g := \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \in \mathbb{R}$$

existiert, dann heißt f differenzierbar in a und g ist die Ableitung von f in a .

Ableitungsregeln

Sei $M \subset \mathbb{R}$ und seien $f, g: M \rightarrow \mathbb{R}$ in $a \in M$ differenzierbar und $h: f(M) \rightarrow \mathbb{R}$ in $f(a)$ differenzierbar. Dann gilt für die Ableitungen:

Produktregel: $(f \cdot g)'(a) = f'(a)g(a) + f(a)g'(a)$;

Quotientenregel: $\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{(g(a))^2}$ falls $g(a) \neq 0$;

Kettenregel: $(g \circ f)'(a) = g'(f(a)) \cdot f'(a)$.

Satz A.22

Sei $M \subset \mathbb{R}$ und sei $f: M \rightarrow \mathbb{R}$ differenzierbar. Sei zudem $f: M \rightarrow f(M)$ invertierbar mit Umkehrfunktion f^{-1} . Sei $y \in f(M)$ derart, dass $f'(f^{-1}(y)) \neq 0$. Dann ist f^{-1} in y differenzierbar, und es gilt

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}.$$

Folgende Ableitungsregeln werden als bekannt vorausgesetzt:

Ableitungen wichtiger Funktionen

$$(ax^b)' = abx^{b-1}$$

$$(e^x)' = e^x$$

$$(a^x)' = a^x \ln a$$

$$(\sin x)' = \cos x$$

$$(\cos x)' = -\sin x$$

$$(\ln x)' = \frac{1}{x}$$

Definition A.23: partielle Ableitung

Seien $n \in \mathbb{N}$, $M \subset \mathbb{R}^n$, $a = (a_1, \dots, a_n) \in M$ und $f: M \rightarrow \mathbb{R}$ gegeben. Wenn für ein $j \in \{1, \dots, n\}$ der Grenzwert

$$\lim_{h \rightarrow 0} \frac{f(a_1, \dots, a_{j-1}, a_j + h, a_{j+1}, \dots, a_n) - f(a)}{h} =: g \in \mathbb{R}$$

existiert, so heißt f *partiell differenzierbar in Richtung von x_j* (oder in der j -ten Komponente) und g heißt *partielle Ableitung von f in Richtung von x_j* .

Notation: $g = \frac{\partial f}{\partial x_j}(a) = \partial_{x_j} f(a) = f_{x_j}(a)$ oder auch verkürzt $\partial_j f(a)$.

Definition A.24

Ist $f: M \rightarrow \mathbb{R}$ differenzierbar in $a \in M$, so heißt der Vektor v mit

$$v^T = \left(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right)$$

Gradient von f in a .

A.4 Integrationsregeln

Während es für eine Abbildung $f: \mathbb{R} \rightarrow \mathbb{R}$ nur einen Begriff für Differenzierbarkeit gibt, existieren verschiedene Integral-Begriffe – die wichtigsten sind das Riemann- und das Lebesgue-Integral. Auf die genauen Definitionen muss hier nicht eingegangen werden, da die Rechenregeln übereinstimmen. Die im Folgenden genannten Regeln und Stammfunktionen sind diejenigen, die als bekannt vorausgesetzt werden.

Definition A.25: Stammfunktion

Seien $I \subset \mathbb{R}$ ein Intervall und $f: I \rightarrow \mathbb{R}$ eine Funktion. Eine differenzierbare Funktion $F: I \rightarrow \mathbb{R}$ heißt *Stammfunktion von f* , wenn $F'(x) = f(x)$ gilt für alle $x \in I$. Das *unbestimmte Integral* $\int f(x)dx$ ist die Menge aller Stammfunktionen von f .

Satz A.26: partielle Integration

Die Funktionen $f, g: [a, b] \rightarrow \mathbb{R}$ seien differenzierbar und $f \cdot g'$ besitze eine Stammfunktion. Dann gilt:

$$\int_a^b f'(x)g(x)dx = - \int_a^b f(x)g'(x)dx + f(b)g(b) - f(a)g(a).$$

Satz A.27: Substitution

Die Funktion $g: [a, b] \rightarrow I \subset \mathbb{R}$ sei differenzierbar und $f: I \rightarrow \mathbb{R}$ besitze eine Stammfunktion. Dann gilt:

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(y)dy.$$

Wichtige Stammfunktionen

Für $n \in \mathbb{Z} \setminus \{-1\}$ und $a > 0$ gilt:

$$\int x^n dx = \frac{1}{n+1}x^{n+1} + C,$$

$$\int \frac{1}{x} dx = \ln|x| + C,$$

$$\int e^x dx = e^x + C,$$

$$\int a^x dx = \frac{a^x}{\ln a} + C,$$

$$\int \sin x dx = -\cos x + C,$$

$$\int \cos x dx = \sin x + C.$$

Es können auch Integrale über 2- oder 3-dimensionalen Mengen auftauchen. Wie diese zu berechnen sind, wiederholen wir jetzt.

Definition A.28

Sei $A \subset \mathbb{R}^n$ für $n \in \{2, 3\}$. Die Menge A heißt *Normalbereich*, falls

$n = 2$: $a_1, b_1 \in \mathbb{R}$ und stetige Funktionen $a_2, b_2: \mathbb{R} \rightarrow \mathbb{R}$ existieren, so dass

$$A = \{(x, y) \in \mathbb{R}^2 \mid a_1 \leq x \leq b_1, a_2(x) \leq y \leq b_2(x)\}.$$

$n = 3$: $a_1, b_1 \in \mathbb{R}$, stetige Funktionen $a_2, b_2: \mathbb{R} \rightarrow \mathbb{R}$ und $a_3, b_3: \mathbb{R}^2 \rightarrow \mathbb{R}$ existieren, so dass

$$A = \{(x, y, z) \in \mathbb{R}^3 \mid a_1 \leq x \leq b_1, a_2(x) \leq y \leq b_2(x), a_3(x, y) \leq z \leq b_3(x, y)\}.$$

Satz A.29

Ist $A \subset \mathbb{R}^n$ ($n \in \{2, 3\}$) ein Normalbereich und $f: A \rightarrow \mathbb{R}$ über A integrierbar, so berechnet man das Integral $\int_A f(x) dx$ wie folgt:

$n = 2$: Ist $A = \{(x, y) \in \mathbb{R}^2 \mid a_1 \leq x \leq b_1, a_2(x) \leq y \leq b_2(x)\}$, so ist

$$\int_A f(x, y) d(x, y) = \int_{a_1}^{b_1} \left(\int_{a_2(x)}^{b_2(x)} f(x, y) dy \right) dx.$$

$n = 3$: Ist $A = \{(x, y, z) \in \mathbb{R}^3 \mid a_1 \leq x \leq b_1, a_2(x) \leq y \leq b_2(x), a_3(x, y) \leq z \leq b_3(x, y)\}$, so ist

$$\int_A f(x, y, z) d(x, y, z) = \int_{a_1}^{b_1} \left(\int_{a_2(x)}^{b_2(x)} \left(\int_{a_3(x, y)}^{b_3(x, y)} f(x, y, z) dz \right) dy \right) dx.$$

Beispiel A.30. Sei $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$. Dies ist ein Normalbereich, denn

$$A = \{(x, y) \in \mathbb{R}^2 \mid -1 \leq x \leq 1, -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}\}.$$

Integrieren wir über A die konstante Funktion $f(x) \equiv 1$, so ist (mit Substitution $x = \sin u$, $dx = \cos u du$)

$$\begin{aligned} \int_A f(x, y) d(x, y) &= \int_{-1}^1 \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1 dy \right) dx \\ &= \int_{-1}^1 \left(\sqrt{1-x^2} - (-\sqrt{1-x^2}) \right) dx \\ &= \int_{-1}^1 2\sqrt{1-x^2} dx \\ &= \int_{\arcsin(-1)}^{\arcsin(1)} 2\sqrt{1-(\sin u)^2} \cos u du \\ &= \int_{\arcsin(-1)}^{\arcsin(1)} 2(\cos u)^2 du \end{aligned}$$

Man berechnet eine Stammfunktion als

$$\int (\cos u)^2 du = \frac{1}{2} (u + \sin u \cos u)$$

und erhält (mit $\arcsin(-x) = -\arcsin(x)$, $\cos(\arcsin x) = \sqrt{1-x^2}$ und $\arcsin(1) = \frac{\pi}{2}$)

$$\begin{aligned} \int_A f(x, y) d(x, y) &= [u + \sin u \cos u]_{\arcsin(-1)}^{\arcsin(1)} \\ &= \arcsin(1) - \arcsin(-1) + 1 \cdot \cos(\arcsin(1)) - (-1) \cdot \cos(\arcsin(-1)) \\ &= 2 \arcsin(1) + \cos(\arcsin(1)) + \cos(\arcsin(-1)) \\ &= 2 \arcsin(1) + \sqrt{1-1^2} + \sqrt{1-(-1)^2} \\ &= \pi. \end{aligned}$$

Dies ist der Flächeninhalt des Einheitskreises.

Satz A.31: Transformationssatz

Sei $B \subseteq \mathbb{R}^d$ eine offene Menge und sei $g: B \rightarrow g(B) \subseteq \mathbb{R}^d$ ein Diffeomorphismus, d.h. eine bijektive, stetig differenzierbare Abbildung mit stetig differenzierbarer Umkehrabbildung. Dann ist eine Abbildung f auf $g(B)$ genau dann integrierbar, wenn $f \circ g \cdot |\det(Dg)|$ auf B integrierbar ist und in diesem Fall ist

$$\int_B f(y) dy = \int_{g(B)} f(g(x)) |\det(Dg(x))| dx.$$

Dieser Satz wird angewendet, wenn bei der Integration die Transformation zu Polar-, Zylinder- oder Kugelkoordinaten gemacht wird.

Koordinatentransformation

Die Funktionaldeterminanten $D := \det(g)$ zu den wichtigsten Transformationen:

Polarkoordinaten: $D = r$ für

$$g(r, \phi) := \begin{pmatrix} r \cos(\phi) \\ r \sin(\phi) \end{pmatrix}, \quad (r, \phi) \in [0, \infty) \times [0, 2\pi)$$

Zylinderkoordinaten: $D = r$ für

$$g(r, \phi, z) := \begin{pmatrix} r \cos(\phi) \\ r \sin(\phi) \\ z \end{pmatrix}, \quad (r, \phi, z) \in [0, \infty) \times [0, 2\pi) \times \mathbb{R}$$

Kugelkoordinaten: $D = r^2 \sin \theta$ für

$$g(r, \theta, \phi) := \begin{pmatrix} r \cos(\phi) \sin \theta \\ r \sin(\phi) \sin \theta \\ r \cos \theta \end{pmatrix}, \quad (r, \theta, \phi) \in [0, \infty) \times [0, \pi) \times [0, 2\pi)$$

Anhang B

Kontroll- und Ergänzungsfragen

B.1 Zu Kapitel 1

1. Können Sie **die** Distributivgesetze für Ereignisse aus Satz 1.4 mit Hilfe von Venn-Diagrammen veranschaulichen?
2. Sind folgende Rechenregeln für beliebige Ereignisse $A, B, C \subseteq \Omega$ korrekt? Mit welchen Rechenregeln aus der Vorlesung können Sie dies begründen?

a) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

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

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

3. Wie zeigt man, dass $\mathcal{P}(\Omega)$ eine σ -Algebra ist?
4. Warum definiert man das W'maß \mathbb{P} nicht als Abbildung $\mathbb{P}: \Omega \rightarrow [0, 1]$?
5. Wie begründet man die Schlussfolgerung aus (1.2.1) im Beweis von Satz 1.15 genau, d.h. warum gilt $\mathbb{P}(\emptyset) = 0$?
6. Unter welcher (notwendigen und hinreichenden) Bedingung gilt für zwei Ereignisse $A, B \subseteq \Omega$ folgende Formel?

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

7. Ist folgende Formel für Ereignisse A_1, \dots, A_n korrekt?

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \left((-1)^{k+1} \sum_{I \subseteq \{1, \dots, n\}, |I|=k} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \right)$$

Tipp: Sehen Sie sich die Siebformel (Satz 1.16) nochmal an.

8. Formulieren Sie die Aussage des binomischen Lehrsatzes (Satz A.4) für $(x, y) = (1, 0)$ und $(x, y) = (1, 1)$.
9. Begründen Sie Satz A.5 für $x, y, n \in \mathbb{N}_0$ mit Hilfe der kombinatorischen Interpretation des Binomialkoeffizienten.
10. Wie viele ganzzahlige Lösungen der Gleichung $\sum_{i=1}^k x_i = n$ gibt es für $k \leq n$, falls

a) $x_i > 0, \forall i \in \{1, \dots, k\}$?

b) $x_i \geq 0, \forall i \in \{1, \dots, k\}$?

B.2 Zu Kapitel 2

1. Gegeben sind die Ereignisse

A : „Neal ist ein Dieb.“

B : „Neal wird verurteilt.“

- a) Beschreiben Sie folgende Wahrscheinlichkeiten $\mathbb{P}(A \cap B)$, $\mathbb{P}(A|B)$ und $\mathbb{P}(B|A)$ mit Worten.
 - b) Peter ist sich zu 90% sicher, dass Neal der Dieb ist und jeder tatsächliche Dieb wird mit einer Wahrscheinlichkeit von 70% schuldig gesprochen und verurteilt. Wie wahrscheinlich ist es, dass Neal ein Dieb ist und verurteilt wird?
2. In einem Museum wird eine Sonderausstellung zu 38% von Einheimischen und zu 62% von Auswärtigen besucht. Von den Einheimischen sind 55% von den Auswärtigen 48% Frauen.
- a) Wie groß ist der Anteil der weiblichen Besucher?
 - b) Der 10 000. Besucher der Ausstellung soll einen Buchpreis bekommen. Falls es eine Frau ist: Mit welcher Wahrscheinlichkeit kommt sie von außerhalb?
 - c) Falls der 20 000. Besucher ein Mann ist: Mit welcher Wahrscheinlichkeit ist er ein Einheimischer?
3. Sehen Sie sich den Beweis von Satz 2.8 nochmal an. Die bedingte Wahrscheinlichkeit wird für zwei Fälle unterschiedlich definiert. Begründen Sie, warum die Formel für $n = 2$ in beiden möglichen Fällen korrekt ist.
4. Überlegen Sie sich für jedes Gleichheitszeichen im Beweis von Satz 2.22, warum es korrekt ist.
5. In Definition 2.24 wird abstrakt angegeben, wie man die stochastischen Unabhängigkeit von endlich oder unendlich vielen Ereignissen überprüft. Für $n = 2$ Ereignisse muss eine Gleichung überprüft werden, für $n = 3$ bereits vier Gleichungen. Wie viele Gleichungen muss man für den Nachweis der stochastischen Unabhängigkeit von n Ereignissen überprüfen?
6. Zeigen Sie: Sind die Ereignisse A, B, C, D unabhängig, so sind auch $A \cap B$ und $(C \cap D)^c$ unabhängig.
7. Für eine Primzahl p sei $\Omega = \{1, 2, \dots, p\}$ der Grundraum zu einem Laplace-Experiment. Die Ereignisse $A, B \subseteq \Omega$ seien unabhängig. Zeigen Sie, dass dann mindestens eines der Ereignisse A und B das sichere oder das unmögliche Ereignis ist.
8. Gehen Sie nochmal die Gleichungsketten im Beweis von Satz 2.29 durch. Können Sie zu jedem Gleichheitszeichen sagen, warum es korrekt ist?

B.3 Zu Kapitel 3

1. Welche der folgenden Zufallsvariablen sind diskret verteilt?
 - a) die Summe der Augenzahlen dreier fairer Würfel
 - b) die Temperatur in Adlershof am nächsten Montag 12:00 Uhr, gerundet auf eine Stelle nach dem Komma
 - c) die exakte Temperatur in Adlershof am nächsten Montag 12:00 Uhr
 - d) die Gesamtanzahl der Kunden an den Kassen in dem Moment, in dem man sich selbst anstellt
2. Überprüfen Sie die am Ende gegebene Formel für $\mathbb{P}(Z = n)$ in Beispiel 3.7 indem Sie die Summen in $\mathbb{P}(Z > n - 1) - \mathbb{P}(Z > n)$ als eine Summe schreiben und vereinfachen.
3. Um zu zeigen, dass die Massenfunktion der hypergeometrischen Verteilung wohldefiniert ist, muss dank Satz 3.10 nur gezeigt werden, dass die Werte der Massenfunktion zu 1 summiert werden.
 - a) Begründen Sie mit den Mitteln der Kombinatorik, warum folgende Formel korrekt ist für $\max\{n, K\} \leq N$:

$$\sum_{k=0}^n \binom{K}{k} \binom{N-K}{n-k} = \binom{N}{n}.$$
 - b) Schließen Sie daraus, dass $\sum_{k=0}^n \mathbb{P}(X = k) = 1$ ist für $X \sim \text{Hyp}(N, K, n)$.
4. Sei $X \sim \text{Geo}(p)$ für ein $p \in (0, 1)$. Überprüfen Sie, dass $\sum_{k \in \mathbb{N}} \mathbb{P}(X = k) = 1$ ist.
5. Gehen Sie die Berechnung von $\mathbb{E}[X^2]$ in Beispiel 3.58 nochmal durch. Können Sie bei jedem Schritt genau sagen, was passiert?
6. Es liegen Karten mit den Zahlen 1 bis 5 verdeckt auf dem Tisch. Ein Spieler bekommt 1 Euro wenn eine ungerade Zahl gezogen wird und er verliert 1 Euro wenn eine gerade Zahl gezogen wird. Wie hoch müsste der Einsatz sein, damit das Spiel fair ist?
7. Damit wir $X: \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ nennen, soll X messbar sein, d.h. $X^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{B}(\mathbb{R})$. Wozu haben wir diese Forderung gestellt?
8. Überprüfen Sie folgende Urbildoperationen für eine Zufallsvariable X und Mengen $A, B \in \mathcal{B}(\mathbb{R})$.
 - a) $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$
 - b) $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$
 - c) $X^{-1}(A^c) = (X^{-1}(A))^c$

B.4 Zu Kapitel 4

1. Sei X gleichverteilt auf $[0, 1]$. Wie lauten dann Verteilungsfunktion und Dichte von $Y = 2X$?
2. T sei die Zeit zwischen zwei aufeinander folgenden Bargeldabholungen an einem Bankautomaten. Wir nehmen an, T sei exponentialverteilt zum Parameter $\lambda = 1$ [min^{-1}].
 - a) Wie groß ist die erwartete Zeit zwischen zwei aufeinanderfolgenden Abhebungen?
 - b) Wie groß ist die Standardabweichung von T ?
 - c) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(T \leq 4)$.
 - d) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(2 \leq T \leq 5)$.

3. Sei X eine reellwertige Zufallsvariable mit $\mathbb{E}[X^2] < \infty$. Zeigen Sie folgende Formel für die mittlere quadratische Abweichung der Zufallsvariablen X von einem beliebigen Wert $a \in \mathbb{R}$:

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

Für welche(s) $a \in \mathbb{R}$ wird die mittlere quadratische Abweichung minimal?

4. Sei $X \sim U([0, 1])$. Bestimmen Sie die Verteilungsfunktion und Dichte der Zufallsvariablen $Y := X^2$.
5. Sei $X \sim U([0, 1])$ und für $\lambda > 0$ sei $Y = -\frac{1}{\lambda} \ln(1 - X)$. Geben Sie den Wertebereich, die Verteilungsfunktion und die Dichte von Y an. Ist Y gemäß einer Ihnen bekannten Verteilung verteilt?
6. Die erreichte Punktzahl in einer Klausur sei annähernd $\mathcal{N}(\mu, \sigma^2)$ -verteilt mit $\mu = 20$ und $\sigma = 4$. Es werden nur die Noten 1 (sehr gut), 2 (gut), 3 (befriedigend), 4 (mangelhaft) und 5 (ungenügend) vergeben, keine Zwischenabstufungen. Etwa 10% der Schüler erhalten die Note 1, etwa 7% die Note 5.
 - a) Bestimmen Sie die Mindestpunktzahl für die Note 'sehr gut'.
 - b) Ab welcher Punktzahl ist ein Schüler durchgefallen, d.h. ab wann bekommt er die Note 'ungenügend'?
7. Sei $X \sim \text{Bin}(n, p)$. Warum gilt für ein $k \in \{0, \dots, n\}$ dann stets $\mathbb{P}(X \leq k) = \mathbb{P}(X \leq k + \frac{1}{2})$?
8. Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Wie groß ist $\mathbb{P}(X = \mu)$?
9. Das Kasino behauptet, dass man bei einem Automaten mit 50% Wahrscheinlichkeit gewinnt und verliert und der Einsatz ist gleich dem möglichen Gewinn, so dass das Spiel fair ist. Ben gewinnt bei 4 850 von 10 000 Spielen. Wie schätzen Sie es ein – hatte er einfach Pech oder hat das Kasino die Gewinnwahrscheinlichkeit als zu hoch angesetzt?

B.5 Zu Kapitel 5

1. Die gemeinsame Verteilung der Zufallsvariablen X und Y sei gegeben durch die folgende Tabelle:

$X \backslash Y$	1	2	3
1	0,2	0,1	0,15
2	0,1	0,05	0,25
3	0	0,15	0

- a) Bestimmen Sie die Randverteilungen von X und Y .
 b) Geben Sie die gemeinsame Verteilung von X und Y an, falls sie stochastisch unabhängig sind und die Randverteilungen aus a) besitzen.
2. Die gemeinsame Dichte zweier Zufallsvariablen X und Y wird angegeben als

$$f(x, y) = \begin{cases} 0,25, & 0 \leq x \leq 4 \text{ und } 0 \leq y \leq 1, \\ 0, & \text{sonst.} \end{cases}$$

- a) Stellen Sie die Funktion an Stelle der Fallunterscheidung mit Hilfe von Indikatorfunktionen dar.
 b) Überprüfen Sie, dass es sich dabei tatsächlich um eine Dichte handelt.
 c) Berechnen Sie $\mathbb{P}(2 \leq X \leq 3, \frac{1}{2} \leq Y \leq 1)$.
3. Seien X, Y und Z drei stochastisch unabhängige Zufallsvariablen, die alle gleichverteilt sind auf $[0, 1]$. Berechnen Sie (unter Verwendung des Ergebnisses aus Beispiel 5.18) die Dichte von $X + Y + Z$.
4. Für Zufallsvariablen X, Y und Z gelte

$$\begin{aligned} \text{Var}(X) &= 4, & \text{Var}(Y) &= 16, & \text{Var}(Z) &= 25, \\ \text{Cov}(X, Y) &= -2, & \text{Cov}(X, Z) &= 0, & \text{Cov}(Y, Z) &= 0,5. \end{aligned}$$

Berechnen Sie $\text{Var}(X + Y + Z)$ und $\text{Var}(X - Y - Z)$.

5. Aus folgt folgende Ungleichung, vorausgesetzt dass die Varianzen und die Kovarianz existieren:

$$(\text{Cov}(X, Y))^2 \leq \text{Var}(X) \cdot \text{Var}(Y). \quad (\text{B.5.1})$$

- a) Beweisen Sie (B.5.1) mit Hilfe von Lemma 5.47.
 b) Formel (B.5.1) wird auch als *Cauchy-Schwarz'sche Ungleichung* bezeichnet. Überlegen Sie sich, dass Cov eine positiv semidefinite Bilinearform auf

$$\mathcal{L}^2(\mathbb{P}) := \{ X : \Omega \rightarrow \mathbb{R} \text{ Zufallsvariable} \mid \mathbb{E}[X^2] < \infty \}$$

definiert, jedoch kein Skalarprodukt.

6. Unter welchen Bedingungen an die Zufallsvariablen X und Y gelten folgende Rechenregeln?
- a) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$,
 b) $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$,
 c) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Anhang C

Python Programme

Ich bin kein Spezialist in Python, daher sind folgende Programme, die im Laufe der Veranstaltung zum Einsatz kamen, möglicherweise nicht optimal programmiert. Für Verbesserungsvorschläge bin ich offen.

Geburtstagsproblem – Beispiel 1.43

```
list = []
list.append(1.0)
proba = 1.0
i = 0
while proba >= 0.3:
    i+=1
    proba = list[i-1]*(365-i+1)/365
    list.append(proba)
    print(i, list[i])
```

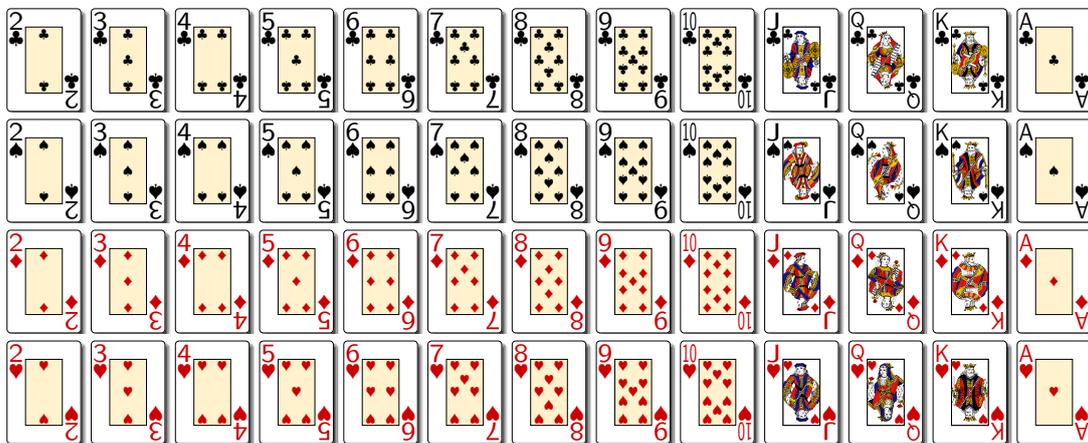
Tabelle der Normalverteilung erzeugen

```
import numpy as np
import pandas as pd
from scipy.stats import norm
a = np.arange(0,4.0,0.1)
b = np.arange(0,0.1,0.01)
X, Y = np.meshgrid(b, a)
tabular = pd.DataFrame(data=norm.cdf(X+Y), columns=b, index=a)
print(tabular)
```


Anhang D

Kartenspiele

Als Referenz für die zahlreichen Aufgaben mit Kartenspielen ist hier eine Übersicht über die Karten eines Spiels mit 52 Karten, auf das wir uns beziehen:



Zeilenweise haben die Karten die gleiche *Farbe* (Kreuz, Pik, Karo, Herz); spaltenweise haben die Karten den gleichen *Wert* (2,3,4,5,6,7,8,9,10, Bube, Dame, König, Ass).

Literaturverzeichnis

- [FLP15] FISCHER, Gerd ; LEHNER, Matthias ; PUCHERT, Angela: *Einführung in die Stochastik*. Springer, 2015
- [Hen08] HENZE, Norbert: *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*. 7. Auflage. Springer-Verlag, 2008
- [Hur08] HURLEY, W. J.: The birthday matching problem when the distribution of birthdays is nonuniform. In: *CHANCE* 21 (2008), Dec, Nr. 4, 20–24. <https://doi.org/10.1007/s00144-008-0035-1>
- [Kö99] KÖNIGSBERGER, Konrad: *Analysis 1*. 4. Auflage. Springer, 1999. – ISBN 3540412824
- [Kre05] KRENGEL, Ulrich: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 8., erweiterte Auflage. Wiesbaden : vieweg, 2005 (Vieweg Studium, Aufbaukurs Mathematik)
- [KS11] KÜTTING, Herbert ; SAUER, J. M. ; PADBERG, Friedhelm (Hrsg.): *Elementare Stochastik. Mathematische Grundlagen und didaktische Konzepte*. 3. Auflage. Berlin, Heidelberg : Springer, 2011 (Mathematik Primarstufe und Sekundarstufe I + II)
- [Mil17] MILLER, Steven J.: *The probability lifesaver: All the tools you need to understand chance*. Princeton University Press, 2017
- [Tim02] TIMMANN, Steffen: *Repetitorium der Analysis. Teil 2*. Binomi, 2002
- [Tim03] TIMMANN, Steffen: *Repetitorium der Analysis. Teil 1*. 2. Auflage. Binomi, 2003