bio
biologie

Matthias Rudolf
Wiltrud Kuhlisch

Biostatistik

Eine Einführung für Biowissenschaftler

Studentengetestet!

WWW.PEARSON-STUDIUM.DE
CWS
COMPANION WEBSITE

PEARSON
Studium

Wahrscheinlichkeitstheorie

3

ÜBERBLICK

In diesem Kapitel werden zufällige Ereignisse und die Wahrscheinlichkeit zufälliger Ereignisse eingeführt sowie Rechenregeln für Wahrscheinlichkeiten anhand von Beispielen veranschaulicht. Die eingeführten Begriffe werden im Grundmodell der Wahrscheinlichkeit zusammengefasst.

Im Mittelpunkt des zweiten Teils dieses Kapitels steht die Beschreibung zufälliger Ereignisse durch Zufallsvariablen und ihre Verteilung. Insbesondere werden die in der Teststatistik (siehe Kapitel 5 bis 8) benötigten Grundbegriffe und Testverteilungen behandelt.

3.1 Grundmodell der Wahrscheinlichkeitstheorie

Während die statistischen Methoden in Kapitel 2 die Beschreibung der erhobenen Daten zum Ziel hatten, liefert die Wahrscheinlichkeitstheorie Modelle für die Gesetzmäßigkeiten beim Eintreten zufälliger Erscheinungen. Umgangssprachlich versteht man unter zufälligen Erscheinungen unerwartete, unkontrollierbare, scheinbar planlose Vorkommnisse, die bemerkt oder unbemerkt unser gesamtes Leben begleiten. In der Biologie äußert sich das Wirken des Zufalls z.B. in Form von phänotypischen Unterschieden zwischen Individuen gleicher Art oder in spontan auftretenden Mutationen. Für eine mathematische Modellierung des Zufalls muss der Begriff eines Zufallsexperiments und der eines zufälligen Ereignisses genauer definiert werden. Im Grundmodell der Wahrscheinlichkeitstheorie wird die Wahrscheinlichkeit als ein Maß definiert, das die Chance für das Eintreten eines zufälligen Ereignisses mit einer Zahl bewertet.

3.1.1 Zufällige Ereignisse und deren Verknüpfung

Grundlage für die folgenden Betrachtungen ist die Definition zufälliger Versuche.

Definition

Man bezeichnet einen Vorgang, dessen Ergebnis im Rahmen bestimmter Möglichkeiten ungewiss ist und der sich bei Beibehaltung der wesentlichen Bedingungen zumindest gedanklich beliebig oft wiederholen lässt, als zufälligen Versuch. Die möglichen einander ausschließenden Ergebnisse eines solchen zufälligen Versuchs werden zu einer Menge Ω, dem Ereignisraum, zusammengefasst.

Ein einfaches Beispiel für einen zufälligen Versuch ist das einmalige Würfeln mit einem Würfel. Der Ereignisraum besteht in diesem Fall aus der Menge der Augenzahlen 1 bis 6:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Würfelt man mit zwei Würfeln, so besteht der Ereignisraum Ω aus 36 geordneten Paaren von Augenzahlen des ersten und des zweiten Würfels (i, j), $i, j = 1,\dots,6$. Solche einfach anmutenden Versuche wie z.B. Würfelexperimente eignen sich gut zur Erforschung des Zufalls, da sie überschaubar sind und die Grundlagen der Wahrscheinlichkeitsrechnung daran besonders anschaulich erläutert werden können, wie etwa der Begriff eines zufälligen Ereignisses.

Definition

Ein Ereignis A ist eine Zusammenfassung von Versuchsergebnissen aus Ω zu einer Teilmenge von Ω. Man sagt, das Ereignis A ist eingetreten, wenn im Ergebnis des Versuchs eines der Elemente von A beobachtet worden ist.

Als Beispiel betrachten wir wieder den Wurf mit einem Würfel. Für diesen Versuch bezeichne $A = \{2,3,4\}$ das Ereignis, eine der Augenzahlen 2, 3 oder 4 zu würfeln, und $B = \{3,4,5\}$ das Ereignis, eine der Augenzahlen 3, 4 oder 5 zu würfeln.

Durch folgende Verknüpfungen der Teilmengen lassen sich aus beliebigen Ereignissen A und B neue Ereignisse bilden:

- Wenn im Ergebnis des betrachteten zufälligen Versuchs beide Ereignisse A und B gleichzeitig eintreten, dann nennt man dieses Ereignis Durchschnitt von A und B und bezeichnet es mit $A \cap B$.
  - Für die oben beschriebenen Ereignisse A und B ist $A \cap B = \{3,4\}$ und dieses Ereignis tritt ein, wenn eine 3 oder 4 gewürfelt wird.
- Wenn Ereignis A oder Ereignis B eintritt, nennt man das Ereignis Vereinigung der Ereignisse A und B und bezeichnet es mit $A \cup B$.
  - Im genannten Beispiel gilt $A \cup B = \{2,3,4,5\}$, d.h. dieses Ereignis tritt ein, wenn eine der Augenzahlen 2, 3, 4 oder 5 gewürfelt wird.
- Das Ereignis $A \setminus B$ tritt ein, wenn Ereignis A, aber nicht Ereignis B eintritt.
  - Im speziellen Beispiel erhält man $A \setminus B = \{2\}$.
- Das Komplementärereignis von A, bezeichnet mit $\bar{A}$, tritt ein, wenn Ereignis A im Ergebnis des Versuchs nicht eintritt, d.h. $\bar{A} = \Omega \setminus A$.
  - Das Komplementärereignis von A im Würfelversuch tritt somit ein, wenn eine der Augenzahlen 1, 5 oder 6 fällt.

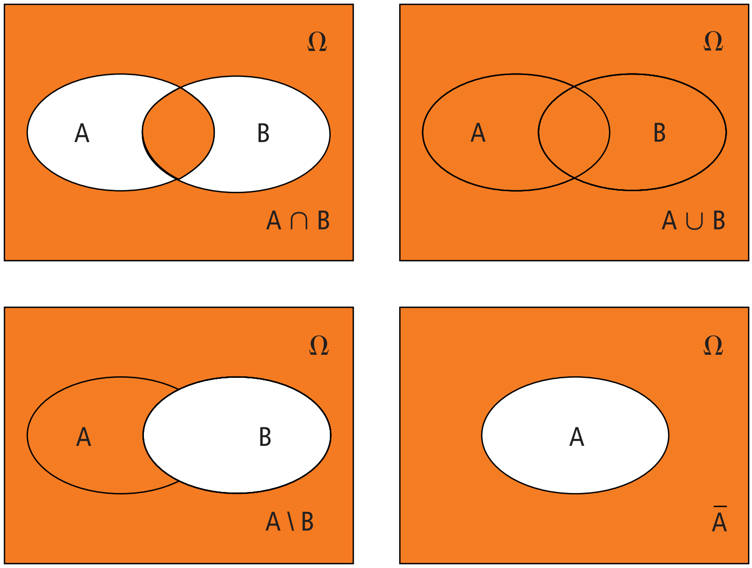

Abbildung 3.1: Veranschaulichung der Ereignisse $A \cap B$, $A \cup B$, $A \setminus B$ und $\overline{A}$ in einem Venn-Diagramm.

- Die Grundmenge Ω bezeichnet das sichere Ereignis und die leere Menge $\varnothing = \overline{\Omega}$ das unmögliche Ereignis. Wenn gilt $A \cap B = \varnothing$, heißen die Ereignisse A und B unvereinbar.
  - Im Beispiel sind die Ereignisse A und B nicht unvereinbar, da $A \cap B = \{3,4\}$ und damit nicht leer ist. Unvereinbare Ereignisse sind z. B. die Ereignisse $G = \{2,4,6\}$ (Würfeln einer geraden Zahl) und $U = \{1,3,5\}$ (Würfeln einer ungeraden Zahl).

Definition

In einem Ereignisfeld $\mathbb{F}$ fasst man alle Ereignisse zusammen, die mit Hilfe der genannten Verknüpfungen von Teilmengen aus Ω gebildet werden können und für die folgende Eigenschaften erfüllt sind:

1. Das sichere Ereignis Ω liegt in $\mathbb{F}$.
2. Für jedes Ereignis A gehört auch das Komplementärereignis $\overline{A}$ zu $\mathbb{F}$.
3. Mit zwei Ereignissen A und B liegt auch die Vereinigung $A \cup B$ in $\mathbb{F}$. Falls $\mathbb{F}$ unendlich viele Ereignisse A_i, $i = 1,2,\ldots$ enthält, dann gehört auch die Vereinigung dieser Mengen $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \ldots$ zu $\mathbb{F}$.

Im einfachen Würfelversuch besteht $\mathbb{F}$ aus allen 64 Teilmengen, die aus Ω gebildet werden können, einschließlich der leeren Menge und der Grundmenge Ω:

$$\mathbb{F} = \{\varnothing, \{1\}, \{2\}, \ldots, \{6\}, \{1,2\}, \ldots \{6,6\}, \{1,2,3\}, \ldots, \{4,5,6\}, \ldots, \{1,2,3,4,5,6\}\} .$$

3.1.2 Klassische Definition der Wahrscheinlichkeit

Für jedes Ereignis aus einem Ereignisfeld $\mathbb{F}$ soll nun eine Wahrscheinlichkeit eingeführt werden als eine Zahl, die die Chance für das Eintreten des Ereignisses beschreibt. Zur Vereinfachung sollen zunächst nur Versuche mit endlich vielen gleichmöglichen Versuchsausgängen betrachtet werden, wie etwa Würfelversuche.

Die Grundmenge Ω enthalte endlich viele gleichmögliche Ereignisse

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}.$$

Für die Wahrscheinlichkeit jedes Elementarereignisses definiert man:

$$\mathrm{P}(\omega_\mathrm{i}) = \frac{1}{n}, \; i = 1, \ldots, n.$$

Für beliebige Ereignisse A aus $\mathbb{F}$ ergibt sich die Wahrscheinlichkeit von A als Quotient der Anzahl der interessierenden Fälle, in denen das Ereignis A eintritt, und der Zahl der möglichen Fälle.

Definition

$$P(A) = \frac{\text{Anzahl der interessierenden Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{\text{Anzahl der Elemente von A}}{\text{Anzahl der Elemente von } \Omega} \tag{3.1}$$

- $P(A)$: Wahrscheinlichkeit des Ereignisses A nach klassischer Definition
- A: Beliebiges Ereignis aus $\mathbb{F}$
- $\mathbb{F}$: Ereignisraum
- Ω: Grundmenge

Die so eingeführte Wahrscheinlichkeit ist eine Zahl, die offensichtlich Eigenschaften einer relativen Häufigkeit für das Eintreten von A in n Versuchen aufweist und somit als ein mathematisches Modell für eine relative Häufigkeit aufgefasst werden kann. Eines der wichtigsten Gesetze der Wahrscheinlichkeitstheorie, das im 17. Jahrhundert von Jacob Bernoulli[1] erkannte Gesetz der großen Zahlen, sagt aus, dass sich in einer sehr langen Reihe von Wiederholungen eines Experiments die relative Häufigkeit des Auftretens eines Ereignisses A mit wachsender Zahl der Beobachtungen stabilisiert. Der Grenzwert ist die Wahrscheinlichkeit $P(A)$.

Aus Definition (3.1) ergibt sich, dass

- die Wahrscheinlichkeit für das sichere Ereignis Ω gleich eins ist, $P(\Omega) = \frac{n}{n} = 1$,
- die Wahrscheinlichkeit für das komplementäre Ereignis von A, wobei A eine Menge mit m Elementen sei, gleich $1 - P(A)$ ist: $P(\bar{A}) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(A)$, und dass
- die Wahrscheinlichkeit für das Eintreten von mindestens einem von zwei unvereinbaren Ereignissen A und B gleich der Summe der Wahrscheinlichkeiten beider Ereignisse ist, d.h. es gilt $P(A \cup B) = P(A) + P(B)$.

Beispiele

1. Beim Zahlenlotto 6 aus 49 gibt es $\binom{49}{6} = 13983816$ mögliche Zahlenreihen. Die Wahrscheinlichkeit für einen Sechser im Zahlenlotto beträgt somit $\mathrm{P(Sechser)} = \frac{1}{13983816}$ und ist also ungefähr $7.151 \cdot 10^{-8}$.

2. Für die Wahrscheinlichkeit, beim Würfeln mit zwei Würfeln eine Augensumme kleiner oder gleich 5 zu erhalten, gilt

$$\mathrm{P(Augensumme} \leq 5) = \frac{10}{36} = 0.2\bar{7}.$$

3. Die Wahrscheinlichkeit, beim Wurf mit vier Würfeln mindestens eine Sechs zu erhalten, beträgt

$$\mathrm{P(Anzahl\ der\ Sechsen} \geq 1) = 1 - \mathrm{P(Anzahl\ der\ Sechsen} = 0) = 1 - \frac{5^4}{6^4} = 0.5177.$$

3.1.3 Axiomatische Definition der Wahrscheinlichkeit

Der klassische Wahrscheinlichkeitsbegriff lässt sich nur auf Versuche mit endlich vielen, gleich möglichen Ausgängen anwenden. Wie kann man einen allgemeingültigen Wahrscheinlichkeitsbegriff ohne diese Einschränkungen einführen? Dieses Problem wurde von dem Mathematiker Kolmogorov[2] gelöst, der 1933 eine axiomatisch aufgebaute Wahrscheinlichkeitstheorie entwickelte. Sein Wahrscheinlichkeitsmodell wird wie folgt definiert, man geht wieder von einem Ereignisfeld $\mathbb{F}$ aus:

1 Jacob Bernoulli, 1654–1705, französischer Mathematiker.

2 Andrei Nikolajevich Kolmogorov, 1903–1987, russischer Mathematiker.

Definition Eine Abbildung P, die jedem Element A aus $\mathbb{F}$ eine reelle Zahl zuordnet, heißt Wahrscheinlichkeit oder Wahrscheinlichkeitsmaß, falls gilt:

1. $0 \leq P(A) \leq 1$,
2. $P(\Omega) = 1$,
3. **a)** für Ereignisräume $\mathbb{F}$ mit endlich vielen Ereignissen und zwei unvereinbare Ereignisse A und B aus $\mathbb{F}$ gilt $P(A \bigcup B) = P(A) + P(B)$,

   b) für Ereignisräume $\mathbb{F}$ mit unendlich vielen Ereignissen fordert man zusätzlich $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, falls $A_i, i \geq 1$, paarweise unvereinbare Ereignisse aus $\mathbb{F}$ sind.

Falls Ω nur endlich viele Elemente ω_i $(i = 1, \ldots, n)$ enthält, dann erfüllt die klassische Wahrscheinlichkeit nach Formel (3.1) die Bedingungen, die in der axiomatischen Definition an eine Wahrscheinlichkeit gestellt werden. Im Grundmodell der Wahrscheinlichkeitstheorie gibt man also die Grundmenge Ω, das Ereignisfeld $\mathbb{F}$ und die Wahrscheinlichkeiten $P(A)$, $A \in \mathbb{F}$, an.

3.1.4 Rechnen mit Wahrscheinlichkeiten

Aus den drei Axiomen der Definition einer Wahrscheinlichkeit ergeben sich die folgenden Rechenregeln für Wahrscheinlichkeiten.

$$\begin{aligned} P(\bar{A}) &= 1 - P(A), \\ P(\varnothing) &= 0, \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned} \tag{3.2}$$

In vielen Versuchen geht man von der Annahme aus, dass ein Ereignis B bereits eingetreten ist oder garantiert eintreten wird. Man interessiert sich unter dieser Vor- oder Teilinformation für die Wahrscheinlichkeit dafür, dass das interessierende Ereignis A eintritt. In der mathematischen Genetik betrachtet man beispielsweise Übergangswahrscheinlichkeiten, die angeben, mit welcher Wahrscheinlichkeit ein Nachkomme einen bestimmten Genotyp aufweist (Ereignis A), wenn der Genotyp der Eltern bekannt ist (Ereignis B).

Eine bedingte Wahrscheinlichkeit wird mit $P_B(A)$ bezeichnet. Sie ist für $A, B \in \mathbb{F}$, $P(B) > 0$ folgendermaßen definiert:

Formel

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \tag{3.3}$$

- $P_B(A)$: Bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B
- $A \cap B$: Durchschnitt der Ereignisse A und B

Für ein fest gewähltes Ereignis B hat die bedingte Wahrscheinlichkeit $P_B(A)$ die Eigenschaften 1 bis 3 einer Wahrscheinlichkeit. In den Anwendungen wird diese Gleichung oft zur Berechnung von $P(A \cap B)$ verwendet. Man erhält aus (3.3) den allgemeinen Multiplikationssatz:

$$P(A \cap B) = P_B(A) \cdot P(B). \tag{3.4}$$

Das folgende Beispiel kann auf viele ähnliche Probleme in der biologischen Forschung übertragen werden.

Beispiel

Man betrachtet zwei in der Natur vorkommende Subspezies von Schnecken, wobei 40 Prozent der Gesamtpopulation auf die erste Art und 60 Prozent der Gesamtpopulation auf die zweite Art entfallen. Es wurde ein Zuordnungsverfahren entwickelt, das aufgrund von Abmessungen wie Gehäusebreite, Gehäusehöhe und Mündungsbreite eine Zuordnung einer Schnecke, deren Herkunft nicht genau bekannt ist, zu einer der beiden Arten mit einer bestimmten Sicherheit erlaubt. Von dem Verfahren sei bekannt, dass die bedingte Wahrscheinlichkeit einer Zuordnung zur ersten Art, wenn die Schnecke tatsächlich zur ersten Art gehört, 96 Prozent beträgt, und die bedingte Wahrscheinlichkeit der Zuordnung zur zweiten Art, wenn die Schnecke wirklich zur zweiten Art gehört, 94 Prozent, d. h. 4 Prozent der Individuen erster Art und 6 Prozent der Individuen zweiter Art werden bei diesem Verfahren im Mittel fälschlicherweise der jeweils anderen Art zugeordnet. Man interessiert sich nach Anwendung des Verfahrens für die Wahrscheinlichkeit, dass ein nach der Klassifikation einer bestimmten Art zugeordnetes Individuum auch tatsächlich zu dieser Art gehört.

Es bezeichnen B_1 bzw. B_2 die Ereignisse, dass ein zufällig ausgewähltes Tier zur ersten oder zweiten Art gehört, mit den Wahrscheinlichkeiten

$$P(B_1) = 0.4 \text{ bzw. } P(B_2) = 0.6.$$

A sei das Ereignis der Zuordnung zur ersten Art. In diesem Beispiel ist

$$P_{B_1}(A) = 0.96 \text{ und } P_{B_2}(\bar{A}) = 0.94.$$

Mit der Definition der bedingten Wahrscheinlichkeit erhält man für die bedingte Wahrscheinlichkeit der Zugehörigkeit zur ersten Art für ein Individuum, das durch das Verfahren dieser Art zugeordnet wurde (Bedingung A),

$$P_A(B_1) = \frac{P(A \cap B_1)}{P(A)}.$$

Das Ereignis A kann in die sich ausschließenden Ereignisse $A \cap B_1$ und $A \cap B_2$ zerlegt werden. Daraus folgt mit Hilfe des Multiplikationssatzes

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) \\ &= P_{B_1}(A) \cdot P(B_1) + P_{B_2}(A) \cdot P(B_2) \end{aligned}$$

und

$$\begin{aligned} P_A(B_1) &= \frac{P(A \cap B_1)}{P(A)} = \frac{P_{B_1}(A) \cdot P(B_1)}{P_{B_1}(A) \cdot P(B_1) + P_{B_2}(A) \cdot P(B_2)} \\ &= \frac{0.96 \cdot 0.4}{0.96 \cdot 0.4 + (1 - 0.94) \cdot 0.6} = 0.914. \end{aligned}$$

Etwa 91 Prozent der bei diesem Verfahren zur ersten Art klassifizierten Individuen sind demnach richtig zugeordnet worden. Hätte man nur die Kenntnis der a-priori-Wahrscheinlichkeit $P(B_1) = 0.4$, würde man eine Schnecke mit 40 Prozent Wahrscheinlichkeit der ersten Art zuordnen. Mit Berücksichtigung der Gehäuseabmessungen kann man a posteriori, also nach Kenntnis der Zusatzinformation, eine Zuordnungswahrscheinlichkeit von 91 Prozent erzielen. Analog kann man die a-posteriori-Wahrscheinlichkeit für die zweite Art bestimmen.

Diese Idee und die in diesem Beispiel hergeleiteten Formeln lassen sich auf den Fall von mehr als zwei Teilpopulationen verallgemeinern. Man erhält die Formel der totalen Wahrscheinlichkeit und die Bayessche Formel:

Formel der totalen Wahrscheinlichkeit

Es sei möglich, den Grundraum Ω in einander ausschließende Ereignisse B_i zu zerlegen: $\Omega = B_1 \cup B_2 \cup \ldots \cup B_n$ mit $P(B_i) > 0, \quad B_i \cap B_j = \varnothing$ für $i \neq j, \quad i, j = 1, \ldots, n$.

Die Wahrscheinlichkeiten $P_{B_i}(A)$ und $P(B_i)$ $(i = 1, \ldots, n)$ seien bekannt.

Dann gilt für die Wahrscheinlichkeit von A

$$P(A) = P_{B_1}(A) \cdot P(B_1) + P_{B_2}(A) \cdot P(B_2) + \ldots + P_{B_n}(A) \cdot P(B_n). \tag{3.5}$$

Kann man also eine Population in mehrere Teilpopulationen (Unterarten, Altersklassen usw.) zerlegen und lässt sich die Wahrscheinlichkeit eines Ereignisses A für jede dieser Teilpopulationen bestimmen, dann erhält man mit Hilfe der Formel der totalen Wahrscheinlichkeit die Wahrscheinlichkeit des Ereignisses A in der Gesamtpopulation.

Bayessche Formel[3]

Wie in der Formel der totalen Wahrscheinlichkeit sei Ω zerlegt in sich ausschließende Ereignisse B_i. Falls $P(A) > 0$, so gilt für $k = 1, \dots, n$

$$P_A(B_k) = \frac{P_{B_k}(A) \cdot P(B_k)}{P_{B_1}(A) \cdot P(B_1) + P_{B_2}(A) \cdot P(B_2) + \dots + P_{B_n}(A) \cdot P(B_n)}. \tag{3.6}$$

Die Bayessche Formel gestattet die Berechnung der a-posteriori-Wahrscheinlichkeiten $P_A(B_k)$ von den Ereignissen B_k unter einer zusätzlichen Information A.

Unabhängige Ereignisse

Definition Wenn für die Ereignisse $A, B \in \mathbb{F}$, $P(B) > 0$, die Beziehung $P_B(A) = P(A)$ gilt, dann heißen die Ereignisse A und B stochastisch unabhängig. Im Fall der Unabhängigkeit von A und B hat das Eintreten von B keinen Einfluss auf die Wahrscheinlichkeit des Eintretens von A und es folgt aus (3.4)

$$P(A \cap B) = P(A) \cdot P(B). \tag{3.7}$$

Im Würfelversuch Wurf mit zwei Würfeln sind zum Beispiel die Ereignisse B (der erste Würfel zeigt eine Sechs) und A (der zweite Würfel zeigt eine Sechs) unabhängige Ereignisse, denn ein Würfel hat kein Gedächtnis. Für diese Ereignisse gilt also

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

und

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

3.2 Zufallsvariablen und ihre Verteilung

Viele zufällige Versuche führen zu Ereignissen, die sich leicht durch reelle Zahlen beschreiben lassen, wie z. B. die größte Augenzahl oder die Summe der Augenzahlen beim Würfeln mit zwei Würfeln oder Messergebnisse bei Längen- oder Gewichtsmessungen. In diesem Kapitel führen wir den Begriff der Zufallsvariablen ein, um

3 Thomas Bayes, 1702–1761, englischer Theologe.

zufällige Ereignisse durch Zahlen geeignet darzustellen. Die Wahrscheinlichkeiten für das Auftreten von bestimmten reellen Werten werden durch die Verteilung der Zufallsvariablen bestimmt. Man unterscheidet zwischen diskreten und stetigen Zufallsvariablen je nachdem, ob die Zufallsvariable abzählbar viele diskrete Werte oder jeden Wert aus einem Intervall annehmen kann. Es gibt Versuche, die stets zu denselben Verteilungen der betrachteten Zufallsvariablen führen und deshalb Modellcharakter haben. Einige dieser Modelle werden in diesem Abschnitt ausführlich beschrieben.

3.2.1 Grundbegriffe

Im vorherigen Abschnitt wurde das Grundmodell der Wahrscheinlichkeitstheorie beschrieben, in dem eine Grundmenge, ein Ereignisfeld und ein Wahrscheinlichkeitsmaß festgelegt werden. Die Einführung von Zufallsvariablen, d.h. einer Abbildung der Menge der Versuchsausgänge in die Menge der reellen Zahlen, bietet, wie man später noch sehen wird, eine Reihe von Vorteilen, insbesondere für praktische Anwendungen.

Definition

Eine Zufallsvariable ist eine Abbildung von Ω in die Menge der reellen Zahlen. Sie ordnet somit jedem Elementarereignis ω aus Ω eine reelle Zahl $X(\omega) = x$ zu.

Neben dem Begriff Zufallsvariable wird für diese Abbildung in der Wahrscheinlichkeitstheorie auch häufig der Begriff Zufallsgröße verwendet.

Zufallsvariablen werden gewöhnlich mit großen Buchstaben und ihre Realisierungen im konkreten Versuch mit kleinen Buchstaben bezeichnet.

In dem Fall, dass die Augensumme beim Würfeln mit zwei Würfeln interessiert, ist diese Zuordnung naheliegend. Die Zufallsvariable X ist die Summe beider Augenzahlen. Das Ereignis $\{\omega \in \Omega : X(\omega) = x\}$, kurz $\{X = x\}$, mit $x = 4$ tritt z. B. ein, wenn eines der Augenzahlpaare (1,3), (3,1) oder (2,2) gewürfelt wird.

Betrachten wir ein anderes Beispiel aus der Genetik: Für einen diallelischen Genlocus mit den Allelen A und a gibt es bei diploiden Lebewesen drei Genotypen AA, Aa und aa. Man kann diesen Genotypen die Zahlen 1, 2 und 3 zuordnen, oder aber auch 2, 1 und 0, so dass im letzteren Fall die Zufallsvariable die A-Allele im Genotyp zählt. Jede der möglichen Zuordnungen beschreibt eine Zufallsvariable, wobei sich beide in ihrem Wertebereich unterscheiden.

Bei einer wiederholten Durchführung eines Versuchs erhält man eine Folge von Realisierungen einer Zufallsvariablen X. Die Verteilung der so gewonnenen Daten kann mit den im Kapitel 2 behandelten Methoden untersucht werden. Die relativen Klassenhäufigkeiten werden z.B. durch Balkengrafik, Histogramm, Polygon oder Summenpolygon grafisch dargestellt. Nach dem Gesetz der großen Zahlen konvergieren die dargestellten Klassenhäufigkeiten gegen die Wahrscheinlichkeit, dass die betrachtete Zufallsvariable X in die jeweilige Klasse fällt. Mit Hilfe dieser Grenzwahrscheinlichkeiten beschreibt man die Verteilung der Zufallsvariable X. Eine Möglichkeit, diese Verteilung zu charakterisieren, bietet die Verteilungsfunktion.

Die Verteilungsfunktion $F_X(x)$ einer Zufallsvariablen X an der Stelle x gibt die Wahrscheinlichkeit für das Ereignis an, dass die Zufallsvariable X einen Wert kleiner oder gleich x annimmt.

Formel

$$F_X(x) = P(X \leq x) \quad (-\infty < x < \infty) \tag{3.8}$$

- F_X: Verteilungsfunktion der Zufallsvariablen X
- X: Zufallsvariable

Die Verteilungen von Zufallsvariablen unterscheiden sich in ihren Eigenschaften, je nachdem, ob die Zufallsvariable nur höchstens abzählbar viele diskrete Werte annimmt oder nicht.[4] Deshalb werden die Eigenschaften der Verteilungen in den folgenden Teilabschnitten getrennt behandelt.

3.2.2 Diskrete Zufallsvariablen

Eine Zufallsvariable heißt diskret, wenn sie höchstens abzählbar unendlich viele Werte $x_1, x_2, \ldots, x_n, \ldots$ annehmen kann.

Die Wahrscheinlichkeiten für das Auftreten dieser Werte in einem zufälligen Versuch $p_i = P(X = x_i), \quad i = 0, 1, 2, \ldots,$ werden als Einzelwahrscheinlichkeiten bezeichnet und können in einer Verteilungstabelle dargestellt werden.

x_i	x_1	x_2	$\cdots$	x_n	$\cdots$
$p_i = P(X = x_i)$	p_1	p_2	$\cdots$	p_n	$\cdots$

Tabelle 3.1: Verteilungstabelle einer diskreten Zufallsvariablen.

Beispiel

Für die Zufallsvariable Augensumme beim Wurf mit zwei Würfeln erhält man die folgende Verteilungstabelle:

4 In der Mathematik nennt man eine Menge abzählbar unendlich, wenn ihre Elemente als unendliche Folge dargestellt werden können.

x_i	2	3	4	5	6	7	8	9	10	11	12
$p_i = P(X = x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Tabelle 3.2: Verteilungstabelle der Zufallsvariablen Augensumme beim Wurf mit zwei Würfeln.

Diese Verteilung wird in ►Abbildung 3.2 durch ein Balkendiagramm veranschaulicht.

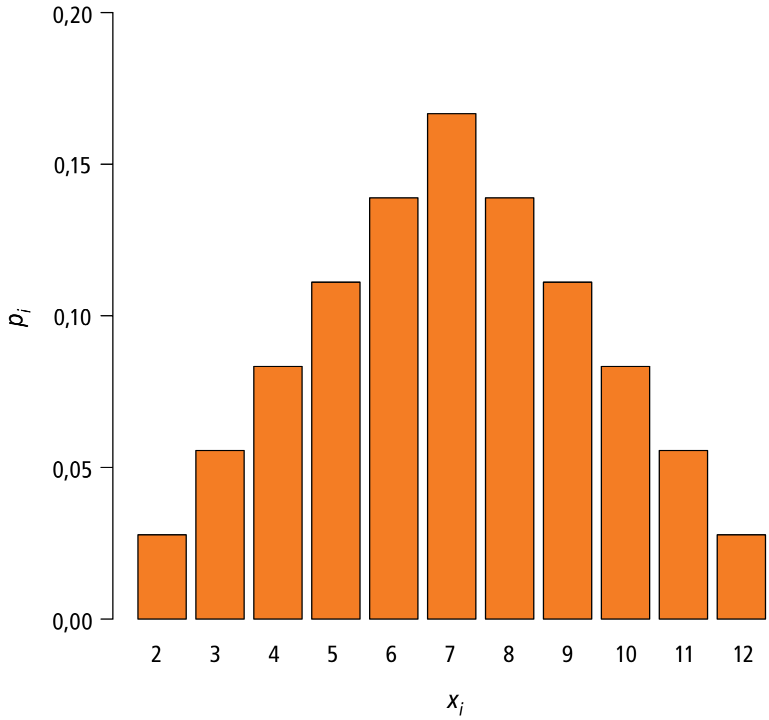

Abbildung 3.2: Darstellung der Einzelwahrscheinlichkeiten der Augensummen beim Wurf mit zwei Würfeln.

Mit Hilfe der Einzelwahrscheinlichkeiten in der Verteilungstabelle und Axiom 3 aus der axiomatischen Definition einer Wahrscheinlichkeit berechnet man die Wahrscheinlichkeit, dass die Zufallsvariable X in ein beliebiges Teilintervall der reellen Zahlen fällt, z. B.

$$P(X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{6}{36},$$

$$P(3 < X \leq 5) = P(X = 4) + P(X = 5) = \frac{3}{36} + \frac{4}{36} = \frac{7}{36}.$$

Aus den Einzelwahrscheinlichkeiten der Verteilungstabelle kann man somit die in (3.8) definierte Verteilungsfunktion für eine diskrete Zufallsvariable folgendermaßen bestimmen:

Formel

$$F_X(x) = P(X \leq x) = \sum_{i:\, x_i \leq x} p_i \quad (-\infty < x < \infty) \tag{3.9}$$

- F_X: Verteilungsfunktion der diskreten Zufallsvariablen X
- p_i $(i \geq 1)$: Einzelwahrscheinlichkeiten

Diese Verteilungsfunktion hat die Gestalt einer Treppenfunktion. Die Sprunghöhen sind gleich den Einzelwahrscheinlichkeiten p_i. ►Abbildung 3.3 zeigt die Verteilungsfunktion für die Zufallsvariable Augensumme beim Wurf zweier Würfel.

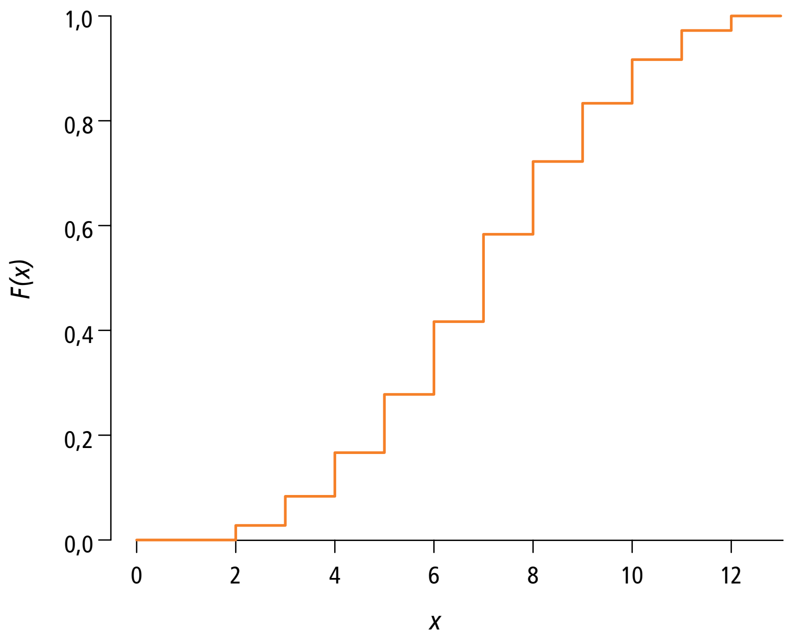

Abbildung 3.3: Verteilungsfunktion der Zufallsvariablen Augensumme beim Wurf mit zwei Würfeln.

Mit Hilfe der Verteilungsfunktion F kann man die Wahrscheinlichkeit, dass die Zufallsvariable X in ein vorgegebenes Intervall $(a, b]$ fällt, folgendermaßen darstellen:

Formel

$$P(a < X \leq b) = F_X(b) - F_X(a) \tag{3.10}$$

- F_X: Verteilungsfunktion der Zufallsvariablen X
- $a, b \in \mathbb{R}$

Charakteristische Eigenschaften einer Verteilung, wie Symmetrie der Einzelwahrscheinlichkeiten oder Lage und Streubreite der Werte, lassen sich durch einzelne Kenngrößen beschreiben (siehe Abschnitt 3.2.4).

3.2.3 Stetige Zufallsvariablen

Die Ergebnisse von zufälligen Versuchen wie Längen- oder Gewichtsmessungen bei Tieren oder Pflanzen können theoretisch jeden Wert aus einem Intervall der reellen Zahlen annehmen. Zufallsvariablen mit solchen Wertebereichen nennt man stetige Zufallsvariablen.

Die sich bei wiederholter Durchführung eines Versuchs ergebende Folge von Realisierungen einer Zufallsvariablen X kann durch ein Histogramm dargestellt werden (siehe Kapitel 2). Wählt man bei dieser Darstellung der relativen Häufigkeiten die Skalierung der x-Achse so, dass die Klassenbreite gleich 1 ist, dann entspricht die Fläche der Balken gerade den relativen Häufigkeiten für die Klassen. Mit wachsender Zahl von Versuchen und wachsender Klassenanzahl approximiert das Histogramm eine Grenzfunktion f_X (siehe ►Abbildung 3.4). Es ist damit naheliegend anzunehmen, dass für ein fest gewähltes Intervall $(a, b]$ auf der Abszisse die Fläche unter dem Graphen von f_X (siehe schraffierte Fläche in Abbildung 3.4) der Wahrscheinlichkeit entspricht, dass die Zufallsvariable X Werte in diesem Intervall annimmt.

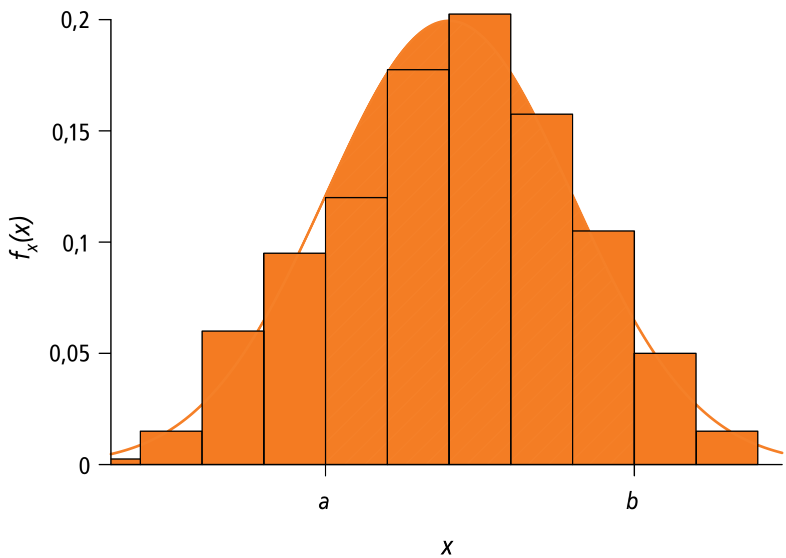

Abbildung 3.4: Darstellung einer Dichtefunktion f_X und der Wahrscheinlichkeit $P(a < X \leq b)$ als Flächeninhalt der schraffierten Fläche.

Für eine stetige Zufallsvariable wird also angenommen, dass eine Funktion f_X, Dichtefunktion genannt, existiert, so dass die Wahrscheinlichkeit des Auftretens von Werten der Zufallsvariablen in einem beliebigen Intervall $(a, b]$ folgendermaßen bestimmt ist:

Formel

$$P(a < X \leq b) = \int_a^b f_X(t)\,dt \tag{3.11}$$

- f_X: Dichtefunktion der stetigen Zufallsvariablen X
- X: Zufallsvariable

Ersetzt man in Formel (3.11) $b = x$, erhält man beim Grenzübergang $a \to -\infty$ den Wert der Verteilungsfunktion $F_X(x)$ an der Stelle $x \in \mathbb{R}$:

$$F_X(x) = P(-\infty < X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

►Abbildung 3.5 zeigt den Graphen der Verteilungsfunktion einer stetigen Zufallsvariablen X.

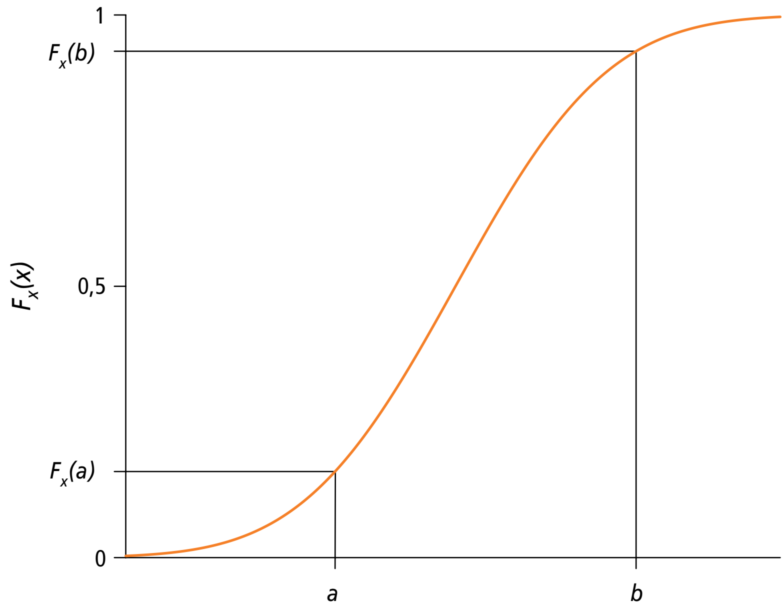

Abbildung 3.5: Darstellung einer stetigen Verteilungsfunktion $F_X(x)$.

Aus den Rechenregeln der Integralrechnung folgt, dass auch für stetige Zufallsvariablen die Formel (3.10) gilt:

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

Beim Grenzübergang $a \to -\infty, b \to \infty$ in Formel (3.11) ergibt sich wegen Axiom 2 für die Gesamtfläche unter der Dichtefunktion der Wert 1. Jede nichtnegative Funktion f mit der Eigenschaft

$$\int_{-\infty}^{\infty} f(t)\,dt = 1$$

kann somit als eine Dichtefunktion einer Zufallsvariablen X aufgefasst werden. Konkrete Beispiele für Dichtefunktionen werden in Abschnitt 3.3.2 betrachtet.

3.2.4 Verteilungsparameter

In praktischen Anwendungen interessiert man sich häufig dafür, welchen Wert die betrachtete Zufallsvariable bei einer langen Reihe von Versuchen „im Mittel" annimmt bzw. welchen Wert man erwarten kann und in welchem Maße die Werte der Zufallsvariablen streuen. Man möchte die Verteilung einer Zufallsvariablen mit wenigen charakteristischen Kennzahlen beschreiben.

Diskrete Verteilungen

Nehmen wir zunächst wieder einen Grundraum mit endlich vielen Ereignissen an. Der Erwartungswert einer diskreten Zufallsvariablen X mit den Werten x_i $(i = 1, \ldots, n)$ ist dann wie folgt definiert:

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i). \tag{3.12}$$

Wenn wir die möglichen Werte x_i $(i = 1, \ldots, n)$ der Zufallsvariablen X als Massepunkte und die zugehörigen Einzelwahrscheinlichkeiten $p_i = P(X = x_i)$ als Massen interpretieren, die entlang einer Linie verteilt sind, dann lässt sich die Größe $E(X)$ physikalisch als Masseschwerpunkt der Masseverteilung deuten.

In der Biologie untersucht man häufig bestimmte Merkmale an den Individuen einer Population. In diesem Fall kann man den Erwartungswert als Populationsmittelwert auffassen. Der Erwartungswert ist somit eine Kennzahl für die Lage der Verteilung. Ein wichtiges Gesetz der Wahrscheinlichkeitsrechnung sagt aus, dass der arithmetische Mittelwert einer empirischen Datenverteilung, der in Abschnitt 2.3 betrachtet worden ist, mit wachsendem Stichprobenumfang gegen den Erwartungswert der zugrunde liegenden Verteilung konvergiert.

Aus Formel (3.12) erhält man für das Beispiel Augensumme beim Wurf mit zwei Würfeln:

$$\begin{aligned} E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} \\ +8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} \\ = 7. \end{aligned}$$

Falls die Zufallsvariable X abzählbar unendlich viele Werte annehmen kann, wird der Erwartungswert in analoger Weise definiert, wobei die Summe durch den Wert der Reihe $\sum_{i=1}^{\infty} x_i \cdot P(X = x_i)$ ersetzt wird, wobei vorausgesetzt wird, dass dieser Grenzwert existiert.

Formel

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i) \qquad (3.13)$$

- $E(X)$: Erwartungswert der diskreten Zufallsvariablen X
- X: diskrete Zufallsvariable
- x_i $(i = 1, 2, \ldots, \infty)$: Werte der Zufallsvariablen X

Um die Streubreite einer Verteilung zu beschreiben, verwendet man die Varianz, das ist die mittlere quadratische Abweichung einer Zufallsvariablen von ihrem Erwartungswert:

Formel

$$Var(X) = E(X - E(X))^2 = \sum_{i=1}^{\infty} (x_i - E(X))^2 \cdot P(X = x_i) \qquad (3.14)$$

- $Var(X)$: Varianz einer diskreten Zufallsvariablen X
- $E(X)$: Erwartungswert der Zufallsvariablen X
- X: diskrete Zufallsvariable
- x_i $(i = 1, 2, \ldots, \infty)$: Werte der Zufallsvariablen X

Man erhält für das Beispiel Augensumme beim Wurf mit zwei Würfeln:

$$\begin{aligned} Var(X) = {} & (2-7)^2 \cdot \frac{1}{36} + (3-7)^2 \cdot \frac{2}{36} + (4-7)^2 \cdot \frac{3}{36} + (5-7)^2 \cdot \frac{4}{36} \\ & + (6-7)^2 \cdot \frac{5}{36} + (7-7)^2 \cdot \frac{6}{36} + (8-7)^2 \cdot \frac{5}{36} + (9-7)^2 \cdot \frac{4}{36} \\ & + (10-7)^2 \cdot \frac{3}{36} + (11-7)^2 \cdot \frac{2}{36} + (12-7)^2 \cdot \frac{1}{36} \\ = {} & 5.83. \end{aligned}$$

Stetige Verteilungen

Für stetige Verteilungen sind die Kenngrößen Erwartungswert $E(X)$ und Varianz $Var(X)$ wie folgt definiert:

Formel

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\,dt$$
$$Var(X) = \int_{-\infty}^{\infty} (t - E(X))^2 \cdot f_X(t)\,dt \qquad (3.15)$$

- $E(X)$: Erwartungswert einer stetigen Zufallsvariablen X
- $Var(X)$: Varianz einer stetigen Zufallsvariablen X
- f_X: Dichte der stetigen Zufallsvariablen X
- X: stetige Zufallsvariable

In praktischen Anwendungen tritt häufig die Frage auf, welcher Wert von der betrachteten Zufallsvariablen mit einer vorgegebenen Wahrscheinlichkeit α, $0 < \alpha < 1$, überschritten wird. Dieser Wert wird als Quantil der Ordnung $1-\alpha$ bezeichnet.

Ein Quantil der Ordnung α, allgemein mit q_α bezeichnet, schneidet den Anteil α von der Gesamtfläche unter der Dichtefunktion am linken Rand ab. Für das Quantil x_α der Zufallsvariablen X gilt also:

Formel

$$F_X(x_\alpha) = \int_{-\infty}^{x_\alpha} f_X(t)\,dt = \alpha \qquad (3.16)$$

- x_α: Quantil der Ordnung α $(0 < \alpha < 1)$
- f_X: Dichte der Zufallsvariablen X
- F_X: Verteilungsfunktion der Zufallsvariablen X
- X: stetige Zufallsvariable

Im folgenden Abschnitt 3.3.2 werden Quantile für mehrere spezielle Beispiele stetiger Verteilungen grafisch veranschaulicht (siehe ►Abbildung 3.8, ►Abbildung 3.12, ►Abbildung 3.14, ►Abbildung 3.16).

Im Spezialfall $\alpha = 0.5$ erhält man eine weitere Kennzahl für die Lage der Verteilung, den Median $Med(X) = x_{0.5}$ mit der Eigenschaft

$$P\big(X \le Med(X)\big) = P\big(X \ge Med(X)\big) = 0.5.$$

Für den besonderen Fall einer symmetrischen Verteilung stimmen die beiden Lageparameter Erwartungswert und Median überein.

Rechenregeln

In den praktischen Anwendungen werden oft die Verteilungen mehrerer Zufallsvariablen gleichzeitig betrachtet.

Analog zu Formel (3.8) kann man die gemeinsame Verteilung von zwei Zufallsvariablen X und Y mit Hilfe der Verteilungsfunktion

$$F(x,y) = P(X \leq x, Y \leq y), \quad (x, y \in \mathbb{R}),$$

definieren.

Man bezeichnet die Zufallsvariablen X und Y als unabhängig, falls die Ereignisse $\{X \leq x\}$ und $\{Y \leq y\}$ für beliebige $x, y \in \mathbb{R}$ stochastisch unabhängig sind. Aus Formel (3.7) folgt

$$F(x,y) = P(X \leq x) \cdot P(Y \leq y), \quad (x, y \in \mathbb{R}).$$

Die Unabhängigkeit von Elementen einer Folge $(X_i)_{i \geq 1}$ wird durch die Unabhängigkeit aller durch die Folgenglieder beschreibbaren Ereignisse definiert.

Sowohl für diskrete als auch für stetige Verteilungen gelten die folgenden Regeln für den Erwartungswert und die Varianz linearer Funktionen von Zufallsvariablen:

$$\begin{aligned} E(aX + bY) &= a \cdot E(X) + b \cdot E(Y), \\ Var(aX + b) &= a^2 \cdot Var(X). \end{aligned} \tag{3.17}$$

Für unabhängige Zufallsvariablen X und Y gilt

$$Var(X + Y) = Var(X) + Var(Y). \tag{3.18}$$

In den statistischen Anwendungen geht man im Allgemeinen davon aus, dass der zufällige Versuch, dem eine Zufallsvariable X zugeordnet worden ist, n-mal unabhängig voneinander und unter gleichen Bedingungen wiederholt wird, so dass die den Einzelversuchen zugeordneten Zufallsvariablen $X_1, X_2, \ldots, X_n$ unabhängig sind und die gleiche Verteilung wie X besitzen. Man nennt dann die bei Durchführung des Experiments angenommenen Werte $x_1, x_2, \ldots, x_n$ dieser Zufallsvariablen Realisierungen der Zufallsvariablen X.

3.3 Spezielle Verteilungen

In diesem Abschnitt werden ausgewählte häufig angewandte Verteilungsmodelle betrachtet, insbesondere die in den Kapiteln 5 bis 8 verwendeten Testverteilungen.

3.3.1 Diskrete Verteilungen

Binomialverteilung

In der biologischen Praxis bestehen Versuche oft darin, dass zufällig ausgewählte Individuen einer Population darauf untersucht werden, ob eine bestimmte Merkmalsausprägung, z. B. eine besondere Farbe, Krankheit oder Mutation, auftritt oder nicht. Solche Versuche lassen sich als Spezialfälle eines sogenannten Bernoulli-Versuchs auffassen, der in der n-fachen Wiederholung eines Experiments unter gleichen Bedingungen besteht, wobei nur die zwei Ausgänge „Erfolg" (1) oder „Nichterfolg" (0) interessieren, die Versuchsausgänge unabhängig voneinander sind und die Erfolgswahrscheinlichkeit für jeden Einzelversuch p beträgt.

Die Elemente des Grundraums Ω zu diesem Versuch können als geordnete Mengen vom Umfang n, sogenannte n-Tupel, die nur aus Nullen (Nichterfolg) und Einsen (Erfolg) bestehen, mathematisch beschrieben werden:

$$\Omega = \left\{(a_1,\dots,a_n) : a_j \in \{0,1\},\, j = 1,2,\dots,n\right\}$$

Das n-Tupel $(1,1,0,\dots,0)$ bezeichnet dann z.B. das Ereignis, dass im Bernoulli-Versuch nach genau zwei Erfolgen noch $n-2$ Misserfolge eingetreten sind.

In solchen Bernoulli-Versuchen interessiert man sich für die Verteilung der zufälligen Anzahl X_n der Erfolge. Wird etwa bei der n-fachen Wiederholung des Experiments „Wurf mit einem Würfel" nur beobachtet, ob eine Sechs fällt oder nicht, und zählt die Zufallsvariable X_n die Anzahl der Sechsen in einer Serie von n Würfen, dann ist die Wahrscheinlichkeit für die möglichen Werte 0,1,…, n dieser Zufallsvariablen zu bestimmen.

Die Zufallsvariable X_n, die jedem Elementarereignis aus Ω die Anzahl der Einsen im n-Tupel, d.h. die Anzahl der Erfolge in n Experimenten, zuordnet, besitzt eine Binomialverteilung mit den Parametern n und p. Es gilt für $0 < p < 1$ und alle natürlichen Zahlen $n \in \mathbb{N}$:

Formel

$$P(X_n = i) = \frac{n!}{i! \cdot (n-i)!} \cdot p^i \cdot (1-p)^{n-i} \quad (i = 0,1\dots,n) \qquad (3.19)$$

- n: Anzahl der Wiederholungen
- i: Anzahl der Erfolge
- p: Erfolgswahrscheinlichkeit in jedem Einzelversuch
- $P(X_n = i)$: Einzelwahrscheinlichkeiten der Verteilung von X_n
- X_n: Binomialverteilte Zufallsvariable

In ►Abbildung 3.6 wird die Verteilung in einem Balkendiagramm für $n = 10$ und zwei verschiedene Werte p dargestellt.

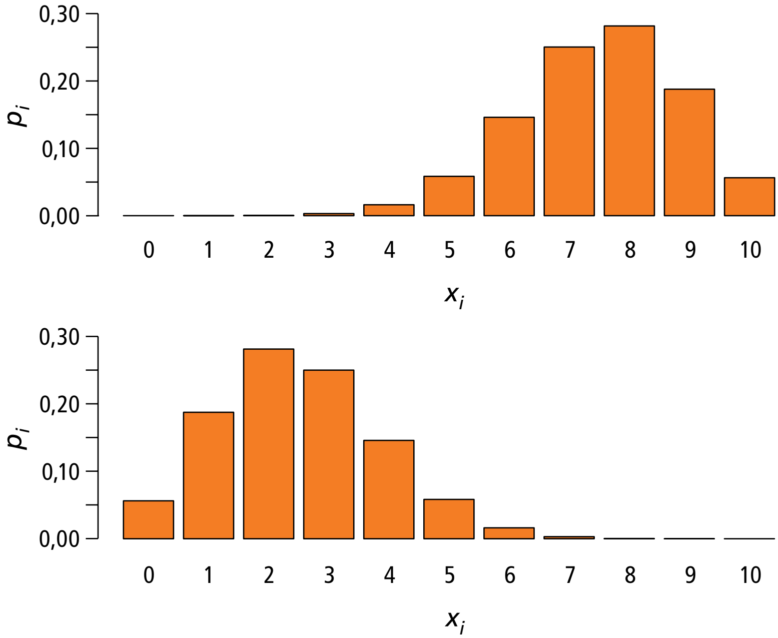

Abbildung 3.6: Einzelwahrscheinlichkeiten der Binomialverteilungen mit n=10 und p=3/4 (oben) sowie n=10 und p=1/4 (unten).

Für den Erwartungswert und die Varianz einer binomialverteilten Zufallsvariablen erhält man nach den Formeln (3.13) und (3.14):

$$E(X) = n \cdot p,$$
$$Var(X) = n \cdot p \cdot (1 - p).$$

Beispiel

Bei seinen Kreuzungsversuchen mit Erbsenpflanzen untersuchte Mendel unter anderem die Blütenfarbe, die in den zwei Ausprägungen rot oder weiß auftrat. Bei der Vererbung dieses Merkmals geht man von den drei Genotypen AA, Aa und aa aus, wobei nur die Individuen mit Genotyp aa eine weiße Farbe produzieren. Mendel beobachtete bei der Kreuzung von gemischterbigen Elternpflanzen mit Genotyp Aa, dass unter den Nachkommen die Farben rot und weiß im Verhältnis 3:1 auftraten. Die zufällige Anzahl von roten Farben bei n Nachkommen von Aa-Eltern kann somit als binomialverteilt mit den Parametern n und $p = 0.75$ angesehen werden.

Poissonverteilung

Die Poissonverteilung beschreibt wie die Binomialverteilung die Verteilung der Anzahl der Beobachtungen eines bestimmten Ereignisses in einem bestimmten Zeitabschnitt oder in einem bestimmten Raum. Wenn man die Anzahl der Wiederholungen im Bernoulli-Versuch nach Unendlich wachsen lässt, erhält man unter zusätzlichen Voraussetzungen die Poissonverteilung als Grenzverteilung. Die Poissonverteilung dient somit als Verteilungsmodell für eine Anzahl von Individuen, für die es keine obere Schranke gibt, wie es z.B. in einer Bakterienpopulation der Fall ist oder für die Anzahl von Todesfällen in einer Robbenpopulation pro Jahr. Sie wird ebenfalls als Verteilungsmodell für zufällige Anzahlen bei der Beschreibung von Verteilungsmustern von Pflanzen und Tieren im Raum verwendet.

Eine Zufallsvariable X heißt poissonverteilt mit dem Parameter λ ($\lambda > 0$), wenn gilt

$$P(X = i) = e^{-\lambda} \cdot \frac{\lambda^i}{i!} \quad (i = 0, 1, \ldots). \tag{3.20}$$

Erwartungswert und Varianz dieser Zufallsvariablen sind gleich. Es gilt $E(X) = Var(X) = \lambda$.

3.3.2 Stetige Verteilungen

Die in diesem Abschnitt eingeführten stetigen Verteilungen spielen insbesondere bei den Schätz- und Testverfahren (siehe Kapitel 4 bis 8) eine zentrale Rolle.

Normalverteilung

Eine Zufallsvariable Z heißt standardnormalverteilt, wenn sie die folgende Dichte hat:

Formel

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \quad (-\infty < x < \infty) \tag{3.21}$$

- φ: Dichte der Standardnormalverteilung

Wegen der glockenförmigen Gestalt wird der Graph der Dichtefunktion $\varphi(x)$ auch als Glockenkurve bezeichnet (siehe ►Abbildung 3.7).

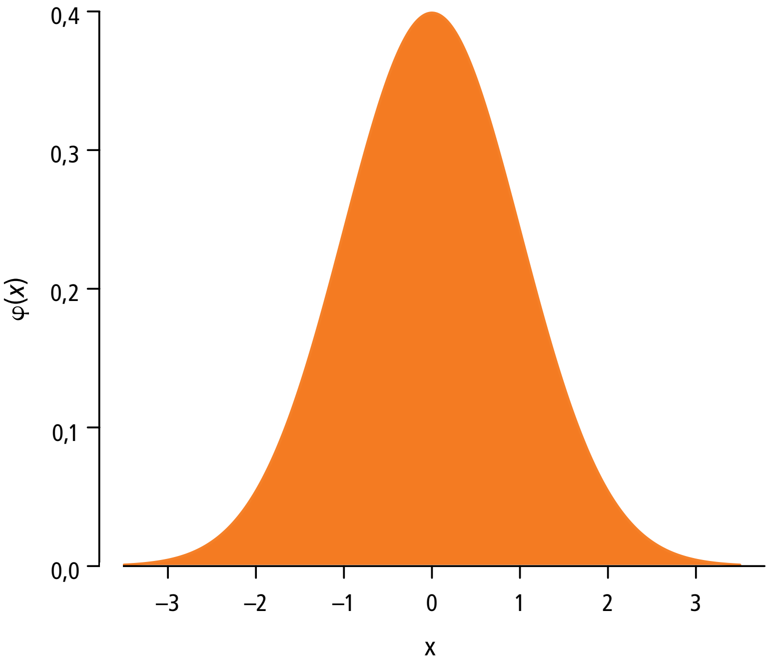

Abbildung 3.7: Dichtefunktion $\varphi(x)$ der Standardnormalverteilung.

Für den Erwartungswert und die Varianz einer standardnormalverteilten Zufallsvariablen Z gilt

$$E(Z) = 0, \quad Var(Z) = 1.$$

Die Standardnormalverteilung ist wegen $\varphi(x) = \varphi(-x)$ für alle $x \in \mathbb{R}$ symmetrisch. Aufgrund dieser Symmetrieeigenschaft gilt für die Verteilungsfunktion $\Phi(x)$, $x \in \mathbb{R}$,

$$\Phi(-x) = 1 - \Phi(x) \quad (-\infty < x < \infty). \tag{3.22}$$

Die Werte der zugehörigen Verteilungsfunktion $\Phi(x)$, $x \in \mathbb{R}$, kann man entweder einer Verteilungstabelle entnehmen (siehe z.B. Anhang B, Tabelle 1) oder mit Hilfe eines geeigneten Statistik-Programms berechnen.

Die zentrale Rolle dieser Verteilung in der schließenden Statistik beruht u.a. auf der Aussage des Grenzwertsatzes von Moivre-Laplace über die Verteilung einer binomialverteilten Zufallsvariablen X_n mit den Parametern n und p. Dieser Grenzwertsatz besagt, dass die Verteilungsfunktion der transformierten (standardisierten) Zufallsvariablen

$$\frac{X_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}},$$

deren Erwartungswert 0 und deren Varianz 1 ist, mit wachsendem n gegen die Verteilungsfunktion $\Phi(x)$ konvergiert. Man nennt diese standardisierte Zufallsvariable deshalb auch asymptotisch normalverteilt.

In vielen statistischen Anwendungen benötigt man Quantile x_α der Ordnung α, die für die Standardnormalverteilung mit z_α bezeichnet werden. Das Quantil $z_{1-\alpha}$ der Standardnormalverteilung schneidet am rechten Rand der Verteilung eine Fläche mit dem Flächeninhalt α von der Gesamtfläche unter der Dichtefunktion $\varphi(x)$ ab (siehe ▶Abbildung 3.8). Durch die Quantile $z_{\alpha/2}$ und $z_{1-\alpha/2}$ werden an beiden Rändern der Verteilung Flächen mit einem Flächeninhalt von jeweils $\alpha/2$ von der Gesamtfläche abgetrennt (siehe Abbildung 3.8).

Quantile der vorgegebenen Ordnung α kann man entweder einer Verteilungstabelle entnehmen (siehe z.B. Anhang B, Tabelle 1) oder mit Hilfe eines geeigneten Statistik-Programms berechnen.

Beispiel

Es sei $\alpha = 0.05$. Dann erhält man $z_{\alpha/2} = -1.96$, $z_\alpha = -1.6449$ und $z_{1-\alpha/2} = 1.96$.

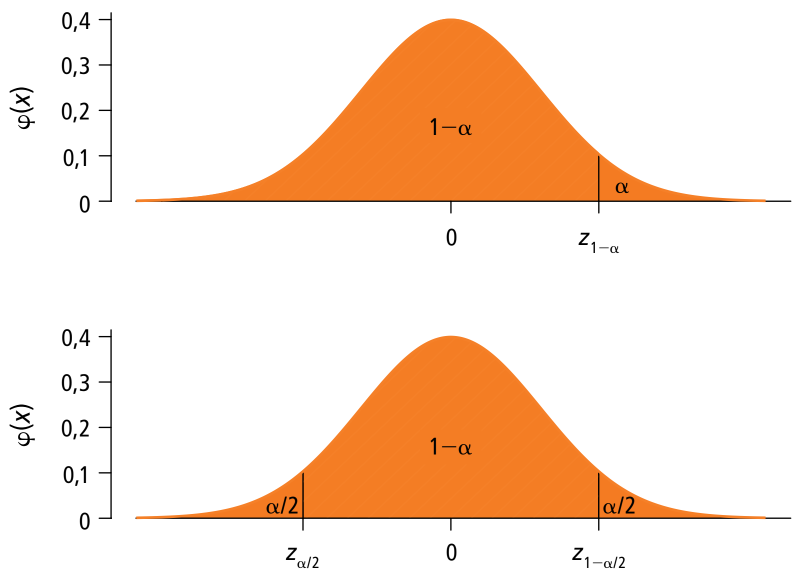

Abbildung 3.8: Quantile einer Standardnormalverteilung.

Bei der Durchführung statistischer Testverfahren interessiert man sich für den p-Wert $p = P(Z \geq z)$ an einer vorgegebenen Stelle z, d.h. für den Flächeninhalt, den der Wert z am rechten Rand von der Gesamtfläche abtrennt (siehe ▶Abbildung 3.9). Im Fall $p < \alpha$ ist diese Fläche und damit die Wahrscheinlichkeit $P(Z \geq z)$ kleiner als α. Die abzutrennende Fläche kann auch auf beide Ränder der Verteilung aufgeteilt werden, dann ist der p-Wert für die Zufallsvariable $|Z|$ gleich der Wahrscheinlichkeit $p = P(|Z| \geq z)$ (siehe Abbildung 3.9).

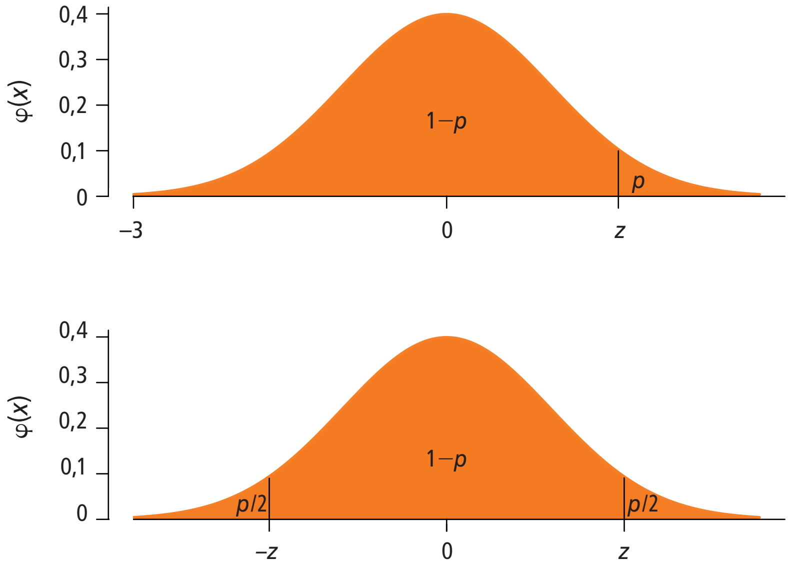

Abbildung 3.9: Veranschaulichung eines einseitigen (oben) bzw. zweiseitigen (unten) p-Werts.

Beispiel

Für $z = 1$ erhält man $p = P(Z \geq 1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$ bzw.

$$\begin{aligned} p &= P(|Z| \geq 1) = 1 - P(-1 \leq Z \leq 1) \\ &= 1 - (\Phi(1) - \Phi(-1)) \\ &= 0.3173. \end{aligned}$$

Für $z = 2$ erhält man analog $p = P(Z \geq 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$ bzw.

$$\begin{aligned} p &= P(|Z| \geq 2) = 1 - P(-2 \leq Z \leq 2) \\ &= 1 - (\Phi(2) - \Phi(-2)) \\ &= 0.0455. \end{aligned}$$

Die transformierte Zufallsvariable $X = \sigma Z + \mu, \quad \sigma > 0, \mu \in \mathbb{R}$, besitzt die Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty). \tag{3.23}$$

Die zugehörige Verteilung der Zufallsvariablen X heißt Normalverteilung mit den Parametern μ,σ^2 und wird kurz als $N(\mu,\sigma^2)$-Verteilung bezeichnet. Mit Formel (3.17) für Erwartungswert und Varianz von transformierten Zufallsvariablen erhält man

$$E(X)=\mu, \quad Var(X)=\sigma^2.$$

Die Familie der Normalverteilungen

$$\{N(\mu,\sigma^2), \quad \mu\in\mathbb{R}, \sigma^2>0\}$$

enthält somit für $\mu=0$ und $\sigma^2=1$ die Standardnormalverteilung als Spezialfall.

Aufgrund der Beziehung $X=\sigma Z+\mu$ kann die Berechnung von Wahrscheinlichkeiten einer $N(\mu,\sigma^2)$-Verteilung folgendermaßen auf die Bestimmung von Wahrscheinlichkeiten der Standardnormalverteilung zurückgeführt werden:

Formel

$$P(a<X\le b)=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma}) \tag{3.24}$$

- Φ: Verteilungsfunktion der Standardnormalverteilung
- μ: Erwartungswert der normalverteilten Zufallsvariablen X
- σ: Standardabweichung der normalverteilten Zufallsvariablen X

Aus diesem Zusammenhang zwischen der Verteilung einer Normalverteilung mit beliebigen Parametern μ und σ und der Standardnormalverteilung folgt für die Quantile der Ordnung α einer $N(\mu,\sigma^2)$-Verteilung:

$$x_\alpha=\sigma z_\alpha+\mu.$$

In ►Abbildung 3.10 werden die Dichtefunktionen von verschiedenen Normalverteilungen mit unterschiedlichen Parametern dargestellt.

Die $N(0,1)$- und $N(2,1)$-Verteilung unterscheiden sich nur in ihrem Lageparameter μ, die $N(0,1)$- und $N(0,2)$-Verteilung unterscheiden sich nur im Streuungsparameter σ.

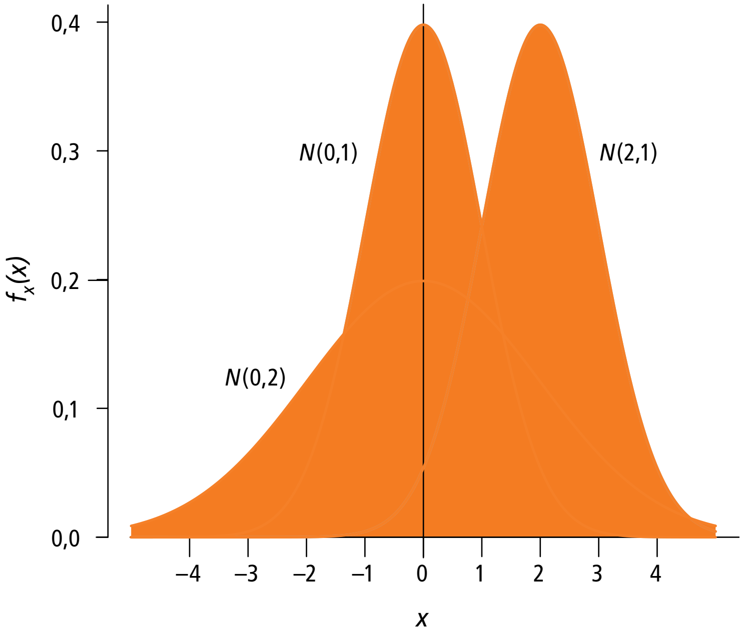

Abbildung 3.10: Dichtefunktionen der $N(0,1)$-, $N(0,2)$- bzw. $N(2,1)$-Verteilungen.

Seien X_1 und X_2 zwei unabhängige normalverteilte Zufallsvariablen, wobei X_1 $N(\mu_1, \sigma_1^2)$-verteilt und X_2 $N(\mu_2, \sigma_2^2)$-verteilt angenommen wird. Dann besitzt die Summe $X_1 + X_2$ eine $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$-Verteilung.

Die Bestimmung der Verteilung von Funktionen mehrerer Zufallsvariablen ist im Allgemeinen nur mit großem Aufwand möglich. In den folgenden Spezialfällen ist jedoch die resultierende Verteilung gut bekannt. Die Verteilungen haben besondere Bedeutung als Verteilungen von Teststatistiken beim Testen statistischer Hypothesen (siehe Kapitel 5, 6, 7 und 8).

Chi-Quadrat-Verteilung

Die Verteilung der Zufallsvariablen

$$Y = \sum_{i=1}^{n} X_i^2,$$

wobei die X_i, $i = 1, \dots, n$, unabhängige, standardnormalverteilte Zufallsvariablen sind, wird als Chi-Quadrat-Verteilung mit n Freiheitsgraden (χ_n^2-Verteilung) bezeichnet. Die Zufallsvariable Y besitzt die Dichte

$$f_Y(x) = \begin{cases} C_n \cdot e^{-\frac{x}{2}} \cdot x^{\frac{n}{2}-1} & x > 0 \text{ mit } C_n = \dfrac{1}{2^{n/2} \cdot \Gamma(n/2)}, \\ 0 & x \leq 0 \end{cases} \tag{3.25}$$

wobei Γ die Gammafunktion bezeichnet. Für den Erwartungswert und die Varianz einer χ^2_n-verteilten Zufallsvariablen Y gilt

$$E(Y) = n, \quad Var(Y) = 2n.$$

Die ►Abbildung 3.11 zeigt die Dichten einer χ^2_5-Verteilung und einer χ^2_{10}-Verteilung.

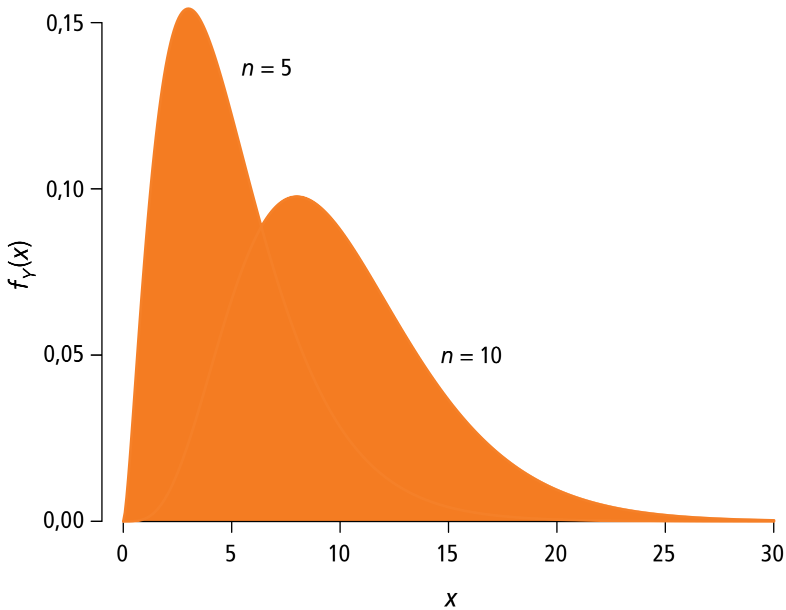

Abbildung 3.11: Dichtefunktionen der Chi-Quadrat-Verteilungen mit n=5 bzw. n=10 Freiheitsgraden.

Die χ^2_n-Verteilung spielt eine wichtige Rolle in der Teststatistik beim Vergleich von Streuungen (siehe Kapitel 6 und Kapitel 8).

Mit $\chi^2_{n,1-\alpha}$ wird das Quantil der Ordnung $1-\alpha$ einer χ^2_n-Verteilung bezeichnet. Es schneidet einen Anteil α von der Gesamtfläche unter der Dichtefunktion am oberen Rand des Wertebereichs der Zufallsvariablen ab. Der p-Wert gibt den Flächenanteil an, der von einem festen Wert χ^2am oberen Rand von der Gesamtfläche abgetrennt wird. Quantile und p-Werte werden in ►Abbildung 3.12 veranschaulicht. Zu ihrer Bestimmung kann eine Verteilungstabelle (siehe Anhang B, Tabelle 2) oder ein geeignetes Statistikprogramm verwendet werden.

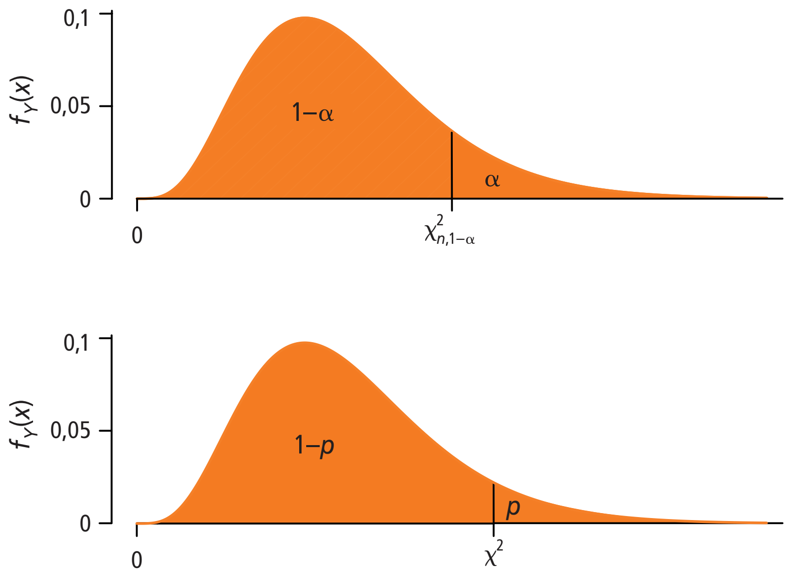

Abbildung 3.12: Quantil der Ordnung $1-\alpha$ und p-Wert einer Chi-Quadrat-Verteilung.

t-Verteilung

Es seien X und Y unabhängige Zufallsvariablen, wobei X einer $N(0,1)$-Verteilung genügt und Y χ^2_n-verteilt ist. Dann heißt die Verteilung des Quotienten

$$X \Big/ \sqrt{\frac{1}{n} \cdot Y} \qquad (3.26)$$

Studentische t-Verteilung mit n Freiheitsgraden.

► Abbildung 3.13 zeigt die Dichtefunktionen für $n = 2$ und $n = 10$.

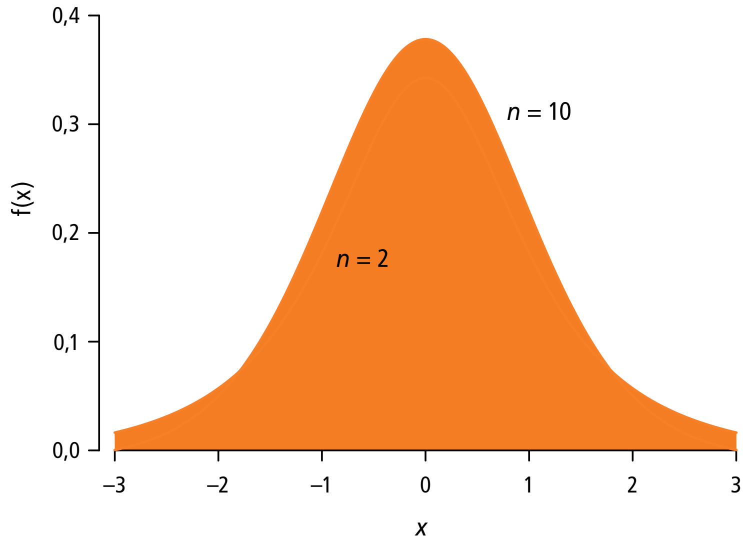

Abbildung 3.13: Dichte der t-Verteilungen mit n=2 bzw. n=10 Freiheitsgraden.

Die Werte einer t-verteilten Zufallsvariablen sind symmetrisch um 0 verteilt. Die Verteilung hat besondere Bedeutung bei Mittelwertvergleichen (siehe Kapitel 6 und 8). Quantile und p-Werte können entweder aus Tabellen (siehe Anhang B, Tabelle 3) oder mit geeigneten Statistikprogrammen bestimmt werden. In ►Abbildung 3.14 werden Quantil und p-Wert veranschaulicht.

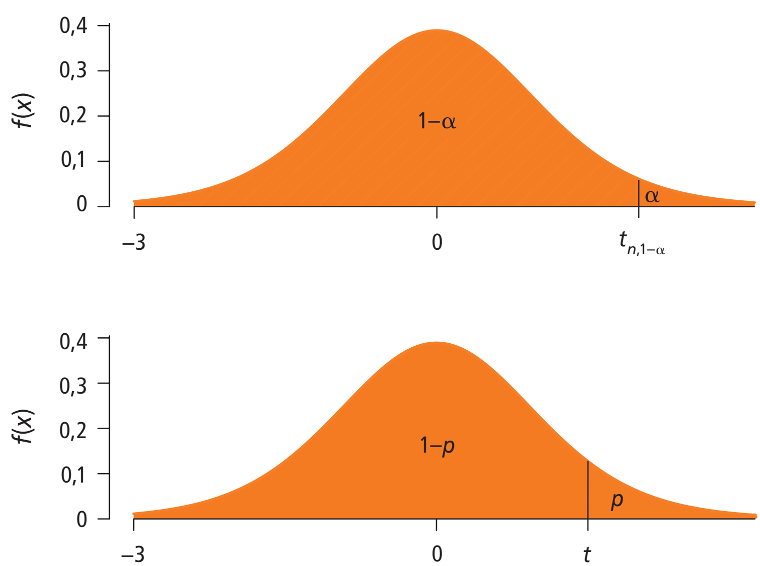

Abbildung 3.14: Quantil der Ordnung $1-\alpha$ und p-Wert einer t-Verteilung.

F-Verteilung

Es seien X und Y unabhängige Zufallsvariablen, wobei X als χ^2_m-verteilt und Y als χ^2_n-verteilt angenommen wird. Dann heißt die Verteilung des Quotienten

$$\frac{X}{m}\Big/\frac{Y}{n} \tag{3.27}$$

F-Verteilung mit m und n Freiheitsgraden und wird kurz als $F_{m,n}$- Verteilung bezeichnet.

F-Verteilungen werden zum Beispiel beim Vergleich von zwei Streuungen angewendet. In diesem Fall sind die Zufallsvariablen X und Y zwei Streuungsschätzer (siehe Kapitel 6, 7 und 8).

►Abbildung 3.15 zeigt die Dichten der $F_{5,10}$- und $F_{10,50}$-Verteilungen.

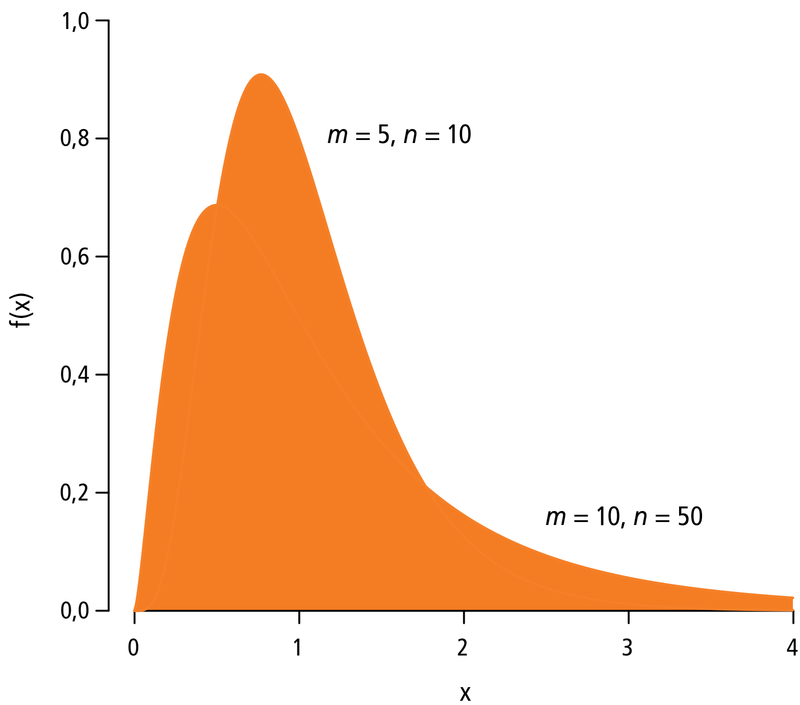

Abbildung 3.15: Dichte der F-Verteilungen mit m=5, n=10 bzw. m=10, n=50 Freiheitsgraden.

Quantile und p-Werte einer F-Verteilung werden in ►Abbildung 3.16 veranschaulicht. Zu ihrer Bestimmung kann eine Verteilungstabelle (siehe Anhang B, Tabelle 4) oder ein Statistikprogramm verwendet werden.

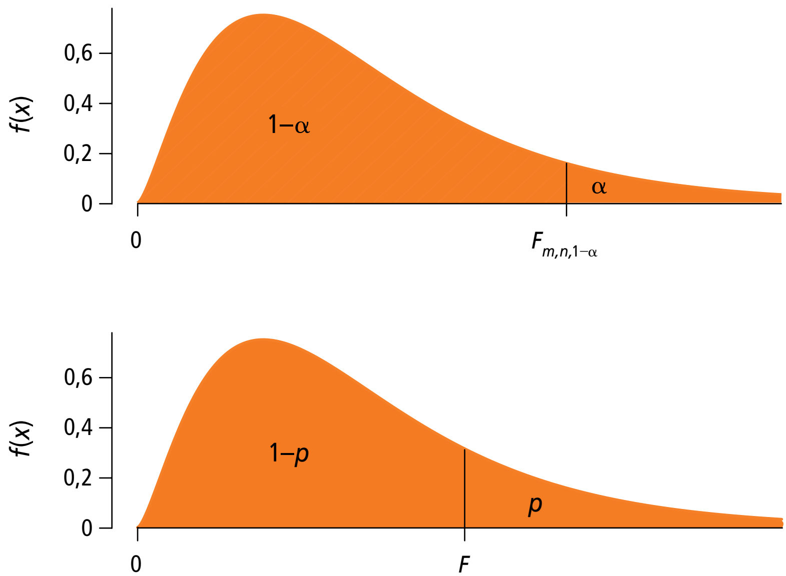

Abbildung 3.16: Quantil der Ordnung $1-\alpha$ und p-Wert einer F-Verteilung.

Zusammenfassung

Das Rechnen mit Wahrscheinlichkeiten beruht auf der Festlegung eines Grundmodells, d.h. der Einführung des Grundraumes der Ereignisse, eines Ereignisfeldes und der Wahrscheinlichkeit.

Die Realisierungen von diskreten Zufallsvariablen sind Zähldaten, während die Werte von stetigen Zufallsvariablen zur Beschreibung von Messdaten verwendet werden.

Mit Hilfe der Verteilungsfunktion ist es möglich, die Wahrscheinlichkeit zu bestimmen, mit der eine Zufallsvariable Werte in bestimmten Intervallen annehmen kann.

Erwartungswert, Varianz und Quantile sind wichtige Kenngrößen einer Verteilung.

Häufig angewendete diskrete Verteilungen sind die Binomial- und Poissonverteilung. In der Teststatistik (Kapitel 6 bis 8) spielen die Normal-, Chi-Quadrat-, t- und F-Verteilung eine wichtige Rolle.

Übungsaufgaben

Aufgabe 3.1

Bei seinen Kreuzungsversuchen mit Erbsenpflanzen untersuchte Mendel unter anderem auch die Form der Samen, die mit zwei Ausprägungen glatt oder runzlig auftraten. Bei der Kreuzung von gemischterbigen Elternpflanzen treten unter den Nachkommen die glatte und runzlige Form der Samen im Verhältnis 3:1 auf. Die zufällige Anzahl von glatten Samen bei n Nachkommen kann somit als binomialverteilt mit den Parametern n und $p = 0.75$ angesehen werden.

a. Wie viele Nachkommen mit glattem Samen erwarten Sie unter $n = 20$ Nachkommen?

b. Wie groß ist die Wahrscheinlichkeit, dass höchstens fünf Nachkommen mit runzligem Samen auftreten?

c. Wie groß ist die Wahrscheinlichkeit, dass mindestens drei Nachkommen mit glattem Samen beobachtet werden?

Aufgabe 3.2

Die Zufallsvariable X besitze eine $N(2,1)$-Verteilung.

a. Berechnen Sie die Wahrscheinlichkeiten $P(1 < X < 3)$ und $P(0 < X < 4)$.

b. Bestimmen Sie die Quantile der Ordnung $\alpha = 0.01$, $\alpha = 0.05$ und $\alpha = 0.95$.

Aufgabe 3.3

Die Zufallsvariable X besitze eine $N(25,4)$-Verteilung.

a. Berechnen Sie die Wahrscheinlichkeiten $P(X < 20)$, $P(X > 30)$ und $P(21 < X < 29)$.

b. Bestimmen Sie die Quantile der Ordnung $\alpha = 0.05$ und $\alpha = 0.99$.

Aufgabe 3.4

Aus einer Menge von zehn Versuchsfeldern werden rein zufällig fünf Felder für eine Behandlung mit einem Düngemittel ausgewählt. Die restlichen Felder bleiben unbehandelt. Angenommen, unter den zehn Feldern gibt es vier Felder mit schlechter Bodenqualität, wodurch der Effekt der Düngergabe auf den Ernteertrag beeinflusst wird. Mit welcher Wahrscheinlichkeit p wird die Anzahl der Felder mit schlechter Qualität in der Behandlungsgruppe überwiegen?

Aufgabe 3.5

In einer Klausuraufgabe werden vier Antwortmöglichkeiten angegeben. Genau eine der Antworten ist richtig. In einer Studentenpopulation beträgt die Wahrscheinlichkeit, die richtige Antwort zu wissen, 0.40. Ein Student, der die richtige Antwort nicht weiß, rät, d.h. er wählt zufällig eine der Antwortmöglichkeiten aus. Wie groß ist die Wahrscheinlichkeit dafür,

a. die richtige Antwort durch Raten zu finden?

b. dass ein Student die richtige Antwort nicht weiß?

c. dass ein Student die richtige Antwort gibt?

d. dass ein Student, der die Frage richtig beantwortet hat, die richtige Antwort weiß?

Ausführliche Lösungen sowie weitere Aufgaben finden Sie auf der Companion Website zum Buch unter **http://www.pearson-studium.de**