

Mathematische und statistische Methoden für Pharmazeuten

Skript zur Vorlesung im Sommersemester 2020 in Regensburg

Florian Strunk

Inhaltsverzeichnis

Allgemeines	1
Elementare mathematische Grundlagen	2
1. Deskriptive Statistik	13
1.1. Grundlegende Begriffe	13
1.2. Eindimensionale Stichproben	15
1.2.1. Absolute und relative Häufigkeiten	15
1.2.2. Gruppierungen	16
1.2.3. Die empirische Verteilungsfunktion	17
1.2.4. Lage- und Streuungsmaße	19
1.3. Zweidimensionale Stichproben	23
1.3.1. Empirische Unabhängigkeit	25
1.3.2. Kovarianz und Korrelationskoeffizient	28
1.3.3. Regressionsrechnung	31
2. Wahrscheinlichkeitstheorie	34
2.1. Grundlegende Begriffe	34
2.2. Elemente der Kombinatorik	38
2.3. Bedingte Wahrscheinlichkeiten	42
2.4. Unabhängige Ereignisse	47
2.5. Zufallsvariablen	48
2.5.1. Endliche Zufallsvariablen	50
2.5.2. Stetige Zufallsvariablen	55
2.6. Der zentrale Grenzwertsatz	61
2.7. Das Gesetz der großen Zahlen	62
3. Induktive Statistik	63
3.1. Grundlegende Begriffe	64
3.2. Punktschätzungen (durch die Maximum-Likelihood Methode)	66
3.3. Konfidenzintervalle	67
3.4. Hypothesentests	69
3.4.1. Hypergeometrischer Test	72
3.4.2. Gauß-Test (Test auf Erwartungswert bei bekannter Varianz)	73
3.4.3. t -Test (Test auf Erwartungswert bei unbekannter Varianz)	73
3.4.4. χ^2 -Test (Test auf Varianz)	74
3.4.5. Approximativer Binomialtest	74
3.4.6. Unabhängigkeits und Homogenitätstests	75
Tabellen	75

Allgemeines

Dieses ist ein kompaktes Skript zu der Vorlesung **Mathematische und statistische Methoden für Pharmazeuten**, die ich im Sommersemester 2020 an der Universität Regensburg halte. Üblicherweise handelt es sich bei dieser Veranstaltung um eine (wöchentlich) zweistündige Vorlesung mit (wöchentlich) zweistündigen Übungen.

Aufgrund der außergewöhnlichen Situation durch das **Coronavirus** in diesem Semester, wird die Vorlesung (zumindest zunächst) nicht „physisch“ stattfinden. Es wird, wie auf der G.R.I.P.S.-Seite zu dieser Vorlesung genauer erläutert ist, einen wöchentlichen Leseauftrag in diesem Skript geben. Zusätzlich wird wöchentlich ein kurzes Video auf der G.R.I.P.S.-Seite mit einer Anleitung zu diesem Leseauftrag bereitgestellt.

Es ist mir klar, dass Ihnen als Pharmaziestudierende nur eine begrenzte Studienzeit für die Mathematik zur Verfügung steht. Ich habe den Eindruck, dass die für Sie als „Anwender“ gedachte mathematische Literatur durch ausführliche schriftliche Erklärungen oft sehr umfangreich gestaltet ist, was zwar einerseits zum Selbststudium ermutigen kann aber erfahrungsgemäß oft zu einer Überfrachtung führt und damit den gegenteiligen Effekt haben könnte. Daher habe ich mich entschlossen ein sehr konzises Skript zu schreiben. Dieses hat den Vorteil, daß Sie am Ende der Vorlesung hoffentlich eine kompakte Zusammenfassung des Vorlesungsinhalts haben werden. Andererseits brauchen Sie sich nicht zu wundern, wenn Sie beim ersten Lesen dieses Skripts nicht so schnell wie gewohnt vorankommen und Abschnitte mehrfach lesen müssen.

Der folgende Pharmazieexperte



Was mache ich hier eigentlich?

wird Ihnen an einigen Stellen auf umgangssprachliche Weise Erklärungs- und Fokussierungshilfen mit auf den Weg zu geben. Außerdem tauchen manchmal in **grüner Schrift** Erklärungen im Text auf, die überlesen werden können und teilweise erst mit Definitionen und Begriffen aus späteren Teilen des Skripts zu verstehen sind.

Das Skript zu dieser Vorlesung beruht hauptsächlich auf dem von Georg Dolzmann im Sommer 2017 zur gleichen Vorlesung angefertigten Skript [1], welches wiederum Skripte zur gleichen Vorlesung von Matthias Wilke [3] und Daniela Schiefeneder [2] als Quellen zitiert. Selbstverständlich liegt die Verantwortung für alle Fehler und Ungenauigkeiten bei mir. Ich freue mich über Korrekturvorschläge an florian.strunk@ur.de.

Elementare mathematische Grundlagen

Bevor wir mit dem statistischen Teil dieser Vorlesung beginnen, möchte ich kurz an einige mathematische Objekte und Tatsachen erinnern, die im Folgenden von großer Wichtigkeit sind.

Definition 1. Eine **Menge** ist eine Zusammenfassung von einzelnen Objekten, deren **Elementen**, zu einem Ganzen. Ist M eine Menge, so bezeichnet $|M|$ die Anzahl deren Elemente. Ist x ein Element von M , so schreibt man $x \in M$. Andernfalls schreibt man $x \notin M$. Zwei Mengen sind genau dann gleich, wenn sie die gleichen Elemente enthalten.

Beispiel 2. Es sind

$$A := \{\text{Auto, Kamel, Rad, Hund}\}, \quad B := \{\Delta, \diamond, \heartsuit\}, \quad C := \{0, 1, 2, 3, \dots\}$$

Beispiele von Mengen. Die erste Menge A hat vier Elemente, nämlich das Element „Auto“, das Element „Kamel“, das Element „Rad“ und das Element „Hund“. Es gilt also $|A| = 4$ und beispielsweise $\text{Rad} \in A$. Die zweite Menge B hat drei Elemente, also $|B| = 3$ und die dritte Menge C unendlich viele, also $|C| = \infty$.



Das Symbol $:=$ anstatt eines normalen Gleichheitszeichens $=$ bedeutet, dass man den Ausdruck auf der linken Seite (also der Seite mit dem Doppelpunkt) erst definiert. Es ist nicht falsch, einfach $B = \{\Delta, \diamond, \heartsuit\}$ zu schreiben. Die Doppelpunkt-Schreibweise dient nur der Verbesserung der Lesbarkeit, da man so direkt merkt, dass das Symbol auf der linken Seite bisher noch keine Bedeutung hatte.

Beispiel 3. In der Beschreibung der Elemente einer Menge kommt es nicht auf Wiederholungen oder die Reihenfolge an. Es ist also beispielsweise

$$\{\Delta, \diamond, \heartsuit\} = \{\diamond, \Delta, \heartsuit\} = \{\diamond, \diamond, \Delta, \heartsuit\}.$$

Definition 4. Es gibt genau eine Menge, die keine Elemente hat. Sie wird die **leere Menge** genannt und mit \emptyset oder $\{\}$ bezeichnet.

Beispiel 5. Die Mengen $\emptyset = \{\}$ und $D := \{\emptyset\}$ sind verschieden. Die Menge \emptyset hat keine Elemente, während die Menge D ein Element hat, nämlich die leere Menge. Es gilt also $\emptyset \in D$. Eine weitere, von diesen beiden Mengen verschiedene Menge ist die zweielementige Menge $E := \{\emptyset, D\}$. Hier gilt $\emptyset \in E$ und $D \in E$.

Bemerkung 6. Es ist (zumindest für uns) kein Problem, dass wir, wie in dem vorherigen Beispiel, Mengen als Elemente von neuen Mengen auffassen. Streng genommen muss man hier etwas vorsichtig sein, denn es kann die Menge, die als Elemente alle Mengen enthält aus logischen Gründen (dieses bedeutet nicht: aus offensichtlichen Gründen) nicht geben. Wieso dieses so ist, möchte ich hier nicht erklären und dieses Problem der Einfachheit halber in diesem Kurs ignorieren.

Man kann aus Mengen durch verschiedene Konstruktionen „neue Mengen“ bilden. Ist M eine Menge und ϕ eine Eigenschaft ihrer Elemente x , sodass $\phi(x)$ entweder wahr oder falsch ist, so können wir die Menge

$$\{x \in M \mid \phi(x) \text{ ist wahr}\}$$

bilden.

Beispiel 7. Betrachten wir wieder die Menge $A = \{\text{Auto, Kamel, Rad, Hund}\}$ aus dem ersten Beispiel. Es sei ϕ die Eigenschaft „ist ein Tier“, so ist $\phi(x)$ für alle Elemente x aus A entweder wahr oder falsch. Es ist beispielsweise $\phi(\text{Auto})$ falsch, da ein Auto kein Tier ist und es gilt

$$\{x \in A \mid \phi(x) \text{ ist wahr}\} = \{\text{Kamel, Hund}\}$$

Definition 8. Sind M und N Mengen, so heißt N eine **Teilmenge** der Menge M , wenn jedes Element von N auch ein Element von M ist. Wir schreiben in diesem Fall $N \subseteq M$.

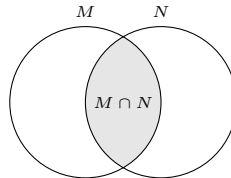
Bemerkung 9. Es gilt also $M = N$ genau dann, wenn $M \subseteq N$ und $N \subseteq M$.

Weitere (und hoffentlich bekannte) Möglichkeiten aus Mengen „neue Mengen“ zu bilden sind die Folgenden. Es seien dafür M und N Mengen.

- Die Menge

$$M \cap N := \{x \mid x \in M \text{ und } x \in N\}$$

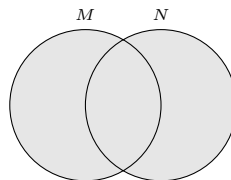
heißt der **Durchschnitt von M und N** .



- Die Menge

$$M \cup N := \{x \mid x \in M \text{ oder } x \in N\}$$

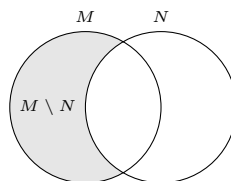
heißt die **Vereinigung von M und N** .



- Die Menge

$$M \setminus N := \{x \in M \mid x \notin N\}$$

heißt das **Komplement von N in M** .



Beispiel 10. Betrachten wir $F := \{0, 1, 2\}$ und $G := \{1, 2, 3, 4\}$, so ist $F \cap G = \{1, 2\}$, $F \cup G = \{0, 1, 2, 3, 4\}$ und $F \setminus G = \{0\}$ und $G \setminus F = \{3, 4\}$.

Die folgenden Mengen werden (zusammen mit ihren üblichen Rechenoperationen wie Addition und Multiplikation) als bekannt vorausgesetzt.

Definition 11.

- Es ist $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ die Menge der **natürlichen Zahlen**.
- Es ist $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ die Menge der **ganzen Zahlen**.
- Es ist $\mathbb{Q} = \{\frac{a}{b} \mid a \in \mathbb{Z} \text{ und } b \in \mathbb{Z} \setminus \{0\}\}$ die Menge der **rationalen Zahlen**.
- Es ist \mathbb{R} die Menge der **reellen Zahlen**.

Es gilt $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$.

Wie wir in Beispiel 3 gesehen haben, nimmt unsere Mengenschreibweise keine Rücksicht auf Reihenfolge und Wiederholungen. Hierfür gibt es den Begriff des Tupels. Ein **Tupel** (oder genauer ein **n -Tupel**) ist eine Liste

$$(x_1, x_2, \dots, x_n)$$

von Elementen x_1, x_2, \dots, x_n aus einer Menge M .

Beispiel 12. Es ist $(5, 0, 0, 1, 1)$ ein 5-Tupel aus natürlichen Zahlen. Man beachte, dass $\{5, 0, 0, 1, 1\} = \{0, 1, 5\}$, aber $(5, 0, 0, 1, 1) \neq (0, 1, 5)$.

Möglicherweise werden wir an einigen Stellen auch Tupel mit mehr als endlich vielen Einträgen betrachten oder Tupel, deren Einträge aus verschiedenen Mengen M stammen. Sind beispielsweise M_1, M_2, \dots, M_n Mengen, so bezeichnet $M_1 \times M_2 \times \dots \times M_n$ die Menge der n -Tupel (x_1, x_2, \dots, x_n) deren i -ter Eintrag x_i aus der Menge M_i stammt.

Als nächstes wollen wir den wichtigen Begriff der Abbildung kennenlernen.

Definition 13. Es seien A und B zwei Mengen. Eine **Abbildung** $f: A \rightarrow B$ von der Menge A in die Menge B ist eine Zuordnung, die jedem Element $x \in A$ genau ein Element $f(x) \in B$ zuordnet. Man schreibt hierfür

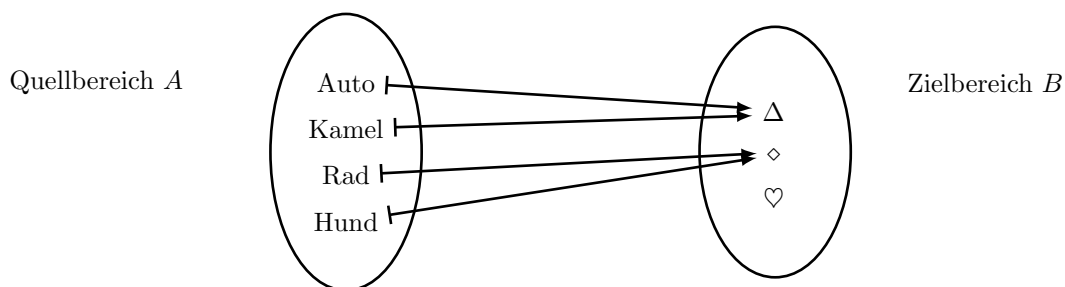
$$x \mapsto f(x).$$

Bemerkung 14. Wenn man eine Abbildung $f: A \rightarrow B$ angeben möchte, gehört also dazu, dass man den **Quellbereich** A und den **Zielbereich** B angibt. Außerdem muss man für jedes Element $x \in A$ sagen, wohin es abgebildet wird, also jeweils ein Element $f(x) \in B$ angeben.



Eine Abbildung ist genau das Gleiche wie eine **Funktion**. Diese Begriffe sind völlig synonym. In der Schule hat man oft nur Funktionen $f: \mathbb{R} \rightarrow \mathbb{R}$ betrachtet. Der Quellbereich oder Zielbereich wurde dort vielleicht Definitionsbereich oder Wertebereich genannt.

Beispiel 15. Wir betrachten die Mengen A und B aus Beispiel 2 und möchten eine Abbildung $f: A \rightarrow B$ angeben. Dafür setzen wir zum Beispiel $f(\text{Auto}) := \Delta$, $f(\text{Kamel}) := \Delta$, $f(\text{Rad}) := \diamond$ und $f(\text{Hund}) := \diamond$, wie die folgende Abbildung veranschaulicht.



Man beachte, dass wir für jedes Element x aus dem Quellbereich A genau ein Element $f(x) \in B$ angegeben haben, worauf dieses abgebildet wird. Allerdings gilt:

- Es können Elemente aus dem Zielbereich „mehrfach getroffen“ werden.
Dieses ist oben beispielsweise mit Δ der Fall.
- Es müssen nicht alle Elemente aus dem Zielbereich „getroffen“ werden.
Dieses ist oben beispielsweise mit \heartsuit der Fall.

Beispiel 16. Ist der Quellbereich unendlich groß, wie zum Beispiel bei einer Abbildung $f: \mathbb{R} \rightarrow \mathbb{R}$, so können wir nicht ohne Weiteres, wie im vorherigen Beispiel, alle Werte $f(x)$ einzeln angeben. Daher benutzen wir zur Angabe von $f(x)$ ein Ausdruck „der von x abhängt“ wie zum Beispiel

$$f: \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) := x^2 - 3.$$

Bei einer solchen Abbildung (Funktion), können wir gut einen **Funktionsgraphen** zeichnen. Hierfür zeichnen wir in der euklidischen Ebene $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, also der Menge aller 2-Tupel, deren Einträge jeweils reelle Zahlen sind, die Punkte $(x, f(x))$ ein. Der Funktionsgraph von f ist also die Teilmenge

$$\{(x, y) \in \mathbb{R}^2 \mid f(x) = y\} \subseteq \mathbb{R}^2.$$

In dem obigen Beispiel erhalten wir das folgende Bild.

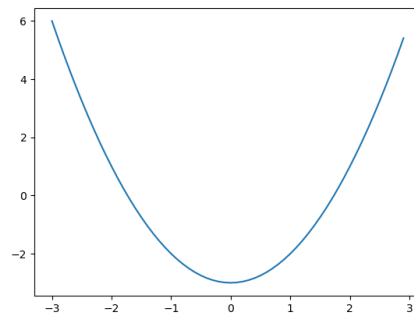
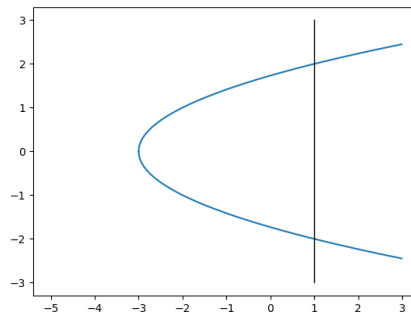


Abbildung 1: Graph der Funktion $f(x) = x^2 - 3$



Es ist wichtig, dass wir einen Unterschied machen zwischen „Funktionen“ und „Funktionswerten“: Es ist f aus dem vorherigen Beispiel 16 eine Funktion von \mathbb{R} nach \mathbb{R} , aber das Symbol $f(x)$ bezeichnet eine reelle Zahl, nämlich das Element aus dem Zielbereich der Abbildung f , auf das das Element x (aus dem Quellbereich) abgebildet wird. Anstatt $x \mapsto f(x)$ für die Abbildung f zu schreiben, werden wir manchmal auch die Abkürzung $f(-)$ verwenden (sofern Quell- und Zielbereich schon angegeben sind).

Beispiel 17. Das folgende (blaue) Bild ist **nicht** Funktionsgraph einer Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$.



Dieses hat zwei Gründe: Zum einen wird den Elementen x mit $x < -3$ im Quellbereich \mathbb{R} (der „X-Achse“) **kein** Funktionswert $f(x)$ zugeordnet. Zum anderen wird den anderen Elementen mit x im Quellbereich **kein eindeutiger** Funktionswert $f(x)$ zugeordnet, wie durch die vertikale Linie angedeutet ist. Das erste Problem könnte man beispielsweise vermeiden, indem man den Quellbereich auf die Teilmenge $\mathbb{R}_{\geq -3} = [-3, \infty) = \{x \in \mathbb{R} \mid x \geq -3\}$ der reellen Zahlen einschränkt.

Die folgende Definition ist aus der Schule bekannt.

Definition 18. Ein **Intervall** I ist eine Teilmenge $I \subseteq \mathbb{R}$ der reellen Zahlen von einer der folgenden Formen für Elemente $a, b \in \mathbb{R}$ mit $a \leq b$,

- $[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$ (Abgeschlossenes Intervall)
- $(a, b) := \{x \in \mathbb{R} \mid a < x < b\}$ (Offenes Intervall)
- $[a, b) := \{x \in \mathbb{R} \mid a \leq x < b\}$ (Halboffenes Intervall)
- $(a, b] := \{x \in \mathbb{R} \mid a < x \leq b\}$ (Halboffenes Intervall)

wobei a oder b auch $\pm\infty$ sein können, wenn sie neben einer runden Klammer stehen.

Nun möchten wir bestimmten Funktionen f eine Ableitungsfunktion f' zuordnen. Um diesen Begriff im Allgemeinen zu definieren, ist viel Formalismus nötig. Daher werden wir uns mit

einer graphischen Interpretation dieses Begriffs zufrieden geben und Ableitungsfunktionen nur für spezielle Funktionen explizit angeben. Betrachten wir also wieder den Funktionsgraphen der obigen Abbildung

$$\begin{aligned} f: I &\rightarrow \mathbb{R} \\ x &\mapsto x^2 - 3, \end{aligned}$$

so ist die **Ableitungsfunktion** (oder einfach nur: die **Ableitung**) eine Abbildung $f': I \rightarrow \mathbb{R}$ mit dem gleichen Quellbereich und Zielbereich wie f und der Eigenschaft, dass die reelle Zahl $f'(x)$ die Steigung der Tangentengeraden an den Funktionsgraphen von f in dem Punkt $(x, f(x))$ ist.

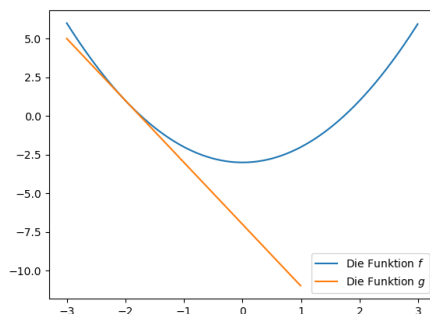
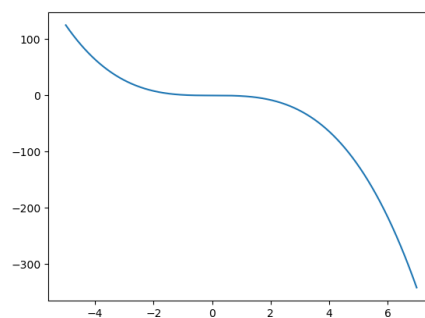
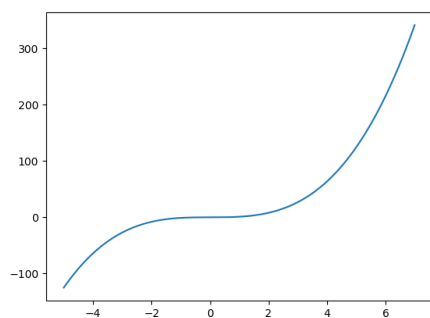


Abbildung 2: Graph der Funktion f mit Tangentengeraden g im Punkt $(-2, 1)$

Beispiel 19. Ohne irgendetwas genauer auszurechnen, können wir mit der obigen graphischen Interpretation sagen, dass die Ableitungsfunktion zu der Funktion mit dem folgenden linken Graphen nur nicht-negative (also positive oder Null) und die Ableitungsfunktion zu der Funktion mit dem folgenden rechten Graphen nur nicht-positive (also negative oder Null) Werte annimmt.



Definition 20. Es sei $I \subseteq \mathbb{R}$ ein Intervall. Eine **Polynomfunktion** ist eine Abbildung $f: I \rightarrow \mathbb{R}$ mit

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

für Elemente $a_i \in \mathbb{R}$, die **Koeffizienten** heißen. Die zugehörige **Ableitung** ist die Funktion $f': I \rightarrow \mathbb{R}$ definiert durch

$$f'(x) = a_n x^{n-1} + a_{n-1} x^{n-2} + \cdots + a_1.$$

Diese Abbildung ist wieder eine Polynomfunktion und besitzt also ebenfalls eine Ableitung $f'' := (f')'$, welche die **zweite Ableitung** der ursprünglichen Polynomfunktion f heißt. Ebenso bildet man alle **höheren Ableitungen** $f^{(k)}: I \rightarrow \mathbb{R}$.

Beispiel 21. Wir betrachten noch einmal die Abbildung f mit $f(x) = x^2 - 3$ aus dem Beispiel 16. Sie ist eine Polynomfunktion vom Grad 2. Ihre Ableitung ist nach der vorherigen Definition die Funktion $f': \mathbb{R} \rightarrow \mathbb{R}$ mit $f'(x) = 2x$, ihre zweite Ableitung ist die Funktion $f'': \mathbb{R} \rightarrow \mathbb{R}$ mit $f''(x) = 2$. Alle weiteren höheren Ableitungen sind die Nullfunktion. Nun wollen wir die Funktion g angeben, deren (orangefarbener) Graph in dem Bildchen 2 eingezeichnet ist. Diese Funktion ist durch $x \mapsto g(x) = mx + b$ gegeben. (Die reelle Zahl m wird in der Schule häufig „Steigung“ und die reelle Zahl b der „Y-Achsenabschnitt“ genannt.) Ihr Graph soll den Graphen von f im Punkt $(-2, f(-2)) = (-2, 1)$ berühren. Mit unserer graphischen Betrachtung oben gilt $m = f'(-2) = -4$. Um b zu bestimmen benutzen wir, dass wir wissen, dass $f(-2) = g(-2) = 1$ gilt, weil beide Graphen durch den Punkt $(-2, 1)$ laufen sollen. Auflösen nach b ergibt dann $g(x) = -4x - 7$.

Ist a eine reelle Zahl und k eine natürliche Zahl, so hat das Symbol a^k eine Bedeutung. Dieses ist die reelle Zahl

$$a^k := a \cdot \dots (k\text{-mal}) \cdot \dots \cdot a,$$

wobei man $a^0 := 1$ setzt. Sei nun a eine reelle Zahl mit $a > 0$. Um den Ausdruck a^k auf ganzzahlige Exponenten auszuweiten, definiert man $a^{-k} := \frac{1}{a^k}$. Wie haben also eine Abbildung

$$\begin{array}{ccc} a^{(-)}: \mathbb{Z} & \rightarrow & (0, \infty) \\ k & \mapsto & a^k \end{array}$$

mit der Eigenschaft $a^{(k+l)} = a^k a^l$ gegeben. Wir möchten diese Abbildung zu einer Abbildung

$$a^{(-)}: \mathbb{R} \rightarrow (0, \infty)$$

erweitern, also zu einer Abbildung, welche die gesamte Menge der reellen Zahlen als Quellbereich hat. Erweitern bedeutet hierbei, dass die neue Abbildung mit der alten Abbildung auf allen $k \in \mathbb{Z}$ übereinstimmen soll. Außerdem möchten wir, dass die neue Abbildung auch „ableitbar“ ist. Es stellt sich heraus, dass dieses möglich ist. Es gibt sogar eine eindeutige bestimmte reelle Zahl $a = e$, die **Eulersche Zahl**, für die diese neue Abbildung $e^{(-)}: \mathbb{R} \rightarrow (0, \infty)$, welche **Exponentialfunktion** (oder **e-Funktion**) genannt wird, als Ableitung ebenfalls die Exponentialfunktion $e^{(-)}: \mathbb{R} \rightarrow (0, \infty)$ hat. Außerdem werden wir die anderen Funktionen $a^{(-)}$ alle durch die Exponentialfunktion beschrieben können.

Definition 22. Die **Exponentialfunktion** $e^{(-)}: \mathbb{R} \rightarrow (0, \infty)$ ist die eindeutige Abbildung mit den Eigenschaften

- (1) $e^0 = 1$ und
- (2) $(e^{(-)})' = e^{(-)}$.

Sie hat die Eigenschaft, dass für alle reellen Zahlen k und l gilt

$$e^{(k+l)} = e^k e^l \quad \text{und} \quad e^{-k} = \frac{1}{e^k}.$$

Man kann die Eulersche Zahl $e := e^1$ numerisch approximieren und bekommt die bekannte Annäherung

$$e \approx 2,7182818 \dots$$

Schaut man sich den folgenden Funktionsgraphen der Exponentialfunktion $e^{(-)}: \mathbb{R} \rightarrow (0, \infty)$ (Die Funktion nimmt nur positive reelle Zahlen $\neq 0$ als Werte (auf der „Y-Achse“) an.)

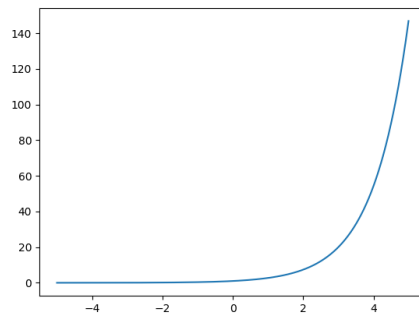


Abbildung 3: Funktionsgraph der Exponentialfunktion

an, so kann man vermuten, dass jede reelle Zahl $y \in (0, \infty)$ im Zielbereich der Exponentialfunktion auch tatsächlich „getroffen“ wird. Mit anderen Worten vermuten wir, dass es zu jedem $y \in (0, \infty)$ (auf der „Y-Achse“) ein $x \in \mathbb{R}$ (auf der „X-Achse“) gibt, mit $e^x = y$. Dieses ist tatsächlich der Fall und es gilt sogar noch mehr: Zu einem beliebig vorgegebenen $y \in (0, \infty)$ finden wir ein **eindeutiges** solches Element $x \in \mathbb{R}$. Dieses Element nennen wir $x = \ln(y)$. Wir erhalten also eine Abbildung $\ln: (0, \infty) \rightarrow \mathbb{R}$.

Definition 23. Die (natürliche) **Logarithmusfunktion** $\ln: (0, \infty) \rightarrow \mathbb{R}$ ist die eindeutige Abbildung mit der Eigenschaft

$$e^{\ln(y)} = y$$

für alle $y \in (0, \infty)$. Sie hat für alle $x \in \mathbb{R}$ die Eigenschaften

$$\ln(e^x) = x \quad \text{und} \quad \ln(1) = 0,$$

sowie für alle positiven reellen Zahlen y und z die Eigenschaften

$$\ln(yz) = \ln(y) + \ln(z) \quad \text{und} \quad \ln\left(\frac{1}{y}\right) = -\ln(y).$$

Die zugehörige **Ableitung** ist die Funktion $\ln': (0, \infty) \rightarrow \mathbb{R}$ gegeben durch

$$\ln'(y) = \frac{1}{y}.$$

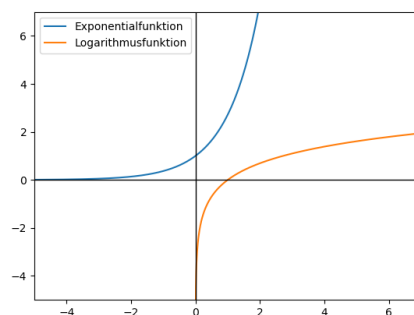


Abbildung 4: Funktionsgraphen der Exponential- und Logarithmusfunktion

Nun können wir die Potenzfunktion für eine beliebige reelle Zahl $a > 0$ definieren. Sie erweitert tatsächlich unsere anfänglich betrachtete Abbildung $a^{(\cdot)}: \mathbb{Z} \rightarrow (0, \infty)$, in die wir nur ganze Zahlen „einsetzen“ konnten.

Definition 24. Sei $a > 0$ eine reelle Zahl. Die **Potenzfunktion** $a^{(\cdot)}: \mathbb{R} \rightarrow (0, \infty)$ ist gegeben durch

$$a^k := e^{\ln(a) \cdot k}$$

für alle $k \in \mathbb{R}$. Sie hat die Eigenschaft, dass für alle reellen Zahlen k, l und $b > 0$ gilt

$$\begin{aligned} a^{(k+l)} &= a^k a^l & \text{und} & & a^{-k} &= \frac{1}{a^k} \\ (ab)^k &= a^k b^k & \text{und} & & (a^k)^l &= a^{kl} \end{aligned}$$

und sie stimmt für $a = e$ mit der Exponentialfunktion überein. Die zugehörige **Ableitung** ist die Funktion $(a^{(-)})' : (0, \infty) \rightarrow \mathbb{R}$ gegeben durch

$$(a^{(-)})'(k) = a^k \ln(a).$$

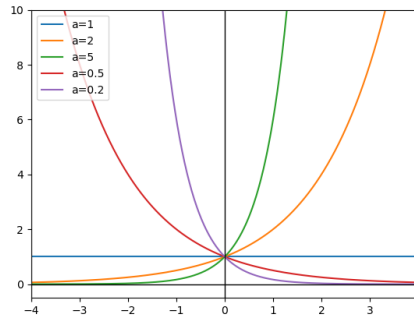


Abbildung 5: Funktionsgraphen einiger Potenzfunktionen

Bemerkung 25. Ist $a > 0$ eine reelle Zahl und $n > 0$ eine natürliche Zahl, so nennt man die reelle Zahl $\sqrt[n]{a} := a^{\frac{1}{n}}$ auch die **n -te Wurzel von a** . Aus den Eigenschaften der Potenzfunktion erhalten wir die Gleichungen $\sqrt[n]{ab} = \sqrt[n]{a} \cdot \sqrt[n]{b}$ und $\sqrt[m]{\sqrt[n]{a}} = \sqrt[nm]{a}$ für eine weitere natürliche Zahl $m > 0$ und eine reelle Zahl $b > 0$.



Wenn wir Wurzelausdrücke immer als Potenzen mit Brüchen als Exponenten schreiben, so brauchen wir uns nur die Eigenschaften der Potenzfunktion aus Definition 24 zu merken und nicht die speziellen Gleichungen aus der vorherigen Bemerkung.

Ist $a > 0$ mit $a \neq 1$ eine reelle Zahl, so sehen wir an den Funktionsgraphen für $a^{(-)}$, genau wie bei der Exponentialfunktion oben, dass jedes Element $y \in (0, \infty)$ im Zielbereich (auf der „Y-Achse“) eindeutig „getroffen“ wird. Mit anderen Worten gibt es also wieder zu einem beliebig vorgegebenen Element $y \in (0, \infty)$ ein eindeutiges $x \in \mathbb{R}$ mit $a^x = y$. Dieses Element nennen wir $x = \log_a(y)$.

Definition 26. Sei $a > 0$ mit $a \neq 1$ eine reelle Zahl. Die **Logarithmusfunktion (zur Basis a)** $\log_a : (0, \infty) \rightarrow \mathbb{R}$ ist die eindeutige Abbildung mit der Eigenschaft

$$a^{\log_a(y)} = y$$

für alle $y \in (0, \infty)$. Sie kann alternativ auch als $\log_a(y) = \frac{\ln(y)}{\ln(a)}$ definiert werden und hat für alle $x \in \mathbb{R}$ die Eigenschaften

$$\log_a(a^x) = x \quad \text{und} \quad \log_a(1) = 0,$$

sowie für alle positiven reellen Zahlen y und z die Eigenschaften

$$\log_a(yz) = \log_a(y) + \log_a(z) \quad \text{und} \quad \log_a\left(\frac{1}{y}\right) = -\log_a(y)$$

und sie stimmt für $a = e$ mit der natürlichen Logarithmusfunktion überein. Die zugehörige **Ableitung** ist die Funktion $\log'_a : (0, \infty) \rightarrow \mathbb{R}$ gegeben durch

$$\log'_a(y) = \frac{1}{y \cdot \ln(a)}.$$

Wir kennen aus der Schule Möglichkeiten, die Ableitungsfunktionen eines Produkts $g(-)h(-)$ und einer Komposition $h(g(-))$ von ableitbaren Funktionen g und h zu bestimmen. Außerdem können wir Konstanten und Summen „beim Ableiten herausziehen“. Diese Regeln sind in dem folgenden Satz aufgelistet.

Satz 27. Ist $f: \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion mit Ableitung f' , so gilt für alle $\lambda \in \mathbb{R}$, dass

$$f'(\lambda x) = \lambda f'(x). \quad (\text{Faktorregel})$$

Ist I ein Intervall und haben die Funktionen $g: I \rightarrow \mathbb{R}$ und $h: I \rightarrow \mathbb{R}$ jeweils Ableitungen, so ist die Ableitung der Funktion $h(-) + g(-): I \rightarrow \mathbb{R}$ gegeben durch

$$(g(-) + h(-))'(x) = g'(x) + h'(x). \quad (\text{Summenregel})$$

Ist I ein Intervall und haben die Funktionen $g: I \rightarrow \mathbb{R}$ und $h: I \rightarrow \mathbb{R}$ jeweils Ableitungen, so ist die Ableitung der Funktion $h(-)g(-): I \rightarrow \mathbb{R}$ gegeben durch

$$(g(-)h(-))'(x) = g'(x)h(x) + g(x)h'(x). \quad (\text{Produktregel})$$

Sind I und J Intervalle und haben die Funktionen $g: I \rightarrow J$ und $h: J \rightarrow \mathbb{R}$ jeweils Ableitungen, so ist die Ableitung der Funktion $h(g(-)): I \rightarrow \mathbb{R}$ gegeben durch

$$h(g(-))'(x) = h'(g(x))g'(x). \quad (\text{Kettenregel})$$

Wählen wir ein $a > 0$, so ist die Potenzfunktion eine Abbildung $a^{(-)}: \mathbb{R} \rightarrow (0, \infty)$. Wir fixieren hier also ein a und „lassen den Exponenten variabel“. Wir haben angegeben, wie man dieser Funktion eine Ableitungsfunktion zuordnet. Andererseits könnten wir auch für einen fixierten Exponenten $k \in \mathbb{R}$ die Funktion $(-)^k: (0, \infty) \rightarrow (0, \infty)$ betrachten. Mit anderen Worten ist dieses (siehe Definition 24) die Funktion $(-)^k = e^{\ln(-) \cdot k}$. Um zu zeigen, dass diese Funktion ableitbar ist und um ihre Ableitung auszurechnen, benutzen wir die Kettenregel aus dem vorherigen Satz 27. Sei also $h(-) := e^{(-)}$ mit Ableitung $h'(-) = e^{(-)}$ und $g(-) := k \ln(-)$ mit Ableitung $g'(-) = k \frac{1}{(-)}$ (siehe Definition 23 und die Faktorregel). Dann gilt mit der Kettenregel, dass

$$\begin{aligned} ((-)^k)'(a) &= h(g(-))'(a) \\ &= h'(g(x))g'(a) \\ &= e^{\ln(a) \cdot k} \cdot k \cdot \frac{1}{a} \\ &= a^k \cdot k \cdot \frac{1}{a} \\ &= k a^{k-1}. \end{aligned}$$

Wir haben also den folgenden Satz gezeigt. Dieser Satz ist uns in dem Spezialfall, dass k eine natürliche Zahl ist, schon bekannt, da wir es in diesem Fall mit einer Polynomfunktion zu tun haben.

Satz 28 (Exponentenregel). Ist k eine beliebige reelle Zahl, so ist die Ableitung der Funktion $(-)^k$ gegeben durch $k \cdot (-)^{k-1}$.

Beispiel 29. Die Exponentenregel macht es einfach, sich die Ableitungen von Wurzelfunktionen zu merken. Möchten wir beispielsweise die Ableitungsfunktion der Funktion $\sqrt[3]{(-)^2}: (0, \infty) \rightarrow (0, \infty)$ berechnen, so ergibt sich mit $\sqrt[3]{(-)^2} = (-)^{\frac{2}{3}}$ die Ableitungsfunktion $\frac{2}{3} \cdot (-)^{(\frac{2}{3}-1)} = \frac{2}{3} \cdot (-)^{-\frac{1}{3}} = \frac{2}{3 \cdot (-)^{\frac{1}{3}}} = \frac{2}{3 \cdot \sqrt[3]{(-)}}$.

Am Ende dieses Einführungskapitels möchten wir uns noch mit Stammfunktionen und Integration beschäftigen.

Definition 30. Sei I ein Intervall und $f: I \rightarrow \mathbb{R}$ eine Funktion. Eine **Stammfunktion** von f ist eine Abbildung $F: I \rightarrow \mathbb{R}$ mit $F' = f$.

Bemerkung 31. Ist F eine Stammfunktion einer Funktion f , so ist für jede Konstante λ die durch $F(x) + \lambda$ definierte Abbildung ebenfalls eine Stammfunktion, da konstante Summanden beim Ableiten „wegfallen“. Deshalb kann man nicht von **der** Stammfunktion von f sondern immer nur von **einer** Stammfunktion von f reden.

Bemerkung 32. Manchmal wird eine Stammfunktion F von f auch mit $\int f(x)dx$ bezeichnet und **unbestimmtes Integral** von f genannt. Dieses ist ein wenig problematisch, da

diese Notation suggeriert, es gäbe (höchstens) eine **eindeutige** Stammfunktion zu einer gegebenen Funktion f . Die vorherige Bemerkung 31 zeigt, dass dieses aber nicht der Fall ist. Trotzdem ist diese Schreibweise praktisch.



Es ist für Sie als Anwender sicherlich manchmal lästig, das Symbol „ dx “ zu schreiben - reicht denn das Integralzeichen nicht aus? Vielleicht sollte man sich das Integralzeichen und das Symbol „ dx “ jeweils als Klammern vorstellen. So bedeutet $\int f(x) + g(x)dx$ beispielsweise $\int (f(x) + g(x))dx$.

Beispiel 33. Hier sind Stammfunktionen zu allen oben betrachteten wichtigen Funktionen.

- Ist $f: I \rightarrow \mathbb{R}$ mit $f(x) = a_n x^n + \dots + a_1 x + a_0$ eine **Polynomfunktion**, so ist eine Stammfunktion durch

$$F(x) = \frac{a_n}{n+1} x^{n+1} + \frac{a_{n-1}}{n} x^n + \dots + \frac{a_1}{2} x^2 + \frac{a_0}{1} x^1$$

gegeben.

- Eine Stammfunktion der **Exponentialfunktion** $e^{(-)}: \mathbb{R} \rightarrow (0, \infty)$ ist die Exponentialfunktion selber.
- Eine Stammfunktion zur (natürlichen) **Logarithmusfunktion** $\ln: (0, \infty) \rightarrow \mathbb{R}$ ist durch

$$y \mapsto \ln(y)y - y$$

gegeben.

- Ist $a > 0$ eine reelle Zahl und $a \neq 1$, so ist eine Stammfunktion zur **Potenzfunktion** $a^{(-)}: \mathbb{R} \rightarrow (0, \infty)$ durch

$$x \mapsto \frac{a^x}{\ln(a)}$$

gegeben. Im Fall $a = 1$, so ist $x \mapsto a^x = 1^x = 1$ eine konstante Funktion. Eine solche Funktion ist eine Polynomfunktion und nach der obigen Regel ist $x \mapsto x$ eine Stammfunktion.

- Ist $a > 0$ eine reelle Zahl und $a \neq 1$, so ist eine Stammfunktion zur **Logarithmusfunktion (zur Basis a)** $\log_a: (0, \infty) \rightarrow \mathbb{R}$ durch

$$y \mapsto y \log_a(y) - \frac{y}{\ln(a)}$$

gegeben.

- (**Exponentenregel für die Integration**) Eine Stammfunktion der Funktion $\frac{1}{(-)} = (-)^{-1}: (0, \infty) \rightarrow (0, \infty)$ ist durch

$$x \mapsto \ln(x)$$

gegeben. Ist $k \neq -1$ eine beliebige reelle Zahl, so ist eine Stammfunktion von $(-)^k: (0, \infty) \rightarrow (0, \infty)$ durch

$$x \mapsto \frac{x^{k+1}}{k+1}$$

gegeben.

Im Folgenden müssen wir für eine präzise Formulierung die Eigenschaft der **Stetigkeit** für eine Funktion $f: I \rightarrow \mathbb{R}$ benutzen. Um diesen Begriff im Allgemeinen zu definieren, ist etwas mehr Formalismus nötig. (Grob gesprochen bedeutet diese Eigenschaft, dass man den zugehörigen Funktionsgraphen in einem Zug zeichnen kann, also „ohne den Stift dabei abzusetzen“.) Daher werden wir auf den Begriff der Stetigkeit nicht eingehen. Allerdings ist jede Funktion $f: I \rightarrow \mathbb{R}$, die eine Ableitung besitzt, automatisch auch stetig. Also sind alle wichtigen Funktionen, die wir oben kennengelernt haben, stetig.

Bemerkung 34. Ist $[a, b] \subseteq I$ ein abgeschlossenes Teilintervall und $f: I \rightarrow \mathbb{R}$ eine stetige Funktion, so ist auch die zugehörige **eingeschränkte Funktion** $f: [a, b] \rightarrow \mathbb{R}$ (also die Funktion $x \mapsto f(x)$, wobei wir für x nur Werte aus $[a, b]$ anstatt aus ganz I „einsetzen“) stetig und wir bezeichnen sie (etwas ungenau) mit dem gleichen Buchstaben. Hat die ursprüngliche Funktion $f: I \rightarrow \mathbb{R}$ eine Stammfunktion F , so ist deren Einschränkung $F: [a, b] \rightarrow \mathbb{R}$ auch

eine Stammfunktion von der eingeschränkten Funktion $f: [a, b] \rightarrow \mathbb{R}$. Man kann übrigens zeigen, dass eine stetige Funktion $f: [a, b] \rightarrow \mathbb{R}$ immer eine Stammfunktion F besitzt.



Die folgenden Sätze sind für Funktionen $f: [a, b] \rightarrow \mathbb{R}$ formuliert. Die vorherige Bemerkung 34 soll rechtfertigen, dass wir sie einfach für unsere oben betrachteten Funktionen $f: I \rightarrow \mathbb{R}$ und deren Stammfunktionen „benutzen“ können.

Definition 35. Ist $f: [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion mit einer Stammfunktion F , so heißt die reelle Zahl

$$\int_a^b f(x) dx := F(b) - F(a)$$

das **Integral von f über $[a, b]$** (oder das **Integral von f mit den Grenzen a, b**). Man schreibt für $F(b) - F(a)$ auch abkürzend $[F(x)]_a^b$.

Das Integral einer stetigen Funktion f über einem Intervall $[a, b]$ hat die folgende graphische Interpretation. Die reelle Zahl $\int_a^b f(x) dx$ gibt genau den **Flächeninhalt (mit Vorzeichen)** der Fläche an, die von dem Graph der Funktion f und der „ X -Achse“ zwischen a und b eingeschlossen wird.

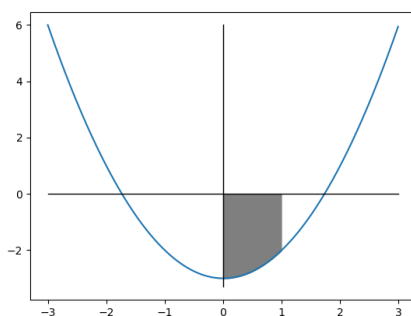


Abbildung 6: Interpretation des Integrals als Fläche

Dabei ist dieser Flächeninhalt (mit Vorzeichen) negativ, wenn sich die Fläche „unterhalb der X -Achse“ befindet.

Beispiel 36. In diesem Beispiel rechnen wir den Flächeninhalt der grauen Fläche im vorherigen Bildchen 6 aus. Sei dazu $a := 0$ und $b := 1$ und wir betrachten die Funktion

$$\begin{aligned} f: I &\rightarrow \mathbb{R} \\ x &\mapsto x^2 - 3, \end{aligned}$$

aus den vorherigen Beispielen. Ihr Graph ist im vorherigen Bildchen 6 blau eingezeichnet. Eine Stammfunktion dieser Polynomfunktion ist nach Beispiel 33 durch $x \mapsto \frac{2}{3}x^3 - 3x$ gegeben. Wir betrachten die Einschränkung $f: [a, b] \rightarrow \mathbb{R}$ der obigen Funktion f . Die reelle Zahl

$$\begin{aligned} \int_0^1 f(x) dx &= \left[\frac{2}{3}x^3 - 3x \right]_0^1 \\ &= \left(\frac{2}{3}1^3 - 3 \cdot 1 \right) - \left(\frac{2}{3}0^3 - 3 \cdot 0 \right) \\ &= \frac{2}{3} - 3 \\ &= -\frac{7}{3} \\ &\approx -2,33 \end{aligned}$$

ist der Flächeninhalt (mit Vorzeichen) des grauen Bereichs im Bildchen 6.

Wir kennen aus der Schule Möglichkeiten, jeweils die Integrale eines Produkts $g(-)h(-)$ und einer Komposition $h(g(-))$ von stetigen Funktionen g und h zu bestimmen. Außerdem können wir Konstanten und Summen „beim Integrieren herausziehen“. Diese Regeln sind in dem folgenden Satz aufgelistet.

Satz 37. Ist $f: [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion, so gilt für alle $\lambda \in \mathbb{R}$

$$\int_a^b \lambda f(x) dx = \lambda \int_a^b f(x) dx \quad (\text{Faktorregel})$$

und ist $g: [a, b] \rightarrow \mathbb{R}$ eine weitere stetige Funktion, so gilt

$$\int_a^b f(x) + g(x) dx = \int_a^b f(x) dx + \int_a^b g(x) dx. \quad (\text{Summenregel})$$

Sind $f, g: [a, b] \rightarrow \mathbb{R}$ Funktionen mit stetigen Ableitungsfunktionen f' und g' , so gilt

$$\int_a^b g'(x)h(x) dx = [g(x)h(x)]_a^b - \int_a^b g(x)h'(x) dx. \quad (\text{Partielle Integration})$$

Ist J ein Intervall, $g: [a, b] \rightarrow J$ eine Funktion mit stetiger Ableitungsfunktion g' und $h: J \rightarrow \mathbb{R}$ eine stetige Funktion, so gilt

$$\int_a^b h(g(x))g'(x) dx = \int_{g(a)}^{g(b)} h(x) dx. \quad (\text{Substitutionsregel})$$

Beispiel 38. Wir wollen die reelle Zahl $\int_2^4 x \ln(x) dx$ „ausrechnen“ und möchten dafür partielle Integration verwenden. Seien $g'(x) := x$ und $h(x) := \ln(x)$ (dieses definiert jeweils eine Funktion $[2, 4] \rightarrow \mathbb{R}$). Es ist durch $g(x) = \frac{1}{2}x^2$ eine Stammfunktion zu g' gegeben. Außerdem ist durch $h'(x) = \frac{1}{x}$ nach Definition 23 die Ableitung von \ln gegeben. Mit der partiellen Integration aus dem vorherigen Satz 37 bekommen wir

$$\begin{aligned} \int_2^4 x \ln(x) dx &= \int_2^4 g'(x)h(x) dx \\ &= [g(x)h(x)]_2^4 - \int_2^4 g(x)h'(x) dx \\ &= \left[\frac{1}{2}x^2 \ln(x)\right]_2^4 - \int_2^4 \frac{1}{2}x^2 \frac{1}{x} dx \\ &= 8 \ln(4) - 2 \ln(2) - \left[\frac{1}{4}x^2\right]_2^4 \\ &= 8 \ln(4) - 2 \ln(2) - 3 \\ &\approx 6,70. \end{aligned}$$



Um die partielle Integration anzuwenden, hoffen wir also, die zu integrierende Funktion f als Produkt $g' \cdot h$ schreiben zu können, wobei g „einfach auflösbar“ und h „einfach ableitbar“ ist.

1. Deskriptive Statistik

Das Ziel der **deskriptiven Statistik** (oder **beschreibenden Statistik**) ist die Aufarbeitung oder Beschreibung von Daten oder Messwerten durch graphische Darstellung oder spezifische Kennzahlen (wie z.B. Mittelwert, Standardabweichung, etc.). Dieses ist beispielsweise wichtig für Medikamententests.

1.1. Grundlegende Begriffe.

Zunächst ein Beispiel:

Beispiel 39. Die aus Regensburg stammenden Teilnehmer*innen einer Statistik-Vorlesung werden befragt nach:

- (1) Studiengang
- (2) Interesse an der Vorlesung (außerordentlich, sehr, normal, kaum, garnicht)
- (3) Anzahl der Geschwister
- (4) Fachsemester
- (5) Körpergröße
- (6) Weglänge zwischen Wohnung und Uni

Die nächste Definition ist nicht von genauer mathematischer Natur und dient der sprachlichen Präzisierung der Dinge, über die wir im Folgenden reden werden.

Definition 40.

- Ein **Merkmalsträger** ist ein Einzelobjekt einer statistischen Untersuchung und Träger einer Information (dem Merkmal).
Im obigen Beispiel ist dieses ein*e Teilnehmer*in der Vorlesung.
- Die **Grundgesamtheit** ist die Zusammenfassung aller in Verbindung mit einer statistischen Untersuchung stehenden Merkmalsträger. Sie ist eine Menge, die häufig als G bezeichnet wird und deren Elemente die Merkmalsträger sind.
Im obigen Beispiel ist dieses die Menge aller Teilnehmer*innen der Vorlesung.
- Die **Untersuchungsmenge** ist eine Teilmenge $U \subseteq G$ der Grundgesamtheit G und besteht aus den Merkmalsträgern, die statistisch untersucht werden.
Im obigen Beispiel sind dieses alle aus Regensburg stammenden Teilnehmer*innen der Vorlesung. Die Vorlesung besteht vielleicht noch aus anderen Teilnehmer*innen, doch diese werden bei der obigen Untersuchung nicht befragt.
- Ein **Merkmal** ist eine zu untersuchende Eigenschaft eines Merkmalsträgers. Es wird häufig mit X oder Y bezeichnet.
Im obigen Beispiel werden sechs verschiedene Merkmale untersucht, zum Beispiel der Studiengang oder die Fachsemester.
- Eine **Merkmalsausprägung** ist ein möglicher Wert, den ein Merkmal annehmen kann. Dieser Wert wird häufig mit a bezeichnet.
Im obigen Beispiel (4) ist dieses eine natürliche Zahl.
- Die **(Merkmals)Ausprägungsmenge** ist die Zusammenfassung aller Merkmalsausprägungen a eines Merkmals. Sie ist eine Menge, die häufig als A bezeichnet wird und deren Elemente die möglichen Merkmalsausprägungen sind.
Im obigen Beispiel (4) ist dieses beispielsweise die Menge $A = \mathbb{N}$.

Werden bei einer statistischen Untersuchung alle Merkmalsträger der Grundgesamtheit vollständig erfasst, d.h. ist $U = G$, spricht man von einer **Totalerhebung**, andernfalls von einer **Teilerhebung**. Die Untersuchung aus dem obigen Beispiel ist also genau dann eine Totalerhebung, wenn alle Teilnehmer*innen der Statistik-Vorlesung aus Regensburg stammen.

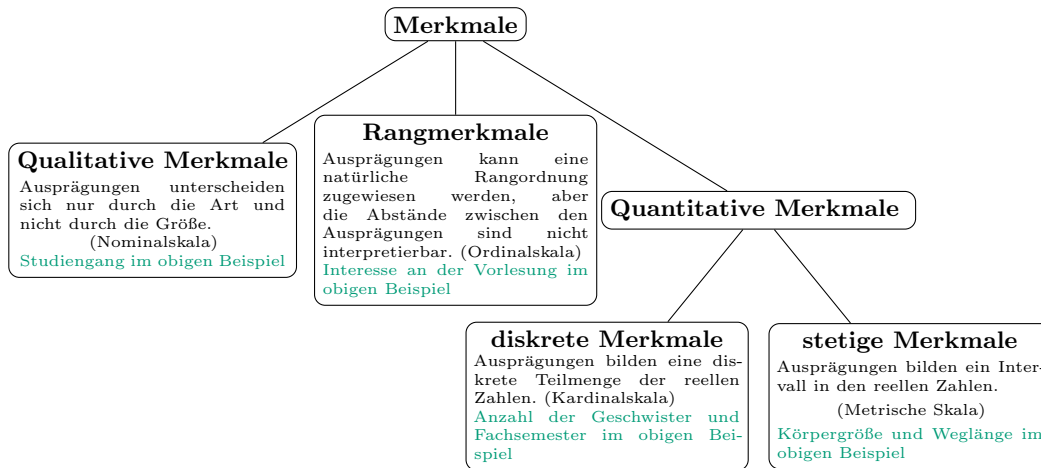
Bemerkung 41. Einen Merkmalsträger nennt man manchmal auch eine **statistische Einheit**. Die Grundgesamtheit nennt man manchmal auch die **statistische Masse**.



Man findet selbst in der Literatur keine einheitlichen Bezeichnungen für diese Begriffe. Es ist also nicht schlimm, wenn man nicht die genauen Fachbegriffe kennt, solange alle wissen, was gemeint ist. In diesem Skript benutzen wir die Wörter aus der obigen Definition.

Bemerkung 42. Hat man einen Text wie in Beispiel 39 gegeben, der eine statistische Untersuchung beschreibt, so sind die in Definition 40 eingeführten Begriffe oft nicht eindeutig festgelegt. Es ist beispielsweise nicht klar, dass die Ausprägungsmenge A des Merkmals X = „Fachsemester“ in dem obigen Beispiel genau die Menge der natürlichen Zahlen ist. Könnte man nicht genauso gut die Menge $A = \{0, 1, \dots, 100000\}$ benutzen, da sicher kein Teilnehmer mehr als 100000 Fachsemester absolviert hat? Oder könnte man nicht $A = \mathbb{R}$ benutzen, auch wenn das Merkmal X nur natürliche Zahlen als Werte annehmen kann? Streng genommen müsste man alle diese Daten bei einer statistischen Untersuchung mit angeben. Dieses wird aber, wie in dem obigen Beispiel 39, oft nicht explizit getan und man muss beispielsweise eine „sinnvolle“ Ausprägungsmenge „raten“. Wir werden im Folgenden Klassen von Ausprägungsmengen betrachten. Es wird innerhalb dieser Klassen praktisch irrelevant sein, welche genaue Wahl einer Ausprägungsmenge man vornimmt.

Wie das obige Beispiel 39 schon andeutet, gibt es verschiedene Arten von Merkmalen. Man teilt Merkmale anhand ihrer möglichen Merkmalsausprägungen auf die folgende Weise in Klassen ein.



Bemerkung 43. Anstatt beispielsweise von einem Qualitativen Merkmal zu reden, sagt man manchmal auch, dass Merkmal (oder genauer: die Merkmalsausprägungen) sei „nominalskaliert“. Diese alternativen Bezeichnungen sind in dem obigen Diagramm in Klammern notiert.

1.2. Eindimensionale Stichproben. Wir möchten eine Grundgesamtheit G auf ein einziges Merkmal X mit Ausprägungsmenge A untersuchen. Es sei $U = \{1, \dots, n\} \subseteq G$ die als endlich angenommene Untersuchungsmenge.

Definition 44. Eine **Stichprobe** x eines Merkmals X ist eine Abbildung $x: U \rightarrow A$, also mit anderen Worten ein Tupel $x = (x_1, \dots, x_n)$, wobei $x_i \in A$ die Merkmalsausprägung des i -ten Merkmalsträgers ist. Es heißt $n = |U|$ der **Stichprobenumfang**.

Bemerkung 45. Eine Stichprobe nennt man manchmal auch eine **Urliste**.

Beispiel 46. Es haben alle $n = 25$ Schüler*innen einer Klasse eine Klausur geschrieben, deren Bewertung anhand der Noten von 1 bis 6 erfolgte. Hier ist $A = \{1, 2, 3, 4, 5, 6\}$. Die gefundene Stichprobe des Merkmals $X =$ „Note in der Klausur“ ist

$$x = (2, 2, 6, 5, 4, 1, 1, 3, 5, 3, 3, 3, 5, 6, 5, 2, 4, 2, 3, 4, 3, 3, 4, 4, 5).$$

Es handelt sich hierbei um eine Totalerhebung, es ist also $U = G$ und X kann als ein quantitatives diskretes Merkmal betrachtet werden. (Die Einordnung als quantitatives diskretes Merkmal ist sicherlich diskussionswürdig: Man könnte X auch als Rangmerkmal betrachten, wenn man annimmt, dass der Abstand zwischen den jeweiligen Noten nicht gleichmäßig ist.)

1.2.1. Absolute und relative Häufigkeiten.

Definition 47. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines Merkmals X mit endlicher Ausprägungsmenge $A = \{a_1, \dots, a_m\}$ und $j \in \{1, \dots, m\}$.

- Die **absolute Häufigkeit** der Merkmalsausprägung a_j ist die Anzahl des Auftretens von a_j in der Stichprobe x . Sie wird häufig mit $h(a_j)$ oder h_j bezeichnet.
- Die **relative Häufigkeit** der Merkmalsausprägung a_j ist $r_j := \frac{h_j}{n}$. Sie wird häufig auch mit $r(a_j)$ bezeichnet.

Bemerkung 48. Es gelten für die absoluten und relativen Häufigkeiten offenbar die folgenden Relationen.

- $0 \leq h_j \leq n$ für alle $j \in \{1, \dots, m\}$
- $\sum_{j=1}^m h_j = n$
- $0 \leq r_j \leq 1$ für alle $j \in \{1, \dots, m\}$
- $\sum_{j=1}^m r_j = 1$

Das Symbol $\sum_{j=1}^m h_j$ bezeichnet hier die Summe $h_1 + h_2 + \dots + h_{m-1} + h_m$.

Beispiel 49. Wir betrachten wieder das vorherige Klausurnoten-Beispiel 46. Hier gilt für die absoluten Häufigkeiten

$$h_1 = 2, \quad h_2 = 4, \quad h_3 = 7, \quad h_4 = 5, \quad h_5 = 5, \quad h_6 = 2$$

und für die relativen Häufigkeiten

$$r_1 = \frac{2}{25}, \quad r_2 = \frac{4}{25}, \quad r_3 = \frac{7}{25}, \quad r_4 = \frac{1}{5}, \quad r_5 = \frac{1}{5}, \quad r_6 = \frac{2}{25}.$$

Falls die Ausprägungsmenge A wenig Elemente hat (also m eine kleine Zahl ist) so bieten sich die folgenden graphischen Darstellungen der absoluten und relativen Häufigkeiten an.

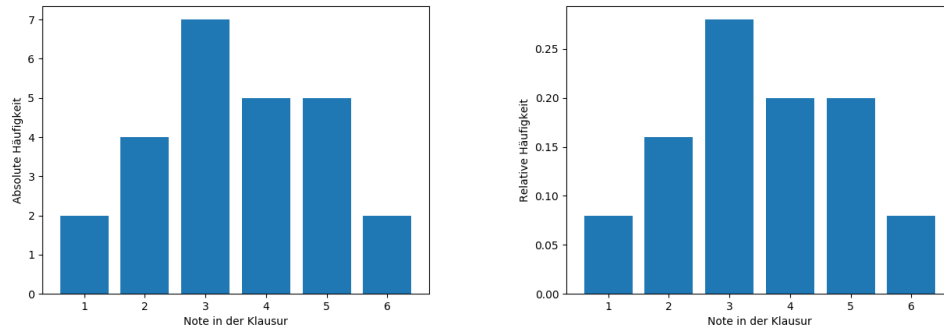


Abbildung 7: Balkendiagramm (Histogramm) der absoluten und relativen Häufigkeiten

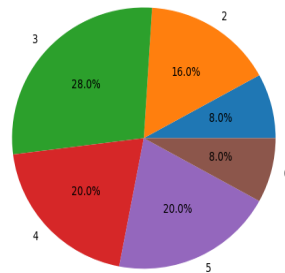


Abbildung 8: Kreisdiagramm (Kuchendiagramm) der relativen Häufigkeiten

1.2.2. Gruppierungen. Falls die Menge A der möglichen Merkmalsausprägungen sehr groß oder sogar unendlich groß ist, wie beispielsweise bei stetigen Merkmalen, so ist es manchmal sinnvoll, „nebeneinanderliegende“ Merkmalsausprägungen zusammenzufassen, obwohl dadurch Informationen „verloren gehen“.



Wollten wir beispielsweise ein Balken- oder Kuchendiagramm zu dem Merkmal „Körpergröße“ aus Beispiel 39 zeichnen, so wäre dieses nicht hilfreich, da vermutlich alle absoluten Häufigkeiten entweder gleich Null oder gleich Eins wären. (Es ist ja sehr unwahrscheinlich, daß zwei der betrachteten Merkmalsträger genau die gleiche Körpergröße haben.)

Man beachte im Folgenden, dass die Ausprägungsmenge A eines quantitativen Merkmals eine Teilmenge der reellen Zahlen ist.

Definition 50. Sei X ein quantitatives Merkmal mit Ausprägungsmenge $A \subseteq \mathbb{R}$. Eine **Gruppeneinteilung** dieses Merkmals ist eine Zerlegung

$$A = G_1 \cup \dots \cup G_M$$

der Ausprägungsmenge in Intervalle $G_j = [a'_j, a'_{j+1})$, die **Gruppen** genannt werden, wobei $j \in \{1, \dots, M\}$ und $a'_1 < \dots < a'_M < a'_{M+1}$.

Bemerkung 51. Hat man eine Gruppeneinteilung eines Merkmals wie in der vorherigen Definition 50 gegeben, so definiert man manchmal die **Gruppenbreite** der Gruppe G_j als die Zahl $d_j := a'_{j+1} - a'_j$ und die **Gruppenmitte** der Gruppe G_j als $m_j := \frac{1}{2}(a'_{j+1} + a'_j)$.

Definition 52. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X mit einer gewählten Gruppeneinteilung wie oben und $j \in \{1, \dots, M\}$.

- Die **absolute Gruppenhäufigkeit** von G_j ist die Anzahl der Merkmalsausprägungen x_i in der Stichprobe, die in G_j enthalten sind. Sie wird häufig mit $h(G_j)$ oder h_j bezeichnet.
- Die **relative Gruppenhäufigkeit** von G_j ist $r_j := \frac{h_j}{n}$. Sie wird häufig auch mit $r(G_j)$ bezeichnet.

Beispiel 53 ([4]). Es liegen fuer 32 europäische Länder als Indikator für den Wohlstand die Zahlen der PKWs pro 1000 Einwohner vor. Die gefundene und bereits geordnete Stichprobe ist

$x = (46, 57, 63, 123, 187, 200, 234, 277, 284, 295, 299, 301, 340, 353, 361, 377, 399, 406, 413, 424, 431, 446, 454, 463, 475, 488, 505, 517, 539, 567, 640, 698)$.

Genauer betrachten wir das quantitative Merkmal X = „Anzahl der PKWs pro 1000 Einwohner“ und $U = \{1, \dots, 32\}$, wobei jedes $i \in U$ eines der untersuchten europäischen Länder repräsentiert. Als Ausprägungsmenge wählen wir $A = [0, 700)$. Wir wählen die Gruppeneinteilung

$G_1 := [0, 200), G_2 := [200, 300), G_3 := [300, 400), G_4 := [400, 500), G_5 := [500, 700)$

und erhalten

Gruppen	absolute Gruppenhäufigkeit	relative Gruppenhäufigkeit
G_1	5	0,15625
G_2	6	0,1875
G_3	6	0,1875
G_4	9	0,28125
G_5	6	0,1875

Nun können wir wieder die obigen graphischen Darstellungen der absoluten und relativen (Gruppen)Häufigkeiten benutzen.

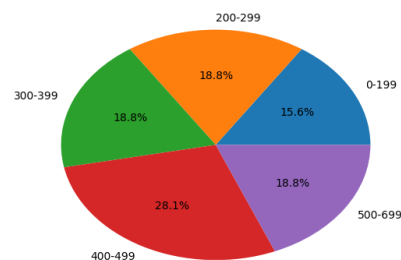
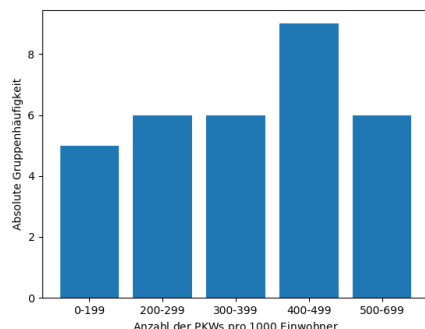


Abbildung 9: Diagramme der absoluten und relativen Gruppenhäufigkeiten



Wir können eine Gruppeneinteilung also dazu benutzen, aus einem quantitativen stetigen Merkmal (oder zumindest einem quantitativem Merkmal mit vielen verschiedenen Ausprägungen) ein quantitativ diskretes Merkmal mit wenigen Ausprägungen zu machen. Eine Gruppeneinteilung ist also sozusagen eine Art „Diskretisierung“.

1.2.3. Die empirische Verteilungsfunktion.

Definition 54. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X mit Ausprägungsmenge $A = \{a_1, \dots, a_m\} \subseteq \mathbb{R}$, wobei wir $a_1 < \dots < a_m$ annehmen, und $j \in \{1, \dots, m\}$. Die **empirische Verteilungsfunktion** dieser Stichprobe ist die Abbildung

$$F: \mathbb{R} \rightarrow [0, 1] \subseteq \mathbb{R}$$

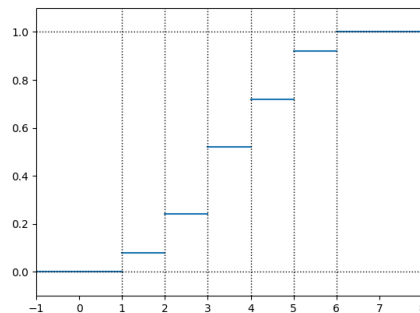
$$t \mapsto \begin{cases} 0 & \text{falls } t < a_1 \\ R_k & \text{falls } a_k \leq t < a_{k+1} \\ 1 & \text{falls } a_m \leq t \end{cases}$$

wobei $R_k := \sum_{j=1}^k r_j$ für $k \in \{1, \dots, m\}$ die Summe der relativen Häufigkeiten r_j der Merkmalsausprägungen a_j ist mit $j \leq k$ ist.

Beispiel 55. Wir betrachten wieder das vorherige Klausurnoten-Beispiel 46 mit Ausprägungsmenge $A = \{1, 2, 3, 4, 5, 6\}$. Hier gilt

$$R_1 = \frac{2}{25}, \quad R_2 = \frac{2}{25} + \frac{4}{25} = \frac{6}{25}, \quad R_3 = \frac{6}{25} + \frac{7}{25} = \frac{13}{25}, \quad R_4 = \frac{18}{25}, \quad R_5 = \frac{23}{25}, \quad R_6 = 1.$$

Die zugehörige empirische Verteilungsfunktion hat den folgenden Funktionsgraphen.



Bemerkung 56. Die empirische Verteilungsfunktion hat an den Stellen a_j (auf der „X-Achse“) Sprünge der Höhe r_j . Diese Funktion ist ein Beispiel einer nicht stetigen und monoton wachsenden Funktion, die nur Werte zwischen 0 und 1 hat.



Auf den ersten Blick sieht die empirische Verteilungsfunktion einer Stichprobe wie ein unnötiges und künstliches Konstrukt aus den relativen Häufigkeiten der Stichprobe aus. Später werden wir jedoch abstrakte Verteilungsfunktionen auf eine natürliche Weise kennenlernen und sie mit empirischen Verteilungsfunktionen „vergleichen“. Es ist übrigens auch kein Zufall, dass die empirischen Verteilungsfunktionen mit einem großen „F“ bezeichnet werden, genau wie Stammfunktionen.

Nun wollen wir auch für eine Stichprobe eines gruppierten Merkmals eine empirische Verteilungsfunktion definieren. Dieses wird fast analog zu der obigen Definition 54 gemacht, wobei man allerdings zwischen den Punkten (a'_{j+1}, R_j) „linear interpoliert“, also eine Gerade zeichnet.

Definition 57. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X mit einer gewählten Gruppeneinteilung mit Zerlegung

$$A = G_1 \cup \dots \cup G_M$$

der Ausprägungsmenge in Intervalle $G_j = [a'_j, a'_{j+1})$ wie in Definition 50. Die **empirische Gruppenverteilungsfunktion** dieser Stichprobe ist die Abbildung

$$F: \mathbb{R} \rightarrow [0, 1] \subseteq \mathbb{R}$$

$$t \mapsto \begin{cases} 0 & \text{falls } t < a'_1 \\ R_{k-1} + \frac{t-a'_k}{d_k} r_k & \text{falls } a'_k \leq t < a'_{k+1} \\ 1 & \text{falls } a'_{M+1} \leq t \end{cases}$$

wobei $R_0 := 0$ und $R_k := \sum_{j=1}^k r_j$ für $k \in \{1, \dots, M\}$ die Summe der relativen Gruppenhäufigkeiten r_j von G_j ist mit $j \leq k$ ist.

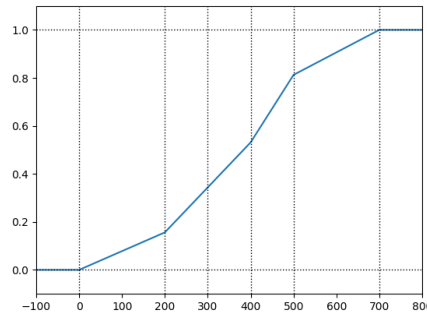
Beispiel 58. Wir betrachten wieder das vorherige PKW-Beispiel 53 mit Gruppeneinteilung

$$G_1 := [0, 200), G_2 := [200, 300), G_3 := [300, 400), G_4 := [400, 500), G_5 := [500, 700).$$

Hier gilt

$$R_0 := 0, \quad R_1 = \frac{5}{32}, \quad R_2 = \frac{5}{32} + \frac{6}{32} = \frac{11}{32}, \quad R_3 = \frac{11}{32} + \frac{6}{32} = \frac{17}{32}, \quad R_4 = \frac{26}{32}, \quad R_5 = 1.$$

Die zugehörige empirische Gruppenverteilungsfunktion hat den folgenden Funktionsgraphen.



1.2.4. Lage- und Streuungsmaße. Neben den graphischen Darstellungen aus dem vorherigen Kapitel, wollen wir einer Stichprobe eines Merkmals in diesem Kapitel Zahlenwerte zuordnen. Für eine Stichprobe x eines quantitativen Merkmals, also eines Merkmals mit Ausprägungsmenge $A \subseteq \mathbb{R}$, unterscheidet man grob zwischen

- **Lagemaßen**, also Zahlen, welche die Lage der Stichprobenwerte x_i in Bezug auf die „Messskala“ \mathbb{R} beschreiben (wie z.B. empirischer Mittelwert, empirischer Median, Quantile,...) und
- **Streuungsmaßen**, also Zahlen, welche die Abweichung der Stichprobenwerte x_i von einem Lagemaß beschreiben (wie z.B. empirische Varianz, empirische Standardabweichung,...).



Diese beiden Begriffe sind (zumindest in diesem Skript) nicht klar definiert und dienen daher nur zur groben sprachlichen Orientierung.

Zunächst führen wir jedoch in der folgenden Definition einen kennzeichnenden Wert für eine Stichprobe eines beliebigen Merkmals ein.

Definition 59. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines Merkmals X . Ein **Modalwert** der Stichprobe ist ein Stichprobenwert $x_{mod} \in \{x_1, \dots, x_n\}$, der am häufigsten in der Stichprobe vorkommt.

Beispiel 60. Im vorherigen Klausurnoten-Beispiel 46 ist $x_{mod} = 3$ ein Modalwert. In diesem Beispiel gibt es nur einen Modalwert. Daher kann man hier von **dem** Modalwert der Stichprobe reden. Im Allgemeinen muss ein Modalwert natürlich nicht eindeutig sein, d.h. es kann mehrere Modalwerte der Stichprobe geben.

Bemerkung 61. Man kann Modalwerte zu einer Stichprobe eines Merkmals einer beliebigen Klasse betrachten, wie zum Beispiel eines qualitativen Merkmals. Insbesondere muss ein Modalwert in diesem Fall nicht unbedingt eine reelle Zahl sein.

Bemerkung 62. Ist $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X mit einer gewählten Gruppeneinteilung wie in Definition 50, so kann man einen **Gruppenmodalwert** der Stichprobe als die Gruppenmitte $x_{mod} = m_j$ einer am häufigsten besetzten Gruppe G_j definieren. Dieses wollen wir in diesem Skript aber nicht weiter betrachten.

Nun führen wir einige wichtige Lagemaße für Stichproben quantitativer Merkmale ein.

Definition 63. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X . Der **empirische Mittelwert** der Stichprobe ist die reelle Zahl

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

Bemerkung 64. Ist $x = (x_1, \dots, x_n)$ eine Stichprobe eines Merkmals X mit endlicher Ausprägungsmenge $A = \{a_1, \dots, a_m\}$, so gilt

$$\bar{x} := \frac{1}{n} \sum_{j=1}^m h_j a_j = \sum_{j=1}^m r_j a_j,$$

wobei die h_j die absoluten und die r_j die relativen Häufigkeiten bezeichnen.

Bemerkung 65. Ist $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X mit einer gewählten Gruppeneinteilung wie in Definition 50, so kann man einen **empirischen Gruppenmittelwert** der Stichprobe als

$$\bar{x} := \frac{1}{n} \sum_{j=1}^m h_j m_j = \sum_{j=1}^m r_j m_j,$$

definieren, wobei die h_j die absoluten und die r_j die relativen Gruppenhäufigkeiten und die m_j die Gruppenmitteln bezeichnen. Dieses wollen wir in diesem Skript aber nicht weiter betrachten.

Beispiel 66. Es werden 10 Personen nach ihrem Brutto-Einkommen befragt und es ergibt sich die Stichprobe

$$x = (1600, 1800, 2150, 2200, 2465, 2745, 3000, 3110, 3500, 16000).$$

Der empirische Mittelwert der Stichprobe ist $\bar{x} = 3857$. Man sieht, dass der empirische Mittelwert „nicht robust gegenüber Ausreißern“ wie dem Stichprobenwert $x_{10} = 16000$ ist.

Das vorherige Beispiel 66 zeigt, dass der Mittelwert nicht die „Mitte der Stichprobe“, sondern den Durchschnitt aller Stichprobeneinträge beschreibt. Daher führt man den folgenden Begriff des empirischen Medians ein.

Definition 67. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X , wobei wir $x_1 \leq \dots \leq x_n$ annehmen. Der **empirische Median** der Stichprobe ist die reelle Zahl

$$\tilde{x} := \begin{cases} \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{falls } n \text{ gerade} \\ x_{\frac{n+1}{2}} & \text{falls } n \text{ ungerade.} \end{cases}$$

Beispiel 68. Im obigen Brutto-Einkommen-Beispiel 66 ist $n = 10$ gerade, der empirische Median also

$$\tilde{x} = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}(2465 + 2745) = 2605.$$

Wir sehen, dass der empirische Mittelwert „robuster gegenüber Ausreißern“ ist als der empirische Mittelwert aus Beispiel 66.

Beispiel 69. Im Klausurnoten-Beispiel 46 ordnen wir zunächst die Stichprobe und erhalten

$$x = (1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6),$$

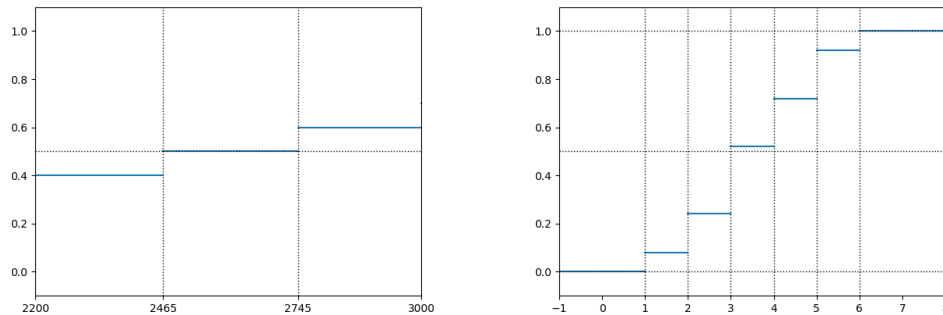
wobei wir (etwas ungenau) den gleichen Buchstaben „ x “ benutzen. Hier ist $n = 25$ ungerade, der empirische Median also

$$\tilde{x} = x_{\frac{26}{2}} = x_{13} = 3.$$



Es gilt also für eine quantitative Stichprobe, dass mindestens die Hälfte der Stichprobenwerte kleiner gleich und mindestens die Hälfte der Stichprobenwerte größer gleich dem zugehörigen empirischen Median sind. Bei einem ungeraden Stichprobenumfang n ist der empirische Median also genau „der Mittlere“ der (geordneten) Stichprobenwerte.

Bemerkung 70. Man kann den empirischen Median auch an der empirischen Verteilungsfunktion $F: \mathbb{R} \rightarrow \mathbb{R}$ ablesen. Betrachten wir dazu die Graphen der empirischen Verteilungsfunktionen der Stichprobe aus dem Brutto-Einkommen-Beispiel 66 (links, Ausschnitt) und der Stichprobe aus dem Klausurnoten-Beispiel 46 (rechts).



Wir möchten gerne für den empirischen Median die Gleichung $F(x) = 0.5$ auflösen, was auf der linken Seite aber für alle $x \in [2465, 2745)$ möglich und auf der rechten Seite nicht lösbar ist. Daher wählen wir auf der linken Seite die „Mitte“ $\tilde{x} = \frac{1}{2}(2465 + 2745) = 2605$ als empirischen Median und auf der rechten Seite $\tilde{x} = 3$ als das „nächst größere x “.

Beispiel 71. Für die Stichprobe $x = (0, 0, 0, 0, 1000)$ ist der zugehörige empirische Mittelwert $\tilde{x} = \frac{1000}{5} = 200$, der empirische Median aber $\tilde{x} = 0$. Dieses Beispiel zeigt, dass der empirische Median zur Interpretation bestimmter Stichproben ungeeignet sein kann – gerade weil er „robust gegenüber Ausreißern“ ist.

Als weiteres Lagemaß lernen wir nun eine Verallgemeinerung des empirischen Medians, die Quantile, kennen. Die Motivation für die folgende Definition ist die graphische Beobachtung aus der vorherigen Bemerkung 70. Wir suchen nun eine „Lösung“ der Gleichung $F(x) = p$ für ein vorgegebene Zahl p mit $0 < p < 1$.

Definition 72. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X , wobei wir $x_1 \leq \dots \leq x_n$ annehmen. Sei weiter $p \in (0, 1)$ eine reelle Zahl. Das **p -Quantil** der Stichprobe ist die reelle Zahl

$$\tilde{x}_p := \begin{cases} \frac{1}{2}(x_{np} + x_{np+1}) & \text{falls } np \in \mathbb{N} \\ x_{\lceil np \rceil} & \text{falls } np \notin \mathbb{N}, \end{cases}$$

wobei $\lceil np \rceil$ die kleinste natürliche Zahl k mit $k \geq np$ bezeichnet.

Bemerkung 73. Das 0,5-Quantil einer Stichprobe ist genau der empirische Median $\tilde{x}_{0,5} = \tilde{x}$. Man nennt das p -Quantil einer Stichprobe auch das $(p \cdot 100)$ -**Perzentil**. So ist beispielsweise das 0,25-Quantil das 25-Perzentil. Man nennt außerdem das 0,25-Quantil auch das **untere Quartil** und das 0,75-Quantil das **obere Quartil**.

Beispiel 74. Wir betrachten das Klausurnoten-Beispiel 46 mit $n = 25$ und möchten das zugehörige untere Quartil, also das p -Quantil mit $p = 0,25$, bestimmen. Es ist die Zahl $np = 25 \cdot 0,25 = 6,25$ nicht natürlich und die nächst größere natürliche Zahl ist $\lceil np \rceil = 7$. Es ist also

$$\tilde{x}_{0,25} = x_7 = 3.$$

Für das zugehörige obere Quartil gilt $25 \cdot 0,75 = 18,75$, also $\tilde{x}_{0,75} = x_{\lceil 18,75 \rceil} = x_{19} = 5$.

Beispiel 75. Wie betrachten das Brutto-Einkommen-Beispiel 66 mit $n = 10$. Hier ist das untere Quartil $\tilde{x}_{0,25} = 2150$, der empirische Median $\tilde{x} = 2605$ und das obere Quartil $\tilde{x}_{0,75} = 3110$. Man veranschaulicht diese Daten oft in einem **Boxplot**.

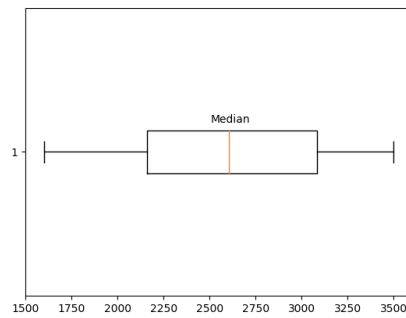


Abbildung 10: Boxpolts zum Brutto-Einkommen-Beispiel (ohne „Ausreißer“)

Hier beschreibt der mittlere Strich den empirischen Median, das linke Ende der Box das untere und das rechte Ende der Box das obere Quartil. Die beiden äußeren vertikalen Striche heißen „whisker“ und markieren das untere und das obere Ende der Stichprobenwerte ohne „Ausreißer“. Die „Ausreißer“ markiert man oft mit Kreisen oder Kreuzen.

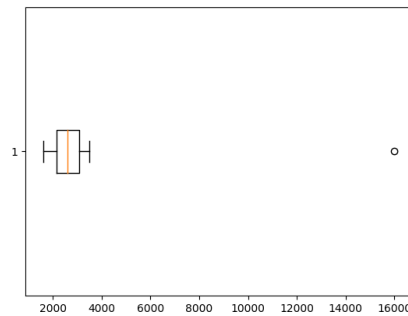


Abbildung 11: Boxpolt zum Brutto-Einkommen-Beispiel (mit „Ausreißern“)

Bemerkung 76. Ist $x = (x_1, \dots, x_n)$ eine Stichprobe eines quantitativen Merkmals X mit einer gewählten Gruppeneinteilung und $p \in (0, 1)$ eine reelle Zahl, so kann man das **p -Gruppenquantil** der Stichprobe als eine Lösung \tilde{x}_p der Gleichung $F(\tilde{x}_p) = p$ definieren, wobei $F: \mathbb{R} \rightarrow \mathbb{R}$ die zugehörige empirische Gruppenverteilungsfunktion ist. Schaut man sich deren Funktionsgraphen (wie im obigen Beispiel 58) an, so sieht man, dass die Gleichung $F(\tilde{x}_p) = p$ eine eindeutige Lösung hat, wenn man fordert, dass die relativen Gruppenhäufigkeiten positiv sind. Dieses wollen wir in diesem Skript aber nicht weiter betrachten.

Abschließend wollen wir noch zwei wichtige und verwandte Streuungsmaße kennenlernen, die jeweils die Abweichung der Stichprobenwerte vom empirischen Mittelwert \bar{x} beschreiben.

Definition 77. Sei $x = (x_1, \dots, x_n)$ eine Stichprobe vom Umfang $n \geq 2$ eines quantitativen Merkmals X mit empirischem Mittelwert \bar{x} . Die **empirische Varianz** der Stichprobe ist die reelle Zahl

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

Die **empirische Standardabweichung** der Stichprobe ist die reelle Zahl

$$s := \sqrt{s^2}.$$

Bemerkung 78. Die empirische Varianz (und auch die empirische Standardabweichung) misst also, wie stark die Stichprobenwerte x_i vom empirischen Mittelwert \bar{x} abweichen. Die Summanden $(x_i - \bar{x})^2$ in der obigen Definition 77 sind (durch das Quadrat) nicht-negative Zahlen, welche die Abweichung von x_i von \bar{x} beschreiben.

Bemerkung 79. Es gilt

$$\begin{aligned}s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\&= \frac{1}{n-1} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\&= \frac{1}{n-1} \left(\left(\sum_{i=1}^n x_i^2 \right) - 2n\bar{x}^2 + n\bar{x}^2 \right) \\&= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)\end{aligned}$$

und der untere Ausdruck ist in der Praxis oft einfacher auszurechnen, indem man zunächst die Zahlen $\sum_{i=1}^n x_i^2$ und \bar{x} einzeln ausrechnet. (Andererseits ist die Zahl $\sum_{i=1}^n x_i^2$ manchmal sehr groß, was wiederum für eine Ausrechnung durch die ursprüngliche Formel aus der Definition 77 spricht.)

Beispiel 80. Wir betrachten das Brutto-Einkommen-Beispiel 66 mit $n = 10$, $\bar{x} = 3857$ und

$$\begin{aligned}\sum_{i=1}^{10} x_i^2 &= 1600^2 + 1800^2 + 2150^2 + 2200^2 + 2465^2 + 2745^2 \\&\quad + 3000^2 + 3110^2 + 3500^2 + 16000^2 \\&= 315795850.\end{aligned}$$

Mit $\bar{x}^2 = 14876449$ und $n\bar{x}^2 = 148764490$ gilt $s^2 = 18559040$ und $s \approx 4308,02$.

Bemerkung 81. Ist $x = (x_1, \dots, x_n)$ eine Stichprobe vom Umfang $n \geq 2$ eines quantitativen Merkmals X mit endlicher Ausprägungsmenge $A = \{a_1, \dots, a_m\}$, so gilt

$$\sum_{i=1}^n x_i^2 = \sum_{j=1}^m h_j a_j^2$$

wobei die h_j die absoluten Häufigkeiten bezeichnen. Diese Beschreibung macht die Ausrechnung der linken Seite manchmal etwas leichter, wie das folgende Beispiel 82 zeigt.

Beispiel 82. Wir betrachten das Klausurnoten-Beispiel 46 mit $n = 25$, $\bar{x} = 3,52$ und

$$\sum_{i=1}^{25} x_i^2 = 2 \cdot 1^2 + 4 \cdot 2^2 + 7 \cdot 3^2 + 5 \cdot 4^2 + 5 \cdot 5^2 + 2 \cdot 6^2 = 358.$$

Mit $\bar{x}^2 = 12,3904$ und $n\bar{x}^2 = 309,76$ gilt $s^2 = 2,01$ und $s \approx 1,42$.

Bemerkung 83. Ist $x = (x_1, \dots, x_n)$ eine Stichprobe vom Umfang $n \geq 2$ eines quantitativen Merkmals X mit empirischem Mittelwert \bar{x} und einer gewählten Gruppeneinteilung, so kann man die **empirische Gruppenvarianz** der Stichprobe als die reelle Zahl

$$s^2 := \frac{1}{n-1} \left(\sum_{j=1}^m h_j m_j^2 - n\bar{x}^2 \right)$$

und die **empirische Gruppenstandardabweichung** als die reelle Zahl $s = \sqrt{s^2}$ definieren, wobei die h_j die absoluten Gruppenhäufigkeiten und die m_j die Gruppenmitteln bezeichnen. Dieses wollen wir in diesem Skript aber nicht weiter betrachten.

1.3. Zweidimensionale Stichproben. Wir möchten eine Grundgesamtheit G auf zwei Merkmale X und Y untersuchen. Dabei hat X die Ausprägungsmenge A und Y die Ausprägungsmenge B . Es sei $U = \{1, \dots, n\} \subseteq G$ die als endlich angenommene Untersuchungs-
menge.

Definition 84. Eine **zweidimensionale Stichprobe** xy zweier Merkmale X und Y ist eine Abbildung $xy: U \rightarrow A \times B$, also mit anderen Worten ein Tupel

$$xy = ((x_1, y_1), \dots, (x_n, y_n)),$$

wobei $x_i \in A$ und $y_i \in B$ die jeweiligen Merkmalsausprägungen des i -ten Merkmalsträgers sind. Es heißt $n = |U|$ der **Stichprobenumfang**.

Bemerkung 85. Eine zweidimensionale Stichprobe nennt man manchmal auch eine **bivariate** Stichprobe.

Bemerkung 86. Es gibt keinen Unterschied zwischen einer zweidimensionalen Stichprobe xy (von X und Y) und einem Paar von Stichproben (x, y) , wobei $x = (x_1, \dots, x_n)$ eine Stichprobe des Merkmals X und $y = (y_1, \dots, y_n)$ eine Stichprobe des Merkmals Y ist. Genauer gesagt kann man aus zwei Abbildungen $x: U \rightarrow A$ und $y: U \rightarrow B$ eine eindeutige Abbildung

$$\begin{aligned} xy: U &\rightarrow A \times B \\ i &\mapsto (x(i), y(i)) \end{aligned}$$

definieren und umgekehrt.



Die vorherige Bemerkung 86 sagt ja, dass es keinen Unterschied zwischen zwei „einzelnen“ Stichproben x und y und einer zweidimensionalen Stichprobe xy gibt – aus dem einen bekommt man das andere und umgekehrt! Wir führen den Begriff der zweidimensionalen Stichprobe aber trotzdem ein, da zum einen die Literatur diesen Begriff benutzt und zum anderen die Einführung eines solchen Begriffs vielleicht einfacher ist, als immer von einem „Paar von Stichproben“ zu reden.

Definition 87. Sei $xy = ((x_1, y_1), \dots, (x_n, y_n))$ eine zweidimensionale Stichprobe zweier Merkmale X und Y mit jeweils endlichen Ausprägungsmengen $A = \{a_1, \dots, a_m\}$ und $B = \{b_1, \dots, b_l\}$. Sei außerdem $j \in \{1, \dots, m\}$ und $k \in \{1, \dots, l\}$.

- Die **absolute Häufigkeit** von (a_j, b_k) ist die Anzahl des Auftretens von (a_j, b_k) in der Stichprobe xy . Sie wird häufig mit $h(a_j, b_k)$ oder h_{jk} bezeichnet.
- Die **relative Häufigkeit** von (a_j, b_k) ist $r_{jk} := \frac{h_{jk}}{n}$. Sie wird häufig auch mit $r(a_j, b_k)$ bezeichnet.

Außerdem definiert man Bezeichnungen für gewisse Summen dieser Häufigkeiten.

- Die **absolute Randhäufigkeit** von a_j ist $h_{j\bullet} := \sum_{k=1}^l h_{jk}$.
- Die **absolute Randhäufigkeit** von b_k ist $h_{\bullet k} := \sum_{j=1}^m h_{jk}$.
- Die **relative Randhäufigkeit** von a_j ist $r_{j\bullet} := \sum_{k=1}^l r_{jk}$.
- Die **relative Randhäufigkeit** von b_k ist $r_{\bullet k} := \sum_{j=1}^m r_{jk}$.

Ist xy eine zweidimensionale Stichprobe zweier Merkmale X und Y mit wenig Ausprägungen (also „kleinen“ Ausprägungsmengen) so kann man zur Darstellung dieser Stichprobe eine **Kontingenztafel** oder **Häufigkeitstabelle** benutzen. Es gibt jeweils eine Kontingenztafel für die absoluten Häufigkeiten

	b_1	\dots	b_l	Randhäufigkeit X
a_1	h_{11}	\dots	h_{1l}	$h_{1\bullet}$
\vdots	\vdots		\vdots	\vdots
a_m	h_{m1}	\dots	h_{ml}	$h_{m\bullet}$
Randhäufigkeit Y	$h_{\bullet 1}$	\dots	$h_{\bullet l}$	$h_{\bullet\bullet} = n$

und für die relativen Häufigkeiten.

	b_1	\dots	b_l	Randhäufigkeit X
a_1	r_{11}	\dots	r_{1l}	$r_{1\bullet}$
\vdots	\vdots		\vdots	\vdots
a_m	r_{m1}	\dots	r_{ml}	$r_{m\bullet}$
Randhäufigkeit Y	$r_{\bullet 1}$	\dots	$r_{\bullet l}$	$r_{\bullet\bullet} = 1$

Beispiel 88. Es werden alle $n = 90$ Teilnehmer*innen einer Vorlesung zu ihrem Mensa- und Kaffeetrinkverhalten befragt. Genauer messen wir das qualitative Merkmal $X =$ „Person isst in der Mensa“ mit Ausprägungsmenge $A = \{\text{Ja, Nein}\}$ und $Y =$ „Art des Kaffeegetränks“ mit Ausprägungsmenge $B = \{\text{Filter, Spezial, Keiner}\}$. Die Kontingenztafel der absoluten Häufigkeiten für die gefundene Stichprobe ist die folgende.

	Filter	Spezial	Keiner	
Ja	$h_{11} = 3$	$h_{12} = 25$	$h_{13} = 19$	$h_{1\bullet} = 47$
Nein	$h_{21} = 11$	$h_{22} = 23$	$h_{23} = 9$	$h_{2\bullet} = 43$
	$h_{\bullet 1} = 14$	$h_{\bullet 2} = 48$	$h_{\bullet 3} = 28$	$h_{\bullet\bullet} = 90$

1.3.1. Empirische Unabhängigkeit. In diesem Kapitel wollen wir ein Maß (dieses wird eine reelle Zahl sein, der sogenannte „Kontingenzkoeffizient“) dafür einführen, wie sehr die Merkmalsausprägungen von X und Y in einer zweidimensionalen Stichprobe xy „voneinander abhängen“.

Beispiel 89. In dem obigen Mensa-Kaffee-Beispiel 88 ist allein durch das Ansehen der Kontingenztafel nicht ersichtlich, ob es einen Zusammenhang zwischen dem Mensabesuch und der Wahl des Kaffeegetränks gibt.

Definition 90. Sei $xy = ((x_1, y_1), \dots, (x_n, y_n))$ eine zweidimensionale Stichprobe zweier Merkmale X und Y mit jeweils endlichen Ausprägungsmengen $A = \{a_1, \dots, a_m\}$ und $B = \{b_1, \dots, b_l\}$. Sei außerdem $j \in \{1, \dots, m\}$ und $k \in \{1, \dots, l\}$, wobei wir annehmen, dass $h_{j\bullet} \neq 0 \neq h_{\bullet k}$ gilt.

- Die bedingte Häufigkeit von $Y = b_k$ unter der Bedingung $X = a_j$ ist

$$f_Y(b_k|a_j) := \frac{h_{jk}}{h_{j\bullet}} = \frac{r_{jk}}{r_{j\bullet}}.$$

- Die bedingte Häufigkeit von $X = a_j$ unter der Bedingung $Y = b_k$ ist

$$f_X(a_j|b_k) := \frac{h_{jk}}{h_{\bullet k}} = \frac{r_{jk}}{r_{\bullet k}}.$$



Bitte erschrecken Sie sich nicht vor dieser Definition! Wir werden sie nicht oft brauchen, da diese Begriffe uns hauptsächlich als Motivation für die Einführung des oben erwähnten „Abhängigkeitsmaßes“ dienen werden.

Bemerkung 91. Es gilt $\sum_{j=1}^m f_X(a_j|b_k) = 1$ und $\sum_{k=1}^l f_Y(b_k|a_j) = 1$.

Beispiel 92. Wir betrachten wieder das Mensa-Kaffee-Beispiel 88. Hier können wir die folgenden zwei Tabellen der jeweiligen bedingten Häufigkeiten aufstellen.

	Filter	Spezial	Keiner
Ja	$f_Y(b_1 a_1) = \frac{3}{47}$	$f_Y(b_2 a_1) = \frac{25}{47}$	$f_Y(b_3 a_1) = \frac{19}{47}$
Nein	$f_Y(b_1 a_2) = \frac{11}{43}$	$f_Y(b_2 a_2) = \frac{23}{43}$	$f_Y(b_3 a_2) = \frac{9}{43}$

	Filter	Spezial	Keiner
Ja	$f_X(a_1 b_1) = \frac{3}{14}$	$f_X(a_1 b_2) = \frac{25}{48}$	$f_X(a_1 b_3) = \frac{19}{28}$
Nein	$f_X(a_2 b_1) = \frac{11}{14}$	$f_X(a_2 b_2) = \frac{23}{48}$	$f_X(a_2 b_3) = \frac{9}{28}$

Definition 93. Sei $xy = ((x_1, y_1), \dots, (x_n, y_n))$ eine zweidimensionale Stichprobe zweier Merkmale X und Y mit jeweils endlichen Ausprägungsmengen $A = \{a_1, \dots, a_m\}$ und $B = \{b_1, \dots, b_l\}$, wobei wir annehmen, dass $h_{j\bullet} \neq 0 \neq h_{\bullet k}$ gilt. Es heißen X und Y **empirisch unabhängig**, falls für alle $k \in \{1, \dots, l\}$ gilt

$$f_Y(b_k|a_1) = \dots = f_Y(b_k|a_m) =: f_Y(b_k)$$

oder äquivalent, falls für alle $j \in \{1, \dots, m\}$ gilt

$$f_X(a_j|b_1) = \dots = f_X(a_j|b_l) =: f_X(a_j).$$



Es sind also X und Y empirisch unabhängig, wenn für alle k der Wert $f_Y(b_k|a_j)$ nicht „von a_j abhängt“. Im vorherigen Mensa-Kaffee-Beispiel 92 sind X und Y also empirisch unabhängig, wenn in der ersten Tabelle jeweils die **Spalten** gleiche Werte aufweisen (oder äquivalent, wenn in der zweiten Tabelle jeweils die **Zeilen** gleiche Werte aufweisen).

Bemerkung 94. Es ist etwas merkwürdig, dass die vorherige Definition eine Eigenschaft der Merkmale X und Y definiert, aber im Wesentlichen von der Stichprobe xy abhängt.

In dem folgenden Satz wollen wir den Begriff der empirischen Unabhängigkeit auf eine andere Weise beschreiben. Diese Charakterisierung erlaubt uns dann zu messen, „wie stark“ zwei Merkmale empirisch unabhängig sind.

Satz 95. Sei eine zweidimensionale Stichprobe xy wie in Definition 93 gegeben. Dann sind die folgenden Aussagen äquivalent.

- (1) X und Y sind empirisch unabhängig.
- (2) Für alle $j \in \{1, \dots, m\}$ und alle $k \in \{1, \dots, l\}$ gilt $h_{jk} = \frac{h_{j\bullet} h_{\bullet k}}{n}$.

Beweis. Zunächst bemerken wir, dass für alle $k \in \{1, \dots, l\}$ gilt

$$r_{\bullet k} = \sum_{j=1}^m r_{jk} = \sum_{j=1}^m \frac{h_{jk}}{n} = \sum_{j=1}^m \left(\frac{h_{jk}}{h_{j\bullet}} \cdot \frac{h_{j\bullet}}{n} \right) = \sum_{j=1}^m \left(f_Y(b_k|a_j) \cdot \frac{h_{j\bullet}}{n} \right).$$

Nun zeigen wir, dass aus der Aussage (1) die Aussage (2) folgt. Unter der Annahme von (1) gilt also für alle $k \in \{1, \dots, l\}$, dass $f_Y(b_k|a_j) = f_Y(b_k)$ für alle $j \in \{1, \dots, m\}$ und damit, dass

$$r_{\bullet k} = \sum_{j=1}^m \left(f_Y(b_k) \cdot \frac{h_{j\bullet}}{n} \right) = f_Y(b_k) \sum_{j=1}^m \frac{h_{j\bullet}}{n} = f_Y(b_k).$$

Daraus folgt für alle $j \in \{1, \dots, m\}$ und alle $k \in \{1, \dots, l\}$, dass

$$\frac{h_{jk}}{h_{j\bullet}} = f_Y(b_k|a_j) = f_Y(b_k) = r_{\bullet k} = \frac{h_{\bullet k}}{n},$$

was nach Umstellung

$$h_{jk} = \frac{h_{j\bullet} h_{\bullet k}}{n}$$

ergibt.

Es ist eine Übung, die andere Implikation zu beweisen. □

Bemerkung 96. Der Beweis des vorherigen Satzes 95 kann dazu benutzt werden zu zeigen, dass die beiden Bedingungen in Definition 93 tatsächlich äquivalent sind.

Wenn wir also messen wollen, wie „stark“ zwei Merkmale X und Y empirisch unabhängig sind, können wir untersuchen, wie groß die Zahl

$$h_{jk} - \frac{h_{j\bullet} h_{\bullet k}}{n}$$

für jedes k und jedes j ist. Ist sie gleich Null (für jedes j und k), so sagt uns der vorherige Satz 95 ja gerade, dass X und Y empirisch unabhängig sind. Um diese Abweichungen in nur einer Zahl (und nicht in jeweils einer für jedes j und k) zu messen, möchten wir die obigen Zahlen alle addieren. Damit sich nicht eine negative dieser Zahlen mit einer positiven zu Null addiert (obwohl vielleicht keine empirische Unabhängigkeit vorliegt), summieren wir die Quadrate der obigen Zahlen, die ja immer positiv sind, und erhalten die Zahl

$$\sum_{j=1}^m \sum_{k=1}^l \left(h_{jk} - \frac{h_{j\bullet} h_{\bullet k}}{n} \right)^2.$$

Diese Zahl „normiert“ man nun noch auf eine bestimmte Weise und erhält die folgende Definition.

Definition 97. Sei eine zweidimensionale Stichprobe xy wie in Definition 93 gegeben.

- Der χ^2 -**Koeffizient** („Chi-Quadrat-Koeffizient“) von xy ist die (nicht-negative) Zahl

$$\chi^2 := \sum_{j=1}^m \sum_{k=1}^l \frac{(h_{jk} - \tilde{h}_{jk})^2}{\tilde{h}_{jk}}, \quad \text{wobei} \quad \tilde{h}_{jk} := \frac{h_{j\bullet} h_{\bullet k}}{n}.$$

- Der (Pearson'sche) **Kontingenzkoeffizient** von xy ist die (nicht negative) Zahl

$$K := \sqrt{\frac{\chi^2}{n + \chi^2}}.$$

- Der **korrigierte Kontingenzkoeffizient** von xy ist die (im Intervall $[0, 1]$ liegende) Zahl

$$K^* := K \cdot \sqrt{\frac{M}{M-1}}, \quad \text{wobei} \quad M := \min\{m, l\}.$$

Bemerkung 98. Es gilt

$$K^* = 0 \Leftrightarrow K = 0 \Leftrightarrow \chi^2 = 0 \Leftrightarrow X \text{ und } Y \text{ sind empirisch unabhängig}$$

und X und Y sind umso „empirisch unabhängiger“, desto kleiner diese Zahlen jeweils sind.

Beispiel 99. Wir möchten im Mensa-Kaffee-Beispiel 88 entscheiden, wie sehr die Merkmale X und Y empirisch unabhängig sind. Hierfür müssen wir den zugehörigen korrigierten Kontingenzkoeffizienten K^* bestimmen, wofür wir zunächst den zugehörigen χ^2 -Koeffizienten ausrechnen wollen. Hierfür wiederum ist es empfehlenswert, zunächst eine „Tabelle der \tilde{h}_{jk} “, eine sogenannte **Unabhängigkeitstabelle**, hinzuschreiben.

	Filter	Spezial	Keiner
Ja	$\tilde{h}_{11} \approx 7,31$	$\tilde{h}_{12} \approx 25,06$	$\tilde{h}_{13} \approx 14,62$
Nein	$\tilde{h}_{21} \approx 6,68$	$\tilde{h}_{22} \approx 22,93$	$\tilde{h}_{23} \approx 13,37$

Es ist beispielsweise $\tilde{h}_{11} = \frac{47 \cdot 14}{90} \approx 7,31$. Bei empirischer Unabhängigkeit wären die Werte in dieser Unabhängigkeitstabelle gleich denen in der Kontingenztafel (siehe Beispiel 88). Nun rechnet man

$$\chi^2 \approx 8,06, \quad K \approx \sqrt{\frac{8,06}{90 + 8,06}} \approx 0,287, \quad K^* = K \cdot \sqrt{\frac{2}{2-1}} \approx 0,405,$$

die beiden Merkmale X und Y sind also nicht besonders stark empirisch abhängig.

Bemerkung 100. Hat man Merkmale X und Y mit jeweils zweielementigen Ausprägungsmengen gegeben, ist also $m = 2$ und $l = 2$, so lässt sich der χ^2 -Koeffizient auch durch

$$\chi^2 = n \cdot \frac{(h_{11}h_{22} - h_{12}h_{21})^2}{h_{1\bullet}h_{2\bullet}h_{\bullet 1}h_{\bullet 2}}$$

ausdrücken, was Rechnungen vereinfachen kann, wie das folgende Beispiel 101 zeigt.

Beispiel 101. Es wurden $n = 100$ Personen zu den Merkmalen X = „Bezug von Kindergeld“ mit Ausprägungsmenge $A = \{\text{Ja}, \text{Nein}\}$ und Y = „Kinder vorhanden“ mit Ausprägungsmenge $B = \{\text{Kinder}, \text{Keine}\}$ befragt. Wir vermuten natürlich, dass diese beiden Merkmale sehr stark empirisch abhängig sind. Die Kontingenztafel der absoluten Häufigkeiten für die gefundene Stichprobe ist die folgende.

	Kinder	Keine	
Ja	$h_{11} = 62$	$h_{12} = 0$	$h_{1\bullet} = 62$
Nein	$h_{21} = 0$	$h_{22} = 38$	$h_{2\bullet} = 38$
	$h_{\bullet 1} = 62$	$h_{\bullet 2} = 38$	$h_{\bullet\bullet} = 100$

Es gilt $\chi^2 = 100 \cdot \frac{(62 \cdot 38 - 0)^2}{62 \cdot 38 \cdot 62 \cdot 38} = 100$. (Insbesondere sehen wir hier noch einmal explizit, dass der χ^2 -Koeffizient stark vom Stichprobenumfang $n = 100$ abhängen kann.) Außerdem ist $K = \sqrt{\frac{100}{100+100}} = \sqrt{\frac{1}{2}}$ und damit $K^* = 1$. Die beiden Merkmale X und Y sind also sehr stark empirisch abhängig.

Bemerkung 102. Eine beliebte Fehlinterpretation einer starken empirischen Abhängigkeit zweier Merkmale ist, dass eine bestimmte Ausprägung des einen Merkmals eine bestimmte Ausprägung des anderen Merkmals „zur Folge hat“. Man spricht bei dieser Fehlinterpretation auch von einer **Scheinkausalität** (welche manchmal irreführenderweise auch „Scheinkorrelation“ genannt wird). Schaut man sich das vorherige Beispiel 101 an, so besteht sicherlich in der Realität die Kausalitätsrelation

$$\text{„Bezug von Kindergeld} = \text{Ja“} \implies \text{„Kinder vorhanden} = \text{Kinder“}$$

wobei der Pfeil „ \implies “ bedeuten soll, dass die linke Seite die rechte Seite impliziert, also „zur Folge hat“. Man könnte also auf die Idee kommen, dass eine hohe empirische Abhängigkeit stets von einer solchen Kausalrelation herkommt. Dieses ist jedoch falsch und führt zu vielen schwerwiegenden Fehlinterpretationen.

Beispiel 103. Als Beispiel einer Scheinkausalität, wie sie in der vorherigen Bemerkung 102 beschrieben wurde, könnte in einer (zweidimensionalen) Stichprobe eine starke empirische Abhängigkeit zwischen den Merkmalen $X = \text{„Existenz einer Tätowierung“}$ und $Y = \text{„Verstorben mit weniger als 60 Jahren“}$ festgestellt werden und anschließend die **falsche** Schlussfolgerung getroffen werden, dass die Existenz einer Tätowierung zu einem frühen Tod führt. Eine Schlagzeile zu dieser falschen Schlussfolgerung wäre: „Tätowieren ist tödlich“. Eine Erklärung für die empirische Abhängigkeit dieser beiden Merkmale könnte aber zum Beispiel eine andere gemeinsame Ursache, wie beispielsweise ein bestimmter Lebensstil, sein.

1.3.2. Kovarianz und Korrelationskoeffizient. In diesem Kapitel wollen wir ein weiteres Maß (dieses wird eine reelle Zahl sein, der sogenannte „Korrelationskoeffizient“) dafür kennenlernen, wie sehr die Merkmalsausprägungen von X und Y in einer zweidimensionalen Stichprobe xy „voneinander abhängen“. Wie wir schon im Kapitel 1.2.2 diskutiert haben, sind die absoluten Häufigkeiten eines quantitativen Merkmals (mit einer großen Ausprägungsmenge) oft Eins oder Null. Daher ist hier das im vorherigen Kapitel eingeführte Maß für eine Abhängigkeit, der (korrigierte) Kontingenzkoeffizient, nutzlos. (Wir könnten uns natürlich auch mit einer Gruppeneinteilung wie in Kapitel 1.2.2 behelfen und den korrigierten Kontingenzkoeffizient benutzen, wollen aber eine andere Methode kennenlernen).



Im Wesentlichen kann man sich merken, dass wir für qualitative oder Rangmerkmale den **(korrigierten) Kontingenzkoeffizienten** und für quantitative Merkmale den (unten eingeführten) **empirischen Korrelationskoeffizienten** benutzen sollten, wenn wir die Stärke der Abhängigkeit zweier Merkmale beschreiben möchten.

Definition 104. Sei $xy = ((x_1, y_1) \dots, (x_n, y_n))$ eine zweidimensionale Stichprobe vom Umfang $n \geq 2$ zweier quantitativer Merkmale X und Y .

- Die (korrigierte) **empirische Kovarianz** von xy ist die reelle Zahl

$$s_{xy} := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

wobei \bar{x} den empirische Mittelwert der Stichprobe $x = (x_1, \dots, x_n)$ und \bar{y} den empirische Mittelwert der Stichprobe $y = (y_1, \dots, y_n)$ bezeichnet.

- Der **empirische Korrelationskoeffizient** von xy ist die (im Intervall $[-1, 1]$ liegende) Zahl

$$r_{xy} := \frac{s_{xy}}{s_x \cdot s_y},$$

wobei s_x die empirische Standardabweichung von x und s_y die empirische Standardabweichung von y bezeichnet.

Bemerkung 105. Durch eine Umformung ergibt sich für den empirischen Korrelationskoeffizienten die Formel

$$r_{xy} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\left(\sqrt{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \right) \cdot \left(\sqrt{\sum_{i=1}^n y_i^2 - n \bar{y}^2} \right)},$$

welche in der Praxis oft einfacher zu verwenden ist.

Bevor wir uns Beispielen von empirischen Korrelationskoeffizienten widmen, wollen wir uns ansehen, was der empirische Korrelationskoeffizient für eine „graphische Bedeutung“ hat. Als graphische Darstellung einer zweidimensionalen Stichprobe $xy = ((x_1, y_1), \dots, (x_n, y_n))$ zweier quantitativer Merkmale X und Y mit Ausprägungsmengen $A \subseteq \mathbb{R}$ und $B \subseteq \mathbb{R}$ bietet sich oft ein sogenannter **Scatterplot** an.

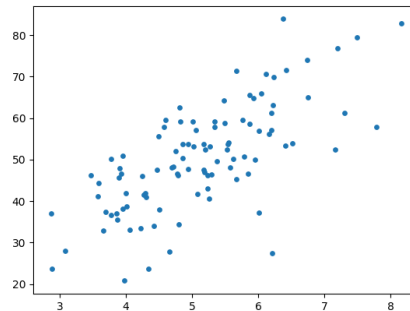


Abbildung 12: Scatterplot

Hierbei zeichnen wir für jede Merkmalsausprägung (x_i, y_i) einen Punkt in die euklidische Ebene $A \times B \subseteq \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. Im obigen Bildchen ist beispielsweise $A = [2, 8]$ (die „X-Achse“) und $B = [20, 100]$ (die „Y-Achse“).

Bemerkung 106.

- Falls der empirische Korrelationskoeffizient r_{xy} nahe bei 1 liegt, deutet dieses auf einen starken **positiven linearen** Zusammenhang hin. Mit anderen Worten bedeutet dieses, man kann durch die Punkte im zugehörigen Scatterplot ungefähr eine Gerade mit positiver Steigung legen.
- Falls der empirische Korrelationskoeffizient r_{xy} nahe bei -1 liegt, deutet dieses auf einen starken **negativen linearen** Zusammenhang hin. Mit anderen Worten bedeutet dieses, man kann durch die Punkte im zugehörigen Scatterplot ungefähr eine Gerade mit negativer Steigung legen.
- Falls der empirische Korrelationskoeffizient r_{xy} nahe bei 0 liegt, deutet dieses auf **keinen linearen** Zusammenhang hin (es könnte aber durchaus einen komplizierteren, wie beispielsweise einen exponentiellen, Zusammenhang geben).

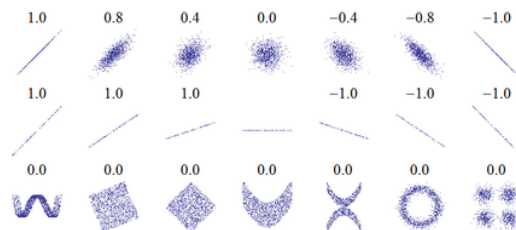


Abbildung 13: Verschiedene Scatterplots mit empirischen Korrelationskoeffizienten [5]

Bemerkung 107. Man sagt die Merkmale X und Y seien

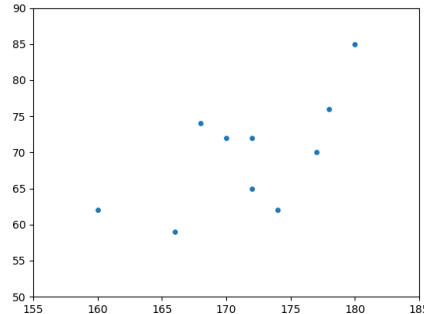
- **(linear) unkorreliert**, falls $0 \leq |r_{xy}| \leq 0,2$,
- **schwach (linear) korreliert**, falls $0,2 < |r_{xy}| \leq 0,5$,
- **mittel (linear) korreliert**, falls $0,5 < |r_{xy}| \leq 0,8$,
- **stark (linear) korreliert**, falls $0,8 < |r_{xy}| \leq 1$,

aber diese Begriffe sind offenbar nicht klar definiert und daher mit Vorsicht zu verwenden.

Beispiel 108. Es wurden $n = 10$ Personen zu den beiden quantitativ stetigen Merkmalen X = „Körpergröße“ und Y = „Gewicht“ befragt und die folgende Stichprobe gefunden

$$xy = ((172, 72), (166, 59), (172, 65), (178, 76), (160, 62), (168, 74), (170, 72), (177, 70), (174, 62), (180, 85)).$$

Der zugehörige Scatterplot sieht wie folgt aus.



Es sind $\bar{x} = \frac{1717}{10} = 171,7$ und $\bar{y} = \frac{697}{10} = 69,7$ die empirischen Mittelwerte von x und y , $s_x \approx \sqrt{36,46} \approx 6,04$ und $s_y \approx \sqrt{62,01} \approx 7,87$ die empirischen Standardabweichungen, sowie $s_{xy} \approx 31,23$ die empirische Kovarianz. Damit gilt für den empirischen Korrelationskoeffizienten

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \approx 0,66$$

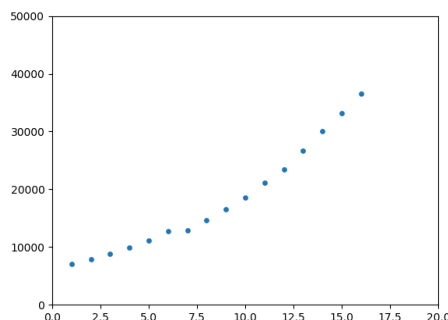
und die beiden Merkmale sind mittel **positiv** (linear) korreliert.

Bemerkung 109. Obwohl bei einem empirischen Korrelationskoeffizienten, der nicht nahe bei ± 1 liegt kein **linearer** Zusammenhang zwischen den Merkmalen X und Y besteht, kann es jedoch trotzdem einen **nichtlinearen** Zusammenhang geben, wie schon in Bemerkung 106 erwähnt wurde. Wenn man die „Art des möglichen Zusammenhangs“ (exponentiell, logarithmisch,...) vermuten kann, kann man diesen Fall durch geeignete Umformungen auf den linearen Fall zurückführen, wie das folgende Beispiel 110 zeigt.

Beispiel 110. Wir betrachten die beiden Merkmale X = „Tag ab dem 15. März 2020“ und Y = „Corona Todesfälle (weltweit)“. Die folgende Tabelle enthält Daten der Weltgesundheitsorganisation (WHO) vom 09.05.2020.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
7030	7833	8779	9840	11186	12784	12944	14653	16594	18573	21124	23477	26628	29997	33215	36522

Der zugehörige Scatterplot sieht wie folgt aus.



Wir vermuten (beispielsweise durch Erfahrung oder Betrachtung des Scatterplots), dass zwischen den Merkmalen X und Y kein linearer, sondern ein exponentieller Zusammenhang besteht. Mit anderen Worten vermuten wir, dass man durch die Punkte im Scatterplot nicht ungefähr eine Gerade, also den Graph einer Funktion $x \mapsto mx + b$, sondern ungefähr den Graph der Funktion $x \mapsto c \cdot a^{\lambda x}$ (mit entsprechendem Quell- und Zielbereich) für gewisse

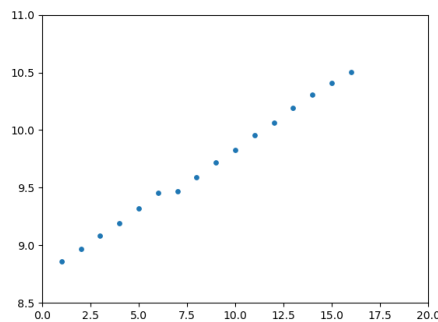
$a > 0$ und $c, \lambda \in \mathbb{R}$ legen kann. Zunächst bemerken wir, dass man wegen $c \cdot a^{\lambda x} = c \cdot e^{(\ln(a)\lambda)x}$ (siehe Definition 24) auch nur Funktionen $x \mapsto c \cdot e^{mx}$ für $c, m \in \mathbb{R}$ betrachten kann (man kann ja einfach $m = \ln(a)\lambda$ setzen). Mit dem gleichen Argument (man kann ja einfach $b = \ln(\pm c)$ setzen), können wir auch nur Funktionen $x \mapsto \pm e^{b+mx}$ für $b, m \in \mathbb{R}$ betrachten. Wendet man (unter der Voraussetzung eines positiven Vorzeichens) die natürliche Logarithmusfunktion hierauf an, bekommt man

$$x \mapsto \ln(e^{b+mx}) = \underbrace{b}_{\text{„Y-Achsenabschnitt“}} + \underbrace{m}_{\text{„Steigung“}} x,$$

also eine „Geradenfunktion“. Mit anderen Worten vermuten wir nach Anwenden des (natürlichen) Logarithmus auf die Merkmalsausprägungen von Y einen linearen Zusammenhang. Das Anwenden von $\ln(-)$ liefert (ungefähr) die folgende Tabelle.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
8,858	8,966	9,080	9,194	9,322	9,456	9,468	9,592	9,717	9,829	9,958	10,064	10,190	10,309	10,411	10,506

Wir bezeichnen das durch Anwenden der Funktion „ln“ modifizierte Merkmal Y als Z . Der zu der (modifizierten) Stichprobe xz gehörige Scatterplot sieht wie folgt aus.



Nun möchten wir den empirischen Korrelationskoeffizienten dieser modifizierten Stichprobe xz ausrechnen. Es sind $\bar{x} = 8,5$ und $\bar{z} \approx 9,623$ die empirischen Mittelwerte von x und z , $s_x \approx 4,761$ und $s_z \approx 0,529$ die empirischen Standardabweichungen, sowie $s_{xz} \approx 2,5$ die empirische Kovarianz. Damit gilt für den empirischen Korrelationskoeffizienten

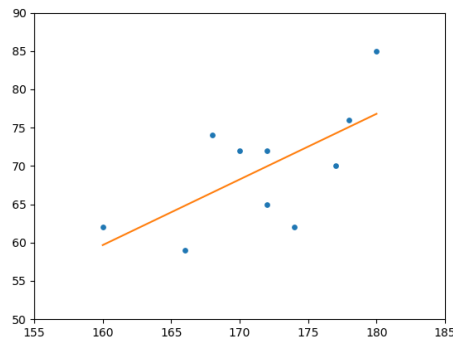
$$r_{xz} = \frac{s_{xz}}{s_x s_z} \approx 0,993$$

und die beiden Merkmale X und Z sind stark **positiv linear** korreliert. Damit besteht also zwischen den beiden ursprünglichen Merkmalen X und Y eine starke **positive exponentielle** Korrelation (wir haben nicht genau definiert, was dieses bedeutet, könnten aber das obige Vorgehen für eine Definition benutzen). Im folgenden Kapitel 1.3.3 versuchen wir, auf eine „möglichst gute Weise“ eine Exponentialfunktion durch diese Stichprobenwerte zu legen. (Mit anderen Worten versuchen wir „möglichst passende“ m und b zu finden.)

1.3.3. Regressionsrechnung. In diesem Kapitel wollen wir zu einer zweidimensionalen Stichprobe xy zweier quantitativer Merkmale X und Y , welche stark linear korreliert sind, ungefähr eine Gerade, die sonenannte **Regressionsgerade**, durch die Punkte

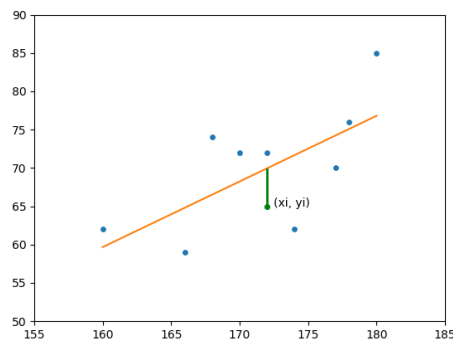
$$\{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq \mathbb{R}^2$$

im zugehörigen Scatterplot legen.



Eine Gerade in der Ebene \mathbb{R}^2 ist der Graph einer linearen Funktion, also einer Funktion der Form $x \mapsto mx + b$ für reelle Zahlen m (die „Steigung“) und b (der „Y-Achsenabschnitt“). Unsere Aufgabe ist es also, die Zahlen m und b mit Hilfe der Stichprobe xy „möglichst gut“ zu bestimmen.

Um dieses „möglichst gut“ zu präzisieren, benutzen wir die **Methode der kleinsten Quadrate**, die wir nun diskutieren werden. Schauen wir uns dafür den Abstand einer Merkmalsausprägung (x_i, y_i) zu der durch die Funktion $x \mapsto mx + b$ beschriebenen Gerade an.



Dieser im obigen Bild als grüne Strecke illustrierte Abstand ist die Zahl $|y_i - (mx_i + b)|$. Um die Güte der Geraden in Bezug auf die Stichprobe zu messen, wollen wir über diese Abstände für alle Merkmalsausprägungen (x_i, y_i) summieren. Damit größere Abstände ein größeres „Gewicht“ beim Maß dieser Güte haben (Diese großen Abstände sollen besonders schlecht sein), quadriert man diese Abstände noch und erhält, abhängig von m und b , die nicht-negative reelle Zahl

$$Q_{xy}(m, b) := \sum_{i=1}^n (y_i - (mx_i + b))^2,$$

welche **Summe der quadratischen Fehler** genannt wird.

Definition 111. Sei $xy = ((x_1, y_1) \dots, (x_n, y_n))$ eine zweidimensionale Stichprobe vom Umfang $n \geq 2$ zweier quantitativer Merkmale X und Y . Eine **Regressionsgerade** für xy ist eine Funktion

$$\begin{aligned} r: \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto mx + b \end{aligned}$$

mit der Eigenschaft, dass die Summe der Quadratischen Fehler $Q_{xy}(m, b)$ minimal ist (unter allen Geraden).

Mit Differenzialrechnung in zwei Variablen kann man den folgenden Satz zeigen.

Satz 112. Sei eine zweidimensionale Stichprobe xy wie in Definition 111 gegeben. Dann existiert eine Regressionsgerade

$$\begin{aligned} r: \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto mx + b \end{aligned}$$

und ist gegeben durch $m := \frac{s_{xy}}{s_x^2}$ und $b := \bar{y} - m\bar{x}$, wobei s_{xy} die empirische Kovarianz von xy , s_x und s_y die jeweilige empirische Standardabweichung und \bar{x} und \bar{y} den jeweiligen empirischen Mittelwert von x und y bezeichnen. Außerdem verläuft der Graph der Regressionsgeraden durch den Punkt (\bar{x}, \bar{y}) .

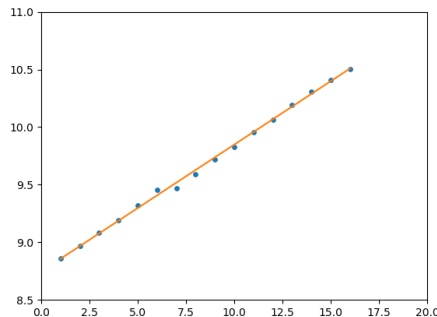
Bemerkung 113. Sei eine zweidimensionale Stichprobe xy wie in Definition 111 gegeben. Wie gut beschreibt die im vorherigen Satz 112 angegebene Regressionsgerade die Stichprobe xy tatsächlich? Wir schauen uns die Summe der quadratischen Fehler $Q_{xy}(m, b)$ an (multipliziert mit der von der Geraden völlig unabhängigen Zahl $\frac{1}{n-1}$).

$$\begin{aligned} \frac{1}{n-1} \cdot Q_{xy}(m, b) &= \frac{1}{n-1} \sum_{i=1}^n ((y_i - (mx_i + b))^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(y_i - \left(\frac{s_{xy}}{s_x^2} x_i + \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x} \right) \right)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n \left((y_i - \bar{y}) - \frac{s_{xy}}{s_x^2} (x_i - \bar{x}) \right)^2 \\ &= s_y^2 \cdot \left(1 - \frac{s_{xy}^2}{s_x^2 s_y^2} \right) \\ &= s_y^2 \cdot (1 - r_{xy}^2) \end{aligned}$$

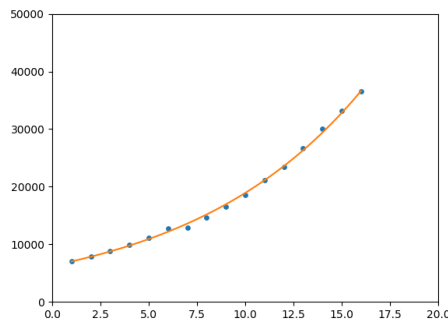
Diese Zahl ist also klein, falls $r_{xy} \approx \pm 1$, die Merkmale X und Y also stark (linear) korreliert sind. In diesem Fall approximiert die Regressionsgerade die Stichprobe xy also gut. Insbesondere sehen wir, dass alle Stichprobenwerte (x_i, y_i) genau dann auf einer Geraden (nämlich der Regressionsgeraden) liegen, falls $r_{xy} = \pm 1$.

Beispiel 114. Wir betrachten das Körpergröße-Gewicht Beispiel 108. Hier gilt für die „Steigung“ $m = \frac{s_{xy}}{s_x^2} \approx 0,86$ und für den „ Y -Achsenabschnitt“ $b = \bar{y} - m\bar{x} \approx -77,29$, die Regressionsgerade ist also (ungefähr) durch die Funktion $x \mapsto 0,86x - 77,29$ gegeben. Die beiden Bildchen am Anfang dieses Kapitels 1.3.3 zeigen den Graph dieser Regressionsgeraden.

Beispiel 115. Wir betrachten das Corona Beispiel 110 und die modifizierte Stichprobe xz . Hier gilt $m = \frac{s_{xz}}{s_x^2} \approx 0,1103$ und $b = \bar{z} - m\bar{x} \approx 8,686$, die Regressionsgerade ist also (ungefähr) durch die Funktion $x \mapsto 0,1103x + 8,686$ gegeben.



Nun wollten wir aber ursprünglich den Graph einer Funktion $x \mapsto \pm e^{b+mx}$ auf eine „möglichst gute Weise“ durch den Scatterplot der Stichprobe xy legen. Die gesuchte Funktion ist also $x \mapsto e^{8,686+0,1103x}$.



Am Ende dieses Beispiels möchten wir noch kurz zwei aktuell diskutierte Begriffe, die Verdopplungszeit und die Replikationsrate erklären. Hierbei nehmen wir stark vereinfacht an, dass die Merkmale X und Y tatsächlich durch die obige Funktion $x \mapsto f(x) := e^{8,686+0,1103x}$ zusammenhängen. Die **Verdopplungszeit** d_x zu einem Zeitpunkt x („Wert auf der X -Achse“) ist ein anderer Zeitpunkt d_x , sodass $f(d_x) = 2f(x)$. Umschreiben, wobei wir $m \neq 0$ voraussetzen, ergibt

$$\begin{aligned} e^{b+md_x} &= 2 \cdot e^{b+mx} \\ \Leftrightarrow b + md_x &= \ln(2) + b + mx \\ \Leftrightarrow d_x &= \frac{\ln(2)}{m} + x \\ \Leftrightarrow d_x &\approx 6.284 + x. \end{aligned}$$

Die **Replikationsrate** R ist eine Zahl R , sodass $f(x+1) = R \cdot f(x)$. (Bisher ist noch nicht klar, dass diese Zahl nicht von x abhängt.) Umschreiben ergibt

$$\begin{aligned} e^{b+m(x+1)} &= R \cdot e^{b+mx} \\ \Leftrightarrow b + mx + m &= \ln(R) + b + mx \\ \Leftrightarrow m &= \ln(R) \\ \Leftrightarrow R &= e^m \\ \Leftrightarrow R &\approx 1,116. \end{aligned}$$

2. Wahrscheinlichkeitstheorie

Das Ziel der Wahrscheinlichkeitstheorie ist die Beschreibung von Zufallsexperimenten, also Experimenten mit unbestimmtem Ausgang. Diese Beschreibung brauchen wir, um uns im nächsten Kapitel 3 mit der induktiven Statistik (oder schließenden Statistik) gut auseinanderzusetzen zu können.

2.1. Grundlegende Begriffe.



Es ist vielleicht nicht schlecht, an dieser Stelle noch einmal einen Blick auf die wesentlichen Begriffe zu dem Thema „Mengen“ am Anfang des Kapitels „Elementare mathematische Grundlagen“ zu werfen.

Definition 116. Sei M eine Menge. Die Menge

$$\mathbb{P}(M) := \{N \mid N \subseteq M\},$$

also die Menge der Teilmengen von M , heißt die **Potenzmenge** von M .

Beispiel 117. Es gilt

- (1) $\mathbb{P}(\{a\}) = \{\emptyset, \{a\}\},$
- (2) $\mathbb{P}(\{1\}) = \{\emptyset, \{1\}\},$
- (3) $\mathbb{P}(\{a, b\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\},$
- (4) $\mathbb{P}(\emptyset) = \{\emptyset\},$ da $\emptyset \subseteq \emptyset.$

Im allgemeinen ist die Potenzmenge einer Menge M „viel größer“ als die Menge M selbst. Ist M eine endliche Menge mit $|M| = n$ Elementen, so hat die Potenzmenge $|\mathbb{P}(M)| = 2^n$ Elemente.

Da die Mathematik in der Wahrscheinlichkeitstheorie ein Zufallsexperiment beschreiben soll, werden für bestimmte mathematische Objekte in der folgenden Definition „neue Bezeichnungen“ eingeführt.

Definition 118. Wir fixieren im Folgenden eine Menge Ω .

- Die Menge Ω heißt **Ergebnismenge**.
- Ein Element $\omega \in \Omega$ heißt **Ergebnis**.
- Eine Teilmenge $A \subseteq \Omega$ (also ein Element $A \in \mathbb{P}(\Omega)$) heißt **Ereignis**.
- Eine einelementige Teilmenge $\{\omega\} \subseteq \Omega$ heißt **Elementarereignis**.

Zwei Ereignisse A und B mit $A \cap B = \emptyset$ heißen **disjunkt**.

Bemerkung 119. Auch führt man für bestimmte mathematische Aussagen mit der Interpretation als Zufallsexperiment die folgenden „neuen Bezeichnungen“ ein. Wieder handelt es sich hierbei nicht um eine präzise mathematische Definition. Wir haben ja beispielsweise nicht einmal genau definiert, was ein „Zufallsexperiment“ ist. Seien Ω eine Ergebnismenge und A und B Ereignisse.

- Ein Ereignis A **tritt ein**, wenn das Zufallsexperiment das Ergebnis $\omega \in A$ liefert.
- Das Ereignis $A \cap B$ bedeutet, dass A **und** B gleichzeitig eintreten.
- Das Ereignis $A \cup B$ bedeutet, dass A **oder** B eintreten. **Achtung:** In der Mathematik bedeutet „oder“ etwas anderes als „entweder oder“.
- Das Ereignis $\bar{A} := \Omega \setminus A$ bedeutet, dass A **nicht** eintritt.
- Das Ereignis $A \setminus B$ bedeutet, dass A , aber nicht B eintritt.
- Sind A und B disjunkt, so bedeutet dieses, dass die Ereignisse A und B **nicht gleichzeitig** eintreten können.

Beispiel 120.

- (1) Das Zufallsexperiment **einfacher Münzwurf** könnte man wie folgt modellieren. Wir setzen $\Omega := \{K, Z\}$, wobei K für „Kopf“ und Z für „Zahl“ steht. Es ist $A = \{Z\}$ ein Beispiels für ein Elementarereignis.
- (2) Das Zufallsexperiment **zweifacher Münzwurf** könnte man durch die Ergebnismenge $\Omega := \{(K, K), (K, Z), (Z, K), (Z, Z)\}$ modellieren. Hier soll der erste Eintrag eines Ergebnisses (a, b) für den ersten Münzwurf und der zweite Eintrag für den zweiten Münzwurf stehen. Ein Beispiel für ein Ereignis ist die Teilmenge

$$A := \{(Z, K), (Z, Z)\}$$

die als das Ereignis „Erster Wurf liefert „Zahl““ interpretiert werden kann. Für dieses mathematische Modell macht es übrigens keinen Unterschied, ob man einmal zwei Münzen gleichzeitig, oder zweimal eine Münze nacheinander wirft.

- (3) Das Zufallsexperiment **einfacher Würfelwurf** (mit einem sechsseitigen Würfel) könnte man durch die Ergebnismenge $\Omega := \{1, 2, 3, 4, 5, 6\}$ modellieren. Beispiele für Ereignisse sind

$$\begin{aligned} A &:= \{1, 3, 5\} &= \text{„Augenzahl ungerade“}, \\ B &:= \{1, 2\} &= \text{„Augenzahl} \leq 2\text{“}, \end{aligned}$$

$$\bar{A} = \{2, 4, 6\}, \quad A \cap B = \{1\}, \quad A \cup B = \{1, 2, 3, 5\}, \quad A \setminus B = \{3, 5\}.$$

- (4) Das Zufallsexperiment **zweifacher Würfelwurf** könnte man durch die Ergebnismenge $\Omega := \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$ mit $|\Omega| = 36$ Elementen modellieren. Wieder macht es für dieses mathematische Modell keinen Unterschied, ob man einmal zwei Würfel gleichzeitig, oder zweimal einen Würfel nacheinander wirft.
- (5) Das Zufallsexperiment **Werfen einer Münze bis „Zahl“ kommt** könnte man durch die Ergebnismenge

$$\Omega := \{Z, (K, Z), (K, K, Z), (K, K, K, Z), \dots\}$$

mit unendlich vielen Elementen modellieren.

- (6) Das Zufallsexperiment **Drehen eines Glücksrades** (mit genauer Winkelmessung) könnte man durch die Ergebnismenge $\Omega := [0, 360)$ modellieren.

- (7) Wir betrachten einen Sack in dem genau drei Kugeln, eine rote („R“), eine grüne („G“) und eine blaue („B“) liegen. Wir wollen das Zufallsexperiment **zweimaliges Ziehen mit Zurücklegen und ohne Reihenfolge** modellieren und wählen als Ergebnismenge

$$\Omega := \{[R, R], [R, G], [R, B], [G, G], [G, B], [B, B]\}.$$

Hier soll „mit Zurücklegen“ bedeuten, dass die beim ersten Ziehen gezogene Kugel wieder in den Sack zurückgelegt wird und nochmal gezogen werden kann. Daher ist also beispielsweise das Ergebnis $[R, R]$ des Zufallsexperiments möglich. Die Forderung „ohne Reihenfolge“ bedeutet, dass wir beispielsweise die Ziehungen (R, B) und (B, R) nicht unterscheiden und miteinander zu dem Ergebnis $[R, B]$ identifizieren – wir achten also nicht auf die Zugreihenfolge.

Bemerkung 121. Wie bereits das erste obige Beispiel 120.(1) zeigt, ist die Modellbildung eines Zufallsexperiment kein „mathematischer Prozess“ und es gibt dabei kein „richtig“ oder „falsch“. So wäre es beispielsweise legitim ein Ergebnis „S“ für „Seite“ mit zu der Ergebnismenge $\Omega = \{K, Z\}$ zu nehmen, da eine Münze ja durchaus auch auf der Seite landen könnte. Noch spitzfindiger könnte man beispielsweise das Ergebnis dafür, dass die Münze überhaupt nicht landet, sondern von geeigneten Windstößen ewig in der Luft herumgeweht wird, mit einbeziehen. Erst wenn wir uns auf ein mathematisches Modell für ein Zufallsexperiment geeinigt haben, können wir präzise „mathematische Aussagen“ treffen. Natürlich ist es eine legitime Frage, wie „gut“ gewisse Modelle sind, wie sehr sie also das tatsächliche Zufallsexperiment abbilden. Diese Fragestellungen haben allerdings nichts mit Mathematik zutun.

Man möchte nun in einem Zufallsexperiment jedem Ereignis $A \subseteq \Omega$ eine Zahl $P(A) \in [0, 1]$ zuordnen, die angibt, wie wahrscheinlich das Eintreten des Ereignisses A ist.

Definition 122. Ein Paar (Ω, P) bestehend aus einer endlichen nichtleeren Menge Ω mit $|\Omega| = n$ Elementen und der Abbildung

$$\begin{aligned} P: \mathbb{P}(\Omega) &\rightarrow [0, 1] \\ A &\mapsto \frac{|A|}{n} \end{aligned}$$

heißt **Laplace Raum** (oder Laplace Experiment). Die reelle Zahl $P(A)$ heißt die **Wahrscheinlichkeit** von A .

Bemerkung 123. Ist (Ω, P) ein Laplace Raum mit $|\Omega| = n$, so gilt für alle $\omega \in \Omega$, dass

$$P(\{\omega\}) = \frac{1}{n}.$$

alle Elementarereignisse sind also gleichwahrscheinlich. Man bezeichnet zur Abkürzung diese Zahl $\frac{1}{n}$ oft als p . Die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$ ist also $P(A) = |A| \cdot p$.

Beispiel 124. (Die Bezeichnungen beziehen sich jeweils auf Beispiel 120)

- (1) Das Zufallsexperiment **einfacher Münzwurf** sollte man als Laplace Experiment modellieren. Hier ist also $P(\{K\}) = \frac{1}{2} = P(\{Z\})$. Man sagt auch etwas ungenau, dass das Zufallsexperiment „einfacher Münzwurf“ eine Laplace Experiment ist, obwohl es sich, wie gesagt, nur um ein bestimmtes Modell handelt. **Es könnte ja beispielsweise sein, dass es durch die physische Beschaffenheit der Münze nicht gleichwahrscheinlich ist, Kopf oder Zahl zu werfen, das mathematische Modell eines Laplace Experiments das Zufallsexperiment also nicht genau beschreibt.**
- (2) Das Zufallsexperiment **zweifacher Münzwurf** sollte man auch als Laplace Experiment modellieren. Es ist dann also die Wahrscheinlichkeit für das Ereignis $A =$ „Erster Wurf liefert „Zahl““ die reelle Zahl

$$P(A) = \frac{|A|}{4} = \frac{|\{(Z, K), (Z, Z)\}|}{4} = \frac{2}{4} = \frac{1}{2}.$$

- (3) Ebenso sollte man das Zufallsexperiment **einfacher Würfelwurf** als Laplace Experiment modellieren. Es gilt dann beispielsweise

$$P(\text{„Augenzahl ungerade“}) = \frac{|\{1, 3, 5\}|}{6} = \frac{3}{6} = \frac{1}{2}.$$

- (4) Das Zufallsexperiment **zweifacher Würfelwurf** sollte man auch als Laplace Experiment modellieren. Hier ist also $p = \frac{1}{36}$.
- (5) Das Zufallsexperiment **Werfen einer Münze bis „Zahl“ kommt** kann man nicht als Laplace Experiment modellieren. Zum einen ist Ω hier keine endliche Menge, zum anderen ist das Eintreten der Ergebnisse Z und (K, K, Z) in der Realität offensichtlich nicht gleichwahrscheinlich.
- (6) Das Zufallsexperiment **Drehen eines Glücksrades** kann man ebenfalls nicht als Laplace Experiment modellieren, da Ω nicht endlich ist.
- (7) Das Zufallsexperiment **zweimaliges Ziehen mit Zurücklegen und ohne Reihenfolge** sollte man ebenfalls nicht als Laplace Experiment modellieren. Das Eintreten des Ergebnisses $[R, G]$ ist in der Realität wahrscheinlicher als das Eintreten des Ergebnisses $[R, R]$, da wir nicht auf die Reihenfolge achten und (G, R) mit (R, G) identifiziert wird, wohingegen es für (R, R) nur eine Möglichkeit gibt.

Um auch die obigen Zufallsexperimente, die nicht durch einen Laplace Experiment modelliert werden sollten wahrscheinlichkeits-theoretisch zu beschreiben, brauchen wir den folgenden allgemeineren Begriff eines Wahrscheinlichkeitsraums, der (in einer noch verallgemeinerten Weise) von Kolmogorov 1933 vorgeschlagen wurde.

Definition 125. Ein endlicher **Wahrscheinlichkeitsraum** ist ein Paar (Ω, P) bestehend aus einer endlichen nichtleeren Menge Ω und einer Abbildung

$$P: \mathbb{P}(\Omega) \rightarrow [0, 1]$$

mit den Eigenschaften

- (1) $P(\emptyset) = 0$ und $P(\Omega) = 1$ und
- (2) für alle $A, B \in \mathbb{P}(\Omega)$ gilt:

$$\text{Aus } A \cap B = \emptyset \text{ folgt } P(A \cup B) = P(A) + P(B).$$

Die Zahl $P(A)$ heißt die **Wahrscheinlichkeit** des Ereignisses A .

Der folgende Satz ist einfach zu zeigen.

Satz 126. Ein Laplace Raum ist ein endlicher Wahrscheinlichkeitsraum.

Wir halten nun erst einmal ein paar „Rechenregeln“ für Wahrscheinlichkeitsräume fest, die einfach mit Hilfe der Definition gezeigt werden können. Nach dem vorherigen Satz 126 gelten diese Regeln insbesondere für Laplace Räume.

Satz 127. Sei (Ω, P) ein Wahrscheinlichkeitsraum. Dann gilt für $A, B \in \mathbb{P}(\Omega)$:

- (1) $P(\bar{A}) = 1 - P(A)$
- (2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- (3) $P(A \setminus B) = P(A) - P(A \cap B)$
- (4) Aus $A \subseteq B$ folgt $P(A) \leq P(B)$
- (5) Für paarweise disjunkte Ereignisse $A_1, A_2, \dots \in \mathbb{P}(\Omega)$ gilt:

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

Es stellt sich die Frage, wie man einen Wahrscheinlichkeitsraum „angibt“. Sicherlich muss man zunächst einmal die Menge Ω angeben, so wie wir es in Beispiel 120.(7) getan haben. Nun muss man den Wert $P(A)$ der Funktion P für jedes Ereignis $A \in \mathbb{P}(\Omega)$ angeben. Im obigen Beispiel 120.(7) mit $|\Omega| = 6$, hat die Potenzmenge $|\mathbb{P}(\Omega)|$ genau $2^6 = 64$ Elemente. Alle diese Funktionswerte einzeln anzugeben, ist anstrengend und auch garnicht nötig, wie der folgende (und ebenfalls einfach zu beweisende) Satz zeigt.

Satz 128. Um einen endlichen Wahrscheinlichkeitsraum (Ω, P) anzugeben, genügt es eine endliche Menge Ω und eine Abbildung

$$f: \Omega \rightarrow [0, 1]$$

mit der Eigenschaft $\sum_{\omega \in \Omega} f(\omega) = 1$ anzugeben, indem man

$$\begin{aligned} P: \mathbb{P}(\Omega) &\rightarrow [0, 1] \\ A &\mapsto \sum_{\omega \in A} f(\omega) \end{aligned}$$

setzt. Insbesondere gilt also $P(\{\omega\}) = f(\omega)$ für $\omega \in \Omega$.

Beispiel 129.

- (7) Wir möchten einen endlichen Wahrscheinlichkeitsraum angeben, der das Zufallsexperiment **zweimaliges Ziehen mit Zurücklegen und ohne Reihenfolge** modelliert. Wir haben in Beispiel 120.(7) schon

$$\Omega := \{[R, R], [R, G], [R, B], [G, G], [G, B], [B, B]\}$$

definiert. Hierbei haben wir von den 9 verschiedenen (und in der Realität gleichwahrscheinlichen) Zugmöglichkeiten

$$(R, R), (R, G), (G, R), (R, B), (B, R), (G, G), (G, B), (B, G), (B, B)$$

einige identifiziert, wie beispielsweise (R, G) und (G, R) . Nach dem vorherigen Satz 128 haben wir durch

$$\begin{aligned} P(\{[R, R]\}) &= P(\{[G, G]\}) = P(\{[B, B]\}) := \frac{1}{9} \\ P(\{[R, G]\}) &= P(\{[R, B]\}) = P(\{[G, B]\}) := \frac{2}{9} \end{aligned}$$

einen Wahrscheinlichkeitsraum definiert, da $\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{2}{9} + \frac{2}{9} + \frac{2}{9} = 1$ gilt. Dieser Wahrscheinlichkeitsraum ist ein geeignetes Modell für das Zufallsexperiment.

2.2. Elemente der Kombinatorik. Um Ergebnisse etwas systematischer abzuzählen, müssen wir uns ein wenig mit elementarer Kombinatorik beschäftigen, was wir in diesem Kapitel machen möchten.

Definition 130. Sei $n \geq 1$ eine natürliche Zahl. Eine **Permutation** von n vielen Objekten x_1, \dots, x_n ist eine Umordnung dieser in einer bestimmten Reihenfolge. Mit anderen Worten ist eine Permutation der Objekte x_1, \dots, x_n ein n -Tupel (y_1, \dots, y_n) , in dem jedes x_i genau einmal vorkommt.

Beispiel 131. Es sind $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$, $(3, 2, 1)$ alle möglichen Permutationen der Objekte 1, 2 und 3.

Satz 132. Sei $n \geq 1$ eine natürliche Zahl. Es gibt genau

$$n! := 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n \quad (\text{sprich „}n \text{ Fakultät“})$$

viele verschiedene Permutationen von n vielen Objekten.

Beweis. Seien n -viele Objekte x_1, \dots, x_n gegeben.

$$\textcircled{x_1} \quad \textcircled{x_2} \quad \textcircled{x_3} \quad \dots \quad \textcircled{x_n}$$

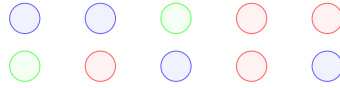
Wir müssen also die Anzahl der n -Tupel (y_1, \dots, y_n) zählen, in denen jedes x_i genau einmal (als ein y_j) vorkommt. Es ist beispielsweise $(y_1, y_2, y_3, \dots, y_n) = (x_2, x_1, x_3, \dots, x_n)$ ein solches Tupel.

$$\textcircled{x_2} \quad \textcircled{x_1} \quad \textcircled{x_3} \quad \dots \quad \textcircled{x_n}$$

Für die erste Stelle y_1 haben wir genau n -viele verschiedene Möglichkeiten, nämlich jedes der Objekte x_1, \dots, x_n . Haben wir die erste Stelle y_1 bereits gewählt, so haben wir für die Belegung der zweiten Stelle y_2 noch genau $(n-1)$ -viele verschiedene Möglichkeiten, nämlich jedes der Objekte x_1, \dots, x_n , bis auf y_1 , also das Objekt, was bereits an der ersten Stelle steht, da ja keines der x_i „doppelt“ vorkommen darf. Insgesamt ergeben sich $n \cdot (n-1)$ verschiedene Möglichkeiten für die Wahl der ersten beiden Plätze $(y_1, y_2, ?, \dots, ?)$. Führt man dieses Argument für die anderen Stellen weiter, ergeben sich also genau $n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1$ -viele Möglichkeiten. \square

Nun wollen wir die Voraussetzung, dass die Objekte x_1, \dots, x_n alle verschieden sind, fallen lassen und annehmen, es gäbe r -viele „Klassen“ aus jeweils n_1, \dots, n_r -vielen Objekten. Die Objekte innerhalb der „Klassen“ seien nicht unterscheidbar. Anstatt von Objekten wollen wir im Folgenden von „Kugeln“ reden.

Beispiel 133. Wir betrachten als Beispiel $n = 5$ Kugeln, die in $r = 3$ Klassen der Größen $n_1 = 2$, $n_2 = 1$, $n_3 = 2$ aufgeteilt sind. Zur besseren Illustration hat die erste Klasse die Farbe Blau \circ , die zweite die Farbe Grün \circ und die dritte die Farbe Rot \circ . Das folgende Bild zeigt zwei Anordnungen dieser Kugeln mit Klasseneinteilung.



Es ist übrigens $n_1 + n_2 + n_3 = n$.

Satz 134. Seien $n, r \geq 1$ natürliche Zahlen. Seien n Kugeln gegeben, die in r Klassen (jeweils „gleicher“ Kugeln) der Größen n_1, \dots, n_r aufgeteilt sind. Es gibt genau

$$\frac{n!}{n_1!n_2! \cdot n_r!}$$

viele verschiedene Anordnungen dieser.

Bemerkung 135. Der erste Satz 132 ist natürlich ein Spezialfall des obigen Satzes 134. Bei einer Permutation (also bei dem ersten Satz Satz 132) haben wir n -viele Klassen mit $n_i = 1$ Elementen, nämlich der einen Kugel x_i .



Es gilt dann mit $1! = 1$, dass

$$\frac{n!}{1 \cdot 1 \cdot \dots \cdot 1} = n!.$$

Das andere Extrembeispiel ist, wenn wir n Kugeln auf nur eine Klasse aufteilen, also $r = 1$ und $n_1 = n$ ist.

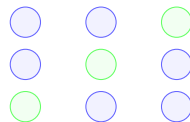


Dann gibt es nach dem obigen Satz genau

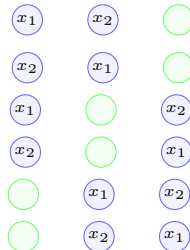
$$\frac{n!}{n!} = 1$$

Anordnungsmöglichkeiten, was offenbar richtig ist.

Beweis von Satz 134 anhand eines Beispiels. Wir schauen uns $n = 3$ Kugeln an, die in $r = 2$ Klassen der Größen $n_1 = 2$ und $n_2 = 1$ aufgeteilt sind. Nun zählen wir alle Anordnungen dieser:



Wir stellen also fest, dass es genau $3 = \frac{6}{2} = \frac{3!}{2!1!}$ verschiedene Anordnungen gibt. Versieht man die beiden Kugeln $\circ \circ$ in der ersten Klasse mit den Bezeichnungen x_1 und x_2 und macht sie damit unterscheidbar, so gibt es nach Satz 132 genau $3! = 6$ verschiedene Anordnungen:

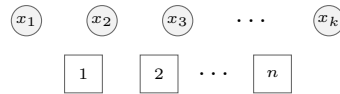


Für die Anordnungen der Kugeln x_1 und x_2 gibt es (nach Satz 132) genau $2! = 2$ verschiedene Anordnungen. Für die Kugel(n) in der zweiten Klasse, gibt es nur $1! = 1$ verschiedene Anordnungen. Diese Permutationen innerhalb der jeweiligen Klassen sollen identifiziert werden. Damit bekommen wir insgesamt

$$\frac{n!}{n_1!n_2!} = \frac{3!}{2!1!} = \frac{6}{2} = 3$$

Anordnungsmöglichkeiten. □

Nun fragen wir uns, auf wieviele Weisen man k -viele Kugeln (also nicht $k = n$ wie oben) auf n -viele „Urnen“ (hier ein anderes Wort für „Boxen“ oder „Schubladen“) verteilen kann. Ein geeignetes Modell zur Beantwortung dieser Frage nennt man **Urnenmodell**.



Dabei unterscheiden wir **Mehrfachbelegung** der Urnen und keine Mehrfachbelegung, sowie **Unterscheidbarkeit** der Kugeln und keine Unterscheidbarkeit. Bevor wir diese Situation untersuchen, führen wir eine Abkürzung ein.

Definition 136. Sind n und k zwei natürliche Zahlen mit $n \geq k \geq 0$, so nennt man die natürliche Zahl

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

den **Binomialkoeffizient** von n und k . (Man setzt $0! := 1$ und $\binom{n}{k} := 0$ für $n < k$.)

Hier ist erst einmal eine wichtige Tabelle, deren Einträge die Anzahl der Möglichkeiten sind, k Kugeln auf n Urnen zu verteilen. Wir werden später Genaueres zu diesen Zahlen sagen. Seien $n \geq 1$ und $k \geq 1$ zwei natürliche Zahlen. Für die rechte innere Spalte der folgenden Tabelle 14 nehmen wir $n \geq k$ an. (Die grünen Bezeichnungen beziehen sich auf ein anderes Modell und werden später behandelt.)

k Kugeln auf n Urnen	mehrfach	nicht mehrfach	
unterscheidbar	n^k	$\frac{n!}{(n-k)!}$	mit Reihenfolge
nicht unterscheidbar	$\binom{n+k-1}{k}$	$\binom{n}{k}$	ohne Reihenfolge
	Zurücklegen	kein Zurücklegen	k Züge aus n vielen Kugeln

Abbildung 14: Urnenmodell

Wir erklären kurz die Zahlen in der obigen Tabelle 14.

- **Unterscheidbar und mehrfach.** Für jede der k -vielen Kugeln hat man genau n -vielen Urnen zur Auswahl und daher gibt es hier

$$n \cdot n \cdot \dots (k\text{-mal}) \cdot \dots \cdot n = n^k$$

viele Möglichkeiten die k unterscheidbaren Kugeln auf die n Urnen zu verteilen.

- **Unterscheidbar und nicht mehrfach.** Für die erste Kugel, habe ich n -viele Möglichkeiten, für die zweite $(n-1)$, und so weiter (siehe den Beweis von Satz 132). Insgesamt ergeben sich

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

viele Möglichkeiten.

- **Nicht unterscheidbar und nicht mehrfach.** Im zugehörigen „unterscheidbaren Fall“ gibt es $\frac{n!}{(n-k)!}$ -viele Möglichkeiten und wir identifizieren, genau wie im Beweis von Satz 134, alle $k!$ -vielen Anordnungen in den Urnen zu einer (da die Kugeln nicht unterscheidbar sein sollen) und erhalten also

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

viele Möglichkeiten.

- **Nicht unterscheidbar und mehrfach.** Diese Variante ist am schwierigsten abzuzählen und wir benutzen den folgenden kleinen Trick. Eine Verteilung der k -vielen Kugeln auf die n -vielen Urnen (mit zugelassener Mehrfachbelegung) interpretieren wir als eine Liste



aus k -vielen Kugeln und $(n-1)$ -vielen Strichen, die wir als „Trennwände zwischen den Urnen“ interpretieren. Eine solche Liste hat also $k + (n-1)$ -viele Einträge und repräsentiert genau eine Verteilung von k -vielen (noch unterscheidbaren) Kugeln auf n -vielen Urnen (mit zugelassener Mehrfachbelegung). Für solche Listen gibt es genau $(k+n-1)! \cdot \frac{1}{(n-1)!}$ -viele Möglichkeiten, da wir die $(n-1)!$ Anordnungen der $(n-1)$ ununterscheidbaren Striche miteinander identifizieren. Ebenso wollen wir die $k!$ -vielen Anordnungen der Kugeln identifizieren, um sie „ununterscheidbar zu machen“, wie im Beweis von Satz 134. Insgesamt erhalten wir

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

viele Möglichkeiten.

Bemerkung 137. Wenn wir uns das Abzählen im obigen letzten Punkt „nicht unterscheidbar und mehrfach“ noch einmal ansehen, könnte man auf die Idee kommen, dass die Zahl $(k+n-1)! \cdot \frac{1}{(n-1)!}$, also die Anzahl der Listen



genau die Anzahl der Möglichkeiten ist, k -viele unterscheidbare Kugeln auf n -vielen Urnen (mit zugelassener Mehrfachbelegung) zu verteilen. Es ist aber

$$n^k \neq (k+n-1)! \cdot \frac{1}{(n-1)!},$$

wo liegt also der Fehler? Der Fehler ist, dass wir bei der obigen Listenabzählung „zu viele“ Listen zählen. Es entsprechen beispielsweise die beiden Listen



der gleichen Weise k -viele unterscheidbare Kugeln mit zugelassener Mehrfachbelegung auf n Urnen zu verteilen. (Mit anderen Worten: **Permutationen der unterscheidbaren Kugeln innerhalb der Urnen werden als eine Möglichkeit gezählt.**)

Bemerkung 138. Überraschenderweise kann das Urnenmodell von oben auch noch auf eine andere Weise interpretiert werden, nämlich als **Zugmodell**. Dabei fragen wir uns, auf wieviele Weisen man k -mal aus n -vielen Kugeln (also nicht k -vielen Kugeln wie beim Urnenmodell) ziehen kann, wobei man unterscheidet zwischen Ziehen **mit Zurücklegen** und ohne Zurücklegen, sowie Betrachtung der **Reihenfolge** und ohne Betrachtung der Reihenfolge. Der Trick zum Vergleich dieser Interpretationen ist, dass man die Urnen im Urnenmodell als die zu ziehenden Kugeln interpretiert und die Kugeln im Urnenmodell als die „Züge“. **Man verteilt also sozusagen die Züge auf die Kugeln.** Sind die Kugeln im Urnenmodell beispielsweise unterscheidbar, so kann man sie mit $1, \dots, k$ nummerieren, was in dem Zugmodell gerade der „Nummer des Zuges“ entspricht. Eine Unterscheidbarkeit der Kugeln im Urnenmodell entspricht also gerade einer Berücksichtigung der Zugreihenfolge im Zugmodell.



Die vorherige Bemerkung 138 ist anfänglich sicherlich nicht leicht zu verstehen. Es ist dringend zu empfehlen, in Ruhe über den Zusammenhang zwischen dem Urnenmodell und dem Zugmodell nachzudenken.

Beispiel 139. Wieviele Autokennzeichen sind in Regensburg möglich, wenn jedes aus genau 2 von 26 Buchstaben und 3 von 10 Ziffern besteht? Für das „Ziehen von $k = 2$ Buchstaben“ aus $n = 26$ Buchstaben mit Reihenfolge und Zurücklegen gibt es 26^2 verschiedene Möglichkeiten. Insgesamt gibt es also

$$26^2 \cdot 10^3 = 676 \cdot 10000 = 67600$$

verschiedene Möglichkeiten für Autokennzeichen in Regensburg.

Beispiel 140. Wieviele k -elementige Teilmengen einer Menge M mit n Elementen gibt es? Fassen wir die Elemente von M als Kugeln auf, so kann man die Auswahl einer k -elementigen Teilmenge als „Ziehen von k -vielen Kugeln“ ohne Reihenfolge und ohne Zurücklegen auffassen. Dafür gibt es $\binom{n}{k}$ -viele Möglichkeiten.

Beispiel 141. Wieviele verschiedene Tipps gibt es beim Lotto „6 aus 49“? Ein Tipp kann man als die Auswahl einer 6-elementigen Teilmenge aus der Menge $M = \{1, 2, \dots, 49\}$ auffassen. Dafür gibt es nach dem vorherigen Beispiel 140 also

$$\binom{49}{6} = 13983816$$

Möglichkeiten.

Beispiel 142. Wieviele Möglichkeiten gibt es, 2 Gummibärchen aus einer Tüte mit Gummibärchen in 3 verschiedenen Farben auszuwählen? Eine solche Möglichkeit interpretieren wir als eine Verteilung von $k = 2$ Kugeln auf $n = 3$ Urnen mit Mehrfachbelegung und nicht unterscheidbaren Kugeln. Dafür gibt es nach der Tabelle

$$\binom{3+2-1}{2} = \binom{4}{2} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 1 \cdot 2} = 1 \cdot 2 \cdot 3 = 6$$

verschiedene Möglichkeiten. Dieses entspricht übrigens genau der Anzahl der Elemente der Ergebnismenge Ω aus Beispiel 120.(7).

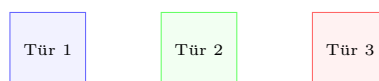
Bemerkung 143. Man kann übrigens bis auf in dem Fall „Mehrfach und nicht unterscheidbar“ (wie das Beispiel 142 zeigt) das Verteilen von k -vielen Kugeln auf n -viele Urnen (bzw. das k -malige Ziehen aus n -vielen Kugeln) als Laplace Experiment modellieren. So wäre beispielsweise die Wahrscheinlichkeit, den Hauptgewinn beim Lotto zu erzielen nach Beispiel 120.(7) durch

$$P(\{\text{Mein Tipp } \omega\}) = \frac{1}{|\Omega|} = \frac{1}{\binom{49}{6}} = \frac{1}{13983816}$$

gegeben. Für diese Modellbildung wird davon ausgegangen, dass alle Tipps „gleichwahrscheinlich“ sind, was in der Realität hoffentlich auch zutreffend ist.

2.3. Bedingte Wahrscheinlichkeiten. In diesem Kapitel reden wir über bedingte Wahrscheinlichkeiten. Hat man für ein bestimmtes Zufallsexperiment ein Modell als Wahrscheinlichkeitsraum gewählt, so kann man in diesem Modell „rechnen“ und anschließend die Ergebnisse wieder als Aussagen über das Zufallsexperiment interpretieren. Überraschenderweise führt dieses Vorgehen schon in einfachen Fällen zu Aussagen über das Zufallsexperiment, die beim ersten Hinsehen wenig intuitiv und somit sehr überraschend sind. Ein klassisches Beispiel ist das folgende.

Beispiel 144 (Ziegenproblem). Bei einer Spielshow im Fernsehen haben Sie die Wahl zwischen drei Türen



wobei hinter einer Tür der Hauptgewinn und hinter den beiden anderen jeweils eine Ziege steht, was einer „Niete“ entsprechen soll. Sie wählen eine Tür, beispielsweise die erste Tür [\[1\]](#), und der Moderator, der weiss, was hinter den Türen steht, öffnet eine andere Tür mit einer Ziege, beispielsweise die dritte Tür [\[3\]](#). Nun stellt der Moderator Sie vor die Wahl, bei der ersten Tür [\[1\]](#) zu bleiben, oder auf die zweite Tür [\[2\]](#) zu wechseln. (Anschließend wird die von Ihnen gewählte Tür geöffnet und Sie erhalten den Preis, der sich dahinter verbirgt.)

Sollten Sie bei der ersten Tür bleiben oder wechseln, wenn Sie den Hauptgewinn bekommen möchten?

Um diese Frage zufriedenstellend zu beantworten, brauchen wir einen neuen Begriff, der die Wahrscheinlichkeit eines Ereignisses A unter der Bedingung, dass ein anderes Ereignis B bereits eingetreten ist, beschreibt.

Definition 145. Sei (Ω, P) ein Wahrscheinlichkeitsraum. Seien A und B Ereignisse mit $P(B) > 0$. Die reelle Zahl

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

heißt die **bedingte Wahrscheinlichkeit von A unter der Bedingung B** .

Beispiel 146. Wir betrachten wieder den einfachen Würfelwurf aus Beispiel 120.(3) mit den Ereignissen

$$A = \{1, 3, 5\} \quad \text{und} \quad B = \{1, 2\}.$$

Was ist die Wahrscheinlichkeit von A (also einer ungeraden Augenzahl) unter der Bedingung B (also Augenzahl ≤ 2)?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{1\})}{P(\{1, 2\})} = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{6} \cdot \frac{6}{2} = \frac{1}{2}.$$

Bemerkung 147. Hier sind zwei Bemerkungen zu der Definition 145 der bedingten Wahrscheinlichkeit.

- Der Nenner $P(B)$ in der Definition von $P(A|B)$ sagt dafür, dass $P(B|B) = 1$.
- Man kann sich leicht überlegen, dass Ω mit der Abbildung

$$P(-|B): \mathbb{P}(\Omega) \rightarrow [0, 1]$$

wieder ein Wahrscheinlichkeitsraum ist. Insbesondere gelten dann also die Aussagen aus Satz 127 wie beispielsweise

$$P(A \cup A'|B) = P(A|B) + P(A'|B)$$

für Ereignisse A und A' mit $A \cap A' = \emptyset$, sowie

$$P(A|B) + P(\bar{A}|B) = 1.$$

In der Zugmodell-Interpretation aus Bemerkung 138 der Zahlen in der Tabelle 14 haben wir unter anderem die Möglichkeiten gezählt, k -viele Kugeln aus einem Sack mit n -vielen Kugeln mit Beachtung der Reihenfolge zu ziehen (mit oder ohne Zurücklegen). Hierbei hatten alle n -viele Kugeln im Sack eine eindeutige „Nummer“, die sie identifizierte. Wie in Bemerkung 143 erwähnt, können wir daraus einen (Laplace) Wahrscheinlichkeitsraum konstruieren.

Was passiert nun aber, wenn die Kugeln im Sack nicht alle eine eindeutige „Nummer“ tragen, sondern zu bestimmten Klassen gehören? Einen Eindruck einer solchen Situation haben wir schon in Beispiel 142 bekommen. Hier wollten wir $(k = 2)$ -mal (mit Zurücklegen) Ziehen aus einem Sack mit $(n = 3)$ -vielen Gummibärchenklassen. Nun wollen wir aber die Gesamtzahl der Gummibärchen mit einbeziehen: Was passiert beispielsweise, wenn nur fünf rote, drei grüne und keine blauen Gummibärchen in dem Sack sind? Unser Ziel ist es, einen Wahrscheinlichkeitsraum (Ω, P) zu definieren, der eine solche Situation modelliert.

Dafür könnte man, wie in Beispiel 129 die Wahrscheinlichkeit aller Elementarereignisse einzeln geeignet angeben. Dieses ist aber nicht so einfach: Was wäre beispielsweise eine geeignete Wahrscheinlichkeit für das Elementarereignis $\{(R, R)\}$ im Gummibärchen-Beispiel, wobei R für „Rot“ steht? Der Trick ist nun, dass man zwar $P(\{(R, R)\})$ nicht direkt angeben kann, wohl aber die Wahrscheinlichkeit von $\{(R, R)\}$ **unter der Bedingung**, dass wir schon den ersten Zug gemacht haben, also $(R, ?)$ bereits gezogen haben. Diese Wahrscheinlichkeit wäre offenbar $\frac{4}{7}$, da man aus 7 Gummibärchen (es wurde ja bereits eines der $5 + 3 = 8$ Gummibärchen gezogen) eines der 4 roten Gummibärchen (es wurde ja bereits eines der 5 roten Gummibärchen gezogen) ziehen möchte.

Satz 148 (Multiplikationssatz). Es sei (Ω, P) ein Wahrscheinlichkeitsraum und A, B Ereignisse mit $P(B) > 0$. Dann gilt

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

Sind allgemeiner A_1, \dots, A_k Ereignisse mit $P(A_1 \cap \dots \cap A_{k-1}) > 0$, so gilt

$$P(A_1 \cap \dots \cap A_k) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap \dots \cap A_{k-1}).$$

Beispiel 149. Betrachten wir wieder das oben angedeutete Gummibären-Beispiel (wobei wir die 0-vielen blauen Gummibärchen ignorieren). Es befinden sich also 5 rote („R“) und 3 grüne („G“) Gummibärchen in einem Sack und wir ziehen ($k = 2$)-mal (ohne Zurücklegen und mit Reihenfolge) jeweils ein Gummibärchen. Wir setzen

$$\Omega := \{(R, R), (R, G), (G, R), (G, G)\}$$

als Ergebnismenge. Nun betrachten wir die folgenden Ereignisse.

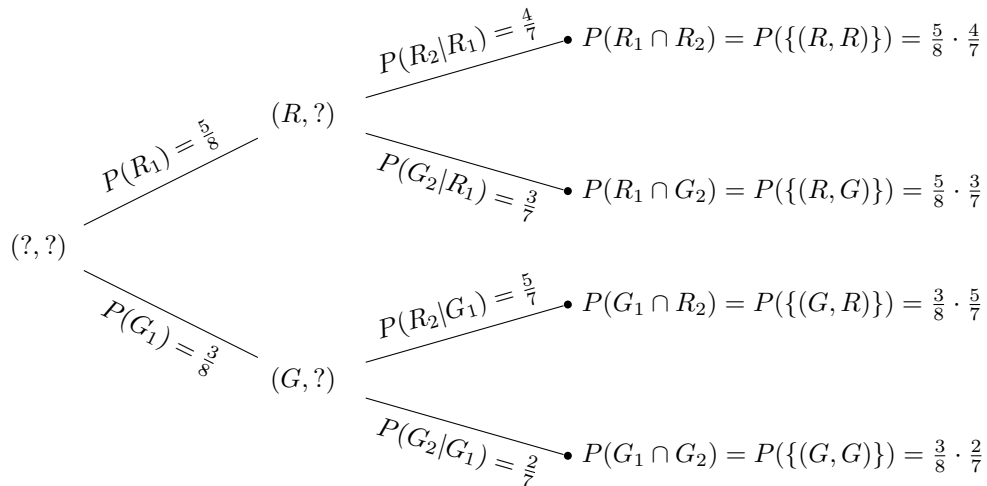
$$\begin{aligned} R_1 &:= \{(R, R), (R, G)\} && \text{„erster Zug ist rot“} \\ G_1 &:= \{(G, R), (G, G)\} && \text{„erster Zug ist grün“} \\ R_2 &:= \{(R, R), (G, R)\} && \text{„zweiter Zug ist rot“} \\ G_2 &:= \{(R, G), (G, G)\} && \text{„zweiter Zug ist grün“} \end{aligned}$$

Nach dem Multiplikationssatz 148 ist also

$$P(\{(R, R)\}) = P(R_1 \cap R_2) = P(R_2|R_1) \cdot P(R_1) = \frac{4}{7} \cdot \frac{5}{8} = \frac{5}{14}.$$

Nun sind wir bei dieser Berechnung **bereits davon ausgegangen**, dass der Wahrscheinlichkeitsraum (Ω, P) (und damit insbesondere die reelle Zahl $P(\{(R, R)\})$) gegeben ist, denn den Multiplikationssatz 148 kann man ja nur dann benutzen, wenn man schon einen Wahrscheinlichkeitsraum „hereinsteckt“.

Wir möchten unten einen „umgekehrten Multiplikationssatz“ formulieren, der genau diese Konstruktion eines Wahrscheinlichkeitsraums liefert. Um dieses präzise zu tun, ist viel Formalismus nötig. Daher werden wir uns mit einem „informellen Satz“ begnügen. Um diesen Satz zu formulieren, schauen wir uns den Begriff des **Baumdiagramms** an. Für unser Gummibärchen-Beispiel sieht dieser wie folgt aus.



Die drei Spalten des obigen Baumdiagramms heißen auch dessen **Stufen**. Die Knoten auf der rechten Seite heißen **Endknoten** und die übrigen Knoten **innere Knoten**. Wir „lesen“ diesen Baum also „von links nach rechts“. Der Multiplikationssatz 148 sagt in dieser graphischen Darstellung nun, dass man alle Wahrscheinlichkeiten auf dem Weg zwischen dem Anfangsknoten links und einem Endknoten (rechts) multiplizieren muss, um auf die Wahrscheinlichkeit am jeweiligen Endknoten zu kommen. Wir stellen noch fest, dass in dem obigen Baumdiagramm für jeden inneren Knoten die Summe der **ausgehenden Wahrscheinlichkeiten**, also der Zahlen auf den Kanten, die von diesem Knoten nach rechts abgehen, genau Eins ist.

Satz 150 (Umgekehrter Multiplikationssatz – informelle Formulierung). Sei eine endliche Ergebnismenge $\Omega = \Omega_1 \times \dots \times \Omega_k$ gegeben und setze für $\omega \in \Omega_i$

$$A_{\omega,i} := \Omega_1 \times \dots \times \underbrace{\{\omega\}}_{i\text{-te Stelle}} \times \dots \times \Omega_k$$

und sei ein k -stufiger Wahrscheinlichkeitsbaum wie oben gegeben, sodass für jeden inneren Knoten die Summe der ausgehenden Wahrscheinlichkeiten gleich Eins ist und sich dort die Wahrscheinlichkeit disjunkter Ereignisse addiert, so kann man einen eindeutigen Wahrscheinlichkeitsraum (Ω, P) definieren, indem man

$$P(\{(\omega_1, \dots, \omega_k)\}) := P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap \dots \cap A_{k-1}).$$

setzt, wobei $A_i := A_{\omega_i,i}$.



Bei dem obigen umgekehrten Multiplikationssatz 150 kommt es nicht darauf an, dass Sie die (ohnehin unpräzise) Formulierung genau verstehen. Wichtig ist nur zu wissen, dass man auf diese Weise in einer konkreten Situation einen Wahrscheinlichkeitsraum angeben kann, wie wir es in dem obigen Beispiel 146 gesehen haben.

Um das Ziegenproblem aus Beispiel 144 zu untersuchen, brauchen wir noch zwei Aussagen über bedingte Wahrscheinlichkeiten.

Satz 151 (Satz von der totalen Wahrscheinlichkeit). Sei (Ω, P) ein Wahrscheinlichkeitsraum und B_1, \dots, B_m Ereignisse, die eine disjunkte Zerlegung von Ω sind (d.h. sie sind paarweise disjunkt und haben Ω als Vereinigung) mit $P(B_j) > 0$ für alle $j \in \{1, \dots, m\}$. Dann gilt für jedes Ereignis A , dass

$$P(A) = \sum_{j=1}^m P(B_j) \cdot P(A|B_j).$$

Beispiel 152. Wir betrachten wieder das Gummibärchen-Beispiel 149. Hier sind R_1 (erster Zug rot) und G_1 (erster Zug grün) eine disjunkte Zerlegung von Ω und es gilt nach dem vorherigen Satz 151 für die Wahrscheinlichkeit von R_2 (zweiter Zug rot), dass

$$P(R_2) = P(R_1) \cdot P(R_2|R_1) + P(G_1) \cdot P(R_2|G_1) = \left(\frac{5}{8} \cdot \frac{4}{7}\right) + \left(\frac{3}{8} \cdot \frac{5}{7}\right) = 0,625.$$

Aus dem Satz von der totalen Wahrscheinlichkeit 151 ergibt sich die folgende wichtige Aussage, die es einem in einer gewissen Weise erlaubt „Ursache und Wirkung umzudrehen“, wie wir später in Beispiel 155 genauer diskutieren werden.

Satz 153 (Satz von Bayes). Sei (Ω, P) ein Wahrscheinlichkeitsraum und B_1, \dots, B_m Ereignisse, die eine disjunkte Zerlegung von Ω sind mit $P(B_j) > 0$ für alle $j \in \{1, \dots, m\}$. Dann gilt für alle $i \in \{1, \dots, m\}$ und jedes Ereignis A , dass

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{j=1}^m P(B_j) \cdot P(A|B_j)}.$$

Beweis.

$$\begin{aligned} P(B_i|A) &= \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i \cap A)}{P(A)} \cdot \frac{P(B_i)}{P(B_i)} = \frac{P(A \cap B_i)}{P(B_i)} \cdot \frac{P(B_i)}{P(A)} = \\ &= P(A|B_i) \cdot \frac{P(B_i)}{\sum_{j=1}^m P(B_j) \cdot P(A|B_j)} = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{j=1}^m P(B_j) \cdot P(A|B_j)}. \quad \square \end{aligned}$$

Beispiel 154. Wir betrachten wieder das Ziegenproblem-Beispiel 144. Unser „Zufallsexperiment“ ist in der Situation, dass wir (ohne Beschränkung der Allgemeinheit – sonst nennen wir die Türen einfach um) die erste Tür 1 gewählt haben und nun der Moderator eine der (anderen) Türen 2 oder 3 öffnet und uns erneut vor die Wahl stellt. Wir betrachten als Ergebnismenge

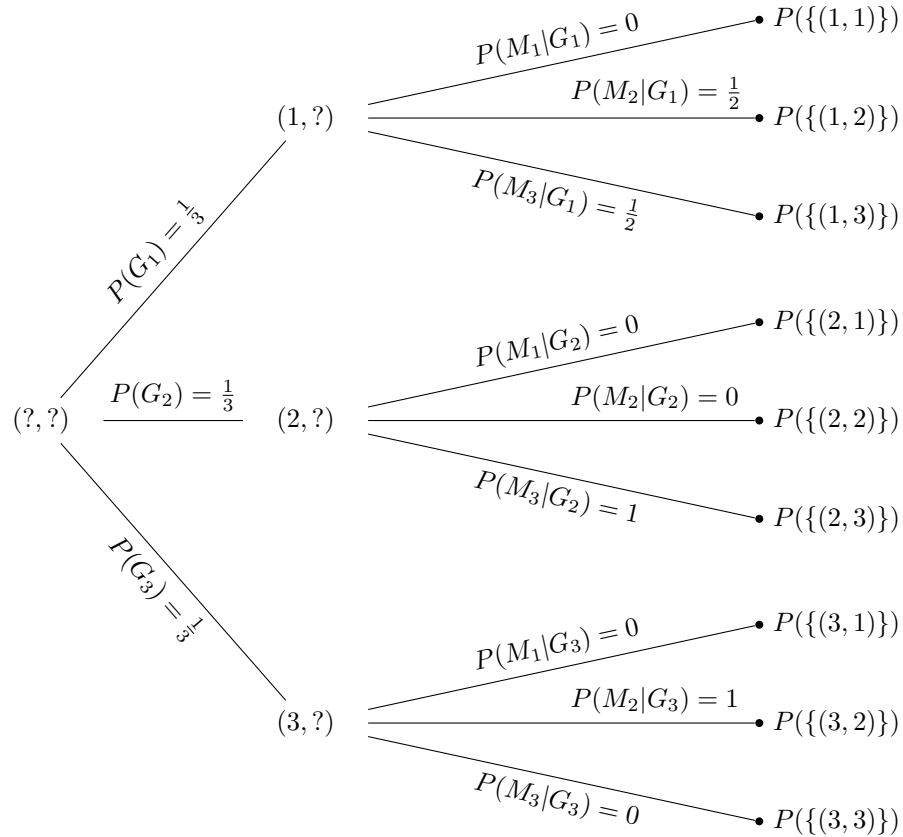
$$\Omega := \{1, 2, 3\} \times \{1, 2, 3\},$$

wobei der erste Eintrag die Tür mit dem Hauptgewinn und der zweite Eintrag die Tür, die der Moderator öffnet, angibt. Für $j \in \{1, 2, 3\}$ definieren wir die folgenden Ereignisse.

$$G_j := \{(j, 1), (j, 2), (j, 3)\} \quad \text{„Der Hauptgewinn ist hinter Tür } j\text{“}$$

$$M_j := \{(1, j), (2, j), (3, j)\} \quad \text{„Der Moderator öffnet Tür } j\text{“}$$

Wir definieren uns einen Wahrscheinlichkeitsraum (Ω, P) mit Hilfe des umgekehrten Multiplikationssatzes 150 angewandt auf den folgenden Wahrscheinlichkeitsbaum.



Nun geht wir (wieder ohne Beschränkung der Allgemeinheit) davon aus, dass der Moderator die dritte Tür 3 öffnet und uns vor die Wahl zwischen der ersten 1 und der zweiten Tür 2 stellt. Wir berechnen mit dem Satz von Bayes 153

$$P(G_1|M_3) = \frac{P(G_1) \cdot P(M_3|G_1)}{P(G_1)P(M_3|G_1) + P(G_2)P(M_3|G_2) + P(G_3)P(M_3|G_3)} = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0} = \frac{1}{3}$$

und

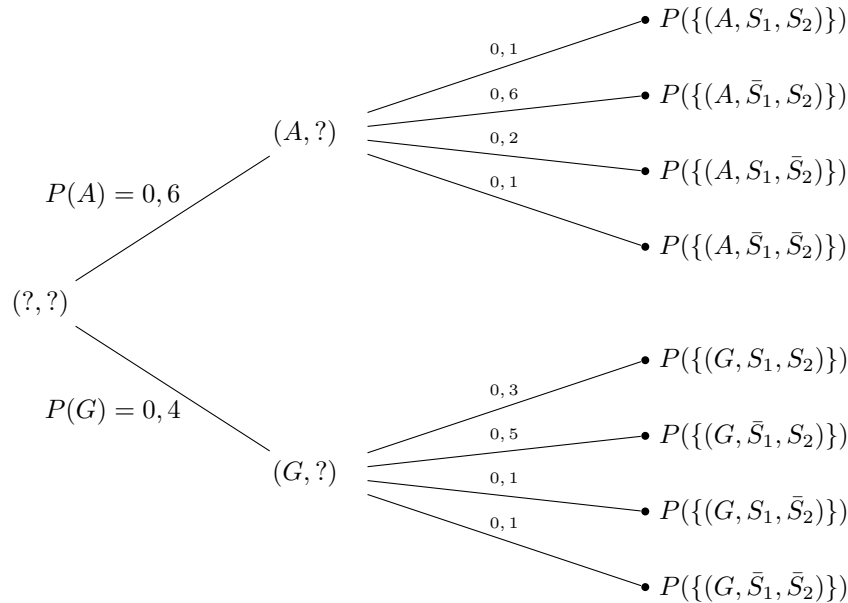
$$P(G_2|M_3) = \frac{P(G_2) \cdot P(M_3|G_2)}{P(G_1)P(M_3|G_1) + P(G_2)P(M_3|G_2) + P(G_3)P(M_3|G_3)} = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0} = \frac{2}{3}$$

und bekommen das Ergebnis, dass es wahrscheinlicher ist, dass der Hauptgewinn hinter der zweiten Tür 2 ist. Wir sollten also die Tür wechseln.

Obwohl dieses im vorherigen Beispiel schon zu sehen war, möchten wir noch einmal darauf eingehen, wie der Satz von Bayes „Ursache und Wirkung umdreht“. Dafür schauen wir uns das folgende Beispiel an.

Beispiel 155. In der Notfallmedizin werden Schlaganfälle mit Blutverdünnern und Gehirnblutungen mit einem genau gegenteilig wirkenden Medikament behandelt. Daher möchte man anhand von Symptomen (in diesem Beispiel betrachten wir nur zwei Symptome S_1 und S_2) die man bei einer Patientin oder einem Patienten beobachtet, schnell entscheiden, ob es sich um einen Schlaganfall A oder eine Gehirnblutung G handelt. Zuvor untersucht man 100 Patient*innen von denen man bereits weiß, ob sie unter A oder G leiden, auf ihre Symptome S_1 und S_2 . Man kann also die bedingten Wahrscheinlichkeiten $P((S_1, S_2)|A)$, $P((\bar{S}_1, \bar{S}_2)|G)$,

und so weiter, angeben. Die Untersuchung der 100 Patient*innen führt zu dem folgenden Wahrscheinlichkeitsbaum.



Nun sieht ein*e Notfallmediziner*in Symptom S_1 , aber nicht das zweite Symptom, also \bar{S}_2 . Sie berechnet (wobei wir etwas ungenau die präzise Definition der auftretenden Ereignismengen unterschlagen)

$$P(A|(S_1, \bar{S}_2)) = \frac{P(A)P((S_1, \bar{S}_2)|A)}{P(A)P((S_1, \bar{S}_2)|A) + P(G)P((S_1, \bar{S}_2)|G)} = \frac{0,12}{0,12 + 0,04} = 0,75$$

und

$$P(G|(S_1, \bar{S}_2)) = \frac{P(G)P((S_1, \bar{S}_2)|G)}{P(A)P((S_1, \bar{S}_2)|A) + P(G)P((S_1, \bar{S}_2)|G)} = \frac{0,04}{0,16} = 0,25$$

und schließt, dass es sich vermutlich um einen Schlaganfall handelt.

2.4. Unabhängige Ereignisse. In diesem Kapitel definieren wir den Begriff der „stochastischen Unabhängigkeit“. Dieser wird im letzten Kapitel 3 mit der empirischen Unabhängigkeit in Verbindung gebracht.

Definition 156. Sei (Ω, P) ein Wahrscheinlichkeitsraum. Zwei Ereignisse A und B mit $P(A) > 0$ und $P(B) > 0$ heißen **stochastisch unabhängig**, falls

$$P(A|B) = P(A) \quad \text{und} \quad P(B|A) = P(B).$$

Bemerkung 157. Aus dem Multiplikationssatz 148 bekommen wir, dass A und B genau dann stochastisch unabhängig sind, wenn gilt

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

Beispiel 158.

- (4) Wir betrachten das Laplace Experiment **zweifacher Würfelwurf** aus Beispiel 124.(4) und die Ereignisse

$$\begin{aligned} A &:= \{(2, 1), (2, 2), \dots, (2, 6)\} && \text{„Erster Wurf ist „2““} \\ B &:= \{(1, 1), (1, 6), (2, 1), (2, 6), \dots\} && \text{„Zweiter Wurf ist „1“ oder „6““} \end{aligned}$$

Dann gilt

$$P(A \cap B) = P(\{(2, 1), (2, 6)\}) = \frac{2}{36} = \frac{1}{18} = \frac{1}{6} \cdot \frac{1}{3} = P(A) \cdot P(B)$$

und A und B sind stochastisch unabhängig. Wir stellen fest, dass es auch in dem Zufallsexperiment „zweifacher Würfelwurf“ keinen kausalen Zusammenhang zwischen diesen beiden Ereignissen zu geben scheint.

- (8) In einem Sack sind $n = 4$ Kugeln (rot, grün, blau, pink) und wir ziehen ($k = 2$)-mal ohne Zurücklegen und mit Reieigenfolge. Wir modellieren dieses Zufallsexperiment als Laplace Experiment mit $|\Omega| = \frac{4!}{2!} = 12$. Wir betrachten die (unpräzise definierten) Ereignisse $A :=$ „Erster Zug ist eine rote Kugel“ und $B :=$ „Zweiter Zug ist eine blaue Kugel“. Es gilt $|A| = 3$ und $|B| = 3$ und

$$P(A \cap B) = \frac{1}{12} \neq \frac{1}{16} = \frac{3}{12} \cdot \frac{3}{12} = P(A) \cdot P(B)$$

und A und B sind nicht stochastisch unabhängig. Wir stellen fest, dass es auch in dem zugehörigen Zufallsexperiment einen kausalen Zusammenhang zwischen diesen beiden Ereignissen gibt.

- (4) Wir betrachten noch einmal den **zweifacher Würfelwurf** mit den beiden Ereignissen $C :=$ „Erster Wurf ist gerade Zahl“ und $D :=$ „Summe beider Würfe ist gerade Zahl“, so gilt $P(C \cap D) = \frac{1}{4}$, sowie $P(C) = \frac{1}{2}$, $P(D) = \frac{1}{2}$ und C und D sind stochastisch unabhängig. In dem zugehörigen Zufallsexperiment besteht aber trotzdem ein kausaler Zusammenhang zwischen diesen beiden Ereignissen, da der erste Wurf die Summe der Augenzahlen „mitbestimmt“.



Das letzte Beispiel 158.(4) soll (wieder einmal) darauf hinweisen, dass die „zufällig Erfüllung“ unserer mathematischen Bedingungen für eine Unabhängigkeit zweier Ereignisse keine kausale Unabhängigkeit zwischen diesen „in der Realität“ impliziert.

2.5. Zufallsvariablen. Mit der Hilfe von Zufallsvariablen lassen sich Ereignisse in einem Wahrscheinlichkeitsraum einfacher angeben und untersuchen.

Definition 159. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum. Eine **Zufallsvariable** ist eine Abbildung $X: \Omega \rightarrow \mathbb{R}$. Das Bild $X(\Omega)$ von X wird oft (etwas ungenau) mit W bezeichnet.

Für eine Zufallsvariable X und $x \in \mathbb{R}$ definiert man

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

und $P(X \neq x) := P(\{\omega \in \Omega \mid X(\omega) \neq x\})$. Analog definiert man $P(X \leq x)$, $P(X < x)$, $P(X \geq x)$, $P(X > x)$ und so weiter.

Bemerkung 160. Ist X eine Zufallsvariable und $x \in \mathbb{R}$ nicht im Bild von X , so gilt also $P(X = x) = P(\emptyset) = 0$.

Außerdem gilt

$$\begin{aligned} P(X \leq x) &= P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \\ &= P(\{\omega \in \Omega \mid X(\omega) = x\} \cup \{\omega \in \Omega \mid X(\omega) < x\}) \\ &= P(X = x) + P(X < x), \end{aligned}$$

(und analog auch $P(X \geq x) = P(X = x) + P(X > x)$) wobei die letzte Gleichung gilt, da die beteiligten Mengen disjunkt sind.

Weiter ist

$$\begin{aligned} P(X > x) &= P(\{\omega \in \Omega \mid X(\omega) > x\}) \\ &= P(\Omega \setminus \{\omega \in \Omega \mid X(\omega) \leq x\}) \\ &= P(\Omega) - P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \quad (\text{siehe Satz 127.(3)}) \\ &= 1 - P(X \leq x) \end{aligned}$$

Beispiel 161.

- (1) Wir betrachten das Laplace Experiment **einfacher Münzwurf** aus Beispiel 124.(1). Es wird auf „Zahl“ gewettet und 1 Euro ausgezahlt, wenn „Zahl“ fällt, sonst 0 Euro. Es ist $\Omega = \{K, Z\}$. Wir geben eine Zufallsvariable

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

durch $X(K) := 0$ und $X(Z) := 1$ an, die den Gewinn (in „Euro“) beschreiben soll. Das Bild von X ist die Menge $W = \{0, 1\}$. Es ist

$$P(X = 1) = P(\{\omega \in \Omega \mid X(\omega) = 1\}) = P(\{Z\}) = \frac{1}{2}$$

und $P(X = 0) = \frac{1}{2}$. Man könnte für ein $x \in \mathbb{R}$ die Zahl $P(X = x)$ als die Wahrscheinlichkeit dafür interpretieren, dass bei dem Zufallsexperiment x Euro ausgezahlt werden.

- (4) Wir betrachten das Laplace Experiment **zweifacher Würfelwurf** aus Beispiel 124.(4). Es ist $\Omega := \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$. Wir möchten eine Zufallsvariable „Summe der Augenzahlen vom ersten und zweiten Wurf“ angeben und definieren

$$\begin{aligned} X: \Omega &\rightarrow \mathbb{R} \\ (i, j) &\mapsto i + j. \end{aligned}$$

Das Bild von X ist die Menge $W = \{2, 3, \dots, 11, 12\}$. Es ist beispielsweise

$$\begin{aligned} P(X = 2) &= P(\{(i, j) \in \Omega \mid X(i, j) = i + j = 2\}) = P(\{(1, 1)\}) = \frac{1}{36} \quad \text{und} \\ P(X = 3) &= P(\{(1, 2), (2, 1)\}) = \frac{2}{36} = \frac{1}{18} \quad \text{und} \\ P(X \leq 4) &= P(\{(1, 1), (1, 2), (2, 1), (1, 3), (3, 1), (2, 2)\}) = \frac{6}{36} = \frac{1}{6}. \end{aligned}$$

Analog zur empirischen Verteilungsfunktion aus Definition 54 definieren wir die Verteilungsfunktion einer Zufallsvariable.

Definition 162. Sei (Ω, P) ein Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Die **Verteilungsfunktion** von X ist die Abbildung

$$\begin{aligned} F: \mathbb{R} &\rightarrow [0, 1] \subseteq \mathbb{R} \\ t &\mapsto P(X \leq t). \end{aligned}$$

Bemerkung 163. Sei $X: \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion F . Diese Funktion ist ein Beispiel einer monoton wachsenden Funktion, die nur Werte zwischen 0 und 1 hat. Es gibt $F(t)$ die Wahrscheinlichkeit dafür an, dass X „höchstens den Wert t “ annimmt.

Beispiel 164.

- (4) Wir betrachten das Laplace Experiment **zweifacher Würfelwurf** und die Zufallsvariable X = „Summe der Augenzahlen“ aus Beispiel 161.(4). Dann ist die zugehörige Verteilungsfunktion die Abbildung

$$F: \mathbb{R} \rightarrow [0, 1] \subseteq \mathbb{R}$$

$$t \mapsto F(t) = \begin{cases} 0 & \text{falls } t < 2 \\ \frac{1}{36} & \text{falls } 2 \leq t < 3 \\ \frac{3}{36} & \text{falls } 3 \leq t < 4 \\ \frac{6}{36} & \text{falls } 4 \leq t < 5 \\ \frac{10}{36} & \text{falls } 5 \leq t < 6 \\ \frac{15}{36} & \text{falls } 6 \leq t < 7 \\ \frac{21}{36} & \text{falls } 7 \leq t < 8 \\ \frac{26}{36} & \text{falls } 8 \leq t < 9 \\ \frac{30}{36} & \text{falls } 9 \leq t < 10 \\ \frac{33}{36} & \text{falls } 10 \leq t < 11 \\ \frac{35}{36} & \text{falls } 11 \leq t < 12 \\ 1 & \text{falls } 12 \leq t \end{cases}$$

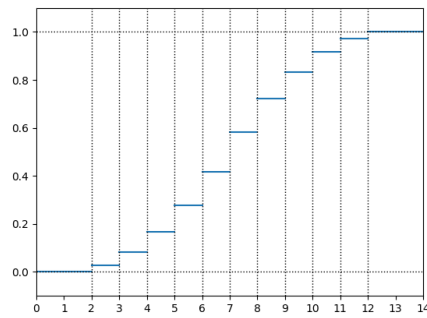


Abbildung 15: Graph der Funktion F

Analog zu der Definition 67 des empirischen Medians und der Definition 72 des p -Quantils definieren wir den Median und das p -Quantil einer Zufallsvariable.

Definition 165. Sei (Ω, P) ein Wahrscheinlichkeitsraum, $X: \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable und $p \in (0, 1)$ eine reelle Zahl. Ein **p -Quantil** von X ist eine reelle Zahl ξ_p mit

$$P(X \leq \xi_p) \geq p \quad \text{und} \quad P(X \geq \xi_p) \geq 1 - p.$$

Ein $p = 0,5$ -Quantil von X heißt auch **Median** von X .

Bemerkung 166. Wir können mit Bemerkung 160 die Bedingung aus der vorherigen Definition umschreiben, wobei F die Verteilungsfunktion von X bezeichnet. Ein p -Quantil von X ist also eine reelle Zahl ξ_p mit

$$F(\xi_p) \geq p \quad \text{und} \quad F(\xi_p) - P(X = \xi_p) \leq p.$$

Beispiel 167.

- (4) Wir betrachten das Laplace Experiment **zweifacher Würfelwurf** und die Zufallsvariable X = „Summe der Augenzahlen“ aus Beispiel 161.(4). Es ist $\xi_{\frac{1}{4}} = 5$, denn

$$P(X \leq \xi) = F(\xi) \geq \frac{1}{4}$$

impliziert $\xi \geq 5$ und

$$P(X \geq \xi) = P(X = \xi) + P(X > \xi) = P(X = \xi) + (1 - P(X \leq \xi))$$

impliziert $\xi = 5$, da

$$P(X \geq \xi) = P(X = \xi) + (1 - P(X \leq \xi)) = \frac{4}{36} + \left(1 - \frac{10}{36}\right) = \frac{5}{6} \geq 0,75 = 1 - \frac{1}{4}.$$

Außerdem gilt $\xi_{\frac{1}{2}} = 7$. Für $p = \frac{5}{6}$ ist jeder Wert $\xi_p \in [9, 10)$ erlaubt.

2.5.1. Endliche Zufallsvariablen. In diesem Kapitel betrachten wir **endliche** Zufallsvariablen, also Abbildungen $X: \Omega \rightarrow \mathbb{R}$, wobei (Ω, P) ein endlicher Wahrscheinlichkeitsraum, also ein Wahrscheinlichkeitsraum mit einer **endlichen** Menge Ω ist. Tatsächlich haben wir bereits in Definition 159 nur von endlichen Zufallsvariablen gesprochen. Wir wollen jedoch im folgenden Kapitel 2.5.2 diese Definition 159 ein wenig verändern und sogenannte **stetige Zufallsvariablen** betrachten, sodass die allgemeinen Begriffe aus dem vorherigen Einführungskapitel 2.5 auch für diese gültig bleiben. Dieses gilt jedoch nicht für die in diesem Kapitel besprochenen Begriffe. Daher werden wir uns in diesem Kapitel lediglich mit **endlichen** Zufallsvariablen beschäftigen.

Definition 168. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbb{R}$ eine endliche Zufallsvariable mit Bild W . Die **Zähldichte** von X ist die Abbildung

$$\begin{aligned} f: \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto P(X = x). \end{aligned}$$

Die Einschränkung von f auf $W \subseteq \mathbb{R}$ wird mit $f_W: W \rightarrow [0, 1]$ bezeichnet und (etwas ungenau aber wegen Bemerkung 169 unproblematisch) ebenfalls **Zähldichte** von X genannt.

Ein **Modalwert** von X ist ein Maximum der Funktion f oder, äquivalent, von f_W .

Bemerkung 169. Eine endliche Zufallsvariable $X: \Omega \rightarrow \mathbb{R}$ hat notwendigerweise ein endliches Bild $W = \{x_1, \dots, x_m\}$. Für die Zähldichte f einer endlichen Zufallsvariablen X gilt, dass $f(x) = 0$, falls $x \notin W$. Die Einschränkung f_W „gehört also zu“ einem eindeutigen f , d.h. f und f_W legen sich also gegenseitig fest. (Die Schreibweise f_W wird nur dafür eingeführt, im folgenden nicht zu viele Fallunterscheidungen machen zu müssen.)

Beispiel 170.

- (4) Wir betrachten wieder das Laplace Experiment **zweifacher Würfelwurf** und die Zufallsvariable X „Summe der Augenzahlen“ aus Beispiel 161.(4). Die Zähldichte f von X hat den folgenden „Graph“,

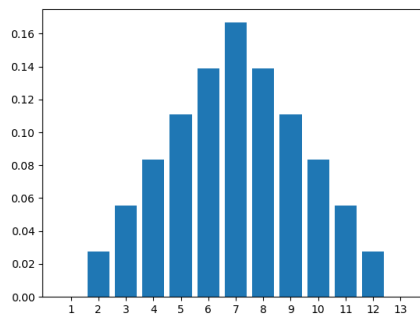


Abbildung 16: „Graph“ der Funktion f

wobei wir Balken für die jeweiligen Funktionswerte gezeichnet haben (Der Graph im Sinn der Definition hat nur Punkte bei $(x, P(X = x))$, falls $x \in W$ und $(x, 0)$ sonst.) und 7 ist ein Modalwert von X .

Bemerkung 171. Sei $X: \Omega \rightarrow \mathbb{R}$ eine endliche Zufallsvariable mit Zähldichte f und Verteilungsfunktion F . Dann gilt

$$F(t) = \sum_{x \leq t} f(x)$$

und die „Sprünge“ im Graph von F (siehe Abbildung 164) an der Stelle x sind von der „Höhe“ $f(x)$.

Analog zum empirischen Mittelwert 63, zur empirischen Varianz und zur empirischen Standardabweichung 77 macht man die folgende Definition.

Definition 172. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbb{R}$ eine endliche Zufallsvariable mit Bild $W = \{x_1, \dots, x_m\}$ und Zähldichte f .

- Der **Erwartungswert** von X ist die reelle Zahl

$$E(X) := \sum_{j=1}^m x_j f(x_j),$$

die manchmal auch mit μ bezeichnet wird.

- Die **Varianz** von X ist die reelle Zahl

$$Var(X) := \sum_{j=1}^m (x_j - \mu)^2 f(x_j),$$

die manchmal auch mit σ^2 bezeichnet wird. Die **Standardabweichung** von X ist die reelle Zahl

$$\sigma := \sqrt{\sigma^2}.$$

Bevor wir zu einem Beispiel zu diesen Zahlen kommen, besprechen wir erst einige Rechenregeln. Sind $X, Y: \Omega \rightarrow \mathbb{R}$ zwei Zufallsvariablen so betrachten wir im folgenden Satz deren Summe $X + Y$ und Produkt XY . Hiermit ist natürlich gemeint, dass man die jeweiligen Werte im Zielbereich der Abbildungen addiert oder multipliziert. So ist beispielsweise die Zufallsvariable $aX + Y: \Omega \rightarrow \mathbb{R}$ für $a \in \mathbb{R}$ gegeben durch $\omega \mapsto aX(\omega) + Y(\omega)$. Man beachte außerdem, dass die Charakterisierung (4) der Varianz im folgenden Satz oft nützlich für praktische Rechnungen ist.

Satz 173 (Eigenschaften von Erwartungswert und Varianz). Es seien X und Y endliche Zufallsvariablen.

- (1) Nimmt X nur den Wert a an, so gilt $E(X) = a$ und $Var(X) = 0$.
- (2) Sind $a, b \in \mathbb{R}$, so $E(aX + b) = aE(X) + b$ und $Var(aX + b) = a^2 Var(X)$.
- (3) Es gilt $E(X + Y) = E(X) + E(Y)$.
- (4) Es gilt $Var(X) = E(X^2) - E(X)^2$. („Verschiebungssatz“)
- (5) Sind X und Y **stochastisch unabhängig**, d.h. gilt für alle Intervalle I und J , dass

$$P(\{\omega \in \Omega \mid X(\omega) \in I\} \cap \{\omega \in \Omega \mid Y(\omega) \in J\}) = P(X \in I) \cdot P(Y \in J).$$

so gilt

$$E(XY) = E(X)E(Y)$$

und

$$Var(X + Y) = Var(X) + Var(Y).$$

Beispiel 174.

- (4) Wir betrachten wieder das Laplace Experiment **zweifacher Würfelwurf** und die Zufallsvariable X „Summe der Augenzahlen“ aus Beispiel 161.(4). X hat das Bild $W = \{2, 3, \dots, 12\} = \{x_1, x_2, \dots, x_m\}$. Dann ist

$$E(X) = \sum_{j=1}^{11} x_j f(x_j) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \dots + 12 \cdot \frac{1}{36} = 7$$

und

$$Var(X) = (2 - 7)^2 \cdot \frac{1}{36} + (3 - 7)^2 \cdot \frac{2}{36} + \dots + (12 - 7)^2 \cdot \frac{1}{36} = \frac{35}{6}$$

$$\text{und } \sigma = \sqrt{\frac{35}{6}} \approx 2,42.$$

Oft ist die Zähldichte einer endlichen Zufallsvariablen von einem der folgenden „Typen“, was es beispielsweise einfacher macht, den zugehörigen Erwartungswert und die zugehörige Varianz zu bestimmen.

Definition 175. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbb{R}$ eine endliche Zufallsvariable mit Bild $W = \{x_1, \dots, x_m\}$ und Zähldichte f .

- Hat X die Zähldichte gegeben durch $f_W(x) = \frac{1}{m}$, so heißt X **gleichmäßig verteilt** (oder **gleichverteilt**) und es gilt

$$E(X) = \frac{1}{m} \sum_{j=1}^m x_j \quad \text{und} \quad Var(X) = \frac{1}{m} \sum_{j=1}^m (x_j - E(X))^2$$

und man schreibt $X \sim \frac{1}{m}$.

- Ist $p \in (0, 1)$ eine reelle Zahl und $W = \{0, 1, \dots, n\}$ für eine natürliche Zahl $n \geq 1$ und hat X die Zähldichte gegeben durch $f_W(x) = \binom{n}{x} p^x \cdot (1 - p)^{n-x}$, so heißt X **binomialverteilt** und es gilt

$$E(X) = np \quad \text{und} \quad Var(X) = np(1 - p)$$

und man schreibt $X \sim Bin(n, p)$.

- Ist $W = \{0, 1, \dots, n\}$ für eine natürliche Zahl n und sind M, N natürliche Zahlen mit $n, M \leq N$ und hat X die Zähldichte gegeben durch

$$f_W(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}},$$

so heißt X **hypergeometrisch** verteilt und es gilt

$$E(X) = n \cdot \frac{M}{N} \quad \text{und} \quad \text{Var}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

und man schreibt $X \sim \text{Hyp}(n, M, N)$.

- Ist $p \in (0, 1)$ eine reelle Zahl und $W = \{1, \dots, n\}$ für eine natürliche Zahl $n \geq 1$ und hat X die Zähldichte gegeben durch $f_W(x) = (1-p)^{x-1} \cdot p$, so heißt X **geometrisch verteilt** und es gilt

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

und man schreibt $X \sim G(p)$.

- Ist $\lambda \in (0, \infty)$ eine reelle Zahl und $W = \{1, \dots, n\}$ für eine natürliche Zahl $n \geq 1$ und hat X die Zähldichte gegeben durch $f_W(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, so heißt X **poissonverteilt** und es gilt

$$E(X) = \lambda \quad \text{und} \quad \text{Var}(X) = \lambda$$

und man schreibt $X \sim P(\lambda)$.

Die folgenden wichtigen Beispiele sind Modelle typischer Zufallsexperimente durch endliche Wahrscheinlichkeitsräume und endliche Zufallsvariablen mit Zähldichten aus der obigen Definition 175.

Beispiel 176.

- (3) Wir betrachten das Laplace Experiment **einfacher Würfelwurf** aus Beispiel 124.(3). Hier war $\Omega = \{1, 2, 3, 4, 5, 6\}$. Die Zufallsvariable X = „Augenzahl“ ist genauer die Abbildung $X: \Omega \rightarrow \mathbb{R}$ mit $X(x) = x$ und dem Bild $W = \{1, 2, 3, 4, 5, 6\} = \Omega$. Es ist X gleichmäßig verteilt und wir lesen aus der obigen Definition 175 ab, dass

$$E(X) = \frac{21}{6} = 3,5 \quad \text{und} \quad \text{Var}(X) = \frac{35}{12}.$$

- (9) Ist $X \sim \text{Bin}(n, p)$ eine binomialverteilte Zufallsvariable, so kann man die reelle Zahl $f(x) = P(X = x)$ als die Wahrscheinlichkeit interpretieren **bei n -vielen Versuchen genau x -mal „Erfolg“** zu haben.

Betrachten wir beispielsweise wieder das Laplace Experiment **einfacher Würfelwurf** mit $\Omega = \{1, 2, 3, 4, 5, 6\}$. Sei X = „Anzahl der 6en“ mit Bild $W = \{0, 1\}$. Wir wissen

$$\begin{aligned} f(1) &= P(X = 1) = \frac{1}{6} = \binom{1}{1} p^1 (1-p)^0 \quad \text{und} \\ f(0) &= P(X = 0) = \frac{5}{6} = \binom{1}{0} p^0 (1-p)^1, \end{aligned}$$

wobei wir $p := \frac{1}{6}$ setzen. Also ist $X \sim \text{Bin}(1, p)$ binomialverteilt und man kann $f(x)$ als die Wahrscheinlichkeit interpretieren, bei $(n = 1)$ -vielen Versuchen, genau x -mal Erfolg zu haben, also eine 6 zu würfeln.

Führen wir dieses Zufallsexperiment n -viele Male durch, so betrachten wir das Laplace Experiment **n -facher Würfelwurf** mit $\Omega = \{1, 2, 3, 4, 5, 6\}^n$. Ein Element von Ω ist also ein n -Tupel mit den Wurfresultaten. Wir betrachten die Zufallsvariable X = „Anzahl der 6en“ mit Bild $W = \{0, 1, \dots, n\}$, welche die Anzahl der 6-en in einem solchen n -Tupel zählt. Die Wahrscheinlichkeit, dass ein spezielles n -Tupel x -mal eine 6 enthält, ist $p^x \cdot (1-p)^{n-x}$ und es gibt $\binom{n}{x}$ -viele Möglichkeiten dafür, wie die 6-en in einem Tupel verteilt sind. Daher ist

$$f(x) = P(X = x) = \binom{n}{x} p^x \cdot (1-p)^{n-x}.$$

Also ist $X \sim \text{Bin}(n, p)$ binomialverteilt und man kann $f(x)$ als die Wahrscheinlichkeit interpretieren, bei n -vielen Versuchen, genau x -mal Erfolg zu haben, also eine 6 zu würfeln.

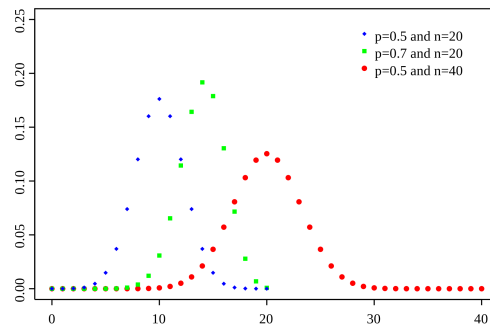


Abbildung 17: Graphen der Zähldichten dreier binomialverteilter Zufallsvariablen (von [6])

- (10) Ist $X \sim \text{Hyp}(n, M, N)$ eine hypergeometrisch verteilte Zufallsvariable, so kann man die reelle Zahl $f(x) = P(X = x)$ als die Wahrscheinlichkeit interpretieren **bei n -vielen Zügen aus einem Sack mit N Kugeln von denen M markiert sind, genau x -viele markierte zu ziehen.**

Betrachten wir beispielsweise das Laplace Experiment **Lotto** aus Beispiel 141, bei dem man versucht, auf $n = 6$ aus $N = 49$ Zahlen zu tippen und dabei x -viele der $M = 6$ markierten Zahlen (der „richtigen“ Tipps) zu erwischen. Sei X „Anzahl der richtigen Tipps“ mit Bild $W = \{0, 1, \dots, 6\}$. Es ist

$$f(x) = P(X = x) = \frac{\binom{6}{x} \binom{49-6}{6-x}}{\binom{49}{6}}.$$

Also ist $X \sim \text{Hyp}(n, M, N)$ hypergeometrisch verteilt und wir lesen aus der obigen Definition 175 ab, dass

$$E(X) = 6 \cdot \frac{6}{49} \approx 0,73 \quad \text{und} \quad \text{Var}(X) \approx 0,58.$$

(Da $E(X) < 1$ kann man beim Lotto also noch nicht einmal erwarten, auch nur eine einzige richtige Zahl getippt zu haben.)

- (11) Ist $X \sim G(p)$ eine geometrisch verteilte Zufallsvariable, so kann man die reelle Zahl $f(x) = P(X = x)$ als die Wahrscheinlichkeit interpretieren, **beim x -ten Versuch erstmals „Erfolg“ zu haben.** (Um tatsächlich in die Situation einer endlichen Zufallsvariable zu kommen, müssten wir theoretisch die Maximalanzahl der Versuche n beschränken, dieses ist aber praktisch nicht relevant und auch anders modellierbar, worauf wir aber nicht eingehen möchten.)

Betrachten wir beispielsweise wieder das Laplace Experiment **n -facher Würfelwurf** (mit einem großen n) und sei X „Wurf mit der ersten 6“ mit Bild $W = \{1, \dots, n\}$. Es ist

$$f(x) = P(X = x) = \left(\frac{5}{6}\right)^{x-1} \cdot \frac{1}{6} = (1-p)^{x-1} \cdot p,$$

wobei wir $p := \frac{1}{6}$ setzen. Also ist $X \sim G(p)$ geometrisch verteilt und wir lesen aus der obigen Definition 175 ab, dass

$$E(X) = 6 \quad \text{und} \quad \text{Var}(X) = 30.$$

Es ist also zu erwarten, beim wiederholten Würfeln die erste 6 im 6-ten Würfelwurf zu würfeln.

- (12) Ist $\lambda = np \in (0, \infty)$ mit $p \in (0, 1)$, so kann man zeigen, dass ungefähr

$$\binom{n}{x} p^x \cdot (1-p)^{n-x} \approx \frac{\lambda^x}{x!} e^{-x} \quad (\text{„Poissonscher Grenzwertsatz“})$$

für kleines p (beispielsweise $p \leq 0,1$) und großes n (beispielsweise $n \geq 100$). Ist also $X \sim P(\lambda)$ eine poissonverteilte Zufallsvariable, so kann man diese als „**Approximation**“ einer binomialverteilten Zufallsvariable bei kleinem p und großem n

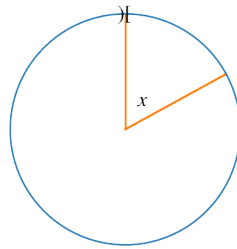
betrachten. Die Zähldichte einer binomialverteilten Zufallsvariable kann sehr schwer numerisch zu bestimmen sein. Daher kann eine solche „Approximation“ hilfreich sein.

2.5.2. Stetige Zufallsvariablen. In diesem Kapitel möchten wir uns mit speziellen nicht-endlichen Zufallsvariablen $X: \Omega \rightarrow \mathbb{R}$ befassen, deren Bild beispielsweise ein Intervall mit nicht endlich vielen Elementen ist. Streng genommen müssen wir dafür die Definition 125 eines endlichen Wahrscheinlichkeitsraums etwas abändern, sodass nicht mehr alle Teilmengen $A \subseteq \Omega$ als Ereignisse in Frage kommen, sondern nur solche aus einer sogenannten „ σ -Algebra“ (welche mit zu den Daten eines nicht-endlichen Wahrscheinlichkeitsraums gehört). Dieses technische Detail werden wir jedoch aus Zeitgründen ignorieren, da es in allen unseren praktischen Beispielen nicht relevant ist.

Außerdem dürfen wir als nicht-endliche Zufallsvariablen X nicht **alle** Abbildungen $\Omega \rightarrow \mathbb{R}$ zulassen, sondern nur solche, die mit der σ -Algebra-Struktur verträglich sind. In diesem Kapitel betrachten wir nur eine spezielle Klasse solcher nicht-endlicher Zufallsvariablen, die sogenannten stetigen Zufallsvariablen. Diese können (und werden) wir definieren, ohne den Begriff „ σ -Algebra“ zu erwähnen. Wir starten mit einem Beispiel, um zu illustrieren, was der Begriff einer „stetigen Zufallsvariable“ leisten soll.

Beispiel 177.

- (6) Wir betrachten das Zufallsexperiment **Drehen eines Glücksrads** aus Beispiel 120.(6). Dabei wird ein Glücksrad gedreht, dann zufällig abgebremst und anschließend der Haltewinkel $x \in [0, 360) = \Omega$ gemessen. Wir betrachten die (sehr einfache) Zufallsvariable $X: \Omega \rightarrow \mathbb{R}$ mit $X(x) := x$, also $X =$ „das Glücksrad hält beim Haltewinkel x “.



Sei $a \in [0, 360]$. Wir möchten die Wahrscheinlichkeit dafür angeben, dass der Haltewinkel des Glücksrads „zwischen 0 und a “ liegt. Es liegt nahe, diese Wahrscheinlichkeit als $P(X \leq a) := \frac{a}{360}$ zu definieren. Mit dieser Definition erhalten wir insbesondere, dass $P(X = 0) = \frac{0}{360} = 0$, die Wahrscheinlichkeit dafür, dass das Glücksrad genau bei dem Winkel $x = 0$ hält, ist also Null. Ebenso sollte dann $P(X = a) = 0$ sein.

Allgemeiner setzen wir für $a, b \in [0, 360]$ mit $a \leq b$

$$\begin{aligned}
 P(a \leq X \leq b) &:= P(\{\omega \in \Omega \mid a \leq X(\omega) \leq b\}) \\
 &= P(\{\omega \in \Omega \mid X(\omega) \leq b\} \setminus \{\omega \in \Omega \mid X(\omega) < a\}) \\
 &= P(X \leq b) - P(\{\omega \in \Omega \mid X(\omega) \leq b\} \cap \{\omega \in \Omega \mid X(\omega) < a\}) \quad (\text{Satz 127.(3)}) \\
 &= P(X \leq b) - P(\{\omega \in \Omega \mid X(\omega) < a\}) \\
 &= P(X \leq b) - P(X < a) \\
 &= P(X \leq b) - (P(X < a) + 0) \\
 &= P(X \leq b) - (P(X < a) + P(X = a)) \\
 &= P(X \leq b) - P(X \leq a) \\
 &= \frac{b}{360} - \frac{a}{360} \\
 &= \frac{1}{360}(b - a).
 \end{aligned}$$

Die Verteilungsfunktion

$$\begin{aligned}
 F: \mathbb{R} &\rightarrow [0, 1] \\
 t &\mapsto P(X \leq t) = \begin{cases} 0 & \text{falls } t < 0 \\ \frac{t}{360} & \text{falls } t \in [0, 360] \\ 1 & \text{falls } 360 < t \end{cases}
 \end{aligned}$$

ist eine stetige Funktion im Gegensatz zum vorherigen Abschnitt, dem endlichen Fall. Im Gegensatz zum endlichen Fall, macht es hier keinen Sinn eine Zähldichte

$$\begin{aligned} f: \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto P(X = x) \end{aligned}$$

zu definieren, da ja $P(X = x) = 0$ gilt und diese Abbildung also die Nullabbildung ist.

Definition 178. Eine **Dichtefunktion** (oder **Wahrscheinlichkeitsdichte**) ist eine (stückweise stetige) Abbildung $f: \mathbb{R} \rightarrow [0, \infty)$ mit

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

Bemerkung 179. Das „Integral“ aus der vorherigen Definition benötigt eine Erklärung, da wir in Definition 35 eines Integrals

$$\int_a^b f(x) dx := F(b) - F(a)$$

zum einen nur stetige Abbildungen $f: [a, b] \rightarrow \mathbb{R}$ (mit einer Stammfunktion F) betrachtet haben und in der vorherigen Definition zum anderen a und b keine reellen Zahlen sind. Die Summenregel aus Satz 37 erlaubt uns zunächst, Integrale auch für „stückweise stetige“ Funktionen $f: [a, b] \rightarrow \mathbb{R}$ (wie beispielsweise die Abbildung aus Beispiel 184 unten) zu definieren. Hat nun $f: \mathbb{R} \rightarrow [0, \infty)$ eine Stammfunktion $F: \mathbb{R} \rightarrow [0, \infty)$, so setzen wir

$$\int_{-\infty}^{\infty} f(x) dx := \lim_{b \rightarrow \infty} \lim_{a \rightarrow -\infty} \int_a^b f(x) dx = \lim_{b \rightarrow \infty} \lim_{a \rightarrow -\infty} (F(b) - F(a)).$$



Die vorherige Bemerkung 179 sieht vielleicht erschreckend kompliziert aus, aber keine Sorge: Wir werden im Folgenden Dichtefunktionen nur explizit angeben und nicht überprüfen müssen, ob diese die obige Integralbedingung tatsächlich erfüllen. Mit anderen Worten können wir uns einfach merken, dass eine Dichtefunktion eine Funktion $f: \mathbb{R} \rightarrow [0, \infty)$ ist, die eine „spezielle Eigenschaft“ erfüllt.

Definition 180. Sei Ω eine Menge.

- Eine **stetige Zufallsvariable** ist eine Abbildung $X: \Omega \rightarrow \mathbb{R}$ zusammen mit einer Dichtefunktion $f: \mathbb{R} \rightarrow [0, \infty)$.
- Für $a, b \in \mathbb{R} \cup \{\pm\infty\}$ mit $a \leq b$ setzen wir

$$P(a \leq X \leq b) := \int_a^b f(x) dx$$

als die **Wahrscheinlichkeit** dafür, dass „ X zwischen a und b liegt“. Wir setzen zur Abkürzung $P(X \leq b) := P(-\infty \leq X \leq b)$ und $P(a \leq X) := P(a \leq X \leq \infty)$.

- Ein **Modalwert** von X ist ein Maximum der Funktion f .
- Die **Verteilungsfunktion** von X ist die Abbildung

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

Bemerkung 181. Ist X eine stetige Zufallsvariable und $a \in \mathbb{R}$, so bekommen wir, analog zum obigen Beispiel 177, dass

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x) dx = F(a) - F(a) = 0.$$

Die Wahrscheinlichkeit dafür, dass X eine einzelne reelle Zahl a „annimmt“ ist also stets Null. Dieses ist ein wesentlicher Unterschied zu den endlichen Zufallsvariablen aus Kapitel 2.5.1.

Bemerkung 182. Die Verteilungsfunktion F einer stetigen Zufallsvariable ist eine stetige, „stückweise differenzierbare“ und monoton wachsende Funktion und es gilt

$$\lim_{t \rightarrow -\infty} F(t) = 0 \quad \text{und} \quad \lim_{t \rightarrow \infty} F(t) = 1,$$

sowie (im Einklang mit Bemerkung 160)

$$P(X \geq a) = P(a \leq X) = \int_a^\infty f(x)dx = \int_{-\infty}^\infty f(x)dx - \int_{-\infty}^a f(x)dx = 1 - P(X \leq a).$$

Bemerkung 183. Ist X eine stetige Zufallsvariable mit einer stetigen Dichtefunktion f und Stammfunktion F , so ist es kein Zufall, dass die Verteilungsfunktion von X ebenfalls mit dem großen Buchstaben „ F “ bezeichnet wird. Bezeichnen wir kurzfristig die Verteilungsfunktion mit \mathbb{F} , so gilt nämlich

$$\begin{aligned} \mathbb{F}(t) &= \int_{-\infty}^t f(x)dx \\ &= \lim_{a \rightarrow -\infty} \int_a^t f(x)dx \\ &= \lim_{a \rightarrow -\infty} (F(t) - F(a)) \\ &= F(t) - \lim_{a \rightarrow -\infty} F(a) \\ &= F(t) - 0 \\ &= F(t), \end{aligned}$$

was sich auf („stückweise stetige“) Dichtefunktionen f überträgt. Daher kann man tatsächlich sagen, dass die Verteilungsfunktion von X eine Stammfunktion von f **ist**.

Beispiel 184.

(6) Im obigen Glücksrad-Beispiel 177 ist X eine stetige Zufallsvariable mit Dichtefunktion

$$\begin{aligned} f: \mathbb{R} &\rightarrow [0, \infty) \\ x &\mapsto \begin{cases} 0 & \text{falls } x < 0 \\ \frac{1}{360} & \text{falls } x \in [0, 360] \\ 0 & \text{falls } 360 < x \end{cases} \end{aligned}$$

und Verteilungsfunktion

$$\begin{aligned} F: \mathbb{R} &\rightarrow [0, \infty) \\ t &\mapsto \begin{cases} 0 & \text{falls } t < 0 \\ \frac{t}{360} & \text{falls } t \in [0, 360] \\ 1 & \text{falls } 360 < t \end{cases} \end{aligned}$$

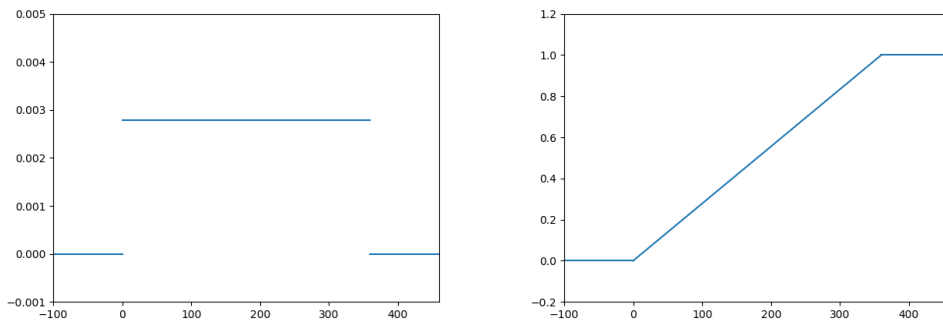


Abbildung 18: Graphen der Dichtefunktion f und Verteilungsfunktion F von X .

Definition 185. Sei $X: \Omega \rightarrow \mathbb{R}$ eine stetige Zufallsvariable mit Dichtefunktion f .

- Sei $p \in (0, 1)$ eine reelle Zahl und F die Verteilungsfunktion von X . Ein **p -Quantil** von X ist eine reelle Zahl ξ_p mit $F(\xi_p) = p$.
- Der **Erwartungswert** von X ist die reelle Zahl

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

die manchmal auch mit μ bezeichnet wird.

- Die **Varianz** von X ist die reelle Zahl

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

die manchmal auch mit σ^2 bezeichnet wird. Die **Standardabweichung** von X ist die reelle Zahl

$$\sigma := \sqrt{\sigma^2}.$$

Satz 186. Der Satz 173 (Eigenschaften von Erwartungswert und Varianz) gilt komplett analog auch für stetige Zufallsvariablen X .

Beispiel 187.

(6) Im obigen Glücksrad-Beispiel 177 gilt

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{360} x \frac{1}{360} dx \\ &= \frac{1}{360} \int_0^{360} x dx = \frac{1}{360} \left[\frac{1}{2} x^2 \right]_0^{360} \\ &= \frac{1}{360} \cdot \frac{1}{2} \cdot (360)^2 = 180 \end{aligned}$$

und (mit dem „Verschiebungssatz“ aus Satz 173)

$$Var(X) = E(X^2) - E(X)^2 = \int_0^{360} x^2 \frac{1}{360} dx - (180)^2 = 10800 \quad \text{also} \quad \sigma \approx 103,92.$$

Ebenso wie für endlichen Zufallsvariablen in Definition 175, schauen wir uns nun Dichtefunktionen speziellen „Typs“ an.

Definition 188. Sei Ω eine Menge und $X: \Omega \rightarrow \mathbb{R}$ eine stetige Zufallsvariable.

- Sind a, b reelle Zahlen mit $a < b$ und hat X die Dichtefunktion f gegeben durch

$$f(x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{1}{b-a} & \text{falls } x \in [a, b] \\ 0 & \text{falls } b < x \end{cases}$$

so heißt X **gleichmäßig verteilt** (oder **gleichverteilt**) und es gilt

$$F(t) = \begin{cases} 0 & \text{falls } t < a \\ \frac{t-a}{b-a} & \text{falls } t \in [a, b] \\ 1 & \text{falls } b < t \end{cases} \quad \text{und} \quad E(X) = \frac{a+b}{2} \quad \text{und} \quad Var(X) = \frac{(b-a)^2}{12}$$

und man schreibt $X \sim \frac{1}{b-a}$.

- Ist $\lambda \in (0, \infty)$ und hat X die Dichtefunktion f gegeben durch

$$f(x) = \begin{cases} 0 & \text{falls } x < 0 \\ \lambda e^{-\lambda x} & \text{falls } x \geq 0 \end{cases}$$

so heißt X **exponentialverteilt** und es gilt

$$F(t) = \begin{cases} 0 & \text{falls } t < 0 \\ 1 - e^{-\lambda t} & \text{falls } t \geq 0 \end{cases} \quad \text{und} \quad E(X) = \frac{1}{\lambda} \quad \text{und} \quad Var(X) = \frac{1}{\lambda^2}$$

und man schreibt $X \sim Ex(\lambda)$.

- Sind $\mu \in \mathbb{R}$ und $\sigma \in (0, \infty)$ reelle Zahlen und hat X die Dichtefunktion f gegeben durch

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

so heißt X **normalverteilt** und es gilt

$$E(X) = \mu \quad \text{und} \quad Var(X) = \sigma^2$$

und man schreibt $X \sim N(\mu, \sigma^2)$.

- Ist $X \sim N(\mu, \sigma^2)$ normalverteilt mit $\mu = 0$ und $\sigma^2 = 1$, so heißt X **standardnormalverteilt** und man schreibt $X \sim N(0, 1)$. In diesem Fall bezeichnet man die Dichtefunktion f gegeben durch

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

auch mit φ (sprich „klein-Fieh“) und die Verteilungsfunktion von X mit Φ (sprich „groß-Fieh“).

Normalverteilte Zufallsvariablen haben insbesondere durch den Zentralen Grenzwertsatz 191 in Kapitel 2.6 eine besondere Bedeutung. Bevor wir zu Beispielen stetiger Zufallsvariablen mit Dichtefunktionen aus der obigen Definition 188 kommen, folgt ersteinmal eine Bemerkung mit einem wichtigen „Umrechnungswerkzeug“ für normalverteilte Zufallsvariablen.

Bemerkung 189 (Umrechnung in Standardnormalverteilung).

- Es ist nicht einfach die numerischen Werte $\varphi(x)$ und $\Phi(t)$ zu bestimmen. Daher benutzen wir die Tabelle im Anhang, um die Werte $\Phi(t)$ „auszurechnen“.
- Wir können eine normalverteilte Zufallsvariable X in eine standardnormalverteilte Zufallsvariable Z „transformieren“ und auf diese Weise die Tabelle für die Standardnormalverteilung im Anhang benutzen, um $F(t)$ zu bestimmen. Ist genauer X eine Zufallsvariable und $N(\mu, \sigma^2)$ -verteilt, so ist (genau dann) die Zufallsvariable

$$Z := \frac{X - \mu}{\sigma}$$

standardnormalverteilt, also $Z \sim N(0, 1)$ und es gilt

$$F(t) = \Phi\left(\frac{t - \mu}{\sigma}\right)$$

(und $f(x) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})$), wobei F die Verteilungsfunktion (und f die Dichtefunktion) von X bezeichnet.

- Der vorherige Punkt ist wichtig für die folgende Rechnung. Sei wieder $X \sim N(\mu, \sigma^2)$ normalverteilt. Für $a, b \in \mathbb{R}$ mit $a \leq b$ interessiert man sich oft für den Wert

$$P(a \leq X \leq b) = F(b) - F(a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

den wir nun „einfach“ aus der Tabelle im Anhang ablesen können.

- In der Tabelle im Anhang sind nur Werte $\Phi(t)$ mit $t \geq 0$ angegeben, es ist aber $\Phi: \mathbb{R} \rightarrow [0, 1]$ eine Abbildung, in die man auch negative reelle Zahlen „einsetzen“ kann. Diese Werte kann man ebenfalls aus der Tabelle ablesen, da (aus Symmetriegründen, die wir nicht weiter erläutern) gilt

$$\Phi(-t) = 1 - \Phi(t).$$

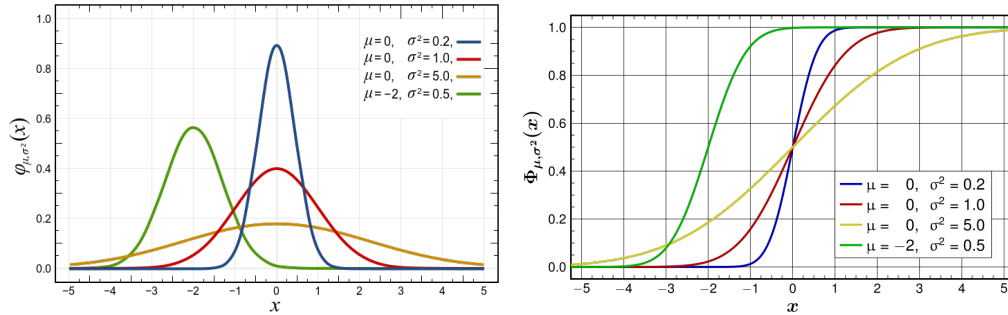


Abbildung 19: Graphen der Dichtefunktion f und Verteilungsfunktion F normalverteilter Zufallsvariablen $X \sim N(\mu, \sigma^2)$ von [7]. Der rote Graph der Dichtefunktion φ der Standardnormalverteilung wird auch **Gauß'sche Glockenkurve** genannt.

- Ist $X \sim N(\mu, \sigma^2)$ -verteilt und $p \in (0, 1)$ eine reelle Zahl, so ist nach Definition 185 das p -Quantil ξ_p von X eine Lösung der Gleichung $F(\xi_p) = p$, wobei F die Verteilungsfunktion von X bezeichnet. Die p -Quantile einer standardnormalverteilten Zufallsvariablen werden auch mit z_p bezeichnet und es gilt $z_p = -z_{1-p}$, sowie $\xi_p = \mu + \sigma z_p$. Aus der Tabelle im Anhang kann man die folgenden (ungefähren) Werte ablesen.

	$p = 0,5$	$p = 0,75$	$p = 0,9$	$p = 0,95$	$p = 0,99$
z_p	0	0,67449	1,28155	1,64485	2,32635

- Sind $X \sim N(\mu, \sigma^2)$ und $X' \sim N(\mu', \sigma'^2)$ zwei normalverteilte Zufallsvariablen und $\lambda, \lambda' \in \mathbb{R}$ reelle Zahlen, so kann man zeigen, dass die Zufallsvariable $\lambda X + \lambda' X'$ ebenfalls normalverteilt ist und es gilt genauer

$$\lambda X + \lambda' X' \sim N(\lambda\mu + \lambda'\mu', \lambda^2\sigma^2 + \lambda'^2\sigma'^2).$$

Beispiel 190.

- (6) Im obigen Glücksrad-Beispiel 177 ist X eine gleichmäßig verteilte Zufallsvariable.
- (13) Exponentialverteilte Zufallsvariablen modellieren oft eine **Lebensdauer** oder einen Zerfallsprozess. Sei beispielsweise die Zufallsvariable $X :=$ „Lebensdauer einer Maschine“ (in Stunden) exponentialverteilt mit Erwartungswert $E(X) = \frac{1}{\lambda} = 500$ (Stunden). Die Wahrscheinlichkeit dafür, dass die Betriebsdauer zwischen $a = 300$ Stunden und $b = 600$ Stunden liegt, ist

$$P(300 \leq X \leq 600) = F(600) - F(300) = (1 - e^{-\frac{600}{500}}) - (1 - e^{-\frac{300}{500}}) \approx 0,25.$$

- (14) Sei die stetige Zufallsvariable $X :=$ „Gewicht einer Medikamentenflasche“ $N(1000, 16)$ -verteilt. Der Erwartungswert von X ist $\mu = 1000$ (Gramm) und die Varianz $\sigma^2 = 16$. Wir fragen uns, mit welcher Wahrscheinlichkeit das Gewicht der Medikamentenflasche zwischen 990 und 1005 Gramm liegt. Dafür berechnen wir

$$\begin{aligned}
 P(990 \leq X \leq 1005) &= F(1005) - F(990) \\
 &= \Phi\left(\frac{1005-1000}{4}\right) - \Phi\left(\frac{990-1000}{4}\right) \\
 &= \Phi(1,25) - \Phi(-2,5) \\
 &= \Phi(1,25) - (1 - \Phi(2,5)) \\
 &\approx 0,8944 - (1 - 0,9938) \\
 &\approx 0,89.
 \end{aligned}$$

Alternativ können wir auch wie folgt vorgehen. Es ist $Z = \frac{X-\mu}{\sigma}$ nach Bemerkung 189 standardnormalverteilt. Umformen ergibt $X = \sigma Z + \mu$. Dann berechnen wir

$$\begin{aligned} P(990 \leq X \leq 1005) &= P(990 \leq \sigma Z + \mu \leq 1005) \\ &= P(990 - \mu \leq \sigma Z \leq 1005 - \mu) \\ &= P(-10 \leq 4Z \leq 5) \\ &= P\left(-\frac{5}{2} \leq Z \leq \frac{5}{4}\right) \\ &= \Phi\left(\frac{5}{4}\right) - \Phi\left(-\frac{5}{2}\right) \\ &= \Phi(1,25) + \Phi(2,5) - 1 \\ &\approx 0,89. \end{aligned}$$

2.6. Der zentrale Grenzwertsatz. (Wir werden im Folgenden, motiviert durch Definition 175 und Definition 188 eine Zähldichte oder eine Dichtefunktion einer Zufallsvariable X auch eine **Verteilung** von X nennen. Achtung: Dieses ist nicht zu verwechseln mit der Verteilungsfunktion einer Zufallsvariablen.)

Die besondere Bedeutung der Normalverteilung ist durch den folgenden überraschenden Satz gerechtfertigt.

Satz 191 (Zentraler Grenzwertsatz – informelle Formulierung).

Es seien X_1, X_2, \dots, X_n unabhängige (stetige oder endliche) Zufallsvariablen, die alle dieselbe Verteilung besitzen (der wesentliche Punkt ist hier: Es ist egal, welche Verteilung) mit Erwartungswert μ und Varianz σ^2 und betrachte die Zufallsvariable $\bar{X}_n := \frac{X_1 + \dots + X_n}{n}$. Dann ist

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{falls } n \gg 0 \text{ groß,}$$

d.h. die Zufallsvariable \bar{X}_n ist „ungefähr“ normalverteilt mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$ und diese Approximation wird umso besser, je größer n ist.

Bemerkung 192. Sei X_1, X_2, \dots eine Folge unabhängiger Zufallsvariablen, die alle dieselbe Verteilung besitzen mit Erwartungswert μ und Varianz σ^2 . Eine präzisere Formulierung der Aussage des Zentralen Grenzwertsatzes ohne das Wort „ungefähr“ ist

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sqrt{n} \cdot \sigma} \leq t\right) = \Phi(t)$$

wobei $S_n := X_1 + \dots + X_n$. Es gibt viele Untersuchungen zu der „Qualität“ der Approximation des Zentralen Grenzwertsatzes, d.h. zu der Frage, wie „schnell“ die obige Konvergenz ist, wie zum Beispiel der Satz von Berry–Esseen, den wir hier aber nicht betrachten werden.

Bemerkung 193. Mit den „Rechenregeln“ aus dem letzten Punkt von Bemerkung 189 können wir die Aussage des Zentralen Grenzwertsatzes auf die folgende Weise etwas anders formulieren. Für die Verteilung der Zufallsvariablen $S_n := X_1 + \dots + X_n$ gilt ungefähr

$$S_n \approx N(n\mu, n\sigma^2) \quad \text{falls } n \gg 0 \text{ groß.}$$

Beispiel 194.

(9') Wir betrachten wieder den n -fachen **Würfelwurf** aus Beispiel 176.(9) mit Ergebnismenge $\Omega = \{1, 2, 3, 4, 5, 6\}^n$ und die Zufallsvariable X = „Anzahl der 6en“, die wir nun mit $S_n := X$ bezeichnen, um die Abhängigkeit von n zu verdeutlichen. Wir haben in Beispiel 176.(9) gezeigt, dass $X \sim \text{Bin}(n, p)$ -verteilt ist mit $p = \frac{1}{6}$. Nun würfeln wir $n = 600$ -mal hintereinander und fragen nach der Wahrscheinlichkeit dafür, mindestens 90 und höchstens 100 Sechsen zu erhalten, also nach der Wahrscheinlichkeit

$$P(90 \leq S_n \leq 100).$$

Dafür müssen wir $P(S_n = 90) + P(S_n = 91) + \dots + P(S_n = 100)$ ausrechnen und wenn wir uns die Zähldichte der Binomialverteilung aus Definition 175 ansehen, ist dieses numerisch enorm anstrengend, da n verhältnismäßig groß ist. Hier machen wir uns den Zentralen Grenzwertsatz zu Nutze!

Dafür definieren wir für $i \in \{1, 2, \dots, 600\}$ jeweils eine Zufallsvariable

$$\begin{aligned} X_i: \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto \begin{cases} 1 & \text{falls } i\text{-ter Wurf eine 6 ist, d.h. falls } \omega_i = 6 \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

Dann ist $S_n := X_1 + \dots + X_{600}$ die obige Zufallsvariable, die die Anzahl der 6en bestimmt. Offenbar sind die X_i unabhängig und besitzen alle die selbe Verteilung mit Erwartungswert $\mu = \frac{1}{6}$ und Varianz $\sigma^2 = E(X_i^2) - E(X_i)^2 = \frac{1}{6} - \frac{1}{36} = \frac{5}{36}$. Nach Bemerkung 193 zum Zentralen Grenzwertsatz ist S_n also ungefähr $N(100, \frac{500}{6})$ -verteilt. Dann gilt mit der Umrechnung in die Standardnormalverteilung aus Bemerkung 189 (Achtung, dass μ und σ für die Umrechnung sind jetzt $\mu' := 100$, $\sigma' := \sqrt{\frac{500}{6}} \approx 9,129$), dass

$$\begin{aligned} P(90 \leq S_n \leq 100) &\approx \Phi\left(\frac{100-\mu'}{\sigma'}\right) - \Phi\left(\frac{90-\mu'}{\sigma'}\right) \\ &= \Phi(0) - \Phi\left(\frac{-10}{\sigma'}\right) \\ &= 0,5 - (1 - \Phi\left(\frac{10}{\sigma'}\right)) \\ &\approx 0,5 + \Phi(1,095) - 1 \\ &\approx 0,5 + 0,8621 - 1 \\ &\approx 0,36. \end{aligned}$$

(Überprüft man mit dem Satz von Berry–Esseen die „Qualität“ dieser Approximation, so stellt man fest, dass der mögliche Fehler $\leq \pm 12,6\% = 0,126$ ist, was eine relativ große Abweichung von den berechneten 0,36 sein kann. Mit einem größeren n wird dieser Fehler kleiner.)

2.7. Das Gesetz der großen Zahlen. Angenommen wir wollen ein Zufallsexperiment modellieren und haben uns bereits für eine Ergebnismenge Ω entschieden. Nun möchten wir die Funktion $P: \mathbb{P}(\Omega) \rightarrow [0, 1]$ angeben. Dazu müssen wir die Wahrscheinlichkeit $p := P(A)$ eines Ereignisses $A \subseteq \Omega$ „bestimmen“.

Hierzu wählen wir eine natürliche Zahl n und machen ein „neues“ Zufallsexperiment mit Ergebnismenge $\Omega' := \{\text{„Ja“}, \text{„Nein“}\}^n$, das die n -malige Ausführung des ursprünglichen Zufallsexperiments unter der Betrachtung „Ist das Ereignis A eingetreten?“ modellieren soll.

Wir definieren für jedes $i \in \{1, \dots, n\}$ eine (endliche) Zufallsvariable

$$\begin{aligned} X_i: \Omega' &\rightarrow \mathbb{R} \\ (\omega'_1, \dots, \omega'_n) &\mapsto \begin{cases} 1 & \text{falls } \omega'_i = \text{„Ja“} \\ 0 & \text{sonst,} \end{cases} \end{aligned}$$

die angibt, ob das Ereignis A bei der i -ten der n -vielen Wiederholungen des ursprünglichen Zufallsexperiments eingetreten ist. Offenbar sind die Zufallsvariablen X_i unabhängig und jeweils $X_i \sim \text{Bin}(1, p)$ -verteilt mit einem (noch unbekannten) $p \in (0, 1)$, d.h. es gilt jeweils

$$P(X_i = 1) = p \quad \text{und} \quad P(X_i = 0) = (1 - p).$$

Die Zufallsvariable $S_n := X_1 + \dots + X_n: \Omega' \rightarrow \mathbb{R}$ beschreibt also die Anzahl der Wiederholungen des ursprünglichen Zufallsexperiments, bei denen das Ereignis A tatsächlich eingetreten ist. Die Zufallsvariable $\bar{X}_n := \frac{X_1 + \dots + X_n}{n}$ gibt die sogenannte **relative Häufigkeit** des Eintretens von A unter den n -vielen Wiederholungen an.

Satz 195 (Bernoulli'sches Gesetz der großen Zahlen). Mit den oben eingeführten Bezeichnungen gilt für jedes $c > 0$, dass

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - p| \geq c) = 0$$

und genauer sogar $P(|\bar{X}_n - p| \geq c) \leq \frac{1}{4nc^2}$ für ein (fixes) $n \geq 1$.

Bemerkung 196. Das Bernoulli'sches Gesetz der großen Zahlen 195 sagt also, dass für viele Wiederholungen (großes $n \gg 0$) des ursprünglichen Zufallsexperiments, die Wahrscheinlichkeit dafür sehr klein ist (und bei größerem n immer kleiner wird), dass die relative Häufigkeit \bar{X}_n des Eintretens von A von der (unbekannten) Wahrscheinlichkeit p abweicht. Dieses bedeutet also, dass man die Wahrscheinlichkeit $p = P(A)$ des Ereignisses A im ursprünglichen Zufallsexperiment durch die relative Häufigkeit \bar{X}_n des Eintretens von A in vielen Wiederholungen des Zufallsexperiments **bestimmen** kann.

Beispiel 197. Die Wahrscheinlichkeit $p := P(A)$ eines Ereignisses $A \subseteq \Omega$ soll durch dessen relative Häufigkeit geschätzt werden. Das zugehörige Zufallsexperiment muss dann mindestens ($n = 50000$)-Mal wiederholt werden, damit diese Schätzung höchstens mit einer Wahrscheinlichkeit von 5% um mehr als $c := 0,01$ von p abweicht, denn

$$P(|\bar{X}_n - p| \geq 0,01) \leq \frac{1}{4 \cdot n \cdot 0,0001}$$

und mit $\frac{1}{4 \cdot n \cdot 0,0001} = 0,05 = 5\%$ folgt $n = 50000$.

Der folgende Satz ist eine (teilweise, da ohne die präzise Abschätzung aus Satz 195) Verallgemeinerung des Bernoulli'schen Gesetzes der großen Zahlen für beliebige (und nicht zwingend $X_i \sim \text{Bin}(1, p)$ -verteilte) unabhängige Zufallsvariablen X_1, X_2, \dots , die alle dieselbe Verteilung besitzen.

Satz 198 ((schwaches) Gesetz der großen Zahlen). Sei X_1, X_2, \dots eine Folge (stetiger oder endlicher) unabhängiger Zufallsvariablen, die alle dieselbe Verteilung besitzen mit Erwartungswert μ und betrachte die Zufallsvariablen $\bar{X}_n := \frac{X_1 + \dots + X_n}{n}$ für $n \geq 1$. Dann gilt für jedes $c > 0$, dass

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq c) = 0.$$

Bemerkung 199. Für die Beziehung zwischen dem Gesetz der großen Zahlen 198 und dem Bernoulli'schen Gesetz der großen Zahlen 195 benutzt man, dass nach Definition 175 der Erwartungswert einer $\text{Bin}(1, p)$ -verteilten Zufallsvariable X_i genau $\mu = p$ ist.

3. Induktive Statistik

Das Ziel der (induktiven, oder „schließenden“) Statistik ist, Eigenschaften neuer und unbekannter Stichproben aus den Daten bereits bekannter und in gewisser Weise ähnlicher Stichproben „vorherzusagen“. Beispielsweise soll aus einer Befragung von 1000 Leuten zu ihrem Wahlverhalten ein Rückschluss auf das Wahlverhalten aller Bürger*innen eines Landes getroffen werden.



Die „Unterschiede“ zwischen Wahrscheinlichkeitstheorie und (induktiver) Statistik sind, grob gesprochen, die Folgenden.

- In der **Wahrscheinlichkeitstheorie** haben wir einen Wahrscheinlichkeitsraum (Ω, P) und eine Zufallsvariable $X: \Omega \rightarrow \mathbb{R}$ gegeben, die ein Zufallsexperiment beschreibt. Wir können Eigenschaften von X wie den Erwartungswert und die Varianz untersuchen und etwas über den (unbekannten) Ausgang des Zufallsexperiments schließen.
- In der (induktiven) **Statistik** beobachtet man in einem Zufallsexperiment eine Stichprobe $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Diese Stichprobe wird als ein Bildpunkt („Realisierung“) von n -vielen Zufallsvariablen $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}$ aufgefasst. Anhand dieser Stichprobe x stellt man eine Vermutung über die unbekannten Dichtefunktionen („Verteilungen“) von X_1, \dots, X_n auf und legt diese fest.
(Sind die Dichtefunktionen von X_1, \dots, X_n bekannt, ist man wieder auf dem Gebiet der Wahrscheinlichkeitstheorie und kann etwas über den (weiteren unbekannten) Ausgang des Zufallsexperiments schließen.)

3.1. Grundlegende Begriffe. Wir betrachten im Folgenden stets n -viele Zufallsvariablen X_1, \dots, X_n von denen wir annehmen, dass sie **unabhängig** sind und alle **dieselbe** (möglicherweise noch unbekannte) **Verteilung** besitzen. Man kann X_1, \dots, X_n also auch als Modellierung einer n -fachen Wiederholung des gleichen Zufallsexperiments interpretieren.



Zu Erinnerung: Der etwas unpräzise Begriff der **Verteilung** bedeutet für uns **Zähldichte** oder **Dichtefunktion**.

Definition 200. Seien $n \geq 1$ und X_1, \dots, X_n (stetige oder endliche) Zufallsvariablen. Dann heißen X_1, \dots, X_n **einfach**, wenn sie unabhängig sind und alle dieselbe Verteilung besitzen. Sind X_1, \dots, X_n einfach, so nennen wir die Verteilung eines jeden X_i auch **gemeinsame Verteilung** der X_1, \dots, X_n .

In der (induktiven) Statistik unterscheidet man grob zwischen zwei Arten der „Bestimmung“ der gemeinsamen unbekannten Verteilung einfacher Zufallsvariablen X_1, \dots, X_n :

- **Parametrische Tests/Schätzungen:** Hier ist die Art der Verteilung bekannt, nur abhängig von einem Parameter $\theta \in \mathbb{R}$ (beispielsweise $X_i \sim N(\theta, 25)$) und man sucht nach dem unbekannten Parameter θ . (Man sagt, dass man den unbekannten Parameter „schätzen“ möchte.)
- **Nicht parametrische Tests/Schätzungen:** Hier ist die Art der Verteilung unbekannt.



Diese beiden Begriffe sind (zumindest in diesem Skript) nicht klar definiert und dienen daher nur zur groben sprachlichen Orientierung.

Definition 201. Seien einfache Zufallsvariablen X_1, \dots, X_n gegeben.

- Eine Abbildung $g: \mathbb{R}^n \rightarrow \mathbb{R}$ nennt man einen **Schätzer** und einen Bildpunkt $g(x_1, \dots, x_n)$ einen **Schätzwert**.
- Ist g ein Schätzer, so nennt man die Abbildung

$$T: \Omega^n \xrightarrow{(X_1, \dots, X_n)} \mathbb{R}^n \xrightarrow{g} \mathbb{R}, \quad (\omega_1, \dots, \omega_n) \mapsto g(X_1(\omega_1), \dots, X_n(\omega_n))$$

eine **Stichprobenfunktion** (oder eine **Schätzvariable** oder eine **Statistik**).

Wir beschäftigen uns im Folgenden mit parametrischen Tests und nehmen also an, dass die gemeinsame (unbekannte) Verteilung einfacher Zufallsvariablen X_1, \dots, X_n nur abhängig ist von einem (unbekannten) Parameter $\theta \in \Theta \subseteq \mathbb{R}$, von dem wir annehmen, dass er aus einer Teilmenge Θ der reellen Zahlen stammt. (Es ist meistens $\Theta := \mathbb{R}$ oder $\Theta := (0, \infty)$.)

Beispiel 202. Wir betrachten ein Zufallsexperiment, das als Ergebnis eine (unbekannte) reelle Zahl liefert. Bei n -vielen unabhängigen Wiederholungen ergab sich die Stichprobe

$$x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Wir nehmen an, dass diese Stichprobe Realisierung, also ein Bildpunkt

$$(x_1, \dots, x_n) = (X_1(\omega_1), \dots, X_n(\omega_n)),$$

von einfachen Zufallsvariablen X_1, \dots, X_n mit gemeinsamer Verteilung $N(\mu, \sigma^2)$ ist mit unbekanntem Parameter $\theta = \mu$. Es ist die Abbildung

$$g: \mathbb{R}^n \rightarrow \mathbb{R} \\ (x_1, \dots, x_n) \mapsto \bar{x} = \frac{x_1 + \dots + x_n}{n}$$

gegeben durch den empirischen Mittelwert aus Definition 63 ein Beispiel für einen Schätzer. Wir werden sehen (und haben bisher nicht gesagt, was diese Aussage bedeutet), dass dieses eine „gute Wahl“ für eine Schätzung des Parameters $\theta = \mu$ ist.

Ist die obige Stichprobe x Realisierung von einfachen Zufallsvariablen X_1, \dots, X_n mit gemeinsamer Verteilung $N(\mu, \sigma^2)$ mit unbekanntem Parameter $\theta = \sigma^2$, so ist die Abbildung

$$g: \mathbb{R}^n \rightarrow \mathbb{R} \\ (x_1, \dots, x_n) \mapsto s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

gegeben durch die empirische Varianz aus Definition 77 ein Beispiel für einen Schätzer. Wir werden sehen, dass dieses eine „gute Wahl“ für eine Schätzung des Parameters $\theta = \sigma^2$ ist.

Konvention 203. Ist die Verteilung f (also die Zähldichte oder die Dichtefunktion) einer Zufallsvariablen abhängig von einem Parameter $\theta \in \Theta$, so schreiben wir dafür f_θ . Da in diesem Fall auch der Wert

$$P(X \leq b) = \begin{cases} \sum_{x \leq t} f_\theta(t) & \text{falls } X \text{ endlich} \\ \int_{-\infty}^t f_\theta(x) dx & \text{falls } X \text{ stetig} \end{cases}$$

von dem Parameter θ abhängt, schreiben wir hierfür $P_\theta(X \leq t)$. Ebenso schreiben wir $E_\theta(X)$, $Var_\theta(X)$ etc.

Definition 204. Seien einfache Zufallsvariablen X_1, \dots, X_n gegeben, deren gemeinsame Verteilung von einem Parameter $\theta \in \Theta$ abhängt. Sei $g: \mathbb{R}^n \rightarrow \mathbb{R}$ ein Schätzer mit zugehöriger Stichprobenfunktion $T: \Omega^n \rightarrow \mathbb{R}$. Dann heißt g (oder T) **erwartungstreu** für θ , falls

$$E_\theta(T) = \theta.$$

Es heißt $Bias_\theta(T) := E_\theta(T) - \theta$ die **Verzerrung** von g (oder T).

Beispiel 205.

- (1) Es ist der durch $g(x_1, \dots, x_n) = \bar{x}$ gegebene Schätzer g aus Beispiel 202 erwartungstreu für den Erwartungswert $\theta = \mu$, denn nach Satz 173 gilt

$$E_\theta(\bar{X}) = E_\theta\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n} \cdot n \cdot E_\theta(X_1) = 1 \cdot \mu = \theta.$$

- (2) Für einfache Zufallsvariablen X_1, \dots, X_n mit gemeinsamer Verteilung $N(\mu, \sigma^2)$ ist der durch $g(x_1, \dots, x_n) = \bar{x}^2$ gegebene Schätzer g nicht erwartungstreu für $\theta = \mu^2$, denn nach Satz 173 gilt

$$E_\theta(\bar{X}^2) = Var_\theta(\bar{X}) + E_\theta(\bar{X})^2 = \frac{1}{n^2} \cdot n \cdot \sigma^2 + \mu^2 = \frac{\sigma^2}{n} + \mu^2,$$

es gilt also $Bias_\theta(\bar{X}^2) = \frac{\sigma^2}{n}$, was nicht im Allgemeinen gleich Null ist. (Dieser Wert ist aber Null für $n \rightarrow \infty$, was man **asymptotisch erwartungstreu** nennt.)

Idealerweise sollten sich die Schätzwerte bei steigendem Stichprobenumfang n dem tatsächlichen gesuchten Parameter ϕ annähern. Diese Bedingung wird in der folgenden Definition präzisiert.

Definition 206. Sei X_1, X_2, \dots eine Folge einfacher Zufallsvariablen, deren gemeinsame Verteilung abhängig ist von einem Parameter $\theta \in \Theta$. Man nennt eine Folge $g_1: \mathbb{R} \rightarrow \mathbb{R}, g_2: \mathbb{R}^2 \rightarrow \mathbb{R}, \dots$ von Schätzern mit zugehörigen Stichprobenfunktionen T_1, T_2, \dots **konsistent** für θ , falls für jedes $c > 0$ gilt, dass

$$\lim_{n \rightarrow \infty} P_\theta(|T_n - \theta| \geq c) = 0.$$

Bemerkung 207. Das (schwache) Gesetz der großen Zahlen 198 sagt genau, daß die durch $g(x_1, \dots, x_n) = \bar{x}$ gegebene Folge g_1, g_2, \dots von Schätzern konsistent ist für den Erwartungswert $\theta = \mu$, wobei es egal ist, welche gemeinsame Verteilung die X_i haben.

Die vorherige Bemerkung 207 liefert uns also einen „guten“ Schätzer für den Erwartungswert. Wollen wir einen anderen Parameter θ als den Erwartungswert schätzen, so ist der folgende Satz hilfreich, um die Konsistenz einer Folge von Schätzern festzustellen.

Satz 208. Sei X_1, X_2, \dots eine Folge einfacher Zufallsvariablen, deren gemeinsame Verteilung abhängig ist von einem Parameter $\theta \in \Theta$ und g_1, g_2, \dots eine Folge von Schätzern mit zugehörigen Stichprobenfunktionen T_1, T_2, \dots . Sind alle Schätzer g_n erwartungstreu für θ und gilt

$$\lim_{n \rightarrow \infty} \text{Var}_\theta(T_n) = 0,$$

so ist die Folge g_1, g_2, \dots von Schätzern konsistent für θ .

Beispiel 209. Obwohl wir dieses sowieso schon durch Bemerkung 207 wissen, zeigen wir durch den vorherigen Satz 208, dass die durch $g(x_1, \dots, x_n) = \bar{x}$ gegebene Folge g_1, g_2, \dots von Schätzern aus Beispiel 205 konsistent ist. Jedes g_n ist erwartungstreu für $\theta = \mu$ nach Beispiel 205.(1) und mit Satz 173 und Beispiel 205.(2) gilt, dass

$$\text{Var}_\theta(T_n) = \text{Var}_\theta(\bar{X}) = E_\theta(\bar{X}^2) - E_\theta(\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \frac{\sigma^2}{n},$$

was bei $n \rightarrow \infty$ gegen Null strebt. Also ist g_1, g_2, \dots eine konsistente Folge von Schätzern für den Erwartungswert.

3.2. Punktschätzungen (durch die Maximum-Likelihood Methode). Die **ML-Methode** („Maximum Likelihood Methode“) ist ein spezieller parametrischer Test, der einen ganz konkreten Wert θ_{ML} liefert. Dieser Wert ist das $\theta \in \Theta$ bei dem die beobachtete Stichprobe $x = (x_1, \dots, x_n)$ die größte Wahrscheinlichkeit hat.

Definition 210. Seien einfache Zufallsvariablen X_1, \dots, X_n gegeben, deren gemeinsame Verteilung f_θ von einem Parameter $\theta \in \Theta$ abhängig ist.

- Die Abbildung

$$\begin{aligned} L: \mathbb{R}^n \times \Theta &\rightarrow \mathbb{R} \\ (x_1, \dots, x_n, \theta) &\mapsto f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n) \end{aligned}$$

heißt **Likelihood-Funktion**.

Sei nun zusätzlich eine Stichprobe $x = (x_1, \dots, x_n)$ gegeben.

- Ein $\theta_{ML} \in \Theta$, sodass die Funktion

$$L(x_1, \dots, x_n, -): \Theta \rightarrow \mathbb{R}$$

maximal wird, heißt **ML-Schätzwert**

- Der Schätzer

$$L(-, \theta_{ML}): \mathbb{R}^n \rightarrow \mathbb{R}$$

heißt **ML-Schätzer**.

Beispiel 211 (Rückfang-Methode). Sei N eine zu schätzende Anzahl von Fischen in einem Teich. Man fängt erst M -viele Fische, markiert sie und setzt sie danach wieder in den Teich. Danach fängt man n -viele Fische. Aus der Anzahl der x_1 -vielen markierten darunter möchten man auf die Gesamtzahl N an Fischen im Teich schließen.

Ein naiver Ansatz zur Bestimmung von N wäre der folgende. Wir setzen

$$(\text{Anzahl der markierten Fische unter den gefangenen}) \cdot \frac{x_1}{n} = \frac{M}{N} \quad (\text{Anzahl der markierten Fische unter allen})$$

und bekommen $N = \frac{M \cdot n}{x_1}$.

Nun benutzen wir die Maximum-Likelihood-Methode zur Bestimmung des unbekannten Parameters $\theta := N \in \Theta$ wobei $\Theta := \{n, n+1, \dots\}$. Wir betrachten die Zufallsvariable $X :=$ „Anzahl der markierten Fische unter den n -vielen gefangenen“. Dann ist (vergleiche Beispiel 176.(10)) die Zufallsvariable $X \sim \text{Hyp}(n, M, N)$ hypergeometrisch verteilt mit Zähldichte

$$f_N(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

Die (einzige) Zufallsvariable $X_1 := X$ ist natürlich einfach. Die zugehörige Likelihood-Funktion ist

$$\begin{aligned} L: \mathbb{R} \times \Theta &\rightarrow \mathbb{R} \\ (x, N) &\mapsto f_N(x) \end{aligned}$$

und eine leichte Rechnung zeigt (dieses ist bei stetigen Zufallsvariablen durch eine Kurvendiskussion oft einfacher), dass die Funktion $L(x_1, -)$ ein Maximum bei $\theta_{ML} = \lceil \frac{M \cdot n}{x_1} \rceil$ hat. Sind also beispielsweise $M = 50$ Fische markiert und $x_1 = 10$ markierte unter den $n = 45$ gefangenen Fischen, so haben wir die Anzahl der Fische im Teich auf $N = \frac{45 \cdot 50}{10} = 225$ geschätzt.

Beispiel 212. Eine unfaire Münze wird n -mal geworfen und liefert eine Stichprobe

$$x = (x_1, \dots, x_n) \in \{0, 1\}^n$$

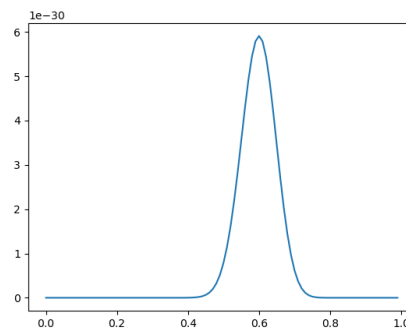
wobei 0 „Kopf“ und 1 „Zahl“ entspricht und wobei s -mal Kopf auftaucht. Diese Stichprobe sei eine Realisierung einfacher $\text{Bin}(1, p)$ -verteilter Zufallsvariablen X_1, \dots, X_n mit Zähldichten

$$f_p(x) = \begin{cases} p & \text{falls } x = 1 \\ 1 - p & \text{falls } x = 0 \\ 0 & \text{sonst,} \end{cases}$$

wobei wir den Parameter $\theta = p$ schätzen möchten, wobei $\Theta := (0, 1)$. Die zugehörige Likelihood-Funktion ist

$$\begin{aligned} L: \mathbb{R}^n \times \Theta &\rightarrow \mathbb{R} \\ (x_1, \dots, x_n, p) &\mapsto f_p(x_1) \cdot \dots \cdot f_p(x_n) \\ &\quad \theta^s \cdot (1 - \theta)^{n-s}. \end{aligned}$$

Im Beispiel $n := 100$ und $s := 60$ hat die Funktion $L(x_1, \dots, x_n, -): \Theta \rightarrow \mathbb{R}$ den Graphen



und ein Maximum an der Stelle $\theta_{ML} = 0,6$. Um dieses genau auszurechnen sei $s \notin \{0, n\}$. Dann können wir die Funktion $L(x_1, \dots, x_n, -)$ (nach θ) ableiten und erhalten

$$s \cdot \theta^{s-1} \cdot (1 - \theta)^{n-s} - \theta^s \cdot (n - s) \cdot (1 - \theta)^{n-s-1} = \left(\frac{s}{\theta} - \frac{n - s}{1 - \theta} \right) \cdot \theta^s \cdot (1 - \theta)^{n-s}$$

und dieses ist gleich Null an der Stelle $\theta = \frac{s}{n}$. Tatsächlich ist dieses auch ein Maximum und also ein ML-Schätzwert $\theta_{ML} = \frac{s}{n}$. Im obigen Zahlenbeispiel ist $\theta_{ML} = \frac{60}{100} = 0,6$.

Beispiel 213. Sei die Stichprobe $x = (x_1, \dots, x_n)$ eine Realisierung einfacher Zufallsvariablen X_1, \dots, X_n , die gemeinsam $N(\mu, \sigma^2)$ -normalverteilt sind. Wenn wir mit der ML-Methode den unbekannten Erwartungswert $\theta = \mu$ schätzen, so bekommen wir $\theta = \bar{x}$, also den empirischen Mittelwert, als ML-Schätzwert. Der zugehörige ML-Schätzer ist genau der durch $g(x_1, \dots, x_n) = \bar{x}$ gegebene Schätzer aus Beispiel 202.

Satz 214 (informelle Formulierung). Ein ML-Schätzer ist zwar selten erwartungstreu aber, unter gewissen Regularitätsannahmen, konsistent.

3.3. Konfidenzintervalle. Wir beginnen mit einem Beispiel.

Beispiel 215. Eine Apotheke erhält eine Lieferung mit $N = 100$ Pillendosen zum Weiterverkauf. Eine Apothekerin entnimmt $n = 15$ Pillendosen und überprüft diese. Es sind $m = 4$ Pillendosen fehlerhaft. Ein ML-Schätzwert für die Anzahl der fehlerhaften Pillendosen unter allen ist $M = 26$. (Vergleiche die Rückfang-Methode 211).

Wir fragen uns, wie „genau“ die Schätzung der Anzahl der fehlerhaften Pillendosen im obigen Beispiel 215 ist. Durch die Konsistenz des ML -Schätzers aus Satz 214 wissen wir nur, dass die Apothekerin bei vielen solchen Lieferungen im Durchschnitt die richtige Zahl schätzen wird.

Nun möchten wir für eine vorgegebene **Irrtumswahrscheinlichkeit** $\alpha \in (0, 1)$ ein „kleinstes“ Intervall I festlegen, sodass bei $(1 - \alpha) \cdot 100\%$ aller Schätzungen, der gesuchte Parameter $\theta \in \Theta \subseteq \mathbb{R}$ tatsächlich im Intervall I liegt. Man nennt ein solches Intervall I ein **Konfidenzintervall**. Genauer nennt man $I = [a, b]$ ein **zweiseitiges Konfidenzintervall** und $[a, \infty)$ bzw. $(-\infty, b]$ ein **einseitiges Konfidenzintervall**. Es heißt $\gamma := (1 - \alpha)$ das **Konfidenzniveau**.

Wir geben in den folgenden Sätzen konkrete Konfidenzintervalle lediglich für den Erwartungswert $\theta = \mu$ und die Varianz $\theta = \sigma^2$ normalverteilter Zufallsvariablen an.

Satz 216 (Konfidenzintervall für μ bei bekannter Varianz).

Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam normalverteilt sind mit Varianz σ^2 . Sei (x_1, \dots, x_n) eine Stichprobe mit empirischen Mittelwert \bar{x} gegeben. Sei $\alpha \in (0, 1)$ eine gewählte Irrtumswahrscheinlichkeit.

- Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma = 1 - \alpha$ für den Erwartungswert $\theta = \mu$ bei bekannter Varianz σ^2 ist

$$\left[\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right].$$

- Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma = 1 - \alpha$ für den Erwartungswert $\theta = \mu$ bei bekannter Varianz σ^2 ist

$$\left(-\infty, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right] \quad \text{bzw.} \quad \left[\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}, \infty \right).$$

Hier bezeichnet z_p das p -Quantil der Standardnormalverteilung (siehe Tabelle).

Bemerkung 217. Im Fall bekannter Varianz ist $l := 2 \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$ die Länge des Konfidenzintervalls aus dem vorherigen Satz 216. Diese hängt erstaunlicherweise nicht von den konkreten Stichprobenwerten ab. Soll eine feste gegebene Länge $L > 0$ nicht überschritten werden, ergibt sich

$$n \geq \frac{4\sigma^2 z_{1-\frac{\alpha}{2}}^2}{L^2}$$

für den Stichprobenumfang.

Beispiel 218. Gegeben sei eine Stichprobe vom Umfang $n = 10000$, die als Realisierung gemeinsam normalverteilter einfacher Zufallsvariablen betrachtet wird, mit empirischem Mittelwert $\bar{x} = 48$ und der bekannten Varianz $\sigma^2 = 400$. Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma := 0,95$ für den Erwartungswert bei bekannter Varianz ist wegen $\alpha = 0,05$, $1 - \frac{\alpha}{2} = 0,975$ und $z_{0,975} \approx 1,96$ durch

$$\left[48 - \frac{20}{\sqrt{10000}} \cdot 1,96, 48 + \frac{20}{\sqrt{10000}} \cdot 1,96 \right] \approx [47,608, 48,392]$$

gegeben. Soll man bestimmen, wie groß der Stichprobenumfang n mindestens sein muss, sodass die Länge des zweiseitigen Konfidenzintervalls zum Niveau $\gamma = 0,99$ höchstens $L := 0,1$ ist, so ergibt sich

$$n \geq \frac{4\sigma^2 z_{1-\frac{\alpha}{2}}^2}{L^2} \approx \frac{4 \cdot 400 \cdot z_{0,995}^2}{(0,1)^2} \approx 1065024.$$

Satz 219 (Konfidenzintervall für μ bei unbekannter Varianz).

Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam normalverteilt sind mit unbekannter Varianz. Sei (x_1, \dots, x_n) eine Stichprobe mit empirischen Mittelwert \bar{x} und empirischer Varianz s^2 gegeben. Sei $\alpha \in (0, 1)$ eine gewählte Irrtumswahrscheinlichkeit.

- Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma = 1 - \alpha$ für den Erwartungswert $\theta = \mu$ bei unbekannter Varianz ist

$$\left[\bar{x} - \frac{s}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x} + \frac{s}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right].$$

- Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma = 1 - \alpha$ für den Erwartungswert $\theta = \mu$ bei unbekannter Varianz ist

$$\left(-\infty, \bar{x} + \frac{s}{\sqrt{n}} t_{n-1, 1-\alpha} \right] \quad \text{bzw.} \quad \left[\bar{x} - \frac{s}{\sqrt{n}} t_{n-1, 1-\alpha}, \infty \right).$$

Hier bezeichnet $t_{n-1, p}$ das p -Quantil der t -Verteilung mit $n - 1$ Freiheitsgraden (siehe Tabelle), welches für $n > 30$ ungefähr dem p -Quantil z_p der Standardnormalverteilung entspricht.

Bemerkung 220. Im Fall unbekannter Varianz ist $l := 2 \frac{s}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}$ die Länge des Konfidenzintervalls aus dem vorherigen Satz 219. Soll eine feste gegebene Länge $L > 0$ nicht überschritten werden, ergibt sich, falls $n > 30$, ungefähr

$$n \geq \frac{4s^2 z_{1-\frac{\alpha}{2}}^2}{L^2}$$

für den Stichprobenumfang.

Beispiel 221. Gegeben sei eine Stichprobe vom Umfang $n = 16$, die als Realisierung gemeinsam normalverteilter einfacher Zufallsvariablen betrachtet wird, mit empirischem Mittelwert $\bar{x} = 37,42$ und empirischer Standardabweichung $s = 2,13$. Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma := 0,95$ für den Erwartungswert bei unbekannter Varianz ist wegen $t_{15, 0,975} \approx 2,131$ durch

$$\left[37,42 - \frac{2,13}{\sqrt{16}} \cdot 2,131, 37,42 + \frac{2,13}{\sqrt{16}} \cdot 2,131 \right] \approx [36,29, 38,55]$$

gegeben.

Satz 222 (Konfidenzintervall für die Varianz σ^2).

Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam normalverteilt sind mit unbekannter Varianz. Sei (x_1, \dots, x_n) eine Stichprobe mit empirischer Varianz s^2 gegeben. Sei $\alpha \in (0, 1)$ eine gewählte Irrtumswahrscheinlichkeit.

- Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma = 1 - \alpha$ für die Varianz $\theta = \sigma^2$ ist

$$\left[\frac{(n-1)s^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)s^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right].$$

- Das zweiseitige Konfidenzintervall zum Konfidenzniveau $\gamma = 1 - \alpha$ für die Varianz $\theta = \sigma^2$ ist

$$\left(-\infty, \frac{(n-1)s^2}{\chi_{n-1, \alpha}^2} \right] \quad \text{bzw.} \quad \left[\frac{(n-1)s^2}{\chi_{n-1, 1-\alpha}^2}, \infty \right).$$

Hier bezeichnet $\chi_{n-1, p}^2$ das p -Quantil der χ^2 -Verteilung mit $n - 1$ Freiheitsgraden (siehe Tabelle).

3.4. Hypothesentests. Hypothesentests (oder „Signifikanztests“) sind das letzte Thema in diesem Skript und von besonderer Wichtigkeit in der Pharmazie und Medizin, wenn man Entscheidungen über die Wirksamkeit eines Medikaments oder eine Therapie treffen möchte. Auch in diesem Abschnitt möchten wir einen Parameter $\theta \in \Theta \subseteq \mathbb{R}$ schätzen.

Definition 223. Sei eine disjunkte Zerlegung $\Theta = \Theta_0 \cup \Theta_1$ gegeben.

- Die Aussage, dass der gesuchte Parameter θ in der Menge Θ_0 liegt, heißt **Nullhypothese**. Man schreibt für diese Aussage auch

$$H_0 : \theta \in \Theta_0.$$

- Die Aussage, dass der gesuchte Parameter θ in der Menge Θ_1 liegt, heißt **Gegenhypothese**. Man schreibt für diese Aussage auch

$$H_1 : \theta \in \Theta_1.$$

Nun möchte man, nach dem „Aufstellen der Hypothesen“ (d.h. nach Angabe der disjunkten Zerlegung $\Theta = \Theta_0 \cup \Theta_1$) anhand des Ausgangs eines Zufallsexperiments (d.h. anhand einer gegebenen Stichprobe $x = (x_1, \dots, x_n)$) entscheiden, ob man die Nullhypothese H_0 **akzeptiert** oder ob man die Nullhypothese H_0 **ablehnt** oder **verwirft** (was gleichbedeutend damit ist, dass man die Gegenhypothese H_1 akzeptiert). Genauer bedeutet dieses, dass wir uns eine Menge $K \subseteq \mathbb{R}^n$ vorgeben, den sogenannten **Ablehnungsbereich** oder **kritischen Bereich** und H_0 genau dann ablehnen, wenn die Stichprobe x im kritischen Bereich liegt.

Definition 224. Seien X_1, \dots, X_n einfache Zufallsvariablen, deren gemeinsame Verteilung von einem Parameter $\theta \in \Theta$ abhängt. Ein **Hypothesentest** besteht aus der Angabe von

- (1) einer disjunkten Zerlegung $\Theta = \Theta_0 \cup \Theta_1$ und
- (2) einer Menge $K \subseteq \mathbb{R}^n$, dem **kritischen Bereich** oder **Ablehnungsbereich**.
(Es wird manchmal die Menge $\mathbb{R}^n \setminus K$ auch der **Annahmebereich** genannt.)

Bemerkung 225. Häufig ist der Ablehnungsbereich $K \subseteq \mathbb{R}^n$ eines Hypothesentests durch einen Schätzer $g: \mathbb{R}^n \rightarrow \mathbb{R}$ und eine Menge $I \subseteq \mathbb{R}$ als das Urbild

$$K := g^{-1}(I) := \{x \in \mathbb{R}^n \mid g(x) \in I\}$$

gegeben. In diesem Fall nennt man die zu g gehörige Stichprobenfunktion $T: \Omega^n \rightarrow \mathbb{R}$ manchmal auch „Teststatistik“.

Ist ein Hypothesentest gegeben, so unterscheidet man zwischen den folgenden „Fehlern“.

	Entscheidung für H_0 (also $x \notin K$)	Entscheidung gegen H_0 (also $x \in K$)
H_0 ist wahr (also $\theta \in \Theta_0$)		Fehler 1. Art
H_0 ist falsch (also $\theta \notin \Theta_0$)	Fehler 2. Art	

Ein **Fehler 1. Art** liegt also vor, wenn man die Nullhypothese ablehnt, obwohl sie richtig ist.

Hypothesentests sind oft so konzipiert, dass die Wahrscheinlichkeit des Eintretens eines Fehlers 1. Art durch eine obere Schranke $\alpha \in [0, 1]$ kontrolliert wird. (Die Wahrscheinlichkeit $\beta \in [0, 1]$ für einen Fehler 2. Art spielt oft keine Rolle.)

Bemerkung 226. Üblicherweise wählt man $\alpha = 0, 1$, $\alpha = 0, 05$ oder $\alpha = 0, 001$.

Da die beiden verschiedenen Fehler durch Hypothesentests oft ungleich behandelt werden, ist es entscheidend, welche Annahme als Nullhypothese und welche als Gegenhypothese formuliert wird. Hierbei ist der folgende „Slogan“ entscheidend.

„Die Nullhypothese sollte die Annahme sein, deren fälschliche Ablehnung man vermeiden möchte.“

Beispiel 227. Ein pharmazeutisches Unternehmen hat ein neues Medikament entwickelt. Es wird nur zugelassen, wenn es in weniger als 10% der Anwendungen Nebenwirkungen zeigt. Sei $\theta = p \in [0, 1] = \Theta$ die zu schätzende Wahrscheinlichkeit der Nebenwirkungen. Wir betrachten die folgenden beiden Wahlen der Hypothesen.

- (1) $\Theta_0 := [0, 1, 1]$ und $\Theta_1 := [0, 0, 1]$ oder etwas illustrativer geschrieben

$$H_0 : p \geq 0,1 \quad \text{und} \quad H_1 : p < 0,1.$$

Hier ist der Fehler 1. Art: Das Medikament ist „schädlich“, obwohl es als „unschädlich“ eingestuft worden ist.

- (2) $\Theta_0 := [0, 0, 1]$ und $\Theta_1 := [0, 1, 1]$ oder etwas illustrativer geschrieben

$$H_0 : p < 0,1 \quad \text{und} \quad H_1 : p \geq 0,1.$$

Hier ist der Fehler 1. Art: Das Medikament ist „unschädlich“, obwohl es als „schädlich“ eingestuft worden ist. (Dieses ist genau der Fehler 2. Art in (1).)

Es ist klar, dass der Fehler 1. Art in (1) gravierendere Folgen hat und unbedingt vermieden werden sollte. Daher sollten wir die Hypothesen aus (1) für einen Hypothesentest wählen.

Für einen gegebenen Hypothesentest mit Θ_0 , Θ_1 und K ist die Zahl

$$\begin{aligned} P_\theta(\text{„Fehler 1. Art“}) &= P_\theta(\{(\omega_1, \dots, \omega_n) \in \Omega^n \mid (X_1(\omega_1), \dots, X_n(\omega_n)) \in K\}) \\ &=: P_\theta(X_1 = x_1, \dots, X_n = x_n \mid (x_1, \dots, x_n \in K)) \\ &=: P_\theta((X_1, \dots, X_n) \in K) \end{aligned}$$

die Wahrscheinlichkeit für einen Fehler 1. Art bei einem gegebenen $\theta \in \Theta_0$.

Definition 228. Sei eine reelle Zahl $\alpha \in (0, 1)$, das **Signifikanzniveau** oder die **Sicherheitswahrscheinlichkeit**, gegeben. Ein Hypothesentest heißt **Hypothesentest zum Signifikanzniveau α** oder **α -Signifikanztest**, falls gilt

$$P_\theta((X_1, \dots, X_n) \in K) \leq \alpha \quad \text{für alle } \theta \in \Theta_0.$$

Bemerkung 229. Man wählt für einen Hypothesentest zum Signifikanzniveau α üblicherweise α sehr klein. Beobachtet man dann eine Stichprobe $x \in K$, so ist dieses unter der Annahme, dass H_0 stimmt, ein sehr unwahrscheinliches Ereignis. Damit kann man also ziemlich sicher sagen, dass H_0 falsch ist.

Ein Hypothesentest zum Signifikanzniveau α sagt nichts über die Wahrscheinlichkeit des Eintretens eines Fehlers 2. Art aus. Normalerweise sucht man unter allen solchen Tests denjenigen, bei dem die Wahrscheinlichkeit

$$P_\theta((X_1, \dots, X_n) \notin K) \quad \text{für } \theta \in \Theta_1$$

eines Fehlers 2. Art minimal wird. Mit anderen Worten wählt man einen Hypothesentest zum Signifikanzniveau α bei dem der Ablehnungsbereich K „maximal groß“ ist. Solche Hypothesentests nennt man **beste** Hypothesentests zum Signifikanzniveau α .

Bemerkung 230. Eine Strategie zur Konstruktion solcher besten Hypothesentests zum Signifikanzniveau α (also zur Wahl des Ablehnungsbereichs K) besteht darin, die oftmals durch eine konkrete Form gegebene Gleichung $P_\theta((X_1, \dots, X_n) \in K) = \alpha$ zu lösen.

Bemerkung 231. (Vorgehen bei der Durchführung eines Hypothesentests) Bei der praktischen „Durchführung“ eines Hypothesentests sollte man sich an die folgende Reihenfolge halten.

- (1) **Verteilungsannahmen** für X_1, \dots, X_n formulieren.
- (2) **Nullhypothese** H_0 formulieren (d.h. disjunkte Zerlegung von Θ angeben).
- (3) **Kritischen Bereich** K wählen, sodass sich ein Hypothesentest zum Signifikanzniveau α ergibt (dieses kann man beispielsweise wie in Bemerkung 225 durch Angabe eines Schätzers g und eines Intervalls I machen).
- (4) **Experiment** zum Erhalten einer Stichprobe $x = (x_1, \dots, x_n)$ durchführen und schauen, ob $x \in K$ oder $x \notin K$ (bzw. $g(x) \in I$ oder $g(x) \notin I$, falls der kritische Bereich durch einen Schätzer g und ein Intervall I gegeben ist).

Beispiel 232. Vor der Wahl des US-Präsidenten im Jahr 2000 wurde vom Wahllager der Republikaner behauptet, G.W. Bush hätte im Bundesstaat New Jersey die Unterstützung von mindestens 50% der Stimmberechtigten. Das demokratische Lager möchte dieses mit einem Hypothesentest zum Signifikanzniveau $\alpha := 0,1$ widerlegen. Es hält sich dabei an die Vorgehensweise aus Bemerkung 231. Sei $\theta = p \in (0,1)$ die (gesuchte) Wahrscheinlichkeit dafür, dass eine zufällig befragte stimmberechtigte Person für Bush stimmt. Wir betrachten die (einzige) Zufallsvariable $X = X_1$, welche die Anzahl der Bush-Stimmen zählt und treffen die (sinnvolle) Verteilungsannahme $X \sim \text{Bin}(n, p)$ (Hier ist n nicht das in diesem Kapitel benutzte, denn wir betrachten nur eine Zufallsvariable. Wir könnten aber auch alternativ n -viele $\text{Bin}(1, p)$ -verteilte Zufallsvariablen betrachten). Wir möchten ($n = 9$)-viele Personen testen. Nun formulieren wir die Hypothesen

$$H_0 : p \geq 0,5 \quad \text{und} \quad H_1 : p < 0,5.$$

Ein Fehler 1. Art wäre hier, dass Bush tatsächlich mehr als 50% der Stimmen hat, der Test jedoch anders entscheidet, was ein besonders peinliches Ergebnis für den demokratischen Tester wäre. Diesen Fehler gilt es also zu vermeiden. Daher wählen wir die Hypothesen auf diese Weise. Nun möchten wir den kritischen Bereich $K \subseteq \mathbb{R}^1$ wählen. Wir betrachten für ein $c \in \{0, \dots, 9\}$ die Zahl

$$P_p(\text{„Fehler 1. Art“}) = P_p(X < c) = \sum_{x \leq c} \binom{9}{x} p^x (1-p)^{9-x} \leq \alpha = 0,1$$

und erhalten die Werte

$c =$	0	1	2	3	4	5	6	7	8	9
$P_p(\text{„Fehler 1. Art“}) \approx$	0,001	0,02	0,09	0,254	0,5	0,746	0,91	0,98	0,998	1

womit wir den kritischen Bereich als $K := \{3, 4, \dots, 9\}$ festlegen. Nun führen wir ein Experiment durch und befragen $n = 9$ stimmberechtigte Personen. Dabei erhalten wir die Stichprobe $x_1 = 3$. Diese liegt im kritischen Bereich. Also muss die Nullhypothese H_0 verworfen werden. Dieses bedeutet, dass das demokratische Lager die Behauptung der Republikaner mit einem Hypothesentest (wie oben) zum Signifikanzniveau $\alpha = 0,1$ **nicht** widerlegen konnte.

Es gibt glücklicherweise, abhängig von der Verteilungsannahme, „konkrete Formeln“ für die jeweiligen Ablehnungsbereiche K eines besten Hypothesentests zu einem Signifikanzniveau α . Im restlichen Teil dieses Skripts werden wir diese „Formeln“ auflisten und mit Beispielen illustrieren.

3.4.1. Hypergeometrischer Test.

Satz 233. Sei $X \sim \text{Hyp}(n, M, N)$ eine hypergeometrisch verteilte Zufallsvariable mit unbekanntem Parameter $\theta = M$. Sei $\alpha \in (0,1)$ ein vorgegebenes Signifikanzniveau und $M_0 \in \{0, \dots, N\}$. Ein bester Hypothesentest zum Signifikanzniveau α zur Prüfung der Nullhypothese

$$H_0 : M \leq M_0$$

hat den Ablehnungsbereich $K := \{m_0 + 1, \dots, n\}$, wobei

$$m_0 := \max \left\{ k \leq n \mid \sum_{m=k}^n \frac{\binom{M_0}{m} \binom{N-M_0}{n-m}}{\binom{N}{n}} > \alpha \right\}.$$

Beispiel 234 (siehe Beispiel 215). Eine Apothekerin bekommt eine Lieferung von $N = 100$ Pillendosen. Ihre Hypothese (Nullhypothese) lautet, dass sich darunter höchstens $M_0 = 10$ fehlerhafte befinden. Um dieses zu testen, möchte sie $n = 15$ Pillendosen auswählen und testen. Wieviele der 15 getesteten müssen fehlerhaft sein, damit sie ihre Hypothese mit einer Sicherheitswahrscheinlichkeit von $\alpha = 0,01$ ablehnen kann? Der vorherige Satz 233 liefert $m_0 = 4$, denn

$$\sum_{m=5}^{15} \frac{\binom{10}{m} \binom{90}{15-m}}{\binom{100}{15}} \approx 0,0063 < \alpha \quad \text{und} \quad \sum_{m=4}^{15} \frac{\binom{10}{m} \binom{90}{15-m}}{\binom{100}{15}} \approx 0,04 > \alpha.$$

Also ist $K = \{5, 6, \dots, 15\}$. Beobachtet die Apothekerin also nun (in einem Experiment) unter den 15 Pillendosen 5 oder mehr fehlerhafte, lehnt sie ihre Nullhypothese $H_0 : M \leq 10$ ab und sendet die Lieferung zurück.

Wir sehen, dass diese Rechnung für großes N und n sehr kompliziert ist. Daher möchten wir wieder die Normalverteilung benutzen, deren zugehörige Tests in den nächsten Kapiteln angegeben werden.

3.4.2. Gauß-Test (Test auf Erwartungswert bei bekannter Varianz).

Satz 235 (Gauß-Test). Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam normalverteilt sind mit Varianz σ^2 und gesuchtem Parameter $\theta = \mu$. Sei $\alpha \in (0, 1)$ ein vorgegebenes Signifikanzniveau und $\mu_0 \in \mathbb{R}$. Wir betrachten jeweils die Hypothesen

$$(1) \quad H_0 : \mu = \mu_0 \quad (2) \quad H_0 : \mu \leq \mu_0 \quad (3) \quad H_0 : \mu \geq \mu_0$$

(„zweiseitiger Test“) („einseitiger Test“) („zweiseitiger Test“)

und sei $g: \mathbb{R}^n \rightarrow \mathbb{R}$ der durch $g(x) = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$ gegebene Schätzer. Dann hat ein bester Hypothesentest zum Signifikanzniveau α der jeweiligen obigen Nullhypothese jeweils den Ablehnungsbereich

$$(1) \quad K := \{x \in \mathbb{R}^n \mid |g(x)| > z_{1-\frac{\alpha}{2}}\} \quad (2) \quad K := \{x \in \mathbb{R}^n \mid g(x) > z_{1-\alpha}\}$$

und (3) $K := \{x \in \mathbb{R}^n \mid g(x) < -z_{1-\alpha}\}$.

Beispiel 236. In einer Produktion soll ein Werkstück der mittleren Länge von 17cm gefertigt werden. Die Zufallsvariable X „Länge des Werkstücks“ sei normalverteilt mit Varianz $\sigma^2 = 2,25$. Aus einer Stichprobe vom Umfang $n = 5$ wurde als empirischer Mittelwert $\bar{x} = 18,1$ bestimmt. Kann damit die Nullhypothese $H_0 : \mu = 17$ zum Signifikanzniveau $\alpha = 0,01$ abgelehnt werden? Mit $z_{1-\frac{\alpha}{2}} = z_{0,995} \approx 2,575830$ sieht man, dass

$$\frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} = \frac{18,1 - 17}{\sqrt{2,25}} \sqrt{5} \approx 1,6398 < 2,575830,$$

die Stichprobe liegt also nicht im kritischen Bereich, also kann die Nullhypothese H_0 nicht abgelehnt werden.

3.4.3. t-Test (Test auf Erwartungswert bei unbekannter Varianz).

Satz 237 (t-Test). Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam normalverteilt sind mit unbekannter Varianz und gesuchtem Parameter $\theta = \mu$. Sei $\alpha \in (0, 1)$ ein vorgegebenes Signifikanzniveau und $\mu_0 \in \mathbb{R}$. Wir betrachten jeweils die Hypothesen

$$(1) \quad H_0 : \mu = \mu_0 \quad (2) \quad H_0 : \mu \leq \mu_0 \quad (3) \quad H_0 : \mu \geq \mu_0$$

(„zweiseitiger Test“) („einseitiger Test“) („zweiseitiger Test“)

und sei $g: \mathbb{R}^n \rightarrow \mathbb{R}$ der durch $g(x) = \frac{\bar{x} - \mu_0}{s} \sqrt{n}$ gegebene Schätzer. Dann hat ein bester Hypothesentest zum Signifikanzniveau α der jeweiligen obigen Nullhypothese jeweils den Ablehnungsbereich

$$(1) \quad K := \{x \in \mathbb{R}^n \mid |g(x)| > t_{n-1, 1-\frac{\alpha}{2}}\} \quad (2) \quad K := \{x \in \mathbb{R}^n \mid g(x) > t_{n-1, 1-\alpha}\}$$

und (3) $K := \{x \in \mathbb{R}^n \mid g(x) < -t_{n-1, 1-\alpha}\}$.

Beispiel 238. Die Messung der mittleren Ozonkonzentration während der Sommermonate bei $n = 26$ Erhebungen ergab den empirischen Mittelwert $\bar{x} = 244$ bei einer empirischen Standardabweichung von $s = 5,1$. Wir nehmen an, dass die Zufallsvariable X „Ozonkonzentration“ normalverteilt ist (mit unbekannter Varianz). Der im Ozongesetz von 1955 festgelegte Warnwert liegt bei $\mu_0 := 240$. Kann dieses Ergebnis (der Messung) auf 1%-Niveau als signifikante Überschreitung des Warnwerts gewertet werden?

Es gilt $t_{25, 0,99} \approx 2,485$ und

$$\frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{244 - 240}{5,1} \sqrt{26} \approx 3,9992.$$

Die Nullhypothese $H_0 : \mu \leq \mu_0$ wird also abgelehnt (da x im kritischen Bereich K liegt) und das Messergebnis kann als signifikante Überschreitung des Warnwerts auf 1%-Niveau gewertet werden.

3.4.4. χ^2 -Test (Test auf Varianz).

Satz 239 (t-Test). Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam normalverteilt sind mit unbekanntem Erwartungswert und gesuchtem Parameter $\theta = \sigma^2$. Sei $\alpha \in (0, 1)$ ein vorgegebenes Signifikanzniveau und $\sigma_0^2 \in \mathbb{R}$. Wir betrachten jeweils die Hypothesen

$$(1) \quad H_0 : \sigma^2 = \sigma_0^2 \quad (2) \quad H_0 : \sigma^2 \leq \sigma_0^2 \quad (3) \quad H_0 : \sigma^2 \geq \sigma_0^2$$

(„zweiseitiger Test“) („einseitiger Test“) („zweiseitiger Test“)

und sei $g: \mathbb{R}^n \rightarrow \mathbb{R}$ der durch $g(x) = \frac{(n-1)s^2}{\sigma_0^2}$ gegebene Schätzer. Dann hat ein bester Hypothesentest zum Signifikanzniveau α der jeweiligen obigen Nullhypothese jeweils den Ablehnungsbereich

$$(1) \quad K := \{x \in \mathbb{R}^n \mid g(x) > \chi_{n-1, 1-\frac{\alpha}{2}}^2 \text{ oder } \chi_{n-1, \frac{\alpha}{2}}^2 > g(x)\}$$

$$(2) \quad K := \{x \in \mathbb{R}^n \mid g(x) > \chi_{n-1, 1-\alpha}^2\}$$

$$\text{und (3) } K := \{x \in \mathbb{R}^n \mid g(x) < \chi_{n-1, \alpha}^2\}.$$

Beispiel 240. Eine Stichprobe x vom Umfang $n = 51$ habe die empirische Varianz $s^2 = 12$ und wir nehmen an, dass die Stichprobe Realisierung von einfachen Zufallsvariablen ist, die gemeinsam normalverteilt sind. Kann die Nullhypothese

$$H_0 : \sigma^2 \leq 10$$

zum Signifikanzniveau $\alpha = 5\%$ abgelehnt werden? Es ist $\chi_{50, 0,95}^2 \approx 67,50$ und

$$\frac{(n-1)s^2}{10} = \frac{50 \cdot 12}{10} = 60.$$

Da die Stichprobe nicht im Ablehnungsbereich K liegt, kann die obige Nullhypothese zum Signifikanzniveau $\alpha = 5\%$ nicht abgelehnt werden.

3.4.5. Approximativer Binomialtest. Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam $\text{Bin}(1, p)$ -verteilt sind, wobei $\theta = p \in (0, 1)$ der gesuchte Parameter ist. Es ist $S_n = X_1 + \dots + X_n \sim \text{Bin}(n, p)$ und nach dem zentralen Grenzwertsatz 191 gilt ungefähr

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \approx N\left(p, \frac{p(1-p)}{n}\right).$$

Satz 241 (Approximativer Binomialtest). Seien X_1, \dots, X_n einfache Zufallsvariablen, die gemeinsam $\text{Bin}(1, p)$ -verteilt sind mit gesuchtem Parameter $\theta = p$. Sei $\alpha \in (0, 1)$ ein vorgegebenes Signifikanzniveau und $p_0 \in (0, 1)$. (Da wir durch eine Normalverteilung approximieren wollen, ist eine „Faustregel“, dass $np_0(1-p_0) > 9$ gelten sollte.) Wir betrachten jeweils die Hypothesen

$$(1) \quad H_0 : p = p_0 \quad (2) \quad H_0 : p \leq p_0 \quad (3) \quad H_0 : p \geq p_0$$

(„zweiseitiger Test“) („einseitiger Test“) („zweiseitiger Test“)

und sei $g: \mathbb{R}^n \rightarrow \mathbb{R}$ der durch $g(x) = \frac{\bar{x} - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$ gegebene Schätzer. Dann hat ein bester Hypothesentest zum Signifikanzniveau α der jeweiligen obigen Nullhypothese (ungefähr) jeweils den Ablehnungsbereich

$$(1) \quad K := \{x \in \mathbb{R}^n \mid |g(x)| > z_{1-\frac{\alpha}{2}}\} \quad (2) \quad K := \{x \in \mathbb{R}^n \mid g(x) > z_{1-\alpha}\}$$

$$\text{und (3) } K := \{x \in \mathbb{R}^n \mid g(x) < -z_{1-\alpha}\}.$$

Beispiel 242. (Wichtiges Medikamentenbeispiel) Ein Medikament besitze die Heilungswahrscheinlichkeit von 0,6. Ein neuentwickeltes Medikament soll nur dann auf den Markt gebracht werden, wenn es statistisch nachgewiesen ist (auf 5%-Niveau), dass die Heilungswahrscheinlichkeit des neuen Medikaments größer als $p_0 = 0,65$ ist.

(1) Das neue Medikament wurde $n = 100$ Personen verabreicht und davon wurden 68 geheilt. Kommt das neue Medikament auf den Markt? Es gilt

$$\frac{\bar{x} - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} = \frac{0,68 - 0,65}{\sqrt{0,65(1-0,65)}} \sqrt{100} \approx 0,62897$$

und $z_{0,95} \approx 1,64485$. Damit wird die Nullhypothese $H_0 : p \leq p_0$ nicht abgelehnt und das neue Medikament kommt nicht auf den Markt.

- (2) Das neue Medikament werde 1000 Personen verabreicht. Wieviele dieser Personen müssten mindestens geheilt werden, damit man die Nullhypothese $H_0 : p \leq p_0$ zum Signifikanzniveau $\alpha = 1\%$ ablehnen kann (und das Medikament somit auf den Markt kommt)?

Umformen der Ungleichung

$$\frac{\bar{x} - 0,65}{\sqrt{0,65(1 - 0,65)}} \sqrt{1000} > z_{0,99} \approx 2,326350$$

liefert $\bar{x} > 0,68509$, d.h. es müssten mindestens 686 Personen geheilt werden, damit man die obige Nullhypothese zum obigen Signifikanzniveau ablehnen kann.

3.4.6. Unabhängigkeits und Homogenitätstests. Es soll nur darauf hingewiesen werden, dass es einen Hypothesentest zur Prüfung der Unabhängigkeit zweier Zufallsvariablen X und Y gibt, der χ^2 -**Unabhängigkeitstest** genannt wird. Außerdem gibt es einen Hypothesentest zur Prüfung, ob unabhängige Zufallsvariablen X_1, \dots, X_n die gleiche Verteilung haben, der χ^2 -**Homogenitätstest** genannt wird.

Verteilungstabellen

1 Standardnormalverteilung

Tabelliert sind die Werte der Verteilungsfunktion $\Phi(z) = P(Z \leq z)$ für $z \geq 0$.

Ablesebeispiel: $\Phi(1.75) = 0.9599$

Funktionswerte für negative Argumente: $\Phi(-z) = 1 - \Phi(z)$

Die z -Quantile ergeben sich genau umgekehrt.

Beispielsweise ist $z(0.9599) = 1.75$ und $z(0.9750) = 1.96$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

2 Students t -Verteilung

Tabelliert sind die Quantile für n Freiheitsgrade.

Für das Quantil $t_{1-\alpha}(n)$ gilt $F(t_{1-\alpha}(n)) = 1 - \alpha$.

Links vom Quantil $t_{1-\alpha}(n)$ liegt die Wahrscheinlichkeitsmasse $1 - \alpha$.

Ablesebeispiel: $t_{0.99}(20) = 2.528$

Die Quantile für $0 < 1 - \alpha < 0.5$ erhält man aus $t_{\alpha}(n) = -t_{1-\alpha}(n)$

Approximation für $n > 30$:

$$t_{\alpha}(n) \approx z_{\alpha} \quad (z_{\alpha} \text{ ist das } (\alpha)\text{-Quantil der Standardnormalverteilung})$$

n	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.2887	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.2767	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.2707	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.2576	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.2573	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.2571	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.2569	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.2567	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.2566	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.2564	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.2563	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.2562	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.2561	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.2560	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.2559	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.2558	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.2557	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
∞	0.2533	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	3.0903	3.2906

3 χ^2 -Verteilung

Tabelliert sind die Quantile für n Freiheitsgrade.

Für das Quantil $\chi^2_{1-\alpha}(n)$ gilt $F(\chi^2_{1-\alpha}(n)) = 1 - \alpha$.

Links vom Quantil $\chi^2_{1-\alpha}(n)$ liegt die Wahrscheinlichkeitsmasse $1 - \alpha$.

Ablesebeispiel: $\chi^2_{0.95}(10) = 18.307$

Approximation für $n > 30$:

$$\chi^2_{\alpha}(n) \approx \frac{1}{2}(z_{\alpha} + \sqrt{2n-1})^2 \quad (z_{\alpha} \text{ ist das } \alpha\text{-Quantil der Standardnormalverteilung})$$

n	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99
1	0.0002	0.0010	0.0039	0.0158	0.4549	2.7055	3.8415	5.0239	6.6349
2	0.0201	0.0506	0.1026	0.2107	1.3863	4.6052	5.9915	7.3778	9.2103
3	0.1148	0.2158	0.3518	0.5844	2.3660	6.2514	7.8147	9.3484	11.345
4	0.2971	0.4844	0.7107	1.0636	3.3567	7.7794	9.4877	11.143	13.277
5	0.5543	0.8312	1.1455	1.6103	4.3515	9.2364	11.070	12.833	15.086
6	0.8721	1.2373	1.6354	2.2041	5.3481	10.645	12.592	14.449	16.812
7	1.2390	1.6899	2.1674	2.8331	6.3458	12.017	14.067	16.013	18.475
8	1.6465	2.1797	2.7326	3.4895	7.3441	13.362	15.507	17.535	20.090
9	2.0879	2.7004	3.3251	4.1682	8.3428	14.684	16.919	19.023	21.666
10	2.5582	3.2470	3.9403	4.8652	9.3418	15.987	18.307	20.483	23.209
11	3.0535	3.8157	4.5748	5.5778	10.341	17.275	19.675	21.920	24.725
12	3.5706	4.4038	5.2260	6.3038	11.340	18.549	21.026	23.337	26.217
13	4.1069	5.0088	5.8919	7.0415	12.340	19.812	22.362	24.736	27.688
14	4.6604	5.6287	6.5706	7.7895	13.339	21.064	23.685	26.119	29.141
15	5.2293	6.2621	7.2609	8.5468	14.339	22.307	24.996	27.488	30.578
16	5.8122	6.9077	7.9616	9.3122	15.338	23.542	26.296	28.845	32.000
17	6.4078	7.5642	8.6718	10.085	16.338	24.769	27.587	30.191	33.409
18	7.0149	8.2307	9.3905	10.865	17.338	25.989	28.869	31.526	34.805
19	7.6327	8.9065	10.117	11.651	18.338	27.204	30.144	32.852	36.191
20	8.2604	9.5908	10.851	12.443	19.337	28.412	31.410	34.170	37.566
21	8.8972	10.283	11.591	13.240	20.337	29.615	32.671	35.479	38.932
22	9.5425	10.982	12.338	14.041	21.337	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	22.337	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	23.337	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	24.337	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	25.336	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	26.336	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	27.336	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	28.336	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892

Literatur

- [1] G. Dolzmann, *Skript zur Vorlesung Mathematische und statistische Methoden für Pharmazeuten* Regensburg, 2017.
- [2] D. Schiefeneder, *Skript zur Vorlesung Mathematische und statistische Methoden für Pharmazeuten* Universität Regensburg.
- [3] M. Wilke, *Skript zur Vorlesung Mathematische und statistische Methoden für Pharmazeuten* Universität Regensburg, 2016.
- [4] Wikipedia, *Histogramm* — Wikipedia, Die freie Enzyklopädie <https://de.wikipedia.org/w/index.php?title=Histogramm>, 2020.
- [5] Wikipedia, *Korrelationskoeffizient* — Wikipedia, Die freie Enzyklopädie <https://de.wikipedia.org/w/index.php?title=Korrelationskoeffizient>, 2020.
- [6] Wikipedia, *Binomialverteilung* — Wikipedia, Die freie Enzyklopädie <https://de.wikipedia.org/w/index.php?title=Binomialverteilung>, 2020.
- [7] Wikipedia, *Normalverteilung* — Wikipedia, Die freie Enzyklopädie <https://de.wikipedia.org/w/index.php?title=Normalverteilung>, 2020.

FAKULTÄT FÜR MATHEMATIK, UNIVERSITÄT REGENSBURG, 93040 REGENSBURG, GERMANY
E-mail address: `florian.strunk@ur.de`