Kapitel 1
Wahrscheinlichkeitsrechnung (Wiederholung)

- **Zufallsvorgänge und Wahrscheinlichkeitsräume**
- **Bedingte Wahrscheinlichkeit und Unabhängigkeit**
- **Zufallsvariablen und Zufallsvektoren**
- **Momente von Zufallsgrößen**
- **Normalverteilung**

Anliegen der Wahrscheinlichkeitstheorie (Stochastik)

Anliegen ist die mathematische Beschreibung von Zufallserscheinungen und die Analyse von Gesetzmäßigkeiten, die ihnen innewohnen.

[Stochastik, griech. Mutmaßung]

Anwendungen

- Analyse von Glücksspielen
- Grundlage der induktiven Statistik und der Ökonometrie
- Grundlage für zahlreiche Ansätze der Entscheidungs- und Spieltheorie, der Finanzmarkttheorie, etc.

1.1 Zufallsvorgänge und Wahrscheinlichkeitsräume

Beispiele für „Zufallsvorgänge“ und „zufällige Ereignisse“

B–1.1 *Werfen eines Würfels.*

Ein Würfel wird geworfen. Wird die Seite ⚅ des Würfels oben liegen?

B–1.2 *Warten auf einen Bus.*

An einer Haltestelle hält alle 10 Minuten ein Omnibus. Eine Person betritt ohne Kenntnis der exakten Zeit die Haltestelle. Muss die Person mehr als 5 Minuten auf den nächsten Bus warten?

B–1.3 *Stichprobenzug einer Glühbirne.*

Wir ziehen aus einem Produktionslos von Glühbirnen (Grundgesamtheit) blind eine Birne. Ist die Birne funktionstüchtig?

D–1.1 Zufallsvorgang, Ergebnis

Ein ***Zufallsvorgang*** oder ***Zufallsexperiment*** ist ein Vorgang, dessen Ausgang im Rahmen verschiedener, prinzipiell bekannter Möglichkeiten ungewiss ist und der sich unter Einhaltung bestimmter Vorschriften beliebig oft (zumindest gedanklich) wiederholen lässt.

Der nach Ablauf des Vorganges tatsächlich eingetretene Ausgang wird als ***Ergebnis*** des Zufallsvorgangs bezeichnet.

Das ***Ziel*** der Wahrscheinlichkeitsrechnung ist es,

(*i*) Gesetzmäßigkeiten, die Zufallsvorgängen innewohnen, zu erkennen, und

(*ii*) die „Neigung“ oder die „Chance“ des Eintretens von zufälligen Ereignissen *vorab* durch Zahlen, die man Wahrscheinlichkeiten nennt, zu messen.

Wahrscheinlichkeitstheoretische Konzepte:

Ω - der Ergebnisraum ; $\mathcal{F}$ - das Ereignisraum ; P - das Wahrscheinlichkeitsmaß .

Der erste Schritt des Studiums von Zufallsvorgängen besteht darin, sich einen Überblick über alle denkbaren Ergebnisse eines Vorgangs zu verschaffen.

D–1.2 Ergebnisraum

Eine Menge Ω heißt *Ergebnisraum* eines Zufallsvorgangs, wenn sie jedes mögliche Ergebnis des Vorgangs als Element enthält.

Weitere Sprechweisen: *Grundraum*, *Stichprobenraum*.

Anmerkung

In der Mengenlehre bezeichnet man eine Menge, die alle in einem bestimmten Kontext relevanten Objekte als Elemente enthält, auch als *Raum*.

B–1.1 *Werfen eines Würfels.*

$\Omega = \{$⚀ ⚁ ⚂ ⚃ ⚄ ⚅$\}$

B–1.2 *Warten auf einen Bus.*

Ω ist die Menge aller Zeitintervalle zwischen 0 und 10 Min.; wir setzen $\Omega = [0\,,10]$.

B–1.3 *Stichprobenzug einer Glühbirne.*

Ω ist die Menge aller Glühbirnen des Produktionsloses; kurz $\Omega = \{\omega_1,\ldots,\omega_N\}$, wobei ω_i die *i*-te Birne symbolisiert.

B–1.4 *Zwei Würfel werden geworfen.*

Mögliche Ergebnisse des Zufallsvorgangs sind alle 2-Tupel der Seiten der beiden Würfel. Kennzeichnet man die Seiten *vereinfachend* durch ihre Augenzahlen 1,2,...,6, so ist der Ergebnisraum eine Menge geordneter Paare

$$\Omega = \{ (1,1), (1,2), (1,3), \ldots , (6,4), (6,5), (6,6)\}$$

mit insgesamt 36 Elementen, d.h. $|\Omega| = 36$.

D–1.3 Ereignis, Ereignisraum

Ein (*zufälliges*) *Ereignis* A eines Zufallsvorgangs ist eine Teilmenge des Ergebnisraumes Ω, d.h. $A \subset \Omega$. Man sagt, ein Ereignis A ist bei der Durchführung eingetreten, wenn das Ergebnis ω des Zufallsvorgangs ein Element von A ist, also $\omega \in A$.

Das Mengensystem $\mathcal{F}$ aller zulässigen Ereignisse eines Zufallsvorgangs heißt *Ereignisraum*.

D–1.3 Ereignis, Ereignisraum

Ein (*zufälliges*) *Ereignis* A eines Zufallsvorgangs ist eine Teilmenge des Ergebnisraumes Ω, d.h. $A \subset \Omega$. Man sagt, ein Ereignis A ist bei der Durchführung eingetreten, wenn das Ergebnis ω des Zufallsvorgangs ein Element von A ist, also $\omega \in A$.

Das Mengensystem $\mathcal{F}$ aller zulässigen Ereignisse eines Zufallsvorgangs heißt *Ereignisraum*.

B–1.1 *Werfen eines Würfels.*

Ereignis "Wurf der Seite ⚅ ". Es gilt $A = \{⚅\} \subset \Omega = \{⚀ ⚁ ⚂ ⚃ ⚄ ⚅\}$.

Ereignis "Wurf der Seite ⚄ oder ⚅ ". Es gilt $B = \{⚄ ⚅\} \subset \Omega = \{⚀ ⚁ ⚂ ⚃ ⚄ ⚅\}$.

Ereignis "Wurf einer ungeraden Augenzahl ". Es gilt $C = \{⚀ ⚂ ⚄\} \subset \Omega = \{⚀ ⚁ ⚂ ⚃ ⚄ ⚅\}$.

B–1.2 *Warten auf einen Bus.*

Ereignis "Wartezeit von mehr als 5 Minuten". Es gilt $A = (5, 10] \subset \Omega = [0, 10]$.

B–1.4 *Zwei Würfel werden geworfen.*

Ereignis "Wurf zweier gleicher Augenzahlen". Es gilt $A = \{(1,1,), (2,2,), \ldots, (6,6,)\} \subset \Omega$
mit $\Omega = \{(1,1), (1,2), \ldots, (5,6), (6,6)\}$.

Definition 1.3 ermöglicht es uns, Ereignisse als Mengen zu behandeln und mit Ereignissen genauso wie mit Mengen zu operieren. Damit dies sinnvoll möglich ist, müssen wir auch die leere Menge $\varnothing$ und Ω selbst als Ereignisse zulassen ($\varnothing \subset \Omega$, $\Omega \subset \Omega$).

D–1.4 Unmögliches und sicheres Ereignis

(i) $\varnothing$ heißt das *unmögliche* Ereignis ($\varnothing$ tritt sicher nicht ein);

(ii) Ω heißt das *sichere* Ereignis (Ω tritt sicher ein).

Charakteristikum der Ereignisse ist, dass sie bei der Durchführung eines Zufallsvorgangs eintreten können, aber nicht eintreten müssen. Welches Ereignis eintreten wird, ist Zufall. Das unmögliche und das sichere Ereignis sind die Ausnahmen von dieser Regel.

D–1.5 Elementarereignisse

E Jedes mehrelementige Ereignis kann als Vereinigungsmenge von einelementige Teilmengen von Ω dargestellt werden. Einelementige Teilmengen $A = \{\omega\}$ von Ω heißen *Elementarereignisse*.

Bei einem Zufallsvorgang soll jetzt die Chance oder Neigung für das Eintreten von Ereignisses A durch *Zahlen* $P(A)$ beschrieben werden, die man *Wahrscheinlichkeit* (*probability*) des Eintretens von A nennt. Wir benötigen dazu eine Abbildung

$$P : \mathcal{F} \to IR \ ,$$

die jedem Ereignis $A \in \mathcal{F}$ eine *Wahrscheinlichkeit* $P(A)$ zuordnet.

Mathematisch wird die Wahrscheinlichkeit *indirekt* durch die Festlegung von gewünschten und als zweckmäßig erachteten Eigenschaften der Abbildung P definiert (*Axiomatisierung*). Alle Sätze über Wahrscheinlichkeit werden auf diesen Eigenschaften aufgebaut.

Heute findet die Axiomatik von **ANDREI N. KOLMOGOROW (1903–1987)** allgemeine Anerkennung.

D–1.6 W-Maß, W-Raum

Eine Abbildung $P : \mathcal{F} \to IR$ heißt *Wahrscheinlichkeitsmaß (W-Maß)*, wenn sie das folgende Axiomsystem von Kolmogorow erfüllt:

Axiom 1 $P(A) \geq 0$ für alle Ereignisse $A \in \mathcal{F}$ (Nichtnegativitätsaxiom)

Axiom 2 $P(\Omega) = 1$ (Normierungsaxiom)

Axiom 3* Für disjunkte Ereignisse $A, B \in \mathcal{F}$ mit $A \cap B = \varnothing$ gelte:

$P(A \cup B) = P(A) + P(B)$ (Additivität)

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

Das Tripel $(\Omega, \mathcal{F}, P)$ wird als *Wahrscheinlichkeitsraum (W-Raum)* des Zufallsvorgangs bezeichnet.

Ist der W-Raum $(\Omega, \mathcal{F}, P)$ eines Zufallsvorgangs bekannt, dann kann jedem beliebigen Ereignis $A \in \mathcal{F}$ die Wahrscheinlichkeit $P(A)$ zugeordnet werden.

Aus dem Axiomsystem lassen sich alle Rechenregeln für Wahrscheinlichkeiten ableiten.

S–1.1

Es seien $(\Omega, \mathcal{F}, P)$ ein W-Raum und $A, B, C \in \mathcal{F}$ Ereignisse. Dann gilt:

(i) $0 \leq P(A) \leq 1$

(ii) $P(\emptyset) = 0$

(iii) $P(\overline{A}) = 1 - P(A)$ (Gegenwahrscheinlichkeit)

(iv) $A \subset B \Rightarrow P(A) \leq P(B)$ (Monotonie)

(v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Additionssatz für 2 Ereignisse)

(vi) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B)$ (Additionssatz für 3 Ereignisse)

$$- P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

1.2 Wahrscheinlichkeitskonzeptionen [Selbststudium]

Aus dem Axiomsystem lassen sich alle Rechenregeln für Wahrscheinlichkeiten ableiten. Die Axiome geben allerdings keinen Aufschluss darüber, wie Wahrscheinlichkeiten für Ereignisse eines konkreten Zufallsvorgangs tatsächlich bestimmt werden.

Diese Lücke schließen

- das *Gleichmöglichkeitsmodell* von Laplace,
- die *statistische Wahrscheinlichkeitskonzeption*,
- die *subjektive Wahrscheinlichkeitskonzeption*.

Gleichmöglichkeitsmodell von PIERRE-SIMON MARQUIS DE LAPLACE (1749–1827)

D–1.7 Laplace-Wahrscheinlichkeiten

Ein Zufallsvorgang erfülle die folgendenVoraussetzungen:

(*i*) Der Ergebnisraum enthält abzählbar-endlich viele Elemente, d.h.

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\} \quad \text{mit} \quad N < \infty \;;$$

(*ii*) Die Elementarereignisse $\{\omega_i\}$ $(i = 1, \ldots, N)$ sind alle gleichmöglich (gleichwahrscheinlich), d.h. sie unterscheiden sich hinsichtlich der Unbestimmtheit ihres Eintretens nicht.

Dann ist für ein beliebiges Ereignis $A \in \mathcal{F}$ die *Laplace-Wahrscheinlichkeit* für das Eintreten von A festgelegt durch das Verhältnis

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{N} = \frac{\text{Anzahl der für } A \text{ günstigen Ausgänge}}{\text{Anzahl aller möglichen Ausgänge}} \;.$$

Speziell folgt für alle Elementarereignisse

$$P(\{\omega_i\}) = \frac{1}{N} \qquad (i = 1, \ldots, N).$$

B–1.1 *Werfen eines Würfels.*

Wir unterstellen einen perfekten (regulären) Würfel und kennzeichnen die Seiten des Würfels vereinfachend durch ihre Augenzahlen 1,2,...,6.

Wegen $|\Omega| = 6$ ist die Wahrscheinlichkeit des Elementarereignisses $A = \{6\}$ nach Laplace gleich $P(A) = 1/6$. Für weitere Ereignisse wie $B = \{5, 6\}$ oder $C = \{1, 3, 5\}$ erhält man

$$P(B) = \frac{2}{6} = \frac{1}{3} \qquad \text{bzw.} \qquad P(C) = \frac{3}{6} = \frac{1}{2} \quad \boxtimes$$

B–1.6 *Lotto 6 aus 49.* (Kombinatorisches Problem)

Aus einer Urne mit 49 durchnummerierten Kugeln werden 6 Kugeln blind und ohne Zurücklegen gezogen. Der Ergebnisraum Ω des Zufallsvorgangs enthält

$$\binom{49}{6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13983816$$

verschiedene Ergebnisse. Die Wahrscheinlichkeit „sechs Richtige“ zu tippen, ist $\frac{1}{13983816}$ $\boxtimes$

S–1.2 Die Laplace-Wahrscheinlichkeiten sind mit den Kolmogorow'schen Axiomen verträglich.

Beweis

(*i*) $|A|, |\Omega| \geq 0 \quad \Rightarrow \quad \frac{|A|}{|\Omega|} = P(A) \geq 0$ für alle $A \in \mathcal{F}$;

(*ii*) $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$;

(*iii*) $A \cap B = \emptyset \quad \Rightarrow \quad |A \cup B| = |A| + |B| \quad \Rightarrow \quad P(A \cup B) = P(A) + P(B)$ für $A, B \in \mathcal{F}$ *q.e.d.*

Satz 1.2 ermöglicht es uns, die Rechenregeln für Wahrscheinlichkeiten aus **Satz 1.1** auf die in Definition 1.7 definierten W-Maße anzuwenden.

Statistische Wahrscheinlichkeiten

D–1.8 Häufigkeiten von Ereignissen

Bei der n-fach wiederholten Durchführung eines Zufallsvorgangs sei ein Ereignis $n(A)$-fach eingetreten. Wir bezeichnen $n(A)$ als die *absolute Häufigkeit* von A und

$$h_n(A) = \frac{n(A)}{n}$$

als die *relative Häufigkeit* von A.

S–1.3

Für die in Definition 1.11 definierten relativen Häufigkeiten von Ereignissen gilt:

(i) $0 \le h_n(A) \le 1$

(ii) $h_n(\Omega) = 1$

(iii) $A \cap B = \varnothing \Rightarrow h_n(A \cup B) = h_n(A) + h_n(B)$.

(*Beweis als Übung*)

Prinzip der großen Zahlen

Führt man einen Zufallsvorgang unter konstanten Rahmenbedingungen wiederholt unabhängig durch, dann kann man immer wieder beobachten, dass die relativen Häufigkeiten $h_n(A)$ eines Ereignisses A sich mit wachsender Wiederholungszahl n um einen festen Wert *stabilisieren*.

GEROLAMO CARDANO (1501–1576), *Liber de ludo aleae* (Buch über das Würfelspiel)

D–1.9 Statistische Wahrscheinlichkeit

Die *statistische Wahrscheinlichkeit* für das Eintreten eines Ereignisses A ist die Zahl $P(A)$, um die sich die relative Häufigkeit $h_n(A)$ bei wachsender Versuchswiederholungszahl n stabilisiert.

Man muss beachten, dass experimentell oder statistisch ermittelte Wahrscheinlichkeiten in ihrem Wesen approximativ sind. Die Zahl $P(A)$ können wir mittels relativer Häufigkeiten *nicht sicher* bestimmen, sondern wir erhalten immer nur einen Näherungs- oder Schätzwert für $P(A)$.

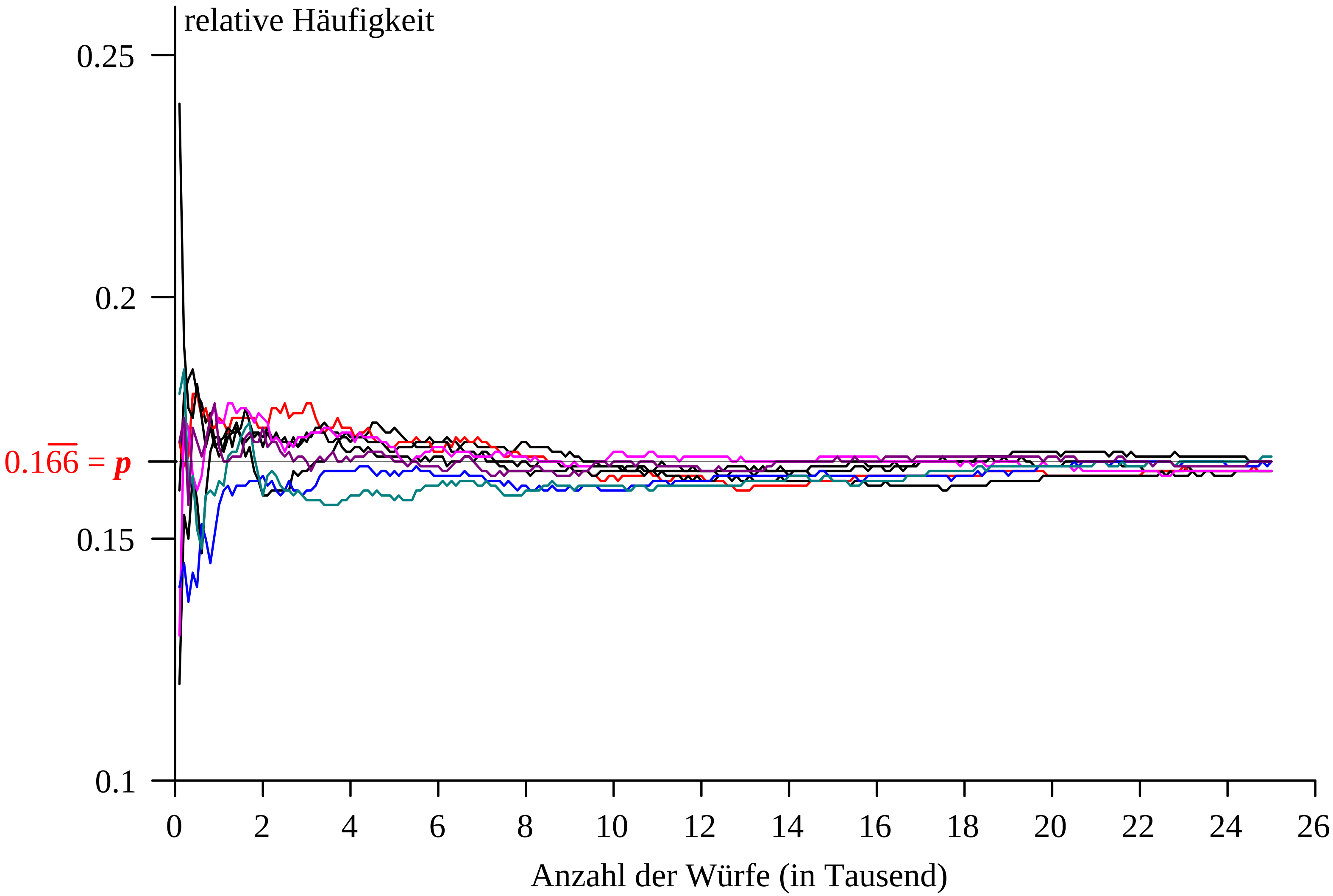
B–1.1 *Werfen eines Würfels.*

Die Wahrscheinlichkeit, dass bei dem Wurf eines regulären Würfels das Ereignis $A = \{6\}$ eintritt, beträgt nach dem Gleichmöglichkeitsmodell $P(A) = p = 1/6$.

Mit Hilfe eines Zufallszahlengenerators wurden auf einem Computer Würfe eines regulären Würfels simuliert. Eine Versuchsserie bestand aus 25000 unabhängigen Würfen.

Es wurden fortlaufend die relativen Häufigkeiten $h_n(A)$ nach n = 1,2,..., 25000 Würfen ermittelt, z.B.

Wurf n	1	2	3	4	5	6	7	8	…
geworfene Augenzahl	5	3	2	**6**	1	1	2	5	…
$h_n(\{6\})$	0	0	0	1/4	1/5	1/6	1/7	1/8	…



Relative Häufigkeiten geworfener „Sechsen“ in **8** Würfelversuchsserien ☒

B–1.7 *Sexualproportion.*

Die Geburt eines Kindes können wir hinsichtlich seines Geschlechts als einen Zufallsvorgang ansehen. Aufgrund langjähriger Beobachtung weiß man, dass die Sexualproportion von Jungen- zu Mädchengeburten in der Bundesrepublik um den Wert 1.06 schwankt. Der hieraus abgeleitete Näherungswert für die Wahrscheinlichkeit des Ereignisses „Mädchengeburt“ (bzw. „Jungengeburt“) ist 0.4854 (bzw. 0.5146) ⊠

Anmerkungen

(*a*) Das Konzept ist natürlich nur auf Massenphänomene anwendbar.

(*b*) Die in den Kolmogorow’schen Axiomen festgelegten Eigenschaften von Wahrscheinlichkeiten sind den Eigenschaften relativer Häufigkeiten (vgl. Satz 1.3) weitgehend nachgebildet.

1.3 Bedingte Wahrscheinlichkeit und Unabhängigkeit

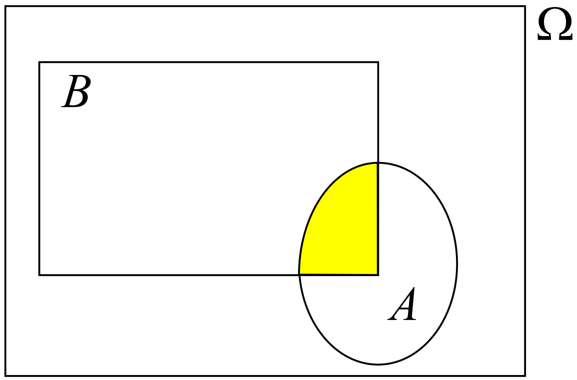
D–1.10 Bedingte Wahrscheinlichkeit

Gegeben seien ein W-Raum $(\Omega, \mathcal{F}, P)$ und die Ereignisse $A, B \in \mathcal{F}$ mit $P(B) > 0$. Dann heißt

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

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

Die Bedingung B bewirkt eine Einschränkung des Ergebnisraums Ω: Bezüglich A sind jetzt nur noch die Elemente von Ω relevant, die auch Element von B sind. $P(A \mid B)$ ist häufig von $P(A)$ verschieden.



B–1.8 *Einfaches Würfelspiel.*

Die Personen X und Y würfeln um einen Jack-Pot. Wer die höchste Augenzahl wirft, gewinnt. Bei Gleichstand gewinnt die Bank. Wir betrachten drei verschiedene Szenarien:

(i) X und Y würfeln gleichzeitig;

(ii) Y würfelt zuerst und erzielt die Augenzahl 2;

(iii) Y würfelt zuerst und erzielt die Augenzahl 6.

Wie hoch ist jeweils die Gewinnwahrscheinlichkeit für X?

Wir bezeichnen mit i bzw. j die gewürfelte Augenzahl von X bzw. Y. Der Ergebnisraum des Zufallsvorgangs ist

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

mit der Mächtigkeit $|\Omega| = 36$. X gewinnt, falls das Ereignis

$$A = \{(2,1),(3,1),(4,1),(5,1),(6,1),(3,2),(4,2),(5,2),\ldots,(6,5)\}$$

mit $|A| = 5 + 4 + 3 + 2 + 1 = 15$ eintritt.

Mit Hilfe des Laplace'schen Gleichmöglichkeitsmodells ermittelt man im Fall

(*i*) X und Y würfeln gleichzeitig

$$P(A)=\frac{|A|}{|\Omega|}=\frac{15}{36}=\frac{5}{12}\ .$$

(*ii*) Y würfelt zuerst und erzielt die Augenzahl 2

Das Ereignis $B=\{(1,2),(2,2),(3,2),(4,2),(5,2),(6,2)\}$ tritt jetzt sicher ein. Der Durchschnitt von A und B ist $A\cap B=\{(3,2),(4,2),(5,2),(6,2)\}$. Es folgt:

$$P(B)=\frac{6}{36}=\frac{1}{6}\quad,\quad P(A\cap B)=\frac{4}{36}=\frac{1}{9}\quad,\quad P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{1/9}{1/6}=\frac{2}{3}.$$

(*iii*) Y würfelt zuerst und erzielt die Augenzahl 6

Das Ereignis $C=\{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\}$ tritt sicher ein. Der Durchschnitt von A und C ist $A\cap C=\varnothing$ und somit folgt

$$P(C)=\frac{6}{36}=\frac{1}{6}\quad,\quad P(A\cap C)=0\quad\text{und}\quad P(A|C)=\frac{P(A\cap C)}{P(C)}=0\ \boxtimes$$

Bedingte Wahrscheinlichkeiten erfüllen die **Axiome der Wahrscheinlichkeitsrechnung**. Somit treffen alle Aussagen für gewöhnliche Wahrscheinlichkeiten auch auf bedingte Wahrscheinlichkeiten zu. Zusätzlich besitzen bedingte Wahrscheinlichkeiten weitere interessante Eigenschaften.

S–1.5 Gemeinsames Eintreten zweier Ereignisse (Multiplikationssatz)

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

Beweis

Aus Definition 1.10 folgt unmittelbar $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B) \cdot P(B)$ *q.e.d.*

Satz 1.5 ist hilfreich bei der Ermittlung von Wahrscheinlichkeiten für Durchschnitte, wenn die bedingten Wahrscheinlichkeiten einfacher als die gesuchten Wahrscheinlichkeiten zu ermitteln sind.

B–1.9 *Ziehen aus einer Urne mit vier Kugeln.*

In einer Urne befinden sich zwei weiße und zwei rote Kugeln, die mit Ausnahme der Farbe identisch sind. Es werden zwei Kugeln ***ohne*** Zurücklegen ***blind*** gezogen. Gesucht ist die Wahrscheinlichkeit, bei zwei Zügen zwei rote Kugeln zu ziehen.

B (bzw. A) bezeichne das Ereignis, dass beim 1. Zug (bzw. 2. Zug) eine rote Kugel gezogen wird. Offensichtlich gilt unter Verwendung des Gleichmöglichkeitsmodells $P(B) = P(A) = 0.5$.

Die Wahrscheinlichkeit beim zweiten Zug eine rote Kugel zu erhalten, wenn das Ereignis B zuvor eingetreten ist, beträgt $P(A \mid B) = 1/3$, denn in der Urne befinden sich unter der Bedingung B eine rote Kugel und zwei weiße Kugeln.

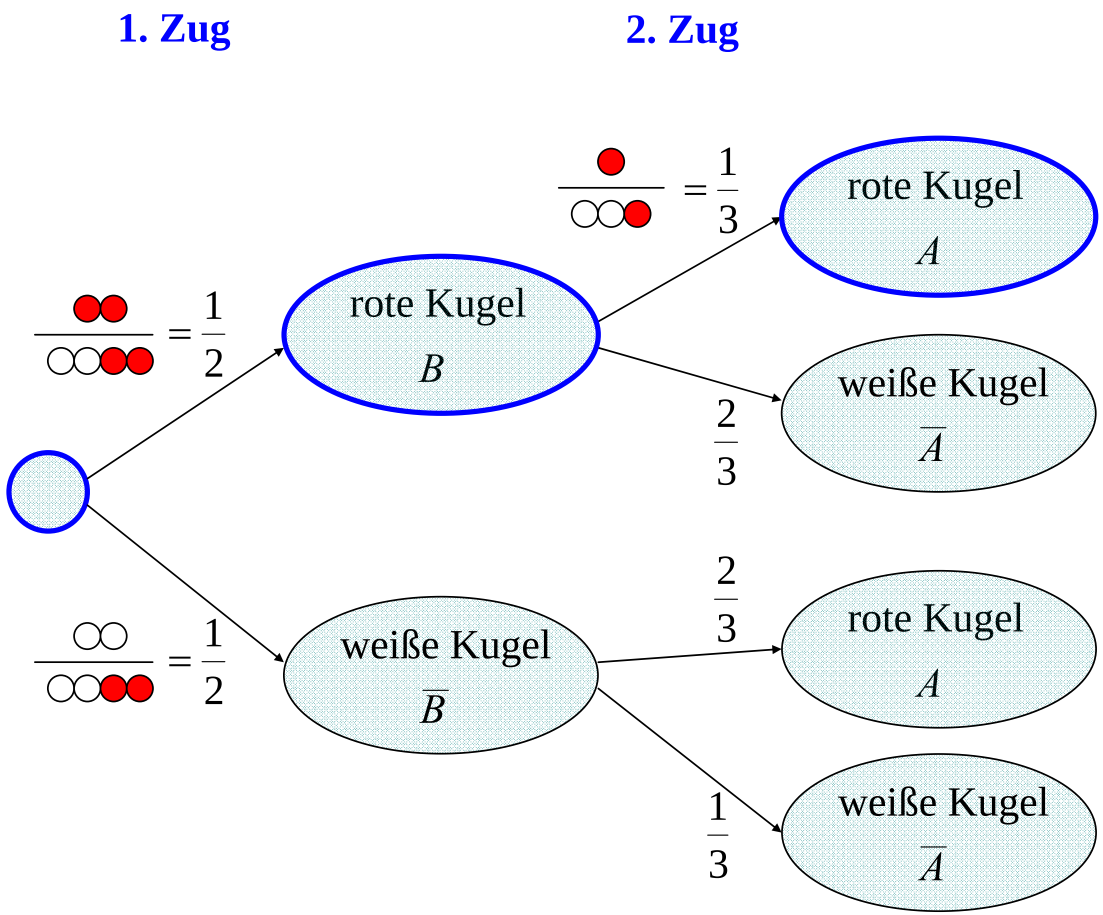
Die Wahrscheinlichkeit bei zwei Zügen zwei rote Kugeln zu ziehen, ist nach Satz 1.5

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

Analog können die Wahrscheinlichkeiten weiterer Durchschnitte ermittelt werden.

Ziehen ohne Zurücklegen aus einer Urne mit vier Kugeln

Baum-Diagramm



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

$$= \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

Der Multiplikationssatz lässt sich auf mehr als zwei Ereignisse verallgemeinern.

S–1.6 Gemeinsames Eintreten dreier Ereignisse (Multiplikationssatz)

$$P(A \cap B \cap C) = P(A \mid B \cap C) \cdot P(B|C) \cdot P(C) \text{ .}$$

Beweis

Die Anwendung von Satz 1.5 liefert $P(B \cap C) = P(B|C) \cdot P(C)$. Setze nun $D = B \cap C$. Wegen Satz 1.5 gilt weiter:

$$\begin{aligned} P(A \cap D) &= P(A|D) \cdot P(D) \\ &= P(A|B \cap C) \cdot P(B \cap C) \\ &= P(A|B \cap C) \cdot P(B \mid C) \cdot P(C) \qquad q.e.d. \end{aligned}$$

In den Beispielen 1.8 und 1.9 haben wir gesehen, dass bedingte Wahrscheinlichkeiten von Ereignissen größer oder kleiner als die unbedingten Wahrscheinlichkeiten sein können. Im Spezialfall können sie auch gleich sein.

D–1.11 Unabhängigkeit von zwei Ereignissen

Gegeben seien ein W-Raum $(\Omega, \mathcal{F}, P)$ und die Ereignisse $A, B \in \mathcal{F}$. Die Ereignisse heißen *stochastisch unabhängig* (voneinander), wenn

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

Gilt $P(A \cap B) = P(A) \cdot P(B)$, dann folgt sofort $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A)$ falls $P(B) > 0$.

Umgekehrt ist auch $P(B \mid A) = \frac{P(A \cap B)}{P(A)} = P(B)$ falls $P(A) > 0$.

B–1.9 *Ziehen aus einer Urne mit vier Kugeln.*

Wir ermittelten bereits

$$P(A \cap B) = \frac{1}{6} \neq P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Die Ereignisse B und A („Rote Kugel beim 1. Zug bzw. 2 Zug“) sind somit *stochastisch abhängig*. Speziell gilt $P(A \mid B) = {}^1\!/\!_3 < P(A) = {}^1\!/\!_2$.

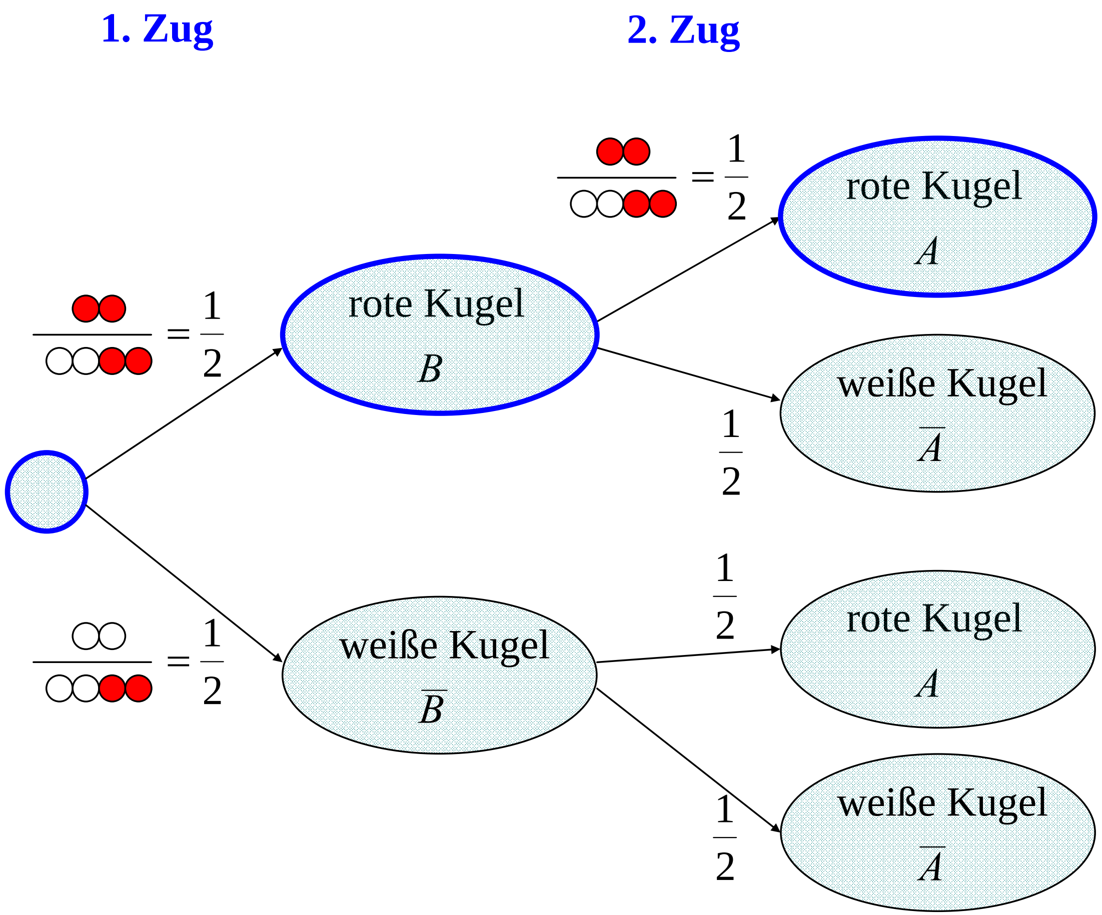
Ziehen wir aus der Urne nicht *ohne*, sondern *mit* Zurücklegen der Kugeln, so erhalten wir ein anderes Ergebnis. Unabhängig davon, wie der 1. Zug endet, befinden sich jetzt vor dem 2. Zug zwei weiße und zwei rote Kugeln in der Urne. Es folgt $P(A \mid B) = P(A) = {}^1\!/\!_2$. Die Wahrscheinlichkeit zwei rote Kugeln hintereinander zu erhalten ist

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{4}.$$

Die Ereignisse A und B sind beim Ziehen mit Zurücklegen stochastisch unabhängig voneinander.

Ziehen mit Zurücklegen aus einer Urne mit vier Kugeln

Baum-Diagramm



$$P(B \cap A) = P(B) \cdot P(A|B)$$
$$= P(B) \cdot P(A)$$
$$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

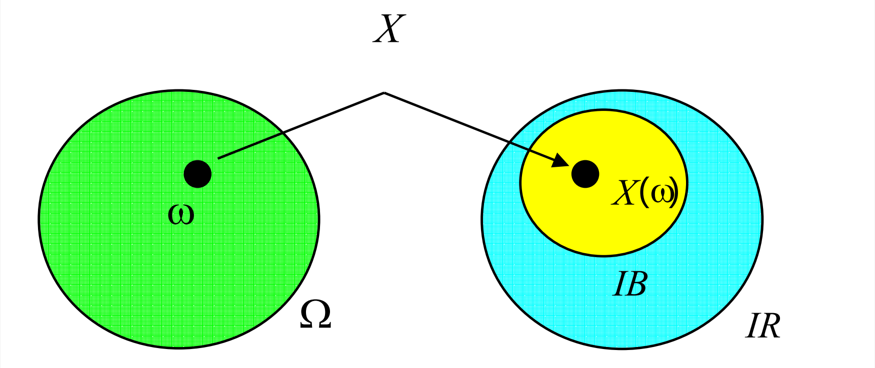
1.4 Zufallsvariable

D–1.12 Zufallsvariable

Gegeben sei ein Zufallsvorgang mit dem W-Raum $(\Omega, \mathcal{F}, P)$. Eine (eindimensionale) *Zufallsvariable* ist eine reellwertige (*messbare*) Abbildung

$$X : \Omega \to IR \ , \quad \omega \mapsto X(\omega) \ ,$$

die jedem Element ω des Ergebnisraums Ω genau eine reelle Zahl $X(\omega)$ zuordnet. Ω ist der Definitionsbereich, IR der Wertebereich und $IB = \{X(\omega) \mid \omega \in \Omega\}$ der Bildbereich (Träger) von X.



Ist ω das Ergebnis der Durchführung des Zufallsvorgangs, dann heißt der zugehörige Wert $X(\omega)$ *Realisierung* oder *Realisation* der Zufallsvariablen X.

B–1.1 *Werfen eines Würfels.*

Der Ergebnisraum des Zufallsexperiments ist Ω = {⚀ ⚁ ⚂ ⚃ ⚄ ⚅}.

Wir haben die Seiten des Würfels vielfach vereinfachend durch ihre Augenzahlen 1,2,...,6 gekennzeichnet. Jetzt betrachten wir dieses Vorgehen nicht mehr als eine Schreibvereinfachung, sondern als eine Abbildung $X : \Omega \to IR$ mit dem Bildbereich $IB = \{1,\ldots,6\} \subset IR$. X ist eine Zufallsvariable.

B–1.2 *Warten auf einen Bus.*

Wir haben den Ergebnisraum Ω des Zufallsexperiments – die Menge aller möglichen Wartezeiten bis zur nächsten Busankunft – in Kapitel 5 durch das Zahlenintervall [0 , 10] beschrieben.

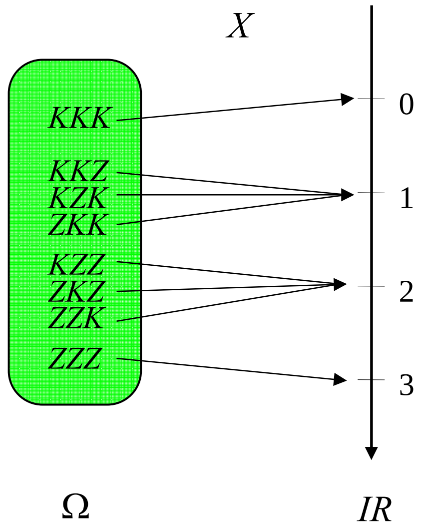
Streng genommen ist Ω nicht mit dem Intervall identisch. Erst wenn wir die Zeit auf einer numerischen Zeitskala messen, ordnen wir einer Wartezeit $\omega \in \Omega$ einen numerischen Wert zu. Das Intervall [0 , 10] ist somit genau genommen der Bildbereich IB einer Zufallsvariablen $X : \Omega \to IR$. Hierbei wird eine beliebig genaue Zeitmessung auf einer Minutenskala unterstellt.

B–1.10 *Münzspiel.*

Eine Münze mit den Seiten K („Kopf") und Z („Zahl") wird dreimal geworfen. Das Spiel besitzt als mögliche Ergebnisse alle Tripel, die aus den Seiten K und Z gebildet werden können:

$$\Omega = \{(KKK),(KKZ),(KZK),(ZKK),(KZZ),(ZKZ),(ZZK),(ZZZ)\} \; .$$

Ein Spieler gewinnt einen Euro-Betrag, der der Anzahl der von ihm geworfenen Seiten Z entspricht. Aus der Sicht des Spielers ist es naheliegend, den Elementen von Ω den zugehörigen Gewinn als numerischen Wert zuzuordnen:



$X : \Omega \to IR$ ist eine Zufallsvariable mit dem Bildbereich $IB = \{0,1,2,3\}$.

Beschreibung von Ereignissen mittels einer Zufallsvariable

- $X = x := \{\omega \in \Omega \mid X(\omega) = x\}$ gesprochen: „X nimmt den Wert x an";
- $X \leq x := \{\omega \in \Omega \mid X(\omega) \leq x\}$ gesprochen: „X nimmt höchstens den Wert x an";
- $X \geq x := \{\omega \in \Omega \mid X(\omega) \geq x\}$ gesprochen: „X nimmt mindestens den Wert x an";
- $x_1 < X \leq x_2 := \{\omega \in \Omega \mid x_1 < X(\omega) \leq x_2\}$ gesprochen: „X nimmt einen Wert im Intervall $(x_1, x_2]$ an"
- u.s.w.

Die Wahrscheinlichkeiten der Ereignisse schreiben wir

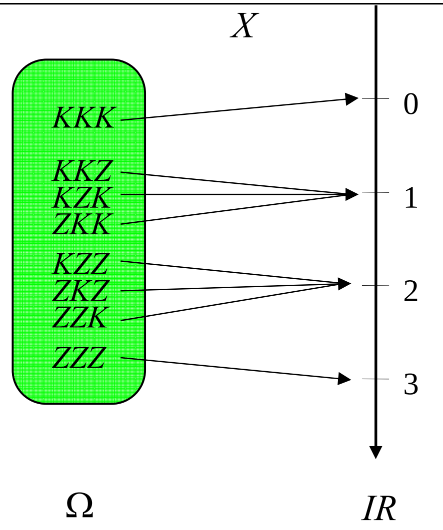
$$P(X = x)\ ,\quad P(X \leq x)\ ,\quad P(X \geq x)\ ,\quad P(x_1 < X \leq x_2)\ ,\quad \ldots\ .$$

B–1.10 *Münzspiel.*

Ergebnisraum

$$\Omega = \{(KKK),(KKZ),(KZK),(ZKK),$$
$$(KZZ),(ZKZ),(ZZK),(ZZZ)\}$$

Abbildung



$$X = 0 = \{\omega \in \Omega \mid X(\omega) = 0\} = \{(KKK)\} \qquad P(X = 0) = \frac{1}{8}$$

$$X \le 1 = \{\omega \in \Omega \mid X(\omega) \le 1\} = \{(KKK),(KKZ),(KZK),(ZKK)\} \qquad P(X \le 1) = \frac{4}{8} = 0.5$$

$$X > 1 = \{\omega \in \Omega \mid X(\omega) > 1\} = \{(KZZ),(ZKZ),(ZZK),(ZZZ)\} \qquad P(X > 1) = \frac{4}{8} = 0.5$$

1.5 Verteilungsfunktion

Eine Konsequenz der Einführung von Zufallsvariablen ist, dass jetzt Wahrscheinlichkeiten von Ereignissen mit Hilfe einer reellen Funktion ausgedrückt werden können. Die Funktion ist ein *theoretisches Analogon zur empirischen Verteilungsfunktion einer Häufigkeitsverteilung*.

D–1.13 Verteilungsfunktion

Es sei X eine Zufallsvariable. Dann heißt die Funktion

$$F : IR \to [\,0\,,1\,] \quad \text{mit} \quad x \mapsto F(x) = P(X \le x)$$

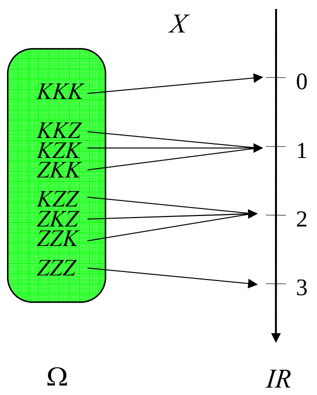
Verteilungsfunktion der Zufallsvariablen X. Sie ordnet jeder reeller Zahl $-\infty < x < +\infty$ die Wahrscheinlichkeit $P(X \le x)$ zu.

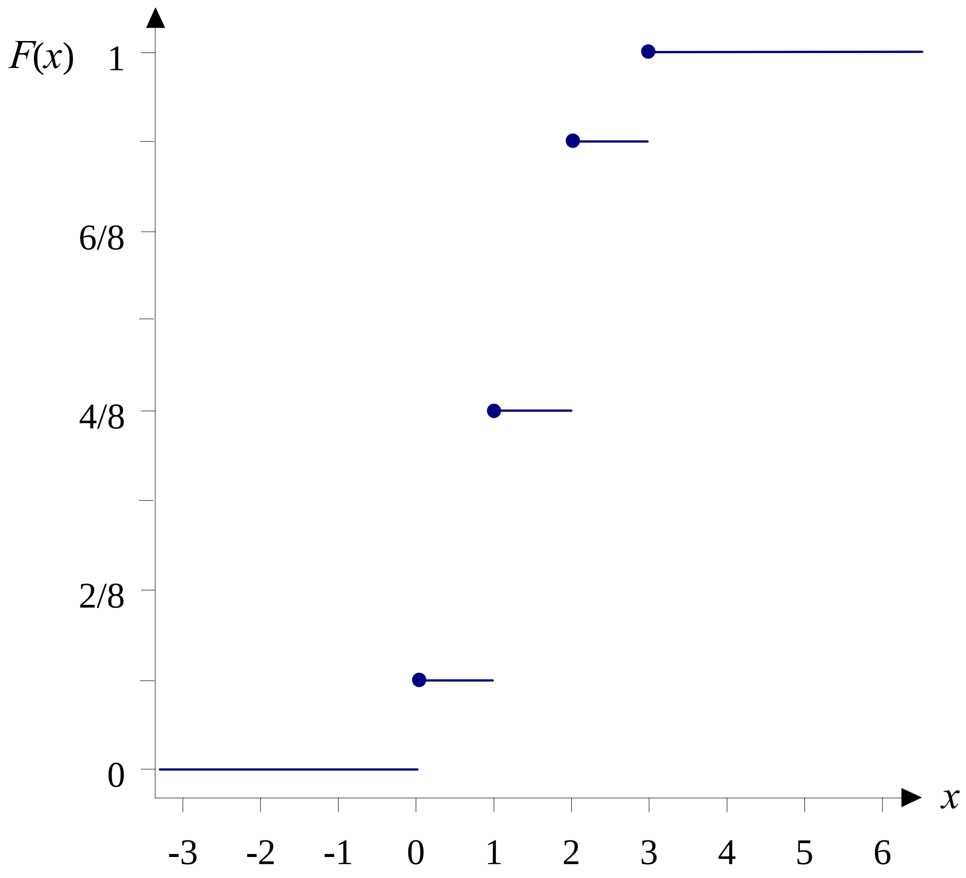
B–1.10 *Münzspiel.*

Die Verteilungsfunktion von X lautet

$$F(x)=\begin{cases} 0 & x<0 \\ 1/8 & 0\le x<1 \\ 4/8 \quad \text{falls} & 1\le x<2 \\ 7/8 & 2\le x<3 \\ 1 & x\ge 3 \end{cases}.$$

Es handelt sich um eine *monoton wachsende Treppenfunktion* mit *Sprungstellen* an den Punkten 0, 1, 2 und 3.





B–1.2 *Warten auf einen Bus.*

Die Verteilungsfunktion der Zufallsvariablen X (= Wartezeit in Minuten) lässt sich mit Hilfe einiger Plausibilitätsüberlegungen herleiten:

Wir wissen, dass X nur Werte im Intervall [0, 10] annehmen kann. Somit gilt $P(0 \leq X \leq 10) = 1$ und wegen $P(X < 0) = 0$ auch $P(X \leq 10) = 1$.

Ferner erscheint es plausibel, dass dem Ereignis $X \leq x$ mit einem beliebigem $x \in (0\,,10]$ eine doppelt so große Wahrscheinlichkeit zukommt wie dem Ereignis $X \leq x/2$.

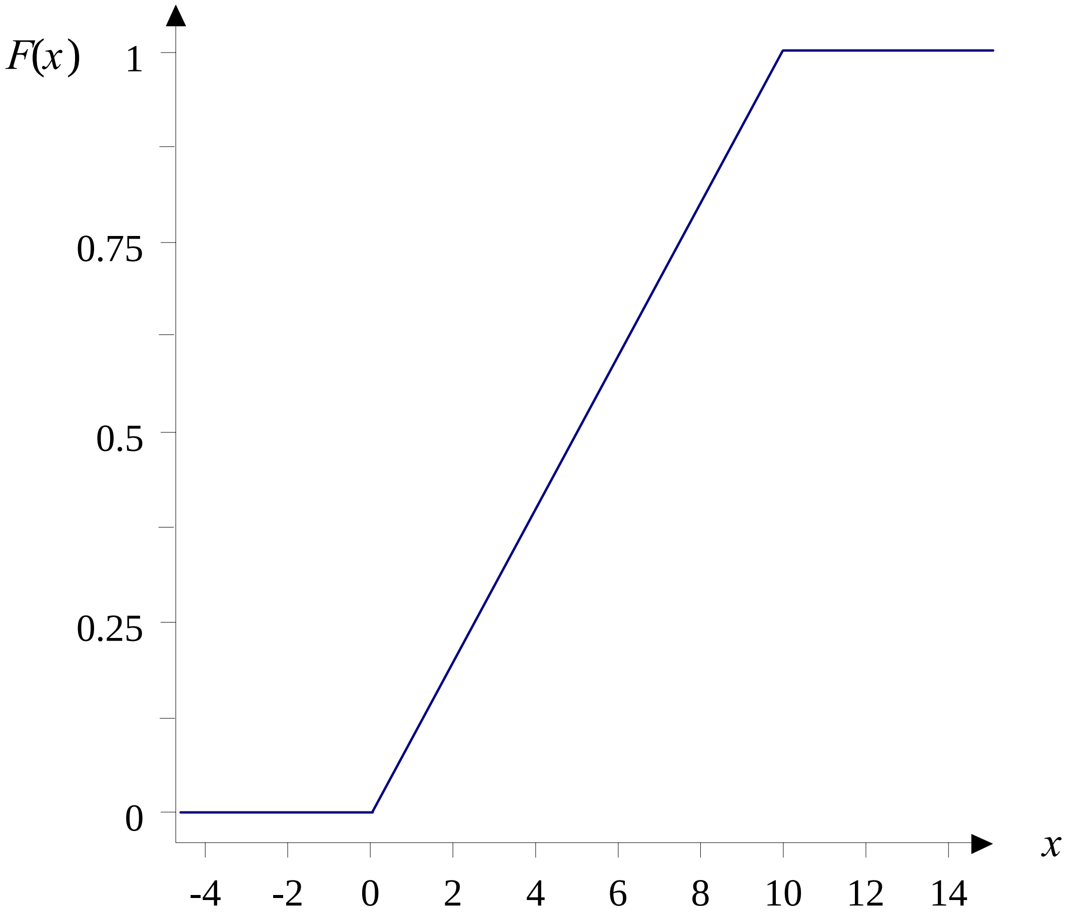
Also setzen wir die Wahrscheinlichkeit von $X \leq x$ proportional zu x:

$P(X \leq x) = k \cdot x$ mit dem Proportionalitätsfaktor $k \in IR$.

Für x = 10 ist nun $P(X \leq 10) = k \cdot 10 = 1$, woraus sofort k = 1/10 folgt. Wir erhalten so die Verteilungsfunktion:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{falls } x < 0 \\ \frac{1}{10}x & \text{falls } 0 \leq x \leq 10 \\ 1 & \text{falls } x > 10 \end{cases} .$$

Bei der Verteilungsfunktion handelt sich um eine *monoton wachsende stetige Funktion*.



$$F(x) = \begin{cases} 0 & \text{falls } x < 0 \\ \frac{1}{10}x & \text{falls } 0 \leq x \leq 10 \\ 1 & \text{falls } x > 10 \end{cases}$$

Aus den Axiomen der Wahrscheinlichkeitsrechnung ergeben sich folgende Eigenschaften einer Verteilungsfunktion:

S–1.7 Eigenschaften von Verteilungsfunktionen

F sei die Verteilungsfunktion einer Zufallsvariablen X. Dann gilt:

(i) F ist monoton wachsend;

(ii) $\lim\limits_{x \to -\infty} F(x) = 0$ und $\lim\limits_{x \to +\infty} F(x) = 1$;

(iii) F ist rechtsstetig.

Mit Hilfe der Verteilungsfunktion F einer Zufallsvariablen X können die Wahrscheinlichkeiten aller interessierenden Ereignisse eines Zufallsexperiments ermittelt werden.

S–1.8

F sei die Verteilungsfunktion einer Zufallsvariablen X und a, b reelle Zahlen mit $a < b$. Es gilt:

(*i*) $P(X > a) = 1 - F(a)$

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

(*iii*) $P(X = a)$ = Höhe des Sprunges von F im Punkt a

(*iv*) $P(X < a) = F(a) - P(X = a)$

(*v*) $P(X \geq a) = 1 - F(a) + P(X = a)$

(*vi*) $P(a \leq X \leq b) = F(b) - F(a) + P(X = a)$

(*vii*) $P(a < X < b) = F(b) - F(a) - P(X = b)$

(*viii*) $P(a \leq X < b) = F(b) - F(a) + P(X = a) - P(X = b)$

(*Beweis als Übung*)

1.6 Arten von Zufallsvariablen

D–1.14 Diskrete Zufallsvariable

Eine Zufallsvariable X heißt *diskret*, wenn ihr Bildbereich IB endlich oder abzählbar unendlich viele Werte enthält. Wir bezeichnen diese Realisationsmöglichkeiten einfach mit $x_1, x_2, x_3, \ldots$. Die Werte nennt man auch *Sprungstellen* und ihre Wahrscheinlichkeiten *Sprunghöhen*.

Die zu den Sprungstellen x_i $(i = 1,2,3,\ldots)$ gehörigen Wahrscheinlichkeiten bezeichnen wir im mit $P(X = x_i) = p_i$. Für alle reellen $x \notin IB$ ist offensichtlich $P(X = x) = 0$. Die Wahrscheinlichkeiten können durch eine Funktion wiedergegeben werden:

D–1.15 Wahrscheinlichkeitsfunktion

Es sei X eine diskrete Zufallsvariable. Die Funktion f: $IR \to [0\ ,\ 1]$ mit

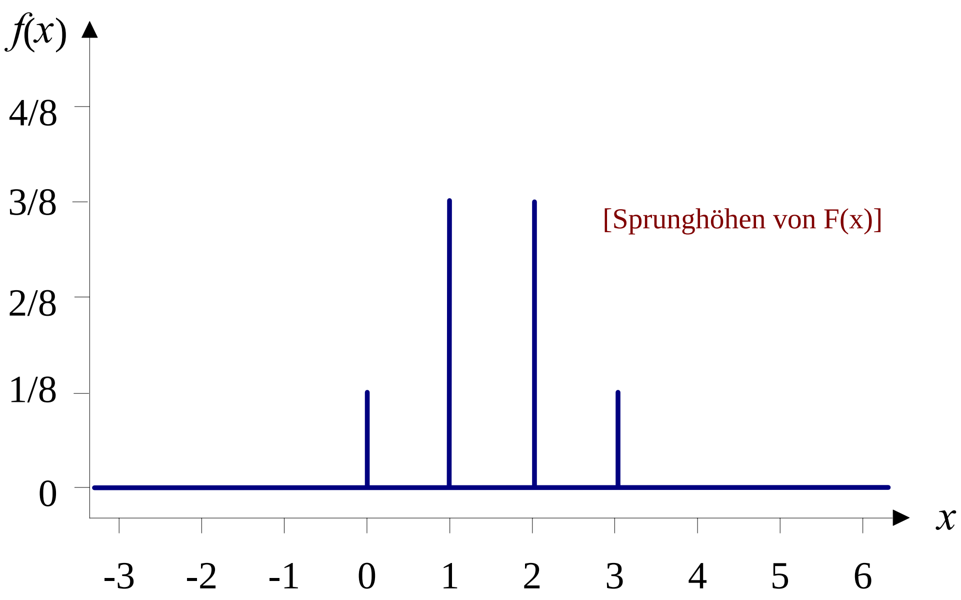
$$f(x) = P(X = x) = \begin{cases} p_i & \text{falls } x = x_i \ (i = 1,2,3,\ldots) \\ 0 & \text{sonst (alle übrigen reellen } x) \end{cases}$$

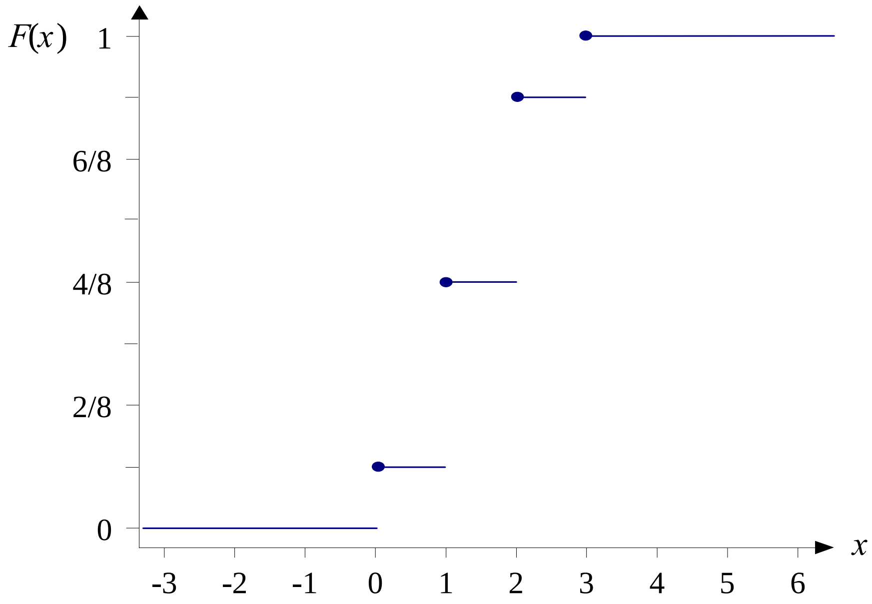
heißt die *Wahrscheinlichkeitsfunktion* von X. (Visualisierung: *Stabdiagramm*)

B–1.10 *Münzspiel.*

Die Zufallsvariable X ist diskret mit den 4 Sprungstellen $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ und $x_4 = 3$. Die Wahrscheinlichkeitsfunktion besitzt die Form

$$f(x) = \begin{cases} 1/8 & \text{falls } x = 0 \text{ oder } x = 3 \\ 3/8 & \text{falls } x = 1 \text{ oder } x = 2 \\ 0 & \text{sonst.} \end{cases}$$





Da X bei der Durchführung des Zufallsexperiments stets irgendeinen Wert aus $I\!B$ annimmt, muss

$$\sum_i f(x_i) = 1$$

gelten.

Zwischen der Wahrscheinlichkeitsfunktion und der Verteilungsfunktion $F(x)$ einer diskreten Zufallsvariablen X besteht der Zusammenhang

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i) \quad ,$$

wobei die Summation über alle Sprungstellen x_i erfolgt, die die Ungleichung $x_i \leq x$ erfüllen.

Ferner gilt für beliebige Teilmengen $B \subset I\!R$

$$P(X \in B) = \sum_{x_i \in B} f(x_i) \quad .$$

B–1.10 *Münzspiel.*

Die Zufallsvariable X ist diskret mit den 4 Sprungstellen $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ und $x_4 = 3$. Die Wahrscheinlichkeitsfunktion besitzt die Form

$$f(x) = \begin{cases} 1/8 & \text{falls } x = 0 \text{ oder } x = 3 \\ 3/8 & \text{falls } x = 1 \text{ oder } x = 2 \\ 0 & \text{sonst}\,. \end{cases}$$

Es gilt:

$$\sum_i f(x_i) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1 \quad ,$$

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i) = \begin{cases} 0 & & x < 0 \\ 1/8 & & 0 \le x < 1 \\ 4/8 & \text{falls} & 1 \le x < 2 \\ 7/8 & & 2 \le x < 3 \\ 1 & & x \ge 3 \end{cases} .$$

Wir betrachten nun Zufallsvariablen *X*, die alle Werte in einem Intervall annehmen können (nicht-abzählbar viele) und die keine Sprungstellen besitzen. Die Verteilungsfunktionen solcher Variablen sind *stetig*. Sie werden deshalb als *stetig* bezeichnet.

D–1.16 Stetige Zufallsvariable, Dichtefunktion

Eine Zufallsvariable *X* heißt *stetig*, wenn eine nicht-negative Funktion $f(x)$ (d.h. $f(x) \geq 0$ für alle $x \in IR$) existiert, die für jede reelle Zahl *x* die Gleichung

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

erfüllt, wobei $F(x) = P(X \leq x)$ die Verteilungsfunktion von *X* ist (*t* = Integrationsvariable).

Die Funktion $f(x)$ heißt *Dichtefunktion* oder kurz *Dichte* von *X*.

S–1.9 Eigenschaften von Dichtefunktionen

Es sei f die Dichtefunktion einer Zufallsvariablen X mit der Verteilungsfunktion F. Dann gilt:

(*i*) $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1;$

(*ii*) $\displaystyle P(a < X \leq b) = \int_a^b f(x)dx$ für beliebige $a, b \in IR$ mit $a < b$;

(*iii*) $\displaystyle\frac{dF(x)}{dx} = F'(x) = f(x)$ falls $f(x)$ stetig im Punkt x $\Leftrightarrow$ $F(x)$ differenzierbar in x.

Beweis

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{x\to\infty} F(x) = 1;$$

$$P(a < X \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_a^b f(x)dx \qquad q.e.d.$$

Anmerkungen

(a) In allen Stetigkeitspunkten einer Dichte gilt offenbar

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x} .$$

Somit ist $f(x) \neq P(X = x)$.

Für sehr kleines Δx gibt der Ausdruck $f(x) \cdot \Delta x$ näherungsweise die Wahrscheinlichkeit des Intervalls $(x < X \leq x + \Delta x)$ an.

(b) Aus der Eigenschaft (ii) in Satz 1.9 folgt für $a \to b$

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

D.h., $(X = a)$ ist für jedes $a \in IR$ ein *fast unmögliches Ereignis*. Bei der Durchführung des Zufallsvorgangs wird X stets einen Wert annehmen, die Wahrscheinlichkeit eines bestimmten Wertes ist jedoch null. Die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen lässt sich ***nicht*** durch Angabe der ***Einzelwahrscheinlichkeiten*** charakterisieren.

Anmerkungen

(c) Man beachte, dass aufgrund der Eigenschaften stetiger Zufallsvariablen stets gilt:

$$P(a < X \le b) = P(a \le X \le b) = P(a \le X < b) = P(a < X < b)\,.$$

(d) Im Gegensatz zu Wahrscheinlichkeitsfunktionen können Dichtefunktionen *Werte größer als eins* annehmen.

B–1.2 *Warten auf einen Bus.*

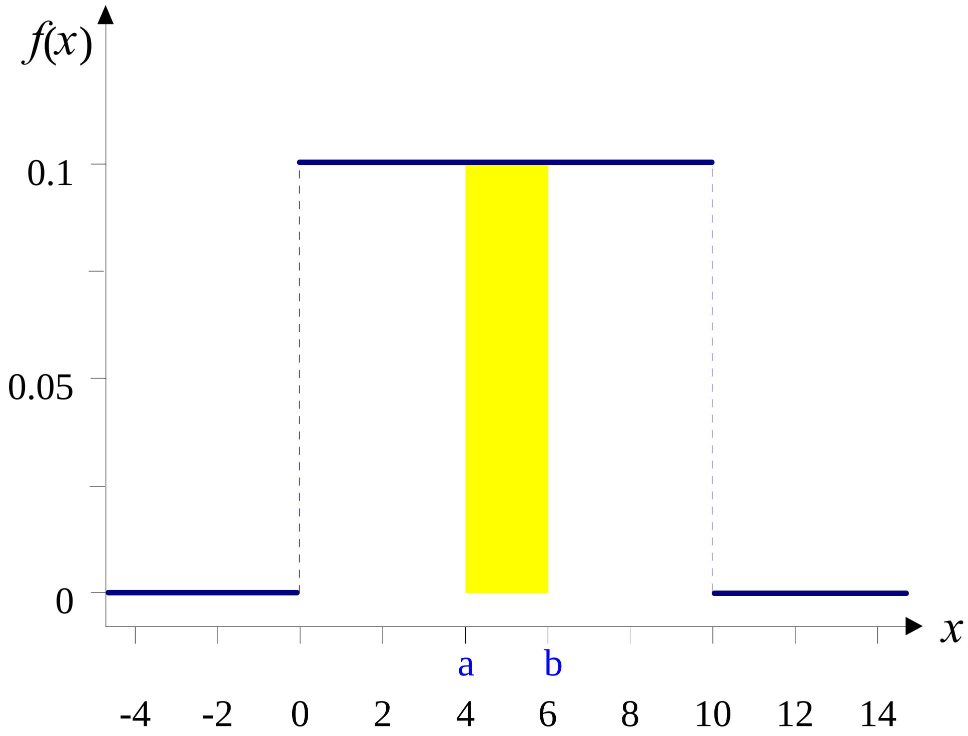
Die Zufallsvariable X besitzt eine überall stetige Verteilungsfunktion

$$F(x) = P(X \le x) = \begin{cases} 0 & \text{falls } x < 0 \\ \frac{1}{10}x & \text{falls } 0 \le x \le 10 \\ 1 & \text{falls } x > 10 \end{cases}$$

und ist somit eine stetige Variable.

$F(x)$ ist für alle reellen Zahlen x mit Ausnahme von $x = 0$ und $x = 10$ differenzierbar. Die zugehörige Dichte erhalten wir für alle $x \ne 0{,}10$ durch Differentiation: $F'(x) = f(x)$. Mit der Vereinbarung $f(0) = f(10) = 1/10$ gilt:

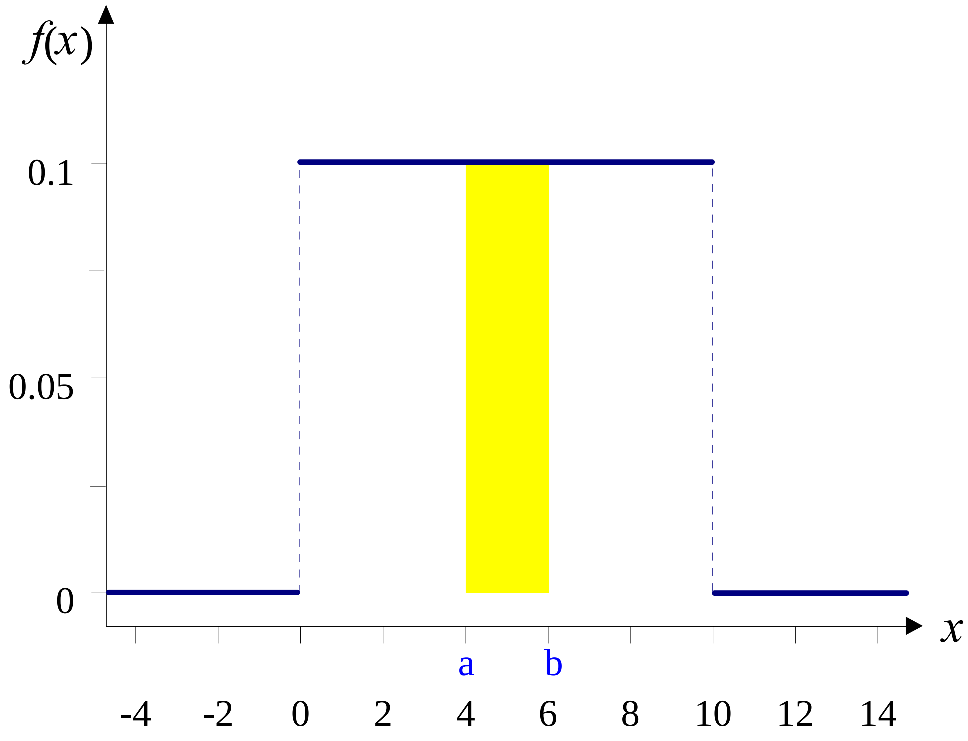
$$f(x) = \begin{cases} \frac{1}{10} & \text{falls } 0 \le x \le 10 \\ 0 & \text{sonst}\,. \end{cases}$$



Die Dichte $f(x)$ ist nicht-negativ.

Die Fläche unter der Funktion entspricht der Fläche des Rechtecks mit der Höhe 1/10 und der Breite 10. Sie ist somit $\frac{1}{10} \cdot 10 = 1$. Also gilt

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



Die Wahrscheinlichkeit eines Ereignisses $a < X \leq b$ entspricht der Fläche unter $f(x)$ zwischen den Abzissenpunkten $x = a$ und $x = b$:

$$P(a < X \leq b) = \int_a^b \frac{1}{10} dx = \frac{1}{10} x \Big|_a^b = \frac{b-a}{10} .$$

Für $a \to b$ strebt die Fläche des Streifens mit der Höhe 1/10 und der Breite $b–a$ gegen null. Da im Intervall [0, 10] nichtabzählbar viele Zeitpunkte liegen, muss die Wahrscheinlichkeit eines einzelnen Zeitpunkts null sein.

Die Wahrscheinlichkeitsverteilung von X in B–1.2 ist ein Spezialfall eines allgemeineren Verteilungsmodells, dass im Folgenden zur Verdeutlichung von Grundkonzepten genutzt werden soll:

D–1.17 Stetige Gleichverteilung oder Rechteckverteilung

Eine stetige Zufallsvariable X heißt *gleich- oder rechteckverteilt* über dem Intervall $[\alpha,\beta] \subset IR$, wenn ihre Dichte die Form

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{falls } \alpha \leq x \leq \beta \\ 0 & \text{sonst} \end{cases}$$

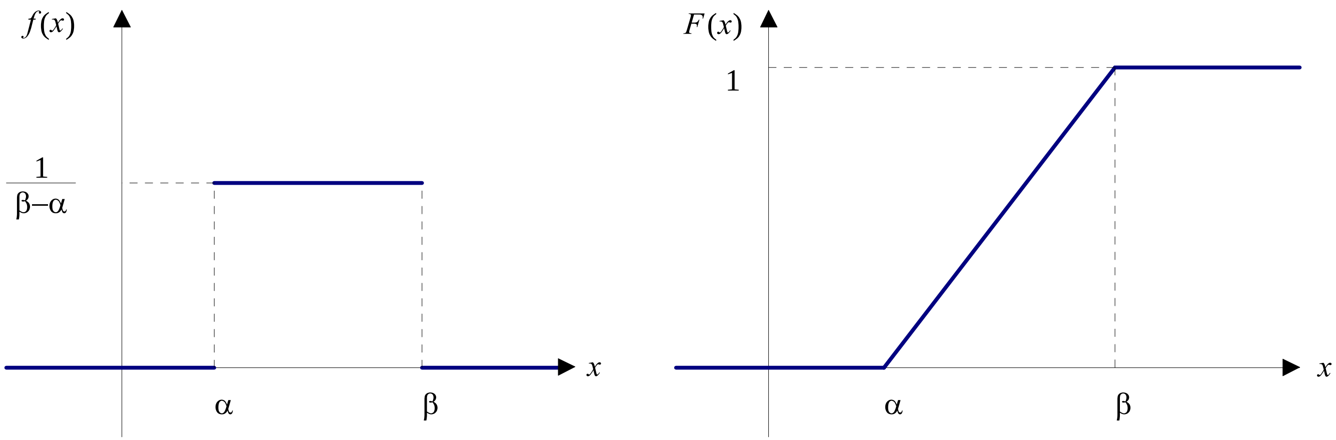
besitzt. Wir schreiben kurz $X \sim R(\alpha,\beta)$ und bezeichnen α, β als Parameter des Verteilungsmodells.

Die Wahrscheinlichkeitsverteilung ist das stetige Analogon zur diskreten Gleichverteilung. Setzen wir $\alpha = 0$ und $\beta = 10$ folgt das Beispiel 1.2 als Spezialfall.

Die Verteilungsfunktion der allgemeinen Rechteckverteilung besitzt die einfache Form

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{falls } x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \text{falls } \alpha \leq x \leq \beta \\ 1 & \text{falls } x > \beta \end{cases}$$

wegen $\int_{-\infty}^{x} f(t)dt = \int_{\alpha}^{x} \frac{1}{\beta-\alpha} dt = \left. \frac{t}{\beta-\alpha} \right|_{\alpha}^{x} = \frac{x}{\beta-\alpha} - \frac{\alpha}{\beta-\alpha} = \frac{x-\alpha}{\beta-\alpha}$ für $\alpha \leq x \leq \beta$.



Graph der Dichte- und Verteilungsfunktion einer rechteckverteilten Zufallsvariable

1.7 Momente (Kennzahlen) einer Wahrscheinlichkeitsverteilung

D–1.18 Mittelwert

Es sei X eine Zufallsvariable mit der Wahrscheinlichkeits- bzw. Dichtefunktion $f(x)$. Dann heißt die reelle Zahl

$$E(X) = \mu = \begin{cases} \sum_i x_i \cdot f(x_i) & \text{falls } X \text{ diskret mit den Sprungstellen } x_i \ (i = 1,2,\ldots) \text{ ist} \\ \int_{-\infty}^{\infty} x \cdot f(x)dx & \text{falls } X \text{ stetig ist} \end{cases}$$

Mittelwert von X oder *Erwartungswert von X*. $E(\bullet)$ heißt *Erwartunsgwertoperator*.

Der Mittelwert μ beschreibt das Lagezentrum einer Wahrscheinlichkeitsverteilung. Im Falle einer diskreten Zufallsvariablen X ist μ das gewogene arithmetische Mittel aller Sprungstellen x_i von X mit deren Einzelwahrscheinlichkeiten $f(x_i) = P(X = x_i)$ als Gewichte. Ist X stetig, dann erfolgt – grob gesprochen – eine Mittlung mit der Dichte von X als Gewichtsfunktion.

D–1.19 Varianz und Standardabweichung

Es sei X eine Zufallsvariable mit der Wahrscheinlichkeits- bzw. Dichtefunktion $f(x)$ und dem Mittelwert μ. Dann heißt die nicht-negative reelle Zahl

$$E[(X-\mu)^2] = \begin{cases} \sum_i (x_i - \mu)^2 \cdot f(x_i) & \text{falls } X \text{ diskret mit den Sprungstellen } x_i \ (i=1,2,\ldots) \text{ ist} \\ \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx & \text{falls } X \text{ stetig ist} \end{cases}$$

Varianz von X. Es ist auch die Schreibweise

$$Var(X) = E[(X-\mu)^2] = \sigma^2$$

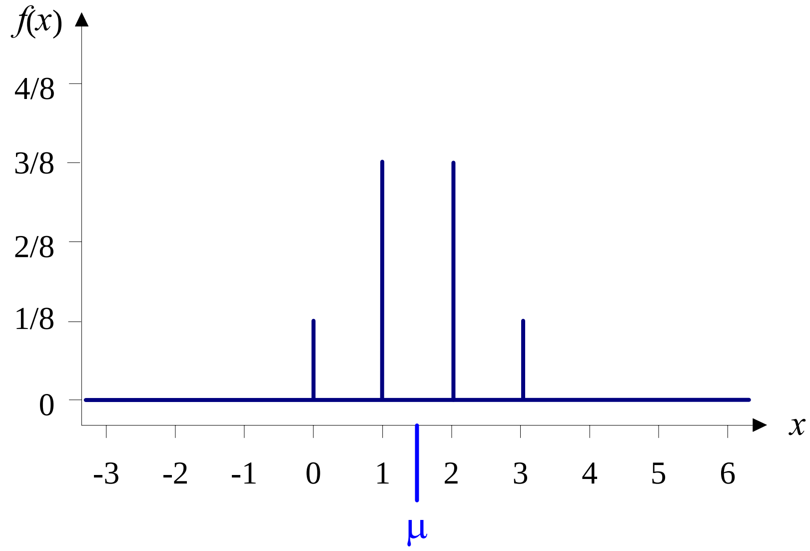
üblich. $Var(\bullet)$ heißt *Varianzoperator*. Die Quadratwurzel der Varianz

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

heißt *Standardabweichung von X.*

B–1.10 *Münzspiel.*

$$f(x) = \begin{cases} 1/8 & \text{falls } x = 0 \text{ oder } x = 3 \\ 3/8 & \text{falls } x = 1 \text{ oder } x = 2 \\ 0 & \text{sonst}\,. \end{cases}$$



Durchschnittliche Gewinnerwartung eines Spielers

$$E(X) = \mu = \sum_{i=1}^{4} x_i \cdot f(x_i)$$

$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5\,[\text{Euro}]$$

Varianz und die Standardabweichung

$$Var(X) = \sigma^2 = \sum_{i=1}^{4} (x_i - \mu)^2 \cdot f(x_i)$$

$$= (0 - 1.5)^2 \cdot \frac{1}{8} + (1 - 1.5)^2 \cdot \frac{3}{8}$$

$$+ (2 - 1.5)^2 \cdot \frac{3}{8} + (3 - 1.5)^2 \cdot \frac{1}{8}$$

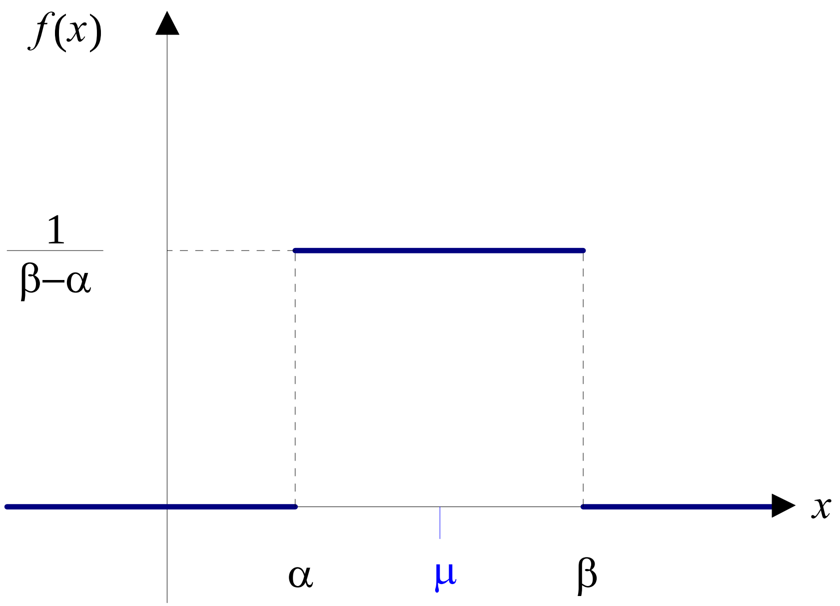
$$= 0.75\,[\text{Euro}^2]$$

$$\sigma = \sqrt{0.75} \cong 0.866 \;\; [\text{Euro}]$$

B–1.11 *Stetige Rechteckverteilung.*

Es ist $X \sim R(\alpha,\beta)$ mit der Dichte:

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{falls } \alpha \le x \le \beta \\ 0 & \text{sonst} \end{cases}$$



Mittelwert

$$E(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta-\alpha}\, dx$$

$$= \frac{1}{\beta-\alpha} \cdot \int_{\alpha}^{\beta} x dx = \frac{1}{\beta-\alpha} \cdot \frac{x^2}{2}\bigg|_{\alpha}^{\beta}$$

$$= \frac{1}{\beta-\alpha} \cdot \left(\frac{\beta^2}{2} - \frac{\alpha^2}{2}\right) = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \frac{\alpha+\beta}{2}$$

Varianz

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

$$= \int_{\alpha}^{\beta} \left(x - \frac{\alpha+\beta}{2}\right)^2 \cdot \frac{1}{\beta-\alpha}\, dx = \frac{(\beta-\alpha)^2}{12}$$

B–1.2 *Warten auf einen Bus.*

Für unser Bus-Beispiel ergibt sich mit $\alpha = 0$ und $\beta = 10$ eine zu erwartende Wartezeit und eine Varianz von

$$\mu = \frac{0+10}{2} = 5 \;\; [\text{ Minuten }]\,, \quad \sigma^2 = \frac{(10-0)^2}{12} = 8.\overline{33} \;\; [\text{ Minuten}^2\,]\,, \quad \sigma = 2.8868 \;\; [\text{Minuten}]\,.$$

S–1.10 Eigenschaften des Erwartungswertoperators

Es seien X eine Zufallsvariable, c, c_1, c_2 reelle Konstanten und g_1, g_2 stetige Funktionen. Dann gilt:

(*i*) $E(c) = c$;

(*ii*) $E(c \cdot X) = c \cdot E(X)$;

(*iii*) $E[c_1 \cdot g_1(X) + c_2 \cdot g_2(X)] = c_1 \cdot E[g_1(X)] + c_2 \cdot E[g_2(X)]$.

(*ohne Beweis*)

S–1.11 Verschiebungssatz der Varianz

Sei X eine Zufallsvariable mit $E(X) = \mu$ und $Var(X) = \sigma^2$. Dann gilt:

$$Var(X) = E(X^2) - E(X)^2 = E(X^2) - \mu^2.$$

Beweis

$$Var(X) = E\left[(X-\mu)^2\right] = E\left(X^2 - 2\mu X + \mu^2\right) = E\left(X^2\right) - 2\mu E(X) + \mu^2 = E\left(X^2\right) - \mu^2 \quad q.e.d.$$

B–1.11 *Stetige Rechteckverteilung.*

Wir wissen

$$E(X) = \mu = \frac{\alpha+\beta}{2}.$$

Ferner ist

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_{\alpha}^{\beta} x^2 \cdot \frac{1}{\beta-\alpha} dx = \frac{1}{\beta-\alpha} \cdot \int_{\alpha}^{\beta} x^2 dx = \frac{1}{\beta-\alpha} \cdot \left. \frac{x^3}{3} \right|_{\alpha}^{\beta} = \frac{1}{\beta-\alpha} \cdot \left(\frac{\beta^3}{3} - \frac{\alpha^3}{3} \right)$$

$$= \frac{\beta^3 - \alpha^3}{3(\beta-\alpha)} \quad .$$

Mit Hilfe von Satz 1.11 erhalten wir

$$Var(X) = E(X^2) - E(X)^2 = \frac{\beta^3 - \alpha^3}{3(\beta-\alpha)} - \frac{(\alpha+\beta)^2}{4} = \frac{(\beta-\alpha)^2}{12} \quad .$$

1.8 Lineartransformationen von Zufallsvariablen

Die Zufallsvariable

$$Y = a_0 + a_1 X \quad \text{mit} \quad a_0, a_1 \in IR \text{ und } a_1 \neq 0$$

heißt *Lineartransformation* der Zufallsvariable X.

S–1.12 Mittelwert und Varianz lineartransformierter Zufallsvariablen

Es sei X eine Zufallsvariable mit dem Mittelwert $E(X) = \mu_X$ und der Varianz $Var(X) = \sigma_X^2$, dann gilt:

$$E(Y) = E(a_0 + a_1 X) = a_0 + a_1 E(X) \qquad \text{bzw.} \qquad \mu_Y = a_0 + a_1 \mu_X \ ,$$

$$Var(Y) = Var(a_0 + a_1 X) = a_1^2 Var(X) \qquad \text{bzw.} \qquad \sigma_Y^2 = a_1^2 \sigma_X^2 \ .$$

D–1.20 Standardisierte Zufallsvariable

Es sei X eine Zufallsvariable mit dem Mittelwert $E(X) = \mu$ und der Varianz $Var(X) = \sigma^2$, dann heißt

$$Z = \frac{X - E(X)}{\sqrt{Var(X)}} \quad \text{bzw.} \quad Z = \frac{X - \mu}{\sigma}$$

die *standardisierte Form von X* oder einfach *standardisierte Zufallsvariable*.

Setzt man $a_0 = -\mu/\sigma$ und $a_1 = 1/\sigma$, dann lässt sich Z auch in der Form

$$Z = a_0 + a_1 X$$

schreiben. Aus Satz 1.12 folgt $E(Z) = 0$ und $Var(Z) = 1$.

1.9 Normalverteilung oder GAUSS-Verteilung

Eine zentrale Rolle in der Statistik spielt die auf ABRAHAM DE MOIVRE (1667-1754) und CARL FRIEDRICH GAUSS (1777-1855) zurückgehende Normalverteilung.

D–1.21 Normalverteilung

Eine stetige Zufallsvariable X heißt *normalverteilt* mit den reellwertigen Parametern μ und $\sigma^2 > 0$, kurz $X \sim N(\mu, \sigma^2)$, wenn sie eine Dichte der Form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (-\infty < x < \infty)$$

besitzt. Im Spezialfall $X \sim N(0,1)$ mit der Dichte

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad (-\infty < x < \infty)$$

bezeichnet man X als *standardnormalverteilt*.

Die Parameter werden durch die Symbole μ und σ^2 gekennzeichnet, weil sie mit dem Mittelwert bzw. mit der Varianz der Verteilung identisch sind.

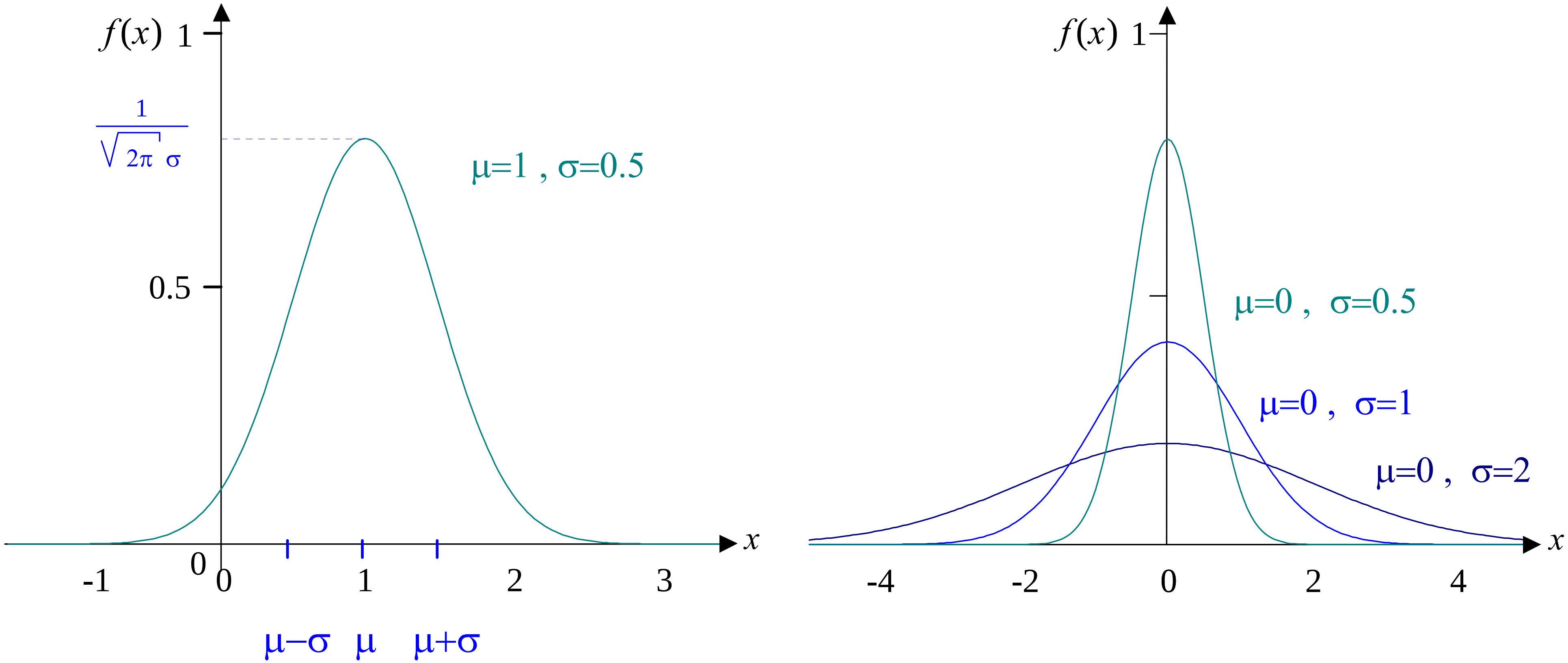
S–1.13 Mittelwert und Varianz der Normalverteilung

Sei $X \sim N(\mu, \sigma^2)$, dann ist $E(X) = \mu$ und $Var(X) = \sigma^2$.

(*ohne Beweis*)

Die Dichte $f(x)$ der Normalverteilung besitzt folgende *Eigenschaften*:

- Sie besitzt ein Maximum $f(x) = \frac{1}{\sigma\sqrt{2\pi}}$ an der Stelle $x = \mu$.
- Sie ist symmetrisch bezüglich $x = \mu$.
- Sie besitzt zwei Wendepunkte an den Stellen $x = \mu \pm \sigma$.



Dichtefunktionen von Normalverteilungen mit verschiedenen Parameterwerten

Die Normalverteilung ist reproduktiv.

S–1.14 Reproduktivität

Es seien $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$ unabhängige Zufallsvariablen, dann gilt für alle reellen Konstanten a, b mit $a, b \neq 0$:

$$aX + bY \sim N\left(a\mu_X + b\mu_Y \ , \ a^2\sigma_X^2 + b^2\sigma_Y^2\right).$$

Speziell gilt für die ungewogene Summe

$$X + Y \sim N\left(\mu_X + \mu_Y \ , \ \sigma_X^2 + \sigma_Y^2\right)$$

(*ohne Beweis*)

Bedeutung der Normalverteilung

(*a*) Viele Zufallsvariablen, die in Experimenten oder Stichprobenerhebungen in der Praxis auftreten, sind zumindest näherungsweise normalverteilt. Dies gilt insbesondere für biologische, physikalische oder technische Variablen wie z.B. die Größe und das Gewicht von Menschen gleichen Geschlechts, der Benzinverbrauch von Autos gleichen Typs, die Abweichung produzierter Schrauben von ihrem Solldurchmesser, etc. Zahlreiche ökonomische Variablen wie z.B. Haushaltseinkommen oder Unternehmensumsätze folgen allerdings nicht einer Normalverteilung.

(*b*) Gewisse nicht normalverteilte Zufallsvariablen lassen sich derart transformieren, dass die resultierenden Variablen normalverteilt sind.

(*c*) Die Normalverteilung ist eine Grenzverteilung zahlreicher anderer Verteilungen, z.B. der Binomialverteilung und der Poisson-Verteilung. Dies ermöglicht unter bestimmten Bedingungen die Approximation dieser Verteilungen durch die Normalverteilung.

(*d*) In statistischen Schätz- und Testverfahren kommen oft Größen vor, die normalverteilt sind oder sich bei Grenzübergängen einer Normalverteilung nähern. Letzteres findet eine theoretische Begründung durch den Zentralen Grenzwertsatz.

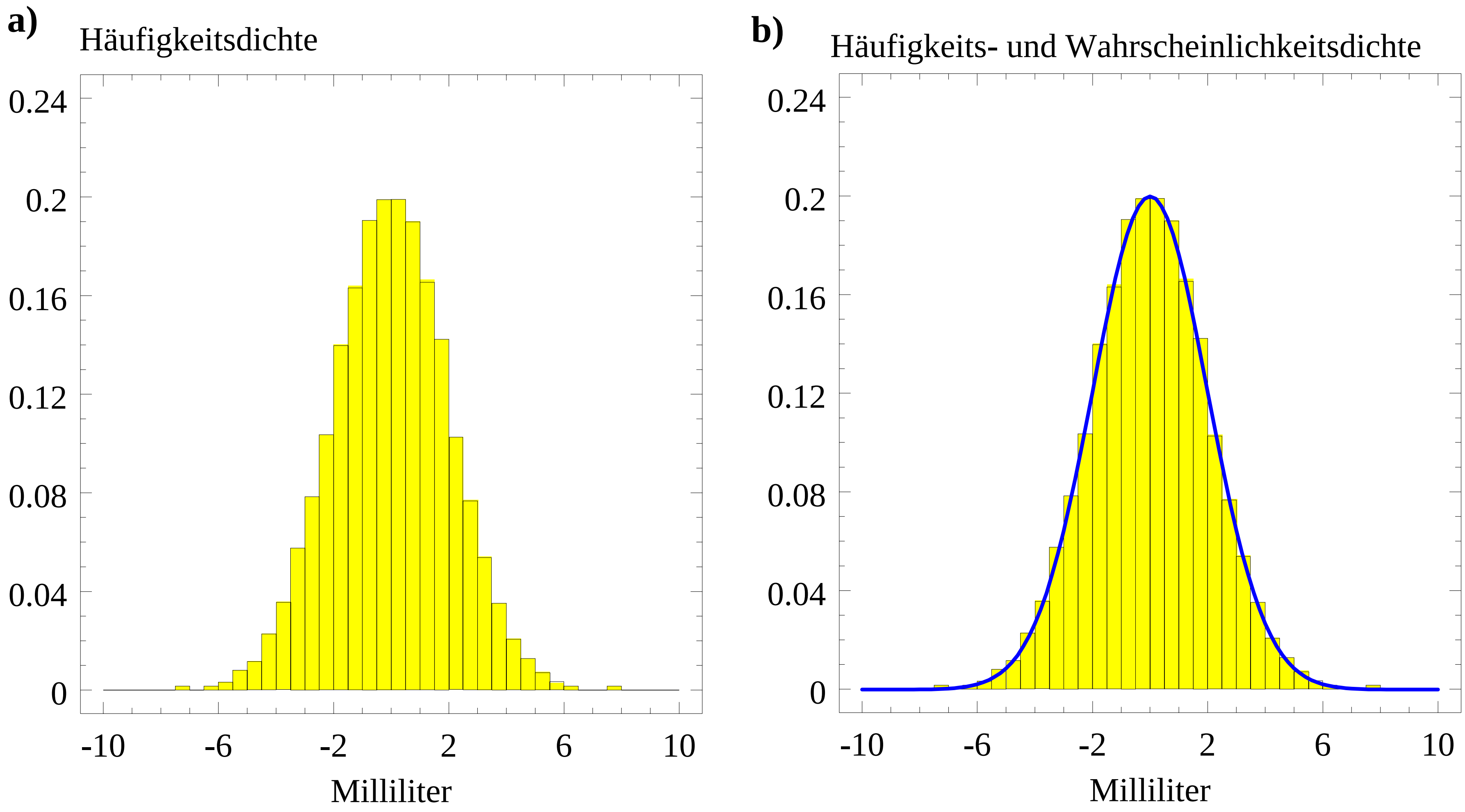
B–1.12 *Bierabfüllung.*

Der Füllinhalt von Bierflaschen wird durch zufällige Effekte bei der Abfüllung beeinflusst (z.B. Raumtemperatur, Luftdruck, etc.).

Der Qualitätssicherungsexperte einer Brauerei hat der laufenden Produktion 1000 Bierflaschen mit einer Sollabfüllung von 0.5 Litern entnommen und jeweils die Abweichung x des Flascheninhalts von der Sollabfüllung in Bruchteilen von Millilitern gemessen. Die Daten weisen den empirischen Mittelwert $\bar{x} \approx 0$ und die empirische Varianz $\tilde{s}^2 \approx 4$ auf. Die folgende Abbildung zeigt das Histogramm der Daten. Hierbei liegen Zählklassen mit einer Breite von 0.5 Millilitern zugrunde.

Die Histogrammhöhen können als eine Funktion von x aufgefasst werden [*Häufigkeitsdichte* $d(x)$]. Die Funktion ist nicht-negativ, die Fläche unter ihrem Graphen ist auf den Flächeninhalt 1 normiert.

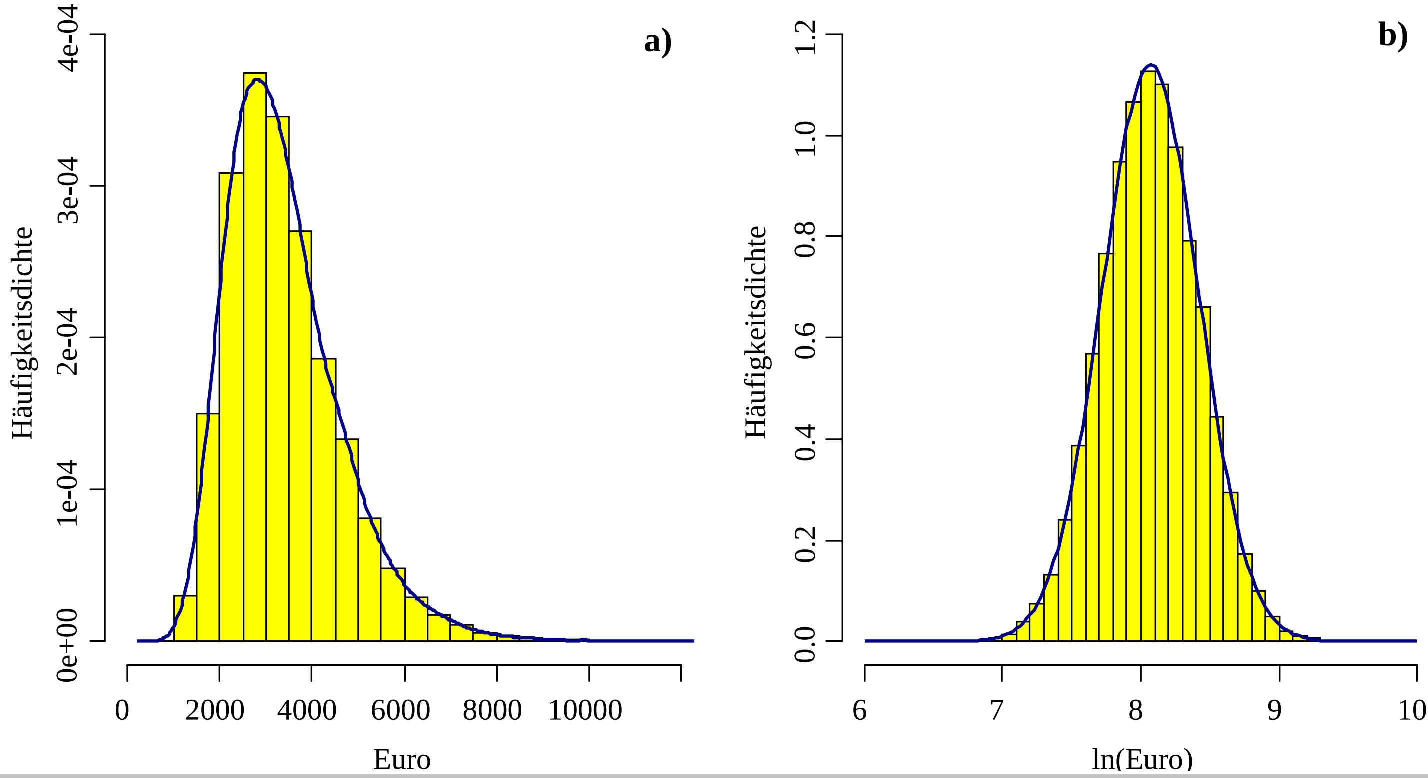
In der Abbildung wird die unstetige Häufigkeitsdichte durch eine stetige Funktion angenähert. Dies ist die (Wahrscheinlichkeits-) Dichte der $N(0,4)$-Verteilung. Die Funktionen stimmen gut überein. Es erscheint begründet, die $N(0,4)$-Verteilung als ein aufgrund von Beobachtungsdaten statistisch gewonnenes Modell für die Füllabweichungen der Bierflaschen zu verwenden.



Histogramm und approximierende Dichtefunktion der $N(0,4)$-Verteilung

B–1.13 *Preise von Gebrauchtwagen.*

Bei ökonomischen Variablen wie Güterpreisen, Unternehmensumsätzen oder Haushaltseinkommen und -vermögen beobachtet man oft eine *rechtsschiefe* Verteilung. Beispielsweise besitzen die im Jahr 2003 erhobenen Verkaufspreise von 5000 Gebrauchtwagen des Typs VW Golf die in der folgenden Abbildung dargestellte Verteilung ($\bar{x} = 3396.54$ Euro, $\tilde{s} = 1246.11$ Euro). Die logarithmischen Verkaufspreise sind näherungsweise normalverteilt.



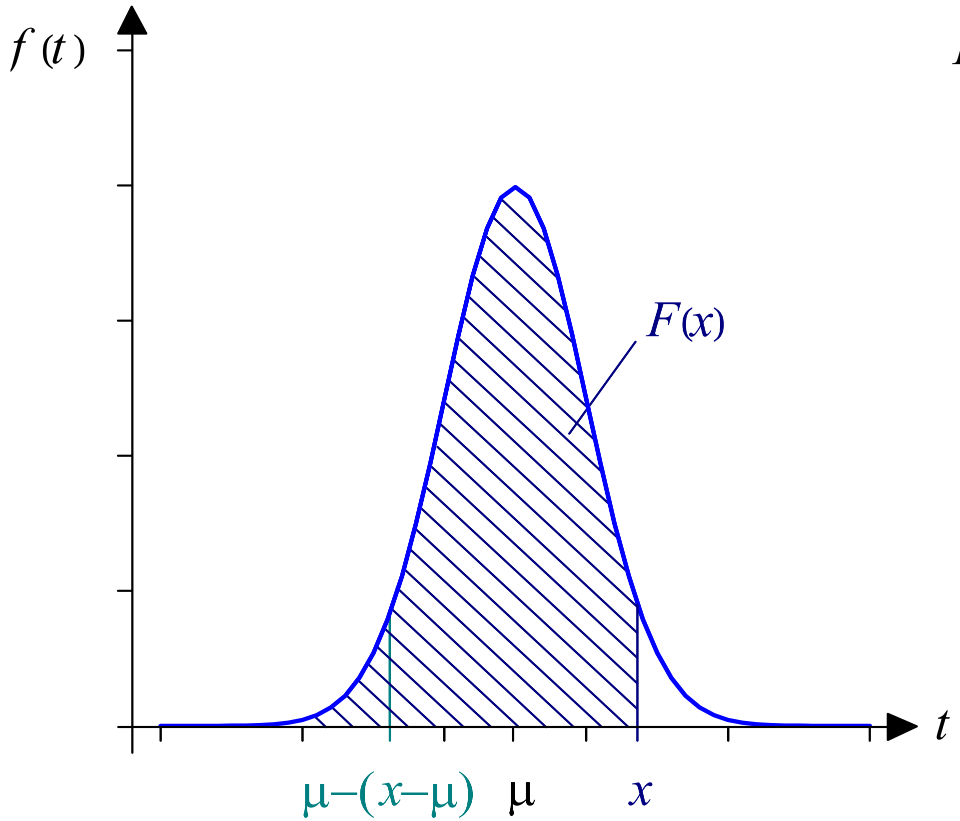
Verteilungsfunktion der Normalverteilung

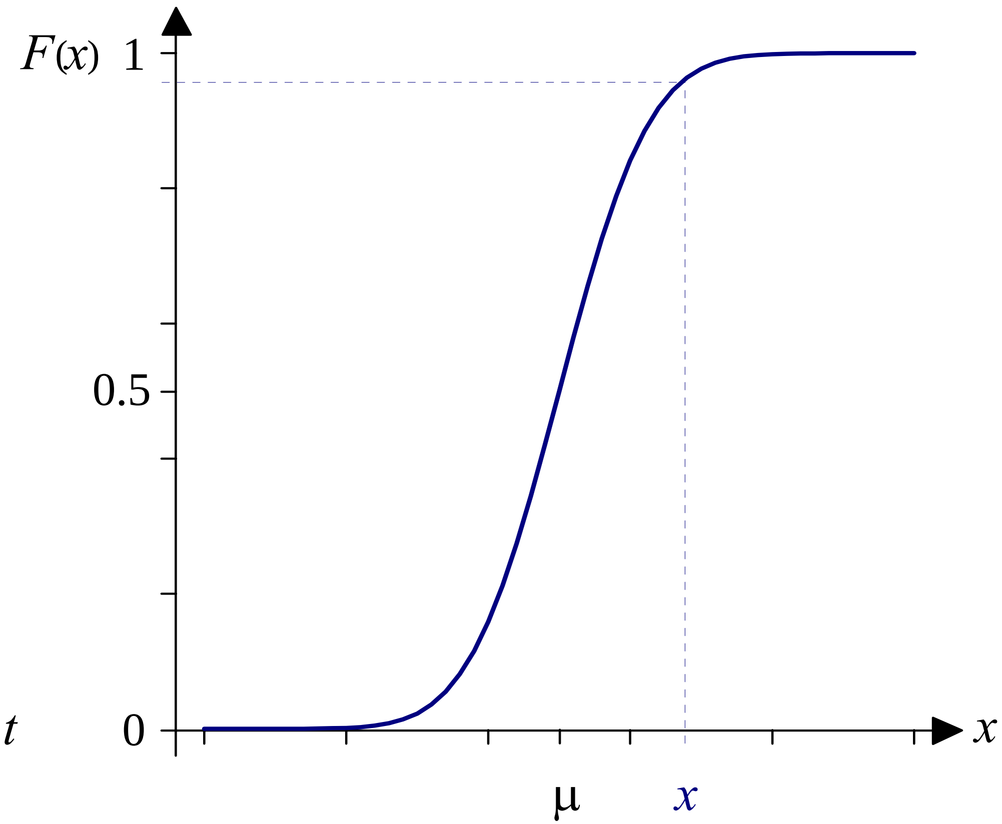
Die Verteilungsfunktion einer Variablen $X \sim N(\mu, \sigma^2)$ ist

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt .$$

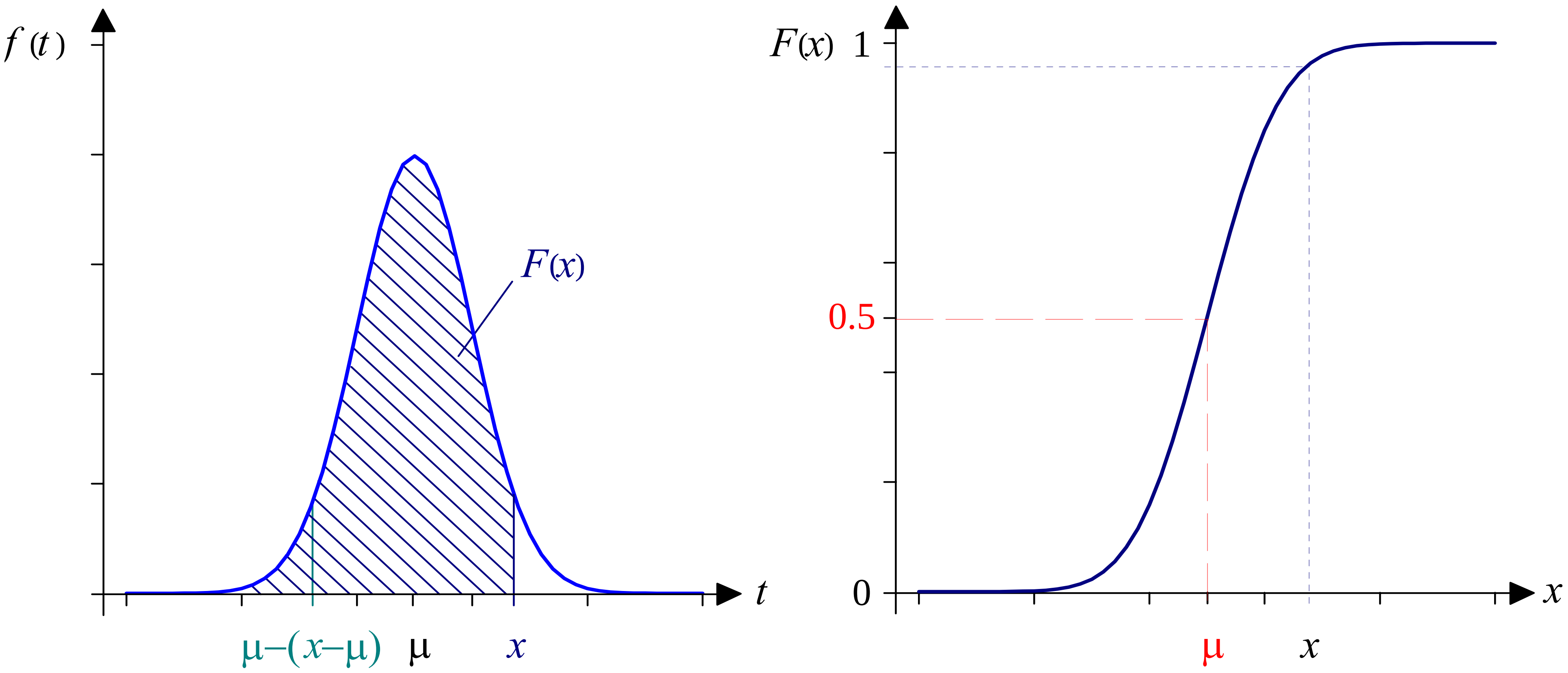
Die Verteilungsfunktion besitzt folgende *Eigenschaften*:

- $F(x) \to 0$ für $x \to -\infty$ und $F(x) \to 1$ für $x \to \infty$;
- Sie ist sigmoid (S-förmig);
- Sie besitzt einen Wendepunkt bei $x = \mu$ mit $F(\mu) = \int_{-\infty}^{\mu} f(x)dx = 0.5$;
- $F(x) = 1 - F(\mu - (x - \mu))$.





Zusammenhang zwischen Dichte- und Verteilungsfunktion einer Normalverteilung



Zusammenhang zwischen Dichte- und Verteilungsfunktion einer Normalverteilung

Bekanntlich können mit Hilfe der Verteilungsfunktion alle interessierenden Wahrscheinlichkeitsaussagen beantworten werden, z.B.

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

$$P(X > x) = 1 - F(x) \ ,$$

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

Hierbei stoßen wir allerdings auf ein ***Problem***:

Es existiert _*keine*_ Stammfunktion der Dichte der Normalverteilung, so dass das Integral $F(x)$ nicht *analytisch* lösbar ist. Das Integral kann nur mit Näherungsverfahren der numerischen Mathematik gelöst werden, wofür es eines Computers bedarf.

Die übliche Lösung dieses Problems besteht darin, dass man anstelle von $X \sim N(\mu,\sigma^2)$ die *standardisierte Zufallsvariable*

$$Z = \frac{X-\mu}{\sigma}$$

mit $\mu = E(X)$ und $\sigma^2 = Var(X)$ betrachtet, für die

$$E(Z) = 0 \qquad \text{und} \qquad Var(Z) = 1$$

gilt. Z besitzt eine $N(0,1)$- oder *Standardnormalverteilung* mit der Verteilungsfunktion

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt \ .$$

Mit Hilfe von $\Phi(z)$ können für alle beliebigen Normalverteilungen die Verteilungsfunktionen bestimmt werden.

Mit Hilfe von $\Phi(z)$ können für alle beliebigen Normalverteilungen die Verteilungsfunktionen bestimmt werden:

$$F\left(x \mid \mu, \sigma^2\right) = P(X \le x) = P(\underbrace{\sigma Z + \mu}_{=X} \le x) = P(Z \le \frac{x-\mu}{\sigma}) = \Phi(\underbrace{\frac{x-\mu}{\sigma}}_{=z}) = \Phi(z).$$

Des Weiteren gilt

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

Die Funktion $\Phi(z)$ ist in allen gängigen Statistik-Lehrbüchern tabelliert (siehe auch *Formelsammlung*, S. 47). Meist ist sie tabelliert für $z \ge 0$. Für $z < 0$ gilt:

$$\Phi(z) = 1 - \Phi(-z) .$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.50399	.50798	.51197	.51596	.51994	.52393	.52791	.53189	.53586
0.1	.53983	.54380	.54776	.55172	.55568	.55962	.56356	.56750	.57143	.57535
0.2	.57926	.58317	.58707	.59096	.59484	.59871	.60257	.60642	.61027	.61410
0.3	.61792	.62172	.62552	.62931	.63308	.63684	.64058	.64431	.64803	.65174
0.4	.65543	.65910	.66276	.66641	.67004	.67365	.67725	.68083	.68439	.68794
0.5	.69147	.69498	.69847	.70195	.70541	.70885	.71227	.71567	.71905	.72241
0.6	.72575	.72907	.73238	.73566	.73892	.74216	.74538	.74858	.75175	.75491
0.7	.75804	.76115	.76424	.76731	.77036	.77338	.77638	.77936	.78231	.78524
0.8	.78815	.79103	.79390	.79674	.79955	.80234	.80511	.80785	.81058	.81327
0.9	.81594	.81859	.82122	.82382	.82640	.82895	.83148	.83398	.83646	.83892
1.0	.84135	.84376	.84614	.84850	.85084	.85315	.85543	.85770	.85993	.86215
1.1	.86434	.86651	.86865	.87077	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88494	.88687	.88877	.89066	.89252	.89436	.89617	.89796	.89973	.90148
1.3	.90320	.90491	.90659	.90825	.90988	.91150	.91309	.91466	.91621	.91774
1.4	.91925	.92074	.92220	.92365	.92507	.92648	.92786	.92922	.93057	.93189
1.5	.93320	.93448	.93575	.93700	.93822	.93943	.94063	.94180	.94295	.94409
1.6	.94521	.94631	.94739	.94845	.94950	.95053	.95155	.95255	.95353	.95449
1.7	.95544	.95637	.95729	.95819	.95908	.95995	.96080	.96164	.96247	.96328
1.8	.96407	.96486	.96563	.96638	.96712	.96785	.96856	.96926	.96995	.97063
1.9	.97129	.97194	.97258	.97320	.97382	.97442	.97501	.97559	.97615	.97671
2.0	.97725	.97779	.97831	.97883	.97933	.97982	.98031	.98078	.98124	.98170

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
2.0	.97725	.97779	.97831	.97883	.97933	.97982	.98031	.98078	.98124	.98170
2.1	.98214	.98258	.98300	.98342	.98383	.98423	.98462	.98500	.98538	.98574
2.2	.98610	.98645	.98680	.98713	.98746	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99062	.99087	.99111	.99135	.99158
2.4	.99181	.99203	.99224	.99246	.99266	.99286	.99306	.99325	.99344	.99362
2.5	.99380	.99397	.99414	.99430	.99446	.99462	.99477	.99492	.99506	.99521
2.6	.99534	.99548	.99561	.99574	.99586	.99598	.99610	.99621	.99632	.99643
2.7	.99654	.99664	.99674	.99684	.99693	.99703	.99711	.99720	.99729	.99737
2.8	.99745	.99753	.99760	.99768	.99775	.99782	.99789	.99795	.99802	.99808
2.9	.99814	.99820	.99825	.99831	.99836	.99842	.99847	.99852	.99856	.99861
3.0	.99866	.99870	.99874	.99878	.99882	.99886	.99890	.99893	.99897	.99900
3.1	.99904	.99907	.99910	.99913	.99916	.99919	.99922	.99924	.99927	.99929
3.2	.99932	.99934	.99936	.99939	.99941	.99943	.99945	.99947	.99949	.99950
3.3	.99952	.99954	.99955	.99957	.99959	.99960	.99962	.99963	.99964	.99966
3.4	.99967	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99979	.99980	.99981	.99982	.99983	.99984	.99985	.99986

B–1.14 *Körpergröße britischer Autofahrer.*

PKW-Ausstattungsdesigner müssen bei ihrer Arbeit die Körpergröße potenzieller Autofahrer berücksichtigen. Eine Untersuchung der Körpergröße X (in mm) männlicher britischer Autofahrer ergab, dass X näherungsweise $N(1738, 68^2)$-verteilt ist. Es gilt dann z. B.

$$P(X \leq 1600) = \Phi\left(\frac{1600 - 1738}{68}\right) \approx \Phi(-2.03)$$

$$= 1 - \Phi(2.03) = 1 - 0.97883 = 0.02117$$

oder

$$P(X \leq 1850) = \Phi\left(\frac{1850 - 1738}{68}\right) \approx \Phi(1.65) = 0.95053.$$

Quantile und zentrale Schwankungsintervalle

Wir betrachten zunächst eine standardisierte Variable $Z = \frac{X-\mu}{\sigma}$ mit $X \sim N(\mu, \sigma^2)$. Bestimmen wir eine Zahl $z_{[p]}$ so, dass

$$\Phi\left(z_{[p]}\right) = P\left(Z \leq z_{[p]}\right) = p \quad \text{mit} \quad 0 < p < 1$$

gilt, dann ist $z_{[p]}$ das *p-Quantil* der standardnormalverteilten Variable Z. Wegen der *Symmetrie der Dichte* $\varphi(z)$ bzgl. $z = 0$ gilt immer

$$z_{[p]} = -z_{[1-p]}\,.$$

Die p-Quantile sind für $p \geq 0.5$ direkt aus der Tabelle ablesbar.

Zwischen dem p-Quantil $z_{[p]}$ der standardisierten Variable Z und dem p-Quantil $x_{[p]}$ der Zufallsvariable $X \sim N(\mu, \sigma^2)$ besteht der Zusammenhang:

$$z_{[p]} = \frac{x_{[p]} - \mu}{\sigma} \quad \Leftrightarrow \quad x_{[p]} = \mu + z_{[p]} \cdot \sigma.$$

B–1.15 *Quantile von Normalverteilungen.*

Sei $Z \sim N(0{,}1)$. Es gilt

$p = 0.975$	$\Rightarrow$	97.5%-Punkt:	$z_{[0.975]} = 1.96$
$p = 0.025$	$\Rightarrow$	2.5%-Punkt:	$z_{[0.025]} = -z_{[0.975]} = -1.96$
$p = 0.95$	$\Rightarrow$	95%-Punkt:	$z_{[0.95]} = 1.645$
$p = 0.05$	$\Rightarrow$	5%-Punkt:	$z_{[0.05]} = -z_{[0.95]} = -1.645$
$p = 0.5$	$\Rightarrow$	Median:	$z_{[0.5]} = 0$.

Sei jetzt $X \sim N(1{,}4)$. Es gilt

$p = 0.975$	$\Rightarrow$	97.5%-Punkt:	$x_{[0.975]} = 1 + 1.96 \cdot 2 = 4.92$
$p = 0.025$	$\Rightarrow$	2.5%-Punkt:	$x_{[0.025]} = 1 - 1.96 \cdot 2 = -2.92$
$p = 0.95$	$\Rightarrow$	95%-Punkt:	$x_{[0.95]} = 1 + 1.645 \cdot 2 = 4.29$
$p = 0.05$	$\Rightarrow$	5%-Punkt:	$x_{[0.05]} = 1 - 1.65 \cdot 2 = -2.29$
$p = 0.5$	$\Rightarrow$	Median:	$x_{[0.5]} = 1 + 0 \cdot 2 = 1 = \mu$.

Wir betrachten jetzt Ereignisse der Form

$$x_{\left[\frac{\alpha}{2}\right]} \leq X \leq x_{\left[1-\frac{\alpha}{2}\right]} \qquad \text{mit} \qquad 0 < \alpha < 1$$

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

Wegen

$$x_{\left[1-\frac{\alpha}{2}\right]} = \mu + z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma$$

und

$$x_{\left[\frac{\alpha}{2}\right]} = \mu + z_{\left[\frac{\alpha}{2}\right]} \cdot \sigma = \mu - z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma$$

können wir die Ereignisse auch schreiben

$$\mu - z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma \leq X \leq \mu + z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma.$$

Das Zahlenintervall

$$\left[\mu - z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma \quad , \mu + z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma\right]$$

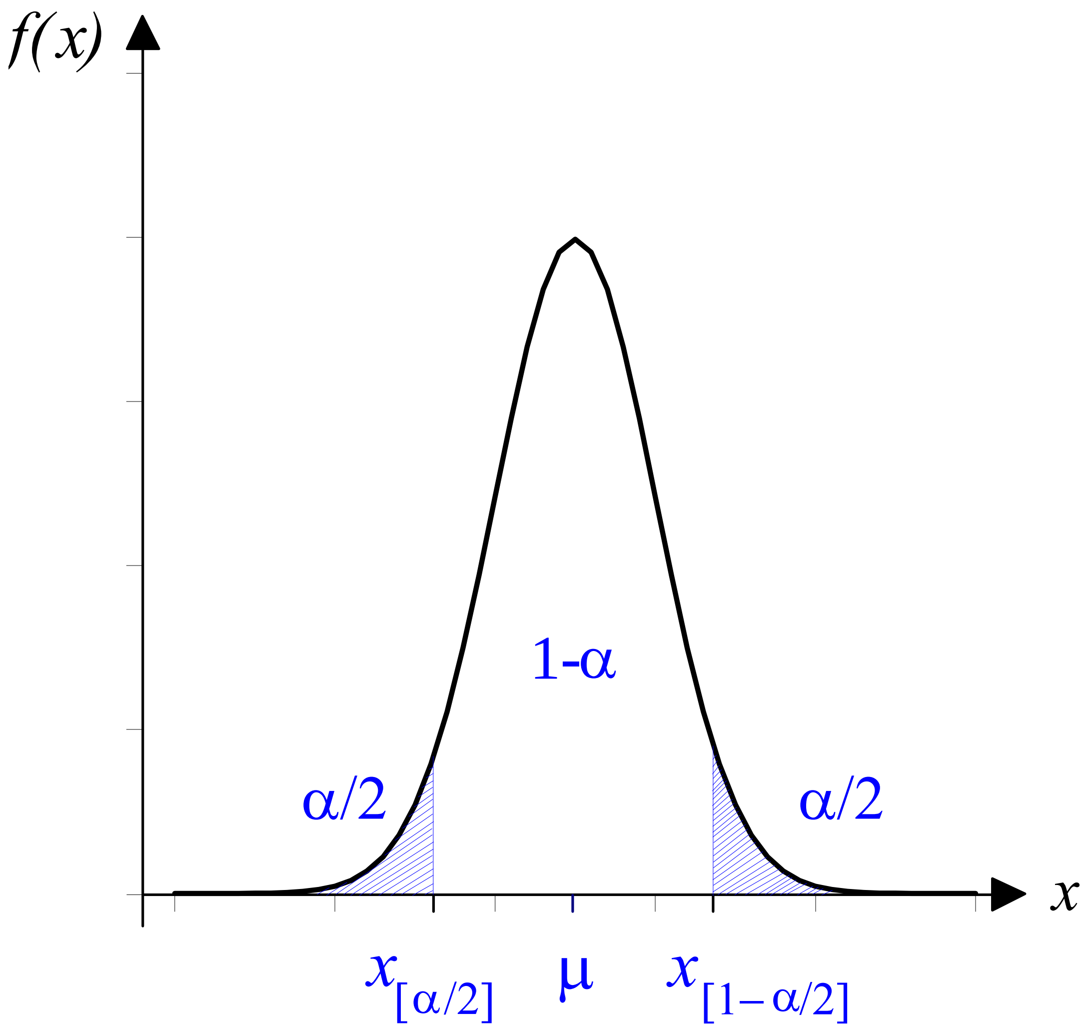
bezeichnet man als ein ***zentrales Schwankungsintervall*** oder als ein ***z–faches σ-Intervall***. Die Wahrscheinlichkeit, dass X sich in einem solchen Intervall realisiert, beträgt

$$P\left(\mu - z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma \le X \le \mu + z_{\left[1-\frac{\alpha}{2}\right]} \cdot \sigma\right) = P\left(-z_{\left[1-\frac{\alpha}{2}\right]} \le \underbrace{\frac{X-\mu}{\sigma}}_{=Z} \le +z_{\left[1-\frac{\alpha}{2}\right]}\right)$$

$$= \Phi\left(z_{\left[1-\frac{\alpha}{2}\right]}\right) - \Phi\left(-z_{\left[1-\frac{\alpha}{2}\right]}\right) = \Phi\left(z_{\left[1-\frac{\alpha}{2}\right]}\right) - \left(1 - \Phi\left(z_{\left[1-\frac{\alpha}{2}\right]}\right)\right)$$

$$= 1 - \frac{\alpha}{2} - \left(1 - \left[1 - \frac{\alpha}{2}\right]\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2}$$

$$= 1 - \alpha \ .$$



Dichtefunktion einer Normalverteilung und zentrales Schwankungsintervall

Man erhält z.B. mit $\alpha = 0.1$ das 1.645-faches σ–Intervall

$$P\left(\mu - z_{[0.95]} \cdot \sigma \leq X \leq \mu + z_{[0.95]} \cdot \sigma\right)$$

$$= P(\mu - 1.645 \cdot \sigma \leq X \leq \mu + 1.645 \cdot \sigma) = 0.9$$

und mit $\alpha = 0.05$ das 1.96-faches σ–Intervall

$$P\left(\mu - z_{[0.975]} \cdot \sigma \leq X \leq \mu + z_{[0.975]} \cdot \sigma\right)$$

$$= P(\mu - 1.96 \cdot \sigma \leq X \leq \mu + 1.96 \cdot \sigma) = 0.95.$$

B–1.14 *Körpergröße britischer Autofahrer.*

PKW-Ausstattungsdesigner müssen bei ihrer Arbeit die Körpergröße potenzieller Autofahrer berücksichtigen. Eine Untersuchung der Körpergröße X (in mm) männlicher britischer Autofahrer ergab, dass X näherungsweise $N(1738,\ 68^2)$-verteilt ist. Es gilt dann

$$x_{[0.05]} = 1738 - 1.645 \cdot 68 = 1626.14,$$

$$x_{[0.95]} = 1738 + 1.645 \cdot 68 = 1849.86.$$

Ein britischer Autofahrer besitzt demnach also mit einer Wahrscheinlichkeit von 90% eine Größe im Intervall [1626.14 , 1849.86].

Zentraler Grenzwertsatz

Der *Zentrale Grenzwertsatz* macht eine Aussage über die Verteilung der Summe von Zufallsvariablen, wenn die Anzahl der n der Summanden über alle Grenzen wächst. Es gibt verschiedene Fassungen des Zentralen Grenzwertsatzes. Wir beschränken uns auf den Satz von *Lindeberg-Lévy*, der ohne Beweis angegeben wird.

Im Folgenden wird die standardisierte Summe von n Zufallsvariablen $X_1,\ldots,X_n$ betrachtet. Bezüglich der Variablen setzen wir voraus:

(i) Die Zufallsvariablen X_i sind *identisch verteilt*, d.h. ihre Verteilungsfunktionen $F_{X_i}(x)$ erfüllen die Bedingung

$$F_{X_1}(x) = \ldots = F_{X_n}(x) \qquad \text{für alle reellen } x.$$

Ferner gelte

$$E(X_i) = \mu \quad \text{und} \quad Var(X_i) = \sigma^2 \quad (i = 1,\ldots,n).$$

Ansonsten können die Variablen beliebig diskret oder stetig verteilt sein.

(ii) Die Zufallsvariablen sind *stochastisch unabhängig*.

Die Summe der Variablen besitzt aufgrund der Voraussetzungen den Mittelwert

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n \cdot \mu$$

und die Varianz

$$Var\left(\sum_{i=1}^{n} X_i\right) \overset{\text{Unabhängigkeit}}{=} \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} \sigma^2 = n \cdot \sigma^2 .$$

Durch Standardisierung der Summe erhält man die Zufallsvariable

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sqrt{n \cdot \sigma^2}} = \frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sqrt{n} \cdot \sigma}$$

mit $E(Z_n) = 0$ und $Var(Z_n) = 1$.

S–1.15 Grenzwertsatz von Lindeberg-Lévy

$X_1,\dots,X_n$ seien unabhängige und identisch verteilte Zufallsvariablen, die alle den Mittelwert μ und die Varianz σ^2 besitzen. Dann konvergiert die Verteilungsfunktion $F_{Z_n}(z) = P(Z_n \leq z)$ der Zufallsvariablen

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sqrt{n} \cdot \sigma}$$

mit wachsender Summandenzahl n gegen die Verteilungsfunktion $\Phi(z)$ der $N(0,1)$-Verteilung:

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z) \qquad (-\infty < z < \infty)\ .$$

Die *praktische Bedeutung* des Satzes liegt darin, dass die Verteilung der Zufallsvariablen Z_n bereits für *endliches n* durch die $N(0,1)$-Verteilung *angenähert* werden kann. Entsprechend kann die Verteilung der nicht standardisierten Summe $\sum_{i=1}^{n} X_i$ für endliches n durch die $N(n\mu\,, n\sigma^2)$–Verteilung approximiert werden.

Die Güte der Näherung hängt natürlich von der Verteilung der Zufallsvariablen X_i ab.

B–1.16 *Würfelexperiment.*

Ein regulärer Würfel werde n-mal geworfen. Wir definieren die Zufallsvariablen

$$X_i = \text{„Augenzahl beim } i\text{-ten Wurf“} \qquad (i = 1, \ldots, n),$$

sowie deren Summe und deren standardisierte Summe

$$Y_n = \sum_{i=1}^{n} X_i = \text{„Summe der Augenzahlen bei } n \text{ Würfen“ },$$

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{Var(Y_n)}} .$$

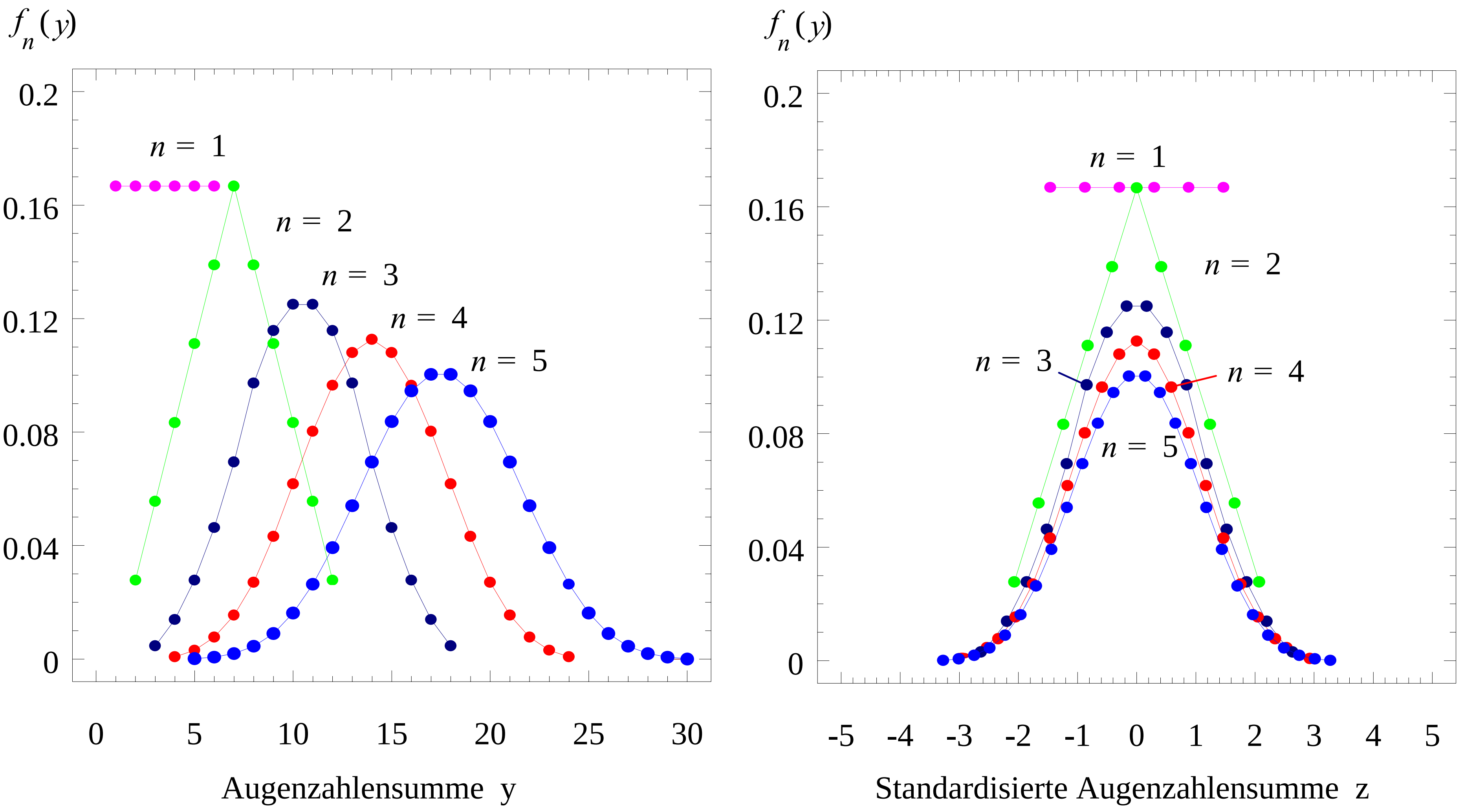
Die Wahrscheinlichkeitsfunktionen $f_n(y)$ der Summenvariable für $n = 1,2,3,4,5$ weist die Tabelle aus.

Die folgende Abbildung gibt ferner eine graphische Darstellung der Wahrscheinlichkeitsfunktionen von Y_n und von Z_n.

y	Wahrscheinlichkeitsfunktion $f_n(y) = P(Y_n = y)$ für $n = 1$	$n = 2$	$n = 3$ multipliziert mit	$n = 4$	$n = 5$
	6	6^2	6^3	6^4	6^5
1	1				
2	1	1			
3	1	2	1		
4	1	3	3	1	
5	1	4	6	4	1
6	1	5	10	10	5
7		6	15	20	15
8		5	21	35	35
9		4	25	56	70
10		3	27	80	126
11		2	27	104	205
12		1	25	125	305
13			21	140	420
14			15	146	540
15			10	140	651
16			6	125	735
17			3	104	780
18			1	80	780

19				56	735
20				35	651
21				20	540
22				10	420
23				4	305
24				1	205
25					126
26					70
27					35
28					15
29					5
30					1
Σ	6	36	216	1296	7776
$E(Y_n)$	3.5	7	10.5	14	17.5
$Var(Y_n)$	2.9167	5.8333	8.7500	11.6667	14.5833

Tab.: Wahrscheinlichkeitsfunktionen der Summenvariable und der standardisierten Summenvariable



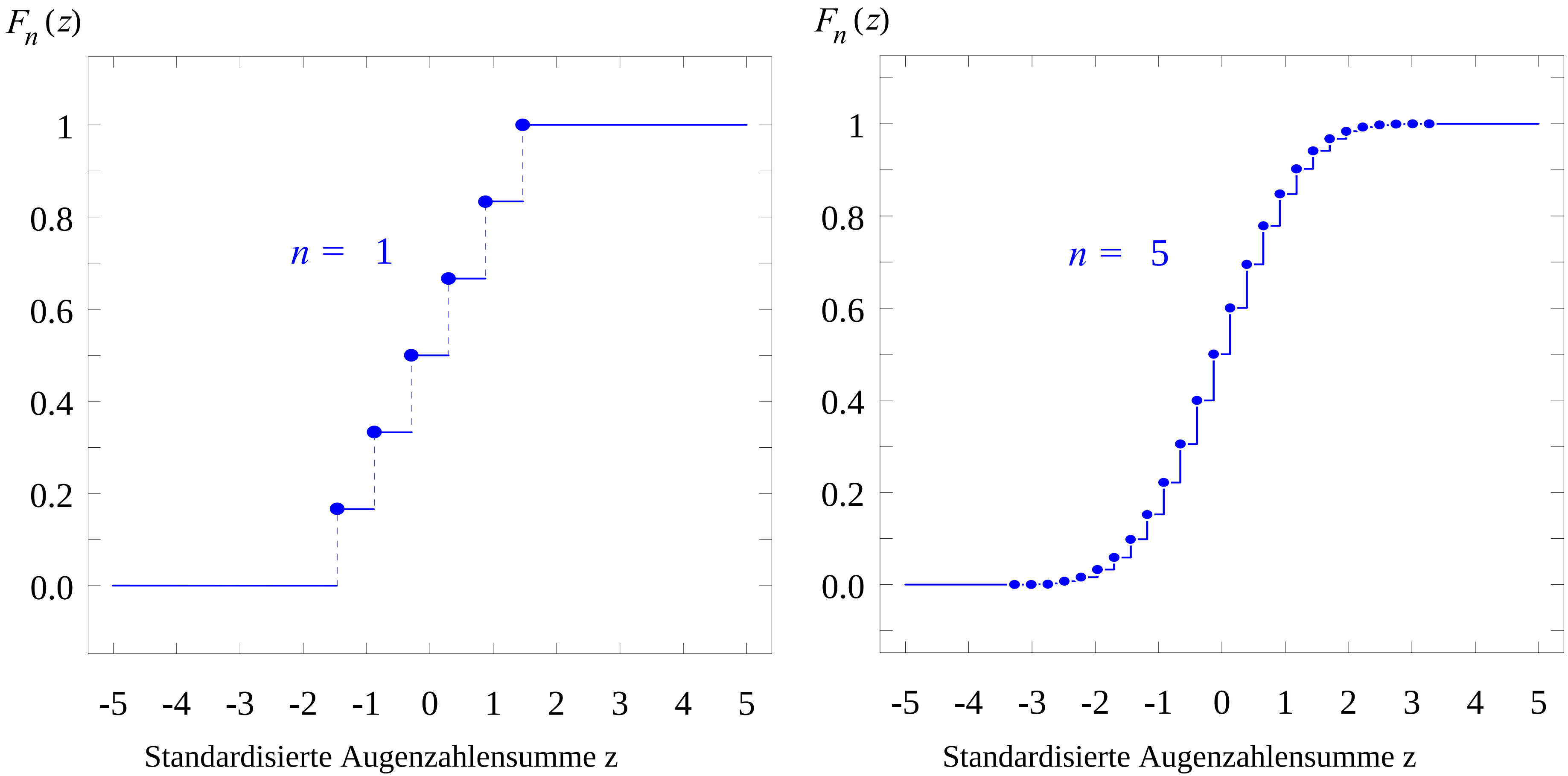
Wahrscheinlichkeitsfunktionen der Augenzahlensummen und der standardisierten Augenzahlensummen

Die Wahrscheinlichkeitsfunktionen weisen bereits für relativ kleines n die typische Glockenform der Normalverteilung auf. Man beachte aber, dass die Wahrscheinlichkeitsfunktionen *nicht* gegen die Dichte der Standardnormalverteilung konvergieren (warum?).

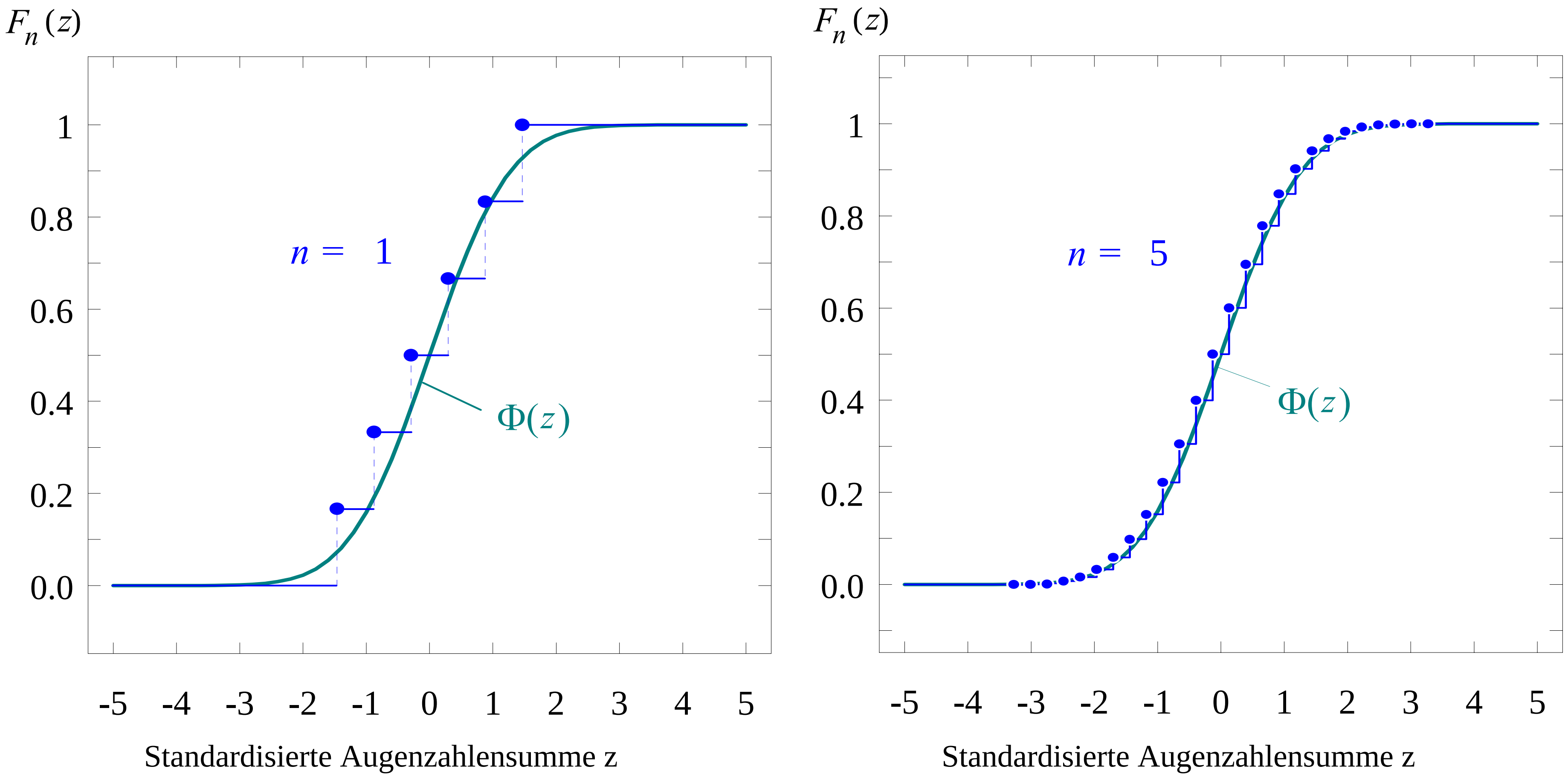
Die Konvergenz vollzieht sich über die Verteilungsfunktionen der standardisierten Summen. Zur Illustration zeigt die nächste Abbildung die Verteilungsfunktionen

$$F_n(z) = P(Z_n \le z) \qquad \text{für} \quad n = 1 \quad \text{bzw.} \quad n = 5$$

sowie die Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung.



Konvergenz der Verteilungsfunktionen gegen die Verteilungsfunktion der Standardnormalverteilung



Konvergenz der Verteilungsfunktionen gegen die Verteilungsfunktion der Standardnormalverteilung

1.10 Zufallsvektoren

D–1.22 Zufallsvektor, Gemeinsame Verteilungsfunktion

Mit $X_1,\dots,X_n$ seien Zufallsvariablen über denselben W-Raum $(\Omega,\mathcal{F},P)$ bezeichnet. Das n-Tupel

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

heißt *n-Zufallsvektor*. Die Funktion $F: I\!R^n \to [\,0\,,1\,]$ der n reellwertigen Veränderlichen $x_1,\dots,x_n$ mit

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = P\big((X_1 \le x_1)\cap\cdots\cap(X_n \le x_n)\big) \equiv P(X_1 \le x_1,\cdots,X_n \le x_n)$$

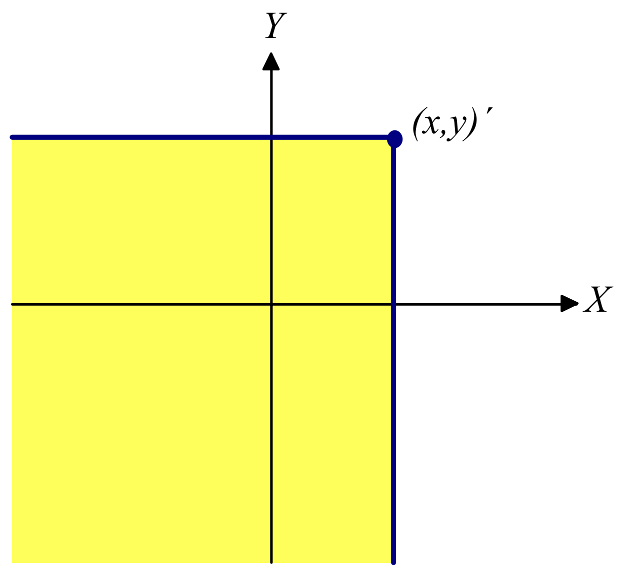
heißt *Verteilungsfunktion des Zufallsvektor* $\boldsymbol{X}$ oder *gemeinsame Verteilungsfunktion der Zufallsvariablen* $X_1,\dots,X_n$.

Mit Hilfe der Verteilungsfunktion können die Wahrscheinlichkeiten aller mit einem Zufallsvektor im Zusammenhang stehenden Ereignisse ausgedrückt werden.

O.B.d.A. betrachten wir nachfolgend nur den *Spezialfall* eines 2-Zufallsvektor $\boldsymbol{X} = (X,Y)'$. Die Verteilungsfunktion

$$F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

gibt die Wahrscheinlichkeit an, dass der Vektor sich in dem in der folgenden Abbildung schraffierten Bereich (einschließlich des oberen und des rechten Randes) realisiert.



D–1.23 Diskreter Zufallsvektor, Wahrscheinlichkeitsfunktion

Ein 2-Zufallsvektor $\boldsymbol{X} = (X, Y)'$ heißt *diskret*, wenn er endlich oder höchstens abzählbar unendlich viele Wertepaare $(x_i, y_j)' \in IR^2$ mit $i = 1,2,...$ und $j = 1,2,...$ annehmen kann.

Die zu solchen Paaren gehörigen Wahrscheinlichkeiten bezeichnen wir mit

$$P\left(X = x_i, Y = y_j\right) = p_{ij}.$$

Die für alle $(x, y)' \in IR^2$ definierte Funktion $f_{XY}(x, y)$ mit

$$f_{X,Y}(x,y) = P(X = x, Y = y) = \begin{cases} p_{ij} & \text{falls } x = x_i, y = y_j \quad (i = 1,2,\ldots\ ;\ j = 1,2,\ldots) \\ 0 & \text{sonst} \end{cases}$$

heißt *Wahrscheinlichkeitsfunktion des Zufallsvektors* oder *gemeinsame Wahrscheinlichkeitsfunktion der Zufallsvariablen X, Y.*

Zwischen der Wahrscheinlichkeitsfunktion $f_{XY}(x,y)$ und der Verteilungsfunktion $F_{XY}(x,y)$ besteht der Zusammenhang:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f_{X,Y}(x_i, y_j).$$

Ferner gilt

$$\sum_i \sum_j f_{X,Y}(x_i, y_j) = 1,$$

wobei über alle Wertepaare summiert wird, für die $f_{XY}(x,y)$ nicht null ist.

B–1.9 *Ziehen aus einer Urne mit vier Kugeln.*

Aus einer Urne mit zwei weißen und zwei roten Kugeln werden zwei Kugeln hintereinander blind gezogen. Das Zufallsexperiment setzt sich aus zwei Teilvorgängen zusammen, denen wir jeweils eine Zufallsvariable zuordnen können. Wir definieren die dichotomen Variablen

X = „Anzahl der roten Kugeln beim 1. Zug"

und Y = „Anzahl der roten Kugeln beim 2. Zug".

Der Zufallsvektor $\boldsymbol{X} = (X, Y)'$ kann vier verschiedene Wertepaare annehmen:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

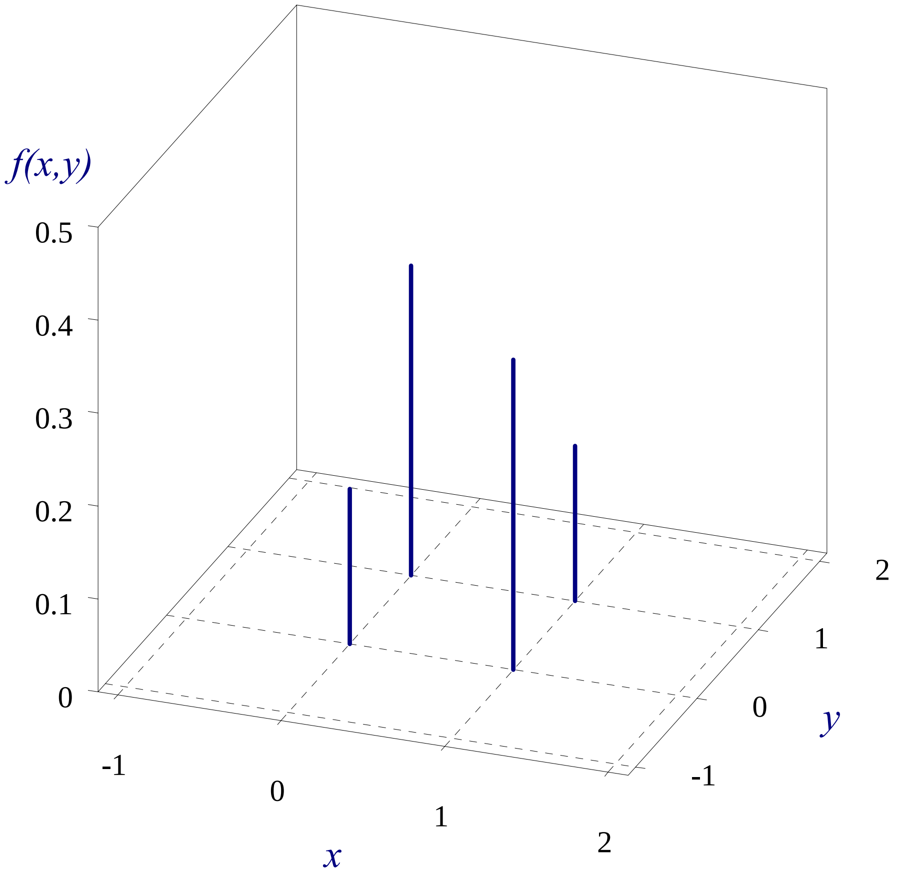
Ziehen wir die Kugeln *ohne Zurücklegen,* dann nimmt die Wahrscheinlichkeitsfunktion $f_{XY}(x,y)$ des Zufallsvektors an obigen Wertepaaren $(x_i, y_j)'$ die folgenden Werte an:

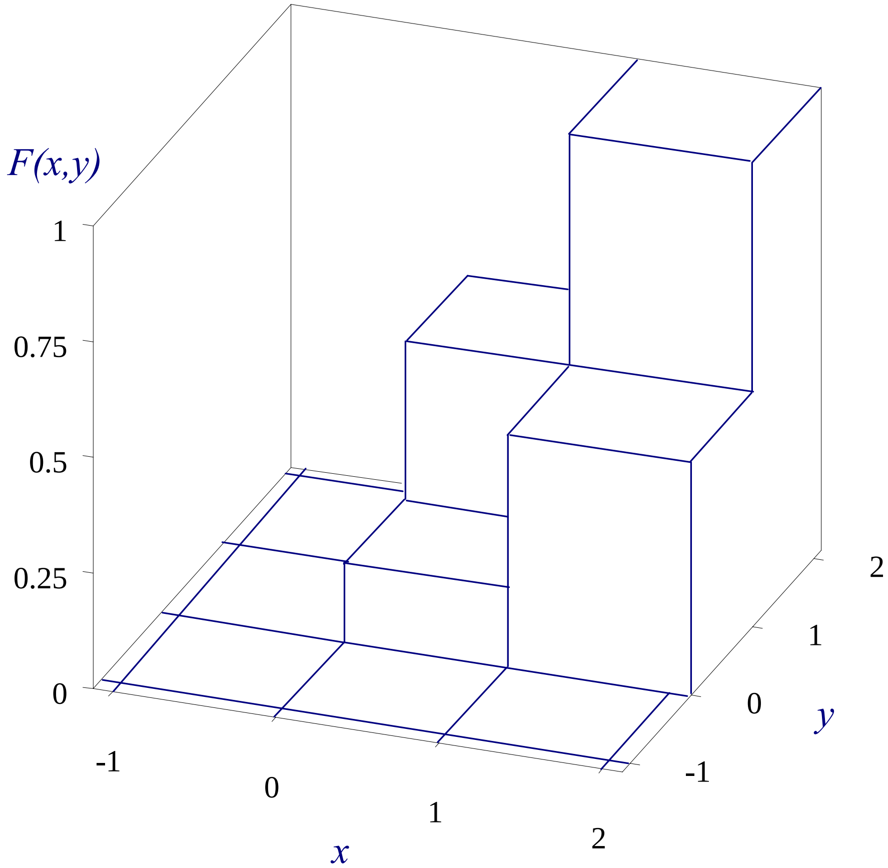
p_{ij}	$y_1 = 0$	$y_2 = 1$
$x_1 = 0$	1/6	2/6
$x_2 = 1$	2/6	1/6

Für alle anderen $(x,y)' \in IR^2$ ist $f_{XY}(x,y) = 0$. Die Verteilungsfunktion des Vektors lautet:

$F_{XY}(x,y)$	$y < 0$	$0 \leq y < 1$	$y \geq 1$
$x < 0$	0	0	0
$0 \leq x < 1$	0	1/6	3/6
$x \geq 1$	0	3/6	1

Die Entwicklung von $f_{XY}(x,y)$ und $F_{XY}(x,y)$ für den Fall, dass die Kugeln *mit Zurücklegen* gezogen werden, sei dem Leser selbst überlassen.





Wahrscheinlichkeits- und Verteilungsfunktion

D–1.24 Stetiger Zufallsvektor, Dichtefunktion

Ein 2-Zufallsvektor $\boldsymbol{X} = (X, Y)'$ heißt *stetig*, wenn sich die zugehörige Verteilungsfunktion durch ein Zweifach-Integral der Form

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t_1, t_2) dt_2 dt_1$$

darstellen lässt (t_1, t_2 = Integrationsvariablen).

Hierbei ist $f_{XY}(x, y)$ eine für alle $(x, y)' \in IR^2$ definierte nicht-negative Funktion und heißt *Dichtefunktion des Zufallsvektors* oder *gemeinsame Dichte der Zufallsvariablen X, Y*.

Die Wahrscheinlichkeit eines „Rechtecks“ $a < X \le b, c < Y \le d$ lässt sich mit Hilfe der Dichtefunktion wie folgt angeben:

$$P(a < X \le b, c < Y \le d) = \int_a^b \int_c^d f_{X,Y}(x,y)\,dy\,dx\,.$$

Analog zum eindimensionalen Fall gilt weiterhin

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = 1.$$

B–1.17 *Zweidimensionale stetige Gleichverteilung.*

Eine Gleichverteilung über einem Rechteck im IR^2

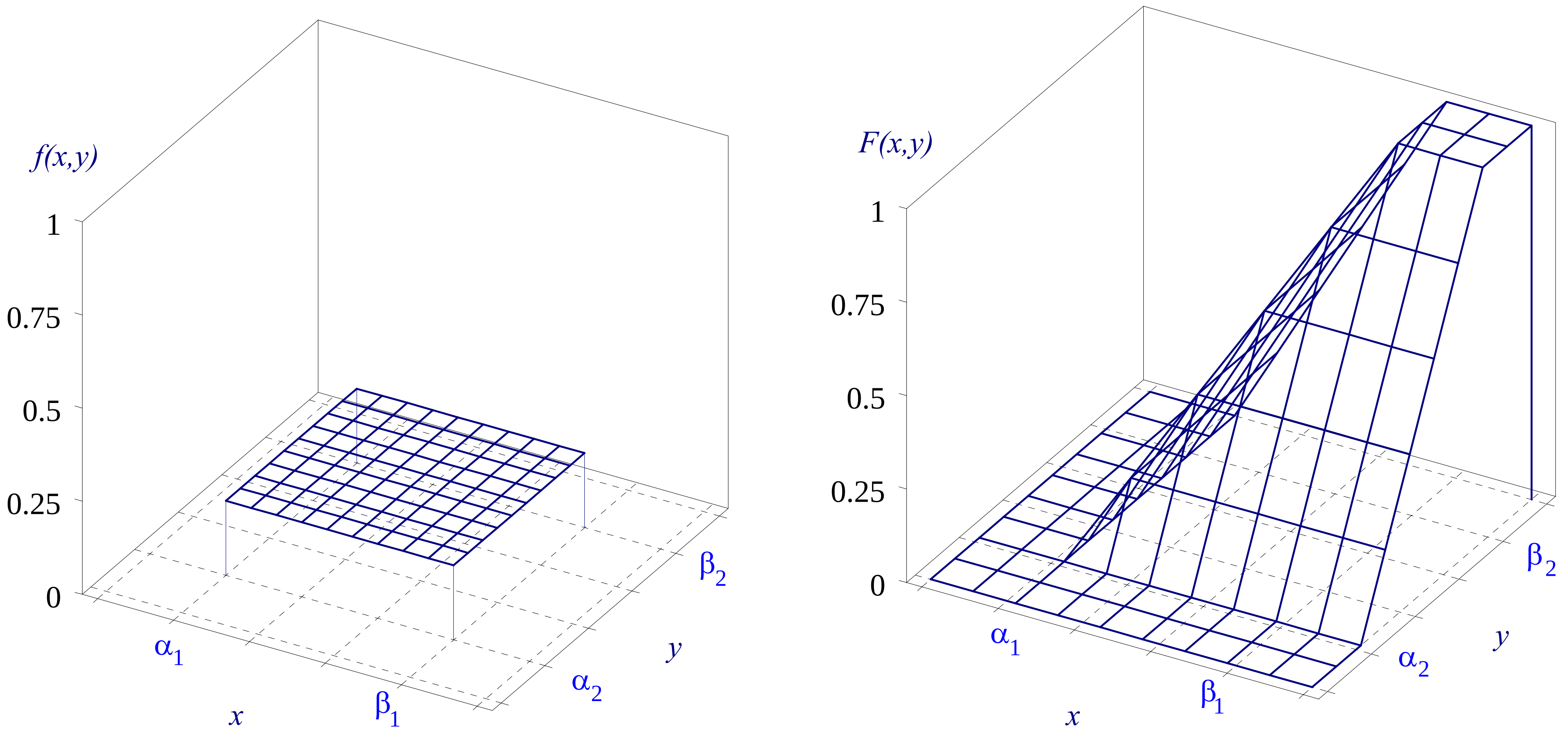
$$R = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \subset IR^2$$

besitzt die Dichte

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{k} & \text{falls } \alpha_1 \le x \le \beta_1, \alpha_2 \le y \le \beta_2 \\ 0 & \text{sonst}. \end{cases}$$

Hierbei ist die Konstante $k = (\beta_1 - \alpha_1) \cdot (\beta_2 - \alpha_2)$ der Flächeninhalt des Rechtecks R. Die Verteilungsfunktion lautet

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t_1,t_2)\,dt_2\,dt_1 = \begin{cases} 0 & \text{falls } x < \alpha_1, y < \alpha_2 \\ \frac{(x-\alpha_1)(y-\alpha_2)}{k} & \text{falls } \alpha_1 \le x \le \beta_1, \alpha_2 \le y \le \beta_2. \\ 1 & \text{falls } x > \beta_1, y > \beta_2. \end{cases}$$



Wahrscheinlichkeits- und Verteilungsfunktion der Rechteck-Verteilung

1.11 Unabhängige Zufallsvariablen

Ist die Verteilungsfunktion $F_{XY}(x,y)$ eines 2-Zufallsvektors $\boldsymbol{X} = (X,Y)'$ bekannt, dann können im Allgemeinen auch die Verteilungsfunktionen $F_X(x)$ und $F_Y(y)$ der eindimensionalen Zufallsvariablen X und Y ermittelt werden. Umgekehrt ist die Kenntnis von $F_X(x)$ und $F_Y(y)$ dann und nur dann ausreichend, um $F_{XY}(x,y)$ zu ermitteln, wenn dieVariablen X und Y stochastisch unabhängig sind.

D–1.25 Unabhängigkeit von Zufallsvariablen

Die Komponenten eines 2-Zufallsvektors $\boldsymbol{X} = (X,Y)'$ heißen *stochastisch unabhängig*, wenn für alle reellen Zahlen x, y gilt:

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \ .$$

Ist dies nicht der Fall, so heißen die Zufallsvariablen *stochastisch abhängig*.

S–1.16

Die Komponenten eines 2-Zufallsvektors $\boldsymbol{X} = (X,Y)'$ sind genau dann stochastisch unabhängig, wenn für ihre gemeinsame Wahrscheinlichkeitsfunktion bzw. Dichte für alle reellen x, y gilt:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \ .$$

B–1.9 *Ziehen aus einer Urne mit vier Kugeln.*

In dem *Urnenbeispiel mit Ziehen ohne Zurücklegen* ergeben sich die Randwahrscheinlichkeiten $p_{i\bullet} = P(X = x_i) = f_X(x_i)$ bzw. $p_{\bullet j} = P(Y = y_j) = f_Y(y_j)$ als Zeilen- bzw. Spaltensummen der Wahrscheinlichkeitstabelle:

p_{ij}	$y_1 = 0$	$y_2 = 1$	$p_{i\bullet}$
$x_1 = 0$	1/6	2/6	3/6
$x_2 = 1$	2/6	1/6	3/6
$p_{\bullet j}$	3/6	3/6	1

Wegen

$$f_{X,Y}(x_i, y_j) \neq f_X(x_i) \cdot f_Y(y_j) \text{ an allen Sprungstellen } (x_i, y_j)' \ (i = 1,2 \,;\, j = 1,2)$$

sind die Zufallsvariablen X und Y *stochastisch abhängig*. Werden die Kugeln *mit* Zurücklegen gezogen, sind X und Y hingegen *stochastisch unabhängig* ⊠

1.12 Momente eines Zufallsvektors

Neben den Momenten der Randverteilungen des 2-Zufallsvektors $\boldsymbol{X} = (X, Y)'$

$$E(X) = \mu_X \;,\quad E(Y) = \mu_Y \;,\quad E[(X-\mu_X)^2] = Var(X) = \sigma_X^2 \;,\quad E[(Y-\mu_Y)^2] = Var(Y) = \sigma_Y^2$$

interessieren auch Momente der gemeinsamen Verteilung von X und Y, insbesondere der Erwartungswert

$$E[(X-\mu_X)(Y-\mu_Y)] = \begin{cases} \sum_i \sum_j (x_i - \mu_X)\cdot(y_j - \mu_Y)\cdot f_{X,Y}(x_i, y_j) & \text{[diskreter Fall]} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)\cdot(y-\mu_Y)\cdot f_{X,Y}(x,y)dydx & \text{[stetiger Fall].} \end{cases}$$

D–1.26 Kovarianz, Korrelationskoeffizient

Es sei $\boldsymbol{X} = (X,Y)'$ ein 2-Zufallsvektor. Die Maßzahl

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = \sigma_{XY}$$

heißt *Kovarianz der gemeinsamen Verteilung von X und Y*. Die normierte Kovarianz

$$Corr(X,Y) = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \rho_{XY}$$

heißt *Korrelationskoeffizient von X und Y*.

Kovarianz und Korrelationskoeffizient sind Maße für die lineare Abhängigkeit der Zufallsvariablen X und Y. Der Korrelationskoeffizient ist eine normierte Größe, für die stets

$$-1 \leq \rho_{XY} \leq +1$$

erfüllt ist. Es gilt $\rho_{XY} = \pm 1$ genau dann, wenn zwischen X und Y der *lineare* Zusammenhang

$$Y = a_0 + a_1 X \qquad \text{mit} \quad a_1 \neq 0$$

besteht. Falls $a_1 > 0$, ist $\rho_{XY} = +1$ und falls $a_1 < 0$, ist $\rho_{XY} = -1$.

Momente von Zufallsvektoren fasst man zweckmäßigerweise in Vektoren und Matrizen zusammen.

D–1.27 Mittelwertvektor, Varianz-Kovarianz-Matrix, Korrelationsmatrix

Es sei $\boldsymbol{X} = (X, Y)'$ ein 2-Zufallsvektor. Dann heißt der 2-Vektor

$$E(\boldsymbol{X}) = \begin{pmatrix} E(X) \\ E(Y) \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} = \mu$$

Mittelwert- oder *Erwartungswertvektor von* $\boldsymbol{X}$. Ferner heißen die (2,2)-Matrizen

$$Var(\boldsymbol{X}) \equiv Cov(\boldsymbol{X}) = E[(\boldsymbol{X} - \mu)(\boldsymbol{X} - \mu)'] = \begin{pmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{YX} & \sigma_{YY} \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{pmatrix} = \Sigma$$

Varianz-Kovarianz-Matrix von $\boldsymbol{X}$ und

$$Corr(\boldsymbol{X}) = \begin{pmatrix} \rho_{XX} & \rho_{XY} \\ \rho_{YX} & \rho_{YY} \end{pmatrix} = \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{YX} & 1 \end{pmatrix} = \mathrm{P}$$

Korrelationsmatrix von $\boldsymbol{X}$. Die Matrizen sind symmetrisch, d.h. es gilt $\Sigma' = \Sigma$ und $\mathrm{P}' = \mathrm{P}$.

Für die Kovarianz existiert eine vereinfachte Berechnungsformel.

S–1.17 Verschiebungssatz für die Kovarianz

$$\sigma_{XY} = E(X \cdot Y) - E(X) \cdot E(Y) = E(X \cdot Y) - \mu_X \cdot \mu_Y \ .$$

Beweis

$$\begin{aligned}\sigma_{XY} &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(X \cdot Y - X \cdot \mu_Y - \mu_X \cdot Y + \cdot \mu_Y \cdot \mu_X) \\ &= E(X \cdot Y) - E(X) \cdot \mu_Y - \mu_X \cdot E(Y) + \mu_Y \cdot \mu_X \\ &= E(X \cdot Y) - \mu_Y \cdot \mu_X \end{aligned}$$

q.e.d.

Folgerung

Hieraus folgt der Verschiebungssatz für die Varianzen als Spezialfall:

$$\sigma_X^2 = \sigma_{XX} = E(X^2) - \mu_X^2 \quad \text{und} \quad \sigma_Y^2 = \sigma_{YY} = E(Y^2) - \mu_Y^2 .$$

Zwischen der stochastischen Abhängigkeit und der Korrelation von Zufallsvariablen besteht der folgende wichtige Zusammenhang.

S–1.18

Wenn die Zufallsvariablen X und Y stochastisch unabhängig sind, dann sind sie auch unkorreliert.

Beweis

Die Zufallsvariablen X und Y seien diskret (stetiger Fall analog). Bei stochastischer Unabhängigkeit gilt wegen Satz 1.16

$$E(X \cdot Y) = \sum_i \sum_j x_i \cdot y_j \cdot f_{X,Y}(x_i, y_j)$$

$$= \sum_i \sum_j x_i \cdot y_j \cdot f_X(x_i) \cdot f_Y(y_j) \qquad [\text{ da } f_{X,Y}(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j) \text{]}$$

$$= \left(\sum_i x_i \cdot f_X(x_i) \right) \cdot \left(\sum_j y_j \cdot f_Y(y_j) \right) = E(X) \cdot E(Y) \text{ .}$$

Für die Kovarianz folgt dann unter Anwendung von Satz 1.17

$$\sigma_{XY} = E(X \cdot Y) - E(X) \cdot E(Y) = E(X) \cdot E(Y) - E(X) \cdot E(Y) = 0$$

und damit ist auch

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = 0 \ . \qquad\qquad q.e.d.$$

Anmerkung

Die Umkehrung des Satzes 1.18 gilt nicht allgemein!

Unkorrelierte Zufallsvariablen können stochastisch abhängig sein, denn es kann ein *nichtlinerarer* Zusammenhang bestehen.

B–1.9 *Ziehen aus einer Urne mit vier Kugeln.*

In unserem Urnenbeispiel gilt beim *Ziehen ohne Zurücklegen*

$$E(X) = \sum_i x_i \cdot f_X(x_i) = 0 \cdot p_{1\bullet} + 1 \cdot p_{2\bullet} = \frac{1}{2} = \mu_X \ ,$$

$$Var(X) = \sum_i (x_i - \mu_X)^2 \cdot f_X(x_i) = \left(0 - \frac{1}{2}\right)^2 p_{1\bullet} + \left(1 - \frac{1}{2}\right)^2 p_{2\bullet} = \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{4} = \sigma_X^2 \ .$$

Analog ergibt sich $\mu_Y = \frac{1}{2}$ und $\sigma_Y^2 = \frac{1}{4}$. Ferner ist

$$E(X \cdot Y) = \sum_i \sum_j x_i y_j f_{X,Y}(x_i, y_j) = 0 \cdot 0 \cdot p_{11} + 0 \cdot 1 \cdot p_{12} + 1 \cdot 0 \cdot p_{21} + 1 \cdot 1 \cdot p_{22} = \frac{1}{6}$$

und damit

$$\sigma_{XY} = E(X \cdot Y) - \mu_x \cdot \mu_Y = \frac{1}{6} - \frac{1}{2} \cdot \frac{1}{2} = -\frac{1}{12} \ .$$

Der Korrelationskoeffizient ist

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{-\frac{1}{12}}{\sqrt{\frac{1}{4} \cdot \frac{1}{4}}} = -\frac{1}{3} \; .$$

Die Zufallsvariablen X und Y sind somit schwach negativ korreliert.

Zeigen Sie, dass $\sigma_{XY} = \rho_{XY} = 0$ ist, wenn die Kugeln *mit Zurücklegen* gezogen werden ☒

1.13 Bedingte Verteilungen [Selbststudium]

Die bedingten Verteilungen zweier Zufallsgrößen informieren über die Verteilung der einen Variablen unter der Nebenbedingung, dass eine andere einen bestimmten Wert als Realisation annimmt.

D–1.28 Bedingte Wahrscheinlichkeitsfunktionen und Dichtefunktionen

Es sei $\boldsymbol{X} = (X, Y)'$ ein 2-Zufallsvektor mit der Wahrscheinlichkeitsfunktion oder Dichte $f_{XY}(x, y)$ sowie den Randfunktionen $f_X(x)$ und $f_Y(y)$. Dann heißt

$$f_{Y|X=x}(y) = \begin{cases} \dfrac{f_{X,Y}(x, y)}{f_X(x)} & \text{falls } f_X(x) > 0 \\ 0 & \text{sonst} \end{cases}$$

die *bedingte Wahrscheinlichkeitsfunktion bzw. Dichte von Y bei gegebenen Wert x von X*. Analog ist

$$f_{X|Y=y}(x) = \begin{cases} \dfrac{f_{X,Y}(x, y)}{f_Y(y)} & \text{falls } f_Y(y) > 0 \\ 0 & \text{sonst} \end{cases}$$

die *bedingte Wahrscheinlichkeitsfunktion bzw. Dichte von X bei gegebenen Wert y von Y.*

Ebenso wie die gemeinsame Verteilung und die Randverteilungen eines Zufallsvektors können die bedingten Verteilungen Momente besitzen. Von besonderer Bedeutung sind die bedingten Erwartungswerte. Wir lernen diese im Kontext der Regressionsanalyse als *Regressionsfunktionen* kennen.

D–1.29 Bedingter Erwartungswert

Es sei $\boldsymbol{X} = (X, Y)'$ ein 2-Zufallsvektor mit der bedingten Wahrscheinlichkeitsfunktion oder Dichte $f_{Y|x=x}(y)$. Dann heißt

$$E(Y \mid X = x) = \begin{cases} \sum_j y_j \cdot f_{Y|X=x}(y_j) & \text{falls } \boldsymbol{X} \text{ diskret ist} \\ \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y)dy & \text{falls } \boldsymbol{X} \text{ stetig ist} \end{cases}$$

bedingter Erwartungswert von Y bei gegebenen Wert x von X.

Analog ist der bedingte Erwartungswert von X bei gegebenen Wert y von Y definiert.

B–1.9 *Ziehen aus einer Urne mit vier Kugeln.*

Wir betrachten weiterhin das *Ziehen ohne Zurücklegen* und untersuchen die bedingten Verteilungen von Y. Für den vorliegenden diskreten Zufallsvektor gilt

$$f_{Y|X=x}(y_j) = P(Y = y_j \mid X = x_i) = \frac{p_{ij}}{p_{i\bullet}} \qquad (j,i = 1,2).$$

Konkret erhalten wir

$$f_{Y|X=0}(0) = \frac{1/6}{1/2} = \frac{1}{3}, \quad f_{Y|X=0}(1) = \frac{2/6}{1/2} = \frac{2}{3}, \quad f_{Y|X=1}(0) = \frac{2/6}{1/2} = \frac{2}{3}, \quad f_{Y|X=1}(1) = \frac{1/6}{1/2} = \frac{1}{3}.$$

Die mittlere Anzahl der roten Kugeln beim 2. Zug, wenn beim 1. Zug eine weiße Kugel gezogen wurde, beträgt

$$E(Y \mid X = 0) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Wurde beim 1. Zug eine rote Kugel gezogen, dann sind beim 2. Zug

$$E(Y \mid X = 1) = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$$

rote Kugeln zu erwarten ⊠

Im Falle stochastisch unabhängiger Zufallsvariablen besitzen bedingte Verteilungen keinen über die Randverteilungen hinausgehenden Erkenntniswert.

S–1.19 Bedingte Verteilungen und Unabhängigkeit

Die Komponenten des 2-Zufallsvektors $\boldsymbol{X} = (X, Y)'$ seien unabhängig, dann gilt

$$f_{Y|X=x}(y) = f_Y(y) \qquad \text{und} \qquad f_{X|Y=y}(x) = f_X(x)$$

für alle reellen x, y.

Beweis

Satz 1.19 ergibt sich aus Satz 1.16. Da im Falle der Unabhängigkeit $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ für alle reellen x, y gilt, folgt sofort

$$\frac{f_{X,Y}(x,y)}{f_X(x)} = f_Y(y) \qquad \text{bzw.} \qquad \frac{f_{X,Y}(x,y)}{f_Y(y)} = f_X(x) \quad q.e.d.$$

1.14 Linearkombinationen von Zufallsvektoren

Später benötigen wir oft die Mittelwerte und Varianzen von Linearkombinationen.

S–1.20

Sei $\boldsymbol{X}$ ein n-Zufallsvektor mit den Momenten $E(\boldsymbol{X}) = \mu$ und $Cov(\boldsymbol{X}) = \Sigma$. Ferner sei $\boldsymbol{A}$ eine reelle (m,n)-Matrix mit $rg(\boldsymbol{A}) = m$ und $\boldsymbol{b}$ ein reeller m-Spaltenvektor. Dann ist die Linearkombination $\boldsymbol{Z} = \boldsymbol{A} \cdot \boldsymbol{X} + \boldsymbol{b}$ ein m-Zufallsvektor mit dem Erwartungswertvektor

$$E(\boldsymbol{AX} + \boldsymbol{b}) = \boldsymbol{A}E(\boldsymbol{X}) + \boldsymbol{b} = \boldsymbol{A}\mu + \boldsymbol{b}$$

und der Varianz-Kovarianz-Matrix

$$Cov(\boldsymbol{AX} + \boldsymbol{b}) = Cov(\boldsymbol{AX}) = \boldsymbol{A}\Sigma\boldsymbol{A}'.$$

Beweis

Wir beweisen nur die Hauptaussage des Satzes. Da der Erwartungswert E ein linearer Operator ist, folgt

$$E(\boldsymbol{Z}) = E(\boldsymbol{AX} + \boldsymbol{b}) = \boldsymbol{A}E(\boldsymbol{X}) + \boldsymbol{b} = \boldsymbol{A}\mu + \boldsymbol{b} \equiv \mu_Z \ .$$

Ferner gilt

$$\begin{aligned} Cov(\boldsymbol{Z}) &= E\left[\{\boldsymbol{Z} - \boldsymbol{\mu}_Z\}\{\boldsymbol{Z} - \boldsymbol{\mu}_Z\}'\right] \\ &= E[\{\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})\}\{\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})\}'] \\ &= E[\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}'] \\ &= \boldsymbol{A}E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})']\boldsymbol{A}' \\ &= \boldsymbol{A}\Sigma\boldsymbol{A}' \end{aligned}$$

q.e.d.

Betrachten wir einen 2-Zufallsvektor $\boldsymbol{X} = (X, Y)'$ und eine Linearkombination mit $\boldsymbol{A} = \boldsymbol{a}' = (a_1, a_2)$ $(m = 1)$, also

$$Z = \boldsymbol{a}'\boldsymbol{X} = a_1 X_1 + a_2 X_2 \, ,$$

dann folgt aus S–1.20

$$\mu_Z = E(\boldsymbol{a}'\boldsymbol{X}) = \boldsymbol{a}'\mu = a_1\mu_X + a_2\mu_Y$$

und

$$\sigma_Z^2 = Var(\boldsymbol{a}'\boldsymbol{X}) = \boldsymbol{a}'\Sigma\boldsymbol{a} = a_1^2\sigma_X^2 + a_2^2\sigma_Y^2 + 2a_1a_2\sigma_{XY} \, .$$

Sind die Zufallsvariablen unkorreliert, dann vereinfacht sich der Varianzterm zu

$$\sigma_Z^2 = a_1^2\sigma_X^2 + a_2^2\sigma_Y^2 \, .$$

B–1.18 *Portfolio-Selektion.*

X und Y seien die zukünftigen Renditen zweier Aktien von Unternehmen aus der Erdölindustrie (X) und der Autoindustrie (Y). Analysten einer Bank haben ermittelt, dass näherungsweise

$$E(X)=0.08,\ \ Var(X)=0.06^2,\ \ E(Y)=0.06,\ \ Var(Y)=0.06^2 \quad \text{und} \quad \rho_{XY}=-0.5$$

gilt. Ein Anleger habe 100000 Euro Kapital und erwägt zwei alternative Anlagestrategien.

Im ersten Fall invertiert er alles in das 1. Wertpapier (*degeneriertes Portfolio*). Kennzeichnen wir mit K_1 den Kapitalrückfluss, so gilt

$$K_1 = 100000 + 100000 \cdot X = 100000 \cdot (1+X),$$

$$E(K_1) = 100000 + 100000 \cdot E(X) = 100000 + 100000 \cdot 0.08 = 108000,$$

$$Var(K_1) = Var(100000 \cdot X) = 100000^2 \cdot Var(X) = 100000^2 \cdot 0.06^2 = 6000^2.$$

Das zweite Portfolio setzt sich in gleichen Teilen aus den beiden Wertpapieren zusammen. Bezüglich des Kapitalrückflusses K_2 können wir feststellen:

$$K_2 = 50000 \cdot (1 + X) + 50000 \cdot (1 + Y),$$

$$E(K_2) = 50000 \cdot (1 + E(X)) + 50000 \cdot (1 + E(Y))$$
$$= 100000 + 50000 \cdot 0.08 + 50000 \cdot 0.06 = 107000,$$

$$Var(K_2) = 50000^2 \cdot Var(X) + 50000^2 \cdot Var(Y) + 2 \cdot 50000 \cdot 50000 \cdot \sigma_{XY}$$
$$= 50000^2 \cdot 0.06^2 + 50000^2 \cdot 0.06^2 + 2 \cdot 50000^2 \cdot (-0.5 \cdot 0.06 \cdot 0.06) = 3000^2$$

(gemäß Satz 1.20).

Hierbei gilt

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = -0.5 \quad \Rightarrow \quad \sigma_{XY} = \rho_{XY} \sigma_X \sigma_Y = -0.5 \cdot 0.06 \cdot 0.06 \ .$$

Seien

$$G_1 = K_1 - 100000 \qquad \text{und} \qquad G_2 = K_2 - 100000$$

die Anlagegewinne der Portfolios. Die Gewinnerwartung $E(G_1) = 8000$ Euro ist im ersten Fall um 1000 Euro höher als im zweiten Fall mit $E(G_2) = 7000$ Euro. Allerdings ist die Gewinnvarianz $Var(G_1) = Var(K_1)$ beträchtlich größer als $Var(G_2) = Var(K_2)$.

Grund für das geringere *Risiko* des zweiten Portfolios ist der Risikoausgleich der beiden Aktien: Die Anlagerenditen X und Y sind negativ korreliert.

Ein *risikoaverser Anleger* wird das zweite Portfolio bevorzugen. Unterstellen wir normalverteilte Gewinne, dann lassen sich die folgenden Verlustwahrscheinlichkeiten ermitteln:

$$P(G_1 \le 0) = \Phi\left(\frac{0-8000}{6000}\right) = \Phi(-1.\overline{33}) = 1 - \Phi(1.\overline{33}) \approx 0.09175 \,,$$

$$P(G_2 \le 0) = \Phi\left(\frac{0-7000}{3000}\right) = \Phi(-2.\overline{33}) = 1 - \Phi(2.\overline{33}) \approx 0.0099 \,.$$

Hierbei ist Φ die Verteilungsfunktion der Standardnormalverteilung ☒

1.15 Multivariate Normalverteilung

Die Normalverteilung ist nicht nur ein oft genutztes Modell für univariate stetige Verteilungen, sie ist auch für multivariate stetige Verteilungen von zentraler Bedeutung.

D–1.30 Multivariate Normalverteilung

Es sei μ ein n-Vektor und Σ sei eine reguläre (n,n)-Matrix (die Inverse Σ^{-1} existiert und $det\,\Sigma \neq 0$). Ein n-Zufallsvektor $\boldsymbol{X}$ heißt *multivariat normalverteilt* mit den Parametern μ und Σ, wenn die gemeinsame Dichte der Komponenten $X_1,\dots,X_n$ von $\boldsymbol{X}$ die Form

$$f_{X_1,\dots,X_n}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\cdot(det\,\Sigma)^{1/2}}\, e^{-\frac{1}{2}(\boldsymbol{x}-\mu)'\Sigma^{-1}(\boldsymbol{x}-\mu)}$$

für alle $\boldsymbol{x} \in IR^n$ besitzt. Wir schreiben kurz $\boldsymbol{X} \sim N_n(\mu,\Sigma)$.

Im eindimensionalen Spezialfall $n = 1$ folgt mit $\Sigma = \sigma^2$ aus obiger Dichte die Dichtefunktion der univariaten Normalverteilung.

Auch im multivariaten Fall sind die Parameter der Normalverteilung zugleich Momente der Verteilung.

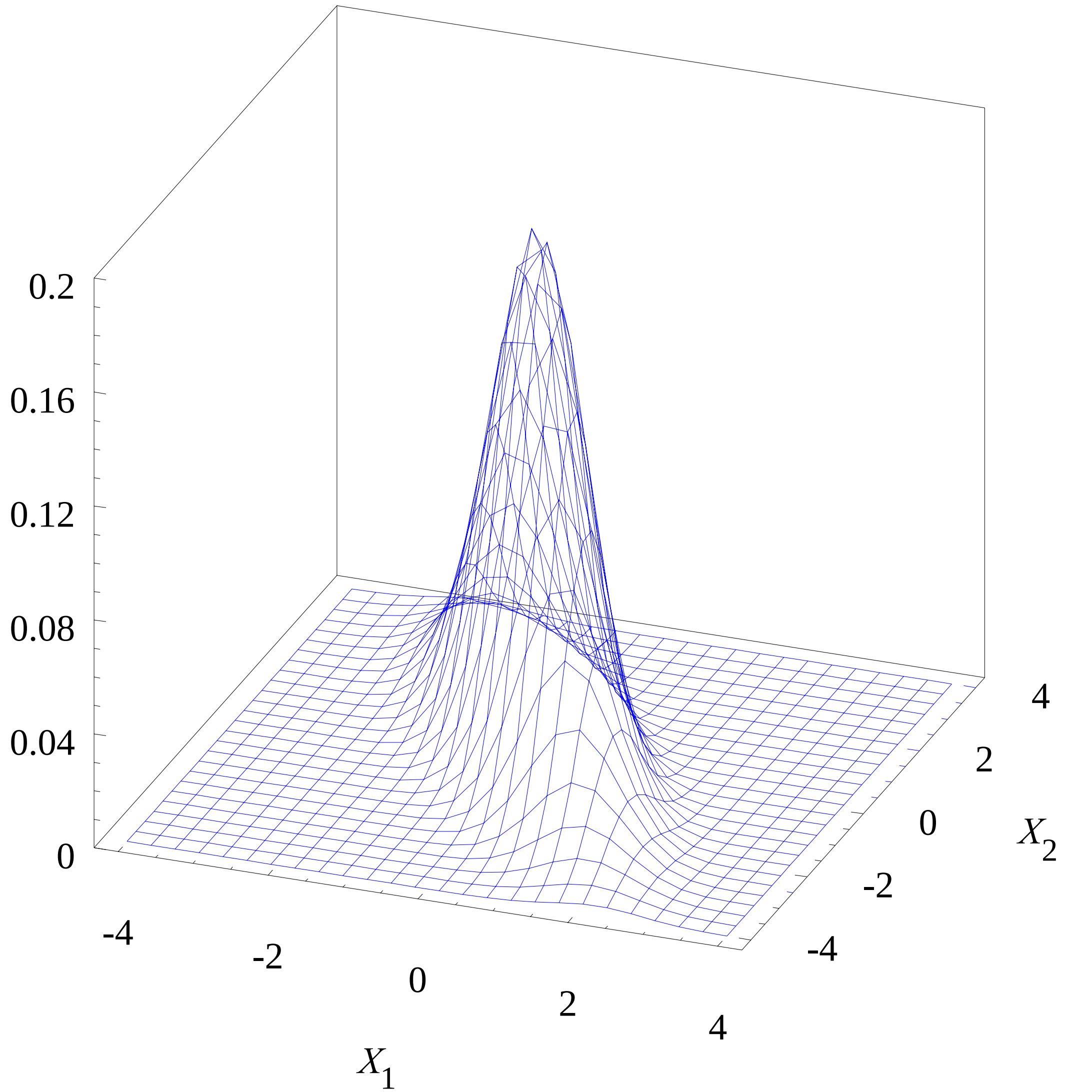
S–1.21 Mittelwert und Varianz der multivariaten Normalverteilung

Sei $\boldsymbol{X} \sim N_n(\mu,\Sigma)$, dann gilt

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = E(\boldsymbol{X}) \qquad \text{und} \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix} = Cov(\boldsymbol{X})$$

mit $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$ und $Cov(X_i, X_j) = \sigma_{ij}$ für $i, j = 1,\ldots,n$.

Zur Veranschaulichung der mehrdimensionalen Verteilung betrachten wir den Spezialfall $n = 2$, die *bivariate Normalverteilung*. Deren Dichte besitzt eine charakteristische Glockenform.



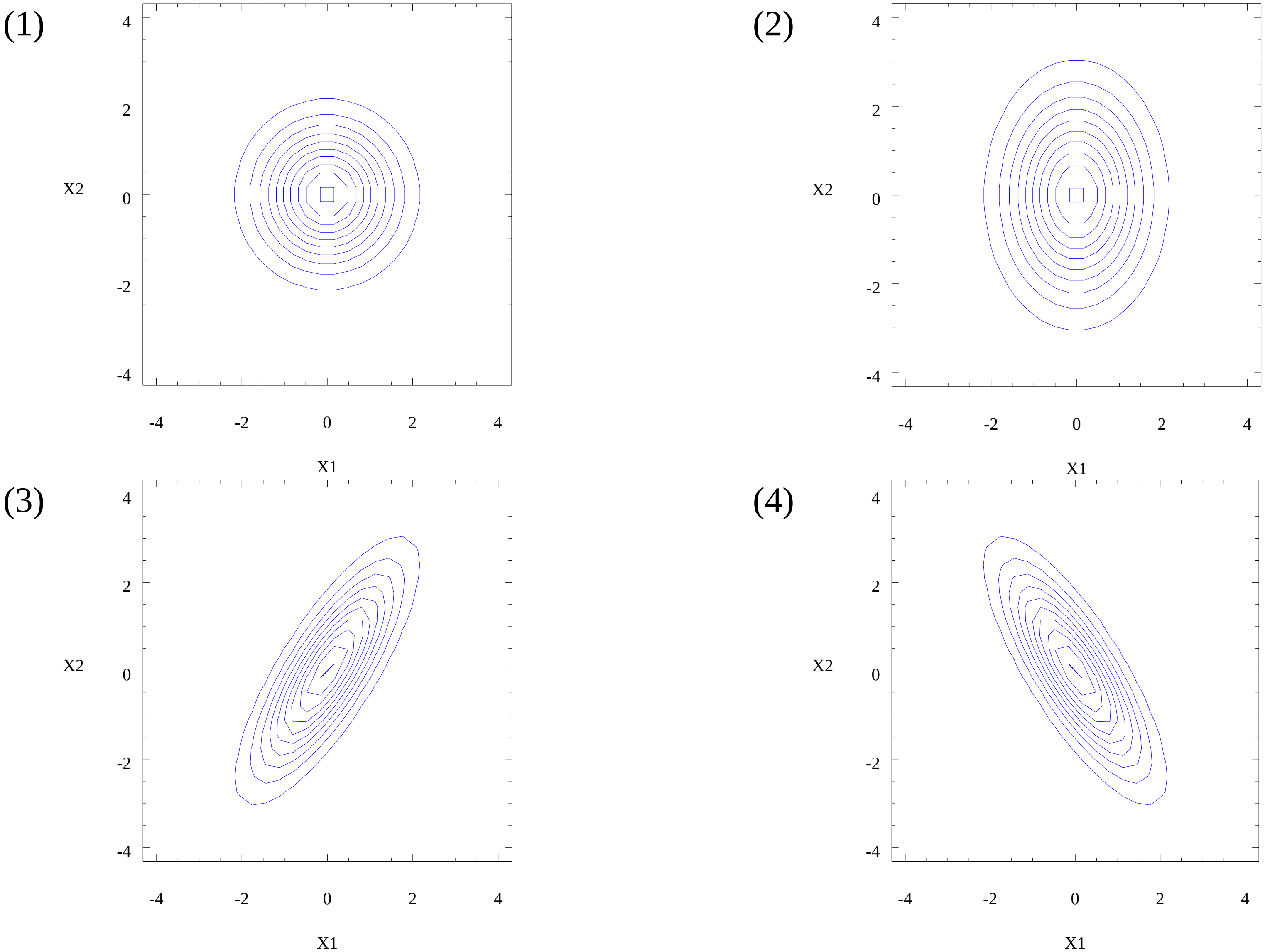
Dichte einer bivariaten Normalverteilung

Die folgende Abbildung zeigt Aufsichten (*Kontur-Graphiken* bestehend aus Iso-Dichte-Linien) der Dichten von vier verschiedenen bivariaten Normalverteilungen. Ihre Parameter sind:

(1) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$ (2) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$

(3) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \Sigma = \begin{pmatrix} 1 & 1.13 \\ 1.13 & 2 \end{pmatrix},$ (4) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \Sigma = \begin{pmatrix} 1 & -1.13 \\ -1.13 & 2 \end{pmatrix}.$

Die Dichte der 4. Verteilung war in der letzten Abbildung perspektivisch dargestellt.


(1)
(2)
(3)
(4)
X1
X2
-4
-2
0
2
4

S–1.22 Linearkombinationen normalverteilter Zufallsvektoren

Sei $\boldsymbol{X} \sim N_n(\mu, \Sigma)$, $\boldsymbol{A}$ eine reelle (m,n)-Matrix mit $rg(\boldsymbol{A}) = m$ und $\boldsymbol{b}$ ein reeller m-dimensionaler Spaltenvektor. Dann ist der m-Zufallsvektor $\boldsymbol{Y}$ mit

$$\boldsymbol{Y} = \boldsymbol{A} \cdot \boldsymbol{X} + \boldsymbol{b}$$

multivariat normalverteilt mit dem Erwartungswertvektor

$$E(\boldsymbol{AX} + \boldsymbol{b}) = \boldsymbol{A}\mu + \boldsymbol{b}$$

und der Varianz-Kovarianz-Matrix

$$Cov(\boldsymbol{AX} + \boldsymbol{b}) = Cov(\boldsymbol{AX}) = \boldsymbol{A}\Sigma\boldsymbol{A}'.$$

Kurz:

$$\boldsymbol{X} \sim N_n(\mu, \Sigma) \;\Rightarrow\; \boldsymbol{AX} + \boldsymbol{b} \sim N_m(\boldsymbol{A}\mu + \boldsymbol{b}, \boldsymbol{A}\Sigma\boldsymbol{A}').$$

(*ohne Beweis*)

Folgerungen

(a) Alle denkbaren Randverteilungen der multivariaten Normalverteilung sind ebenfalls Normalverteilungen. Speziell sind die Randverteilungen der einzelnen Komponenten X_i von $\boldsymbol{X}$ univariate Normalverteilungen:

Mit dem n-dimensionalen Zeilenvektor $\boldsymbol{A} = (0,\ldots,0,1,0,\ldots,0)$ und dem m-Nullvektor $\boldsymbol{b} = \boldsymbol{0}$ folgt

$$Y = X_i \quad \text{und} \quad X_i \sim N\left(\mu_i, \sigma_i^2\right).$$

Eine Umkehrung der Aussage ist nicht generell zulässig. Univariat normalverteilte Zufallsgrößen $X_1,\ldots,X_n$ sind nicht notwendig gemeinsam multivariat normalverteilt.

(b) Mit einem n-dimensionalen Zeilenvektor $\boldsymbol{A} = \boldsymbol{a}'$ und dem m-Nullvektor $\boldsymbol{b} = \boldsymbol{0}$ folgt

$$Y = \boldsymbol{a}'\boldsymbol{X} = a_1 X_1 + \ldots + a_n X_n \quad \text{und} \quad Y \sim N(\boldsymbol{a}'\mu\,, \boldsymbol{a}'\Sigma\boldsymbol{a}).$$

Jede eindimensionale Linearkombination multivariat normalverteilter Zufallsvariablen ist univariat normalverteilt.

Es wurde gezeigt, dass aus der stochastischen Unabhängigkeit von Zufallsvariablen stets deren Unkorreliertheit folgt (vgl. Satz 1.18). Die Umkehrung der Aussage gilt hingegen meist nicht. Eine bemerkenswerte Ausnahme liegt im Falle der multivariaten Normalverteilung vor.

S–1.23

Sei $\boldsymbol{X} \sim N_n(\mu, \Sigma)$. Die Komponenten $X_1, \ldots, X_n$ von $\boldsymbol{X}$ sind genau dann stochastisch unabhängig voneinander, wenn sie unkorreliert sind, d.h., wenn Σ eine Diagonalmatrix ist.

Beweis

Die Komponenten $X_i \ (i = 1,\dots,n)$ des Zufallsvektors $\boldsymbol{X} \sim N_n(\mu,\Sigma)$ seien unkorreliert. Dann ist Σ eine Diagonalmatrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{pmatrix} = Cov(\boldsymbol{X}) \ ,$$

woraus folgt

$$(\boldsymbol{x}-\mu)'\Sigma^{-1}(\boldsymbol{x}-\mu) = (x_1-\mu_1, x_2-\mu_2, x_n-\mu_n) \cdot \begin{pmatrix} \frac{1}{\sigma_1^2} & & & 0 \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_n^2} \end{pmatrix} \cdot \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_n-\mu_n \end{pmatrix} = \sum_{i=1}^{n} \left(\frac{x_i-\mu_i}{\sigma_i} \right)^2$$

und

$$det\,\Sigma = \sigma_1^2 \cdot \sigma_2^2 \cdot \ldots \cdot \sigma_n^2 \qquad \text{bzw.} \qquad (det\,\Sigma)^{1/2} = \sigma_1 \cdot \sigma_2 \cdot \ldots \cdot \sigma_n \ .$$

Aufgrund der Unkorreliert lässt sich die Dichte von $\boldsymbol{X}$ als Produkt der univariaten Dichten f_{X_i} der Vektorkomponenten X_i schreiben:

$$f_{X_1,\ldots,X_n}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2} \cdot (det\,\Sigma)^{1/2}}\, e^{-\frac{1}{2}(\boldsymbol{x}-\mu)'\Sigma^{-1}(\boldsymbol{x}-\mu)}$$

$$= \frac{1}{(2\pi)^{n/2} \cdot \sigma_1 \cdot \ldots \cdot \sigma_n}\, e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2} = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

mit

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2} \qquad (i = 1,\ldots,n).$$

Gemäß Satz 1.16 sind die unkorrelierten Zufallsvariablen X_i $(i = 1,\ldots,n)$ stochastisch unabhängig.

Durch Umkehrung obiger Argumentation lässt sich zeigen, dass aus der Unabhängigkeit normalverteilter Zufallsgrößen auch deren Unkorreliertheit folgt. *q.e.d.*

1.16 Stichprobenverteilungen: χ^2-, *t*- und *F*-Verteilung

Die in der Statistik vorkommenden Wahrscheinlichkeitsverteilungen kann man nach ihrem Verwendungszweck in zwei Klassen einteilen:

(*i*) Verteilungen, die im Zusammenhang mit der *mathematischen Modellierung von Zufallsexperimenten* auftreten;

(*ii*) Verteilungen, die die Grundlage statistischer Methoden wie der *Intervallschätzung* und der *Testverfahren* bilden. Man spricht hier auch von *Stichproben-* oder *Testverteilungen*.

Manche Verteilungen gehören beiden Klassen an, allen voran die Normalverteilung.

In diesem Abschnitt werden drei Verteilungen skizziert, die eng mit der Normalverteilung verwandt sind und in die zweite Klasse fallen.

D–1.31 Chi-Quadrat-Verteilung

$Z_1,\dots,Z_n$ seien n stochastisch unabhängige, jeweils $N(0,1)$-verteilte Zufallsvariablen. Die Verteilung der Quadratsumme

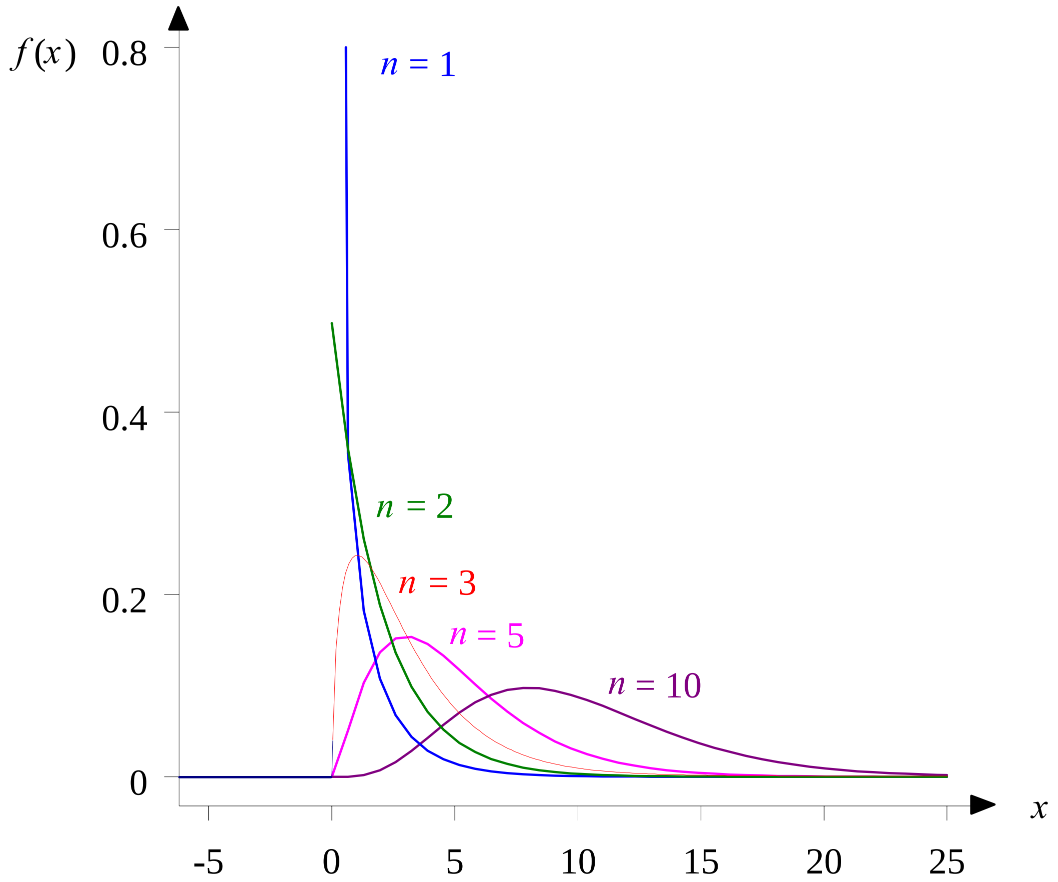
$$\chi^2 = \sum_{i=1}^{n} Z_i^2$$

heißt *Chi-Quadrat-Verteilung mit n Freiheitsgraden* oder kurz $\chi^2(n)$-*Verteilung*.

Die sogenannte Anzahl n der *Freiheitsgrade* ist der Parameter der Verteilung.

Eigenschaften der $\chi^2(n)$-Verteilung

- Die Zufallsvariable χ^2 nimmt stets nicht-negative Werte an. Die Dichte $f(x)$ der $\chi^2(n)$-Verteilung zeigt für verschiedene Freiheitsgrade n folgende Abbildung.



- Eine $\chi^2(n)$-verteilte Zufallsvariable besitzt den Mittelwert $\mu = n$ und die Varianz $\sigma^2 = 2n$.

- Aufgrund des Zentralen Grenzwertsatzes 1.15 lässt sich die $\chi^2(n)$-Verteilung für hinreichend großes n brauchbar durch die $N(n,2n)$-Verteilung annähern (*Faustregel*: $n > 100$).

- Die p-Quantile der Verteilung bezeichnen wir mit $\chi^2_{[p;n]}$.

  Sie sind für ausgewählte p und $n \leq 50$ in der Formelsammlung vertafelt. Für $n > 50$ können sie auch über die Beziehung

$$\chi^2_{[p;n]} \approx \frac{1}{2} \cdot \left(z_{[p]} + \sqrt{2n-1}\right)^2$$

  angenähert werden, wobei $z_{[p]}$ das p-Quantil der Standardnormalverteilung ist.

D–1.32 *t*-Verteilung oder STUDENT-Verteilung

Z sei $N(0,1)$-verteilt, χ^2 besitze eine $\chi^2(n)$-Verteilung, und beide Zufallsvariablen seien unabhängig voneinander. Die Verteilung der Zufallsvariable

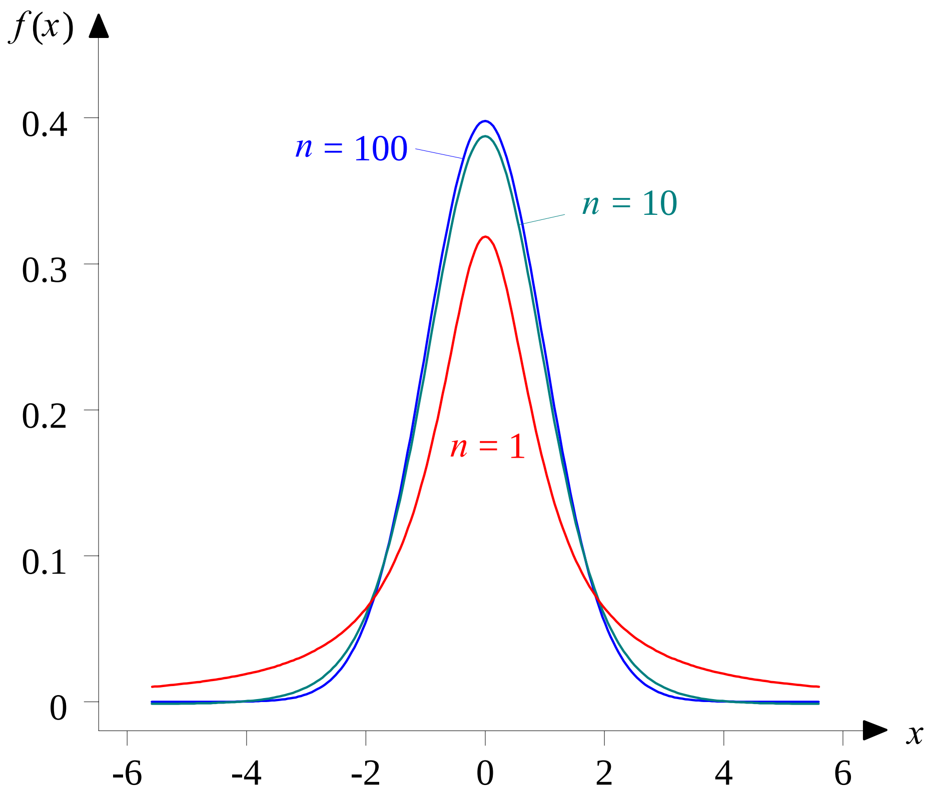
$$T = \frac{Z}{\sqrt{\chi^2 / n}}$$

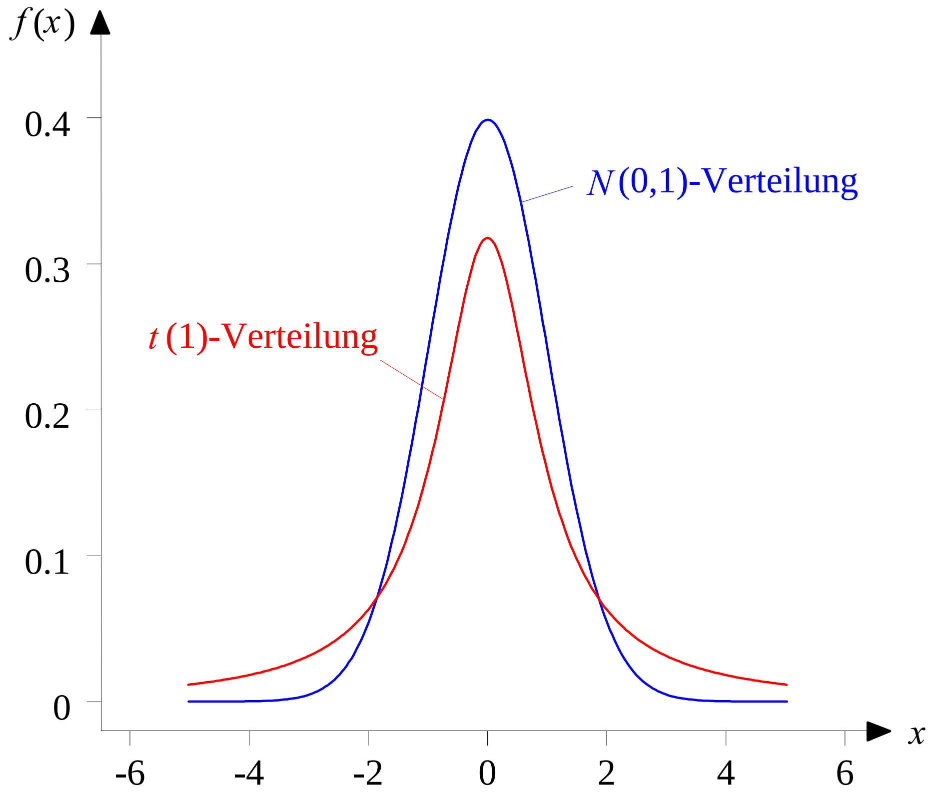
heißt *t-Verteilung mit n Freiheitsgraden* oder kurz *t(n)-Verteilung* (auch *STUDENT-Verteilung*).

Die Anzahl n der Freiheitsgrade ist der Parameter der Verteilung.

Eigenschaften der t-Verteilung

- Die Dichte $f(x)$ der $t(n)$-Verteilung zeigt für verschiedene Freiheitsgrade n folgende Abbildung. Die Dichte ist symmetrisch um den Wert $x = 0$.





- Eine $t(n)$-verteilte Variable besitzt für n = 2,3,4,... den Mittelwert $\mu = 0$. Für $n = 1$ existiert der Mittelwert nicht. Die Varianz ist für n = 3,4,5,...

$$\sigma^2 = \frac{n}{n-2} \quad .$$

Für n = 1,2 existiert die Varianz der t-Verteilung nicht.

- T ist asymptotisch standardnormalverteilt. Für $n \to \infty$ konvergiert die Dichte und die Verteilungsfunktion der $t(n)$-Verteilung gegen die Dichte und die Verteilungsfunktion der $N(0,1)$-Verteilung.

Für kleines n weist die $t(n)$-Verteilung eine größere Wahrscheinlichkeitsmasse an den Rändern auf als die $N(0,1)$-Verteilung (engl.: *heavy tail distribution*).

Für $n > 30$ kann die $t(n)$-Verteilung brauchbar durch die $N(0,1)$-Verteilung angenähert werden (*Faustregel*).

- Die p-Quantile der Verteilung bezeichnen wir mit $t_{[p;n]}$.

  Sie sind für ausgewählte p und $n \leq 50$ in der Formelsammlung vertafelt. Wegen der Symmetrie der Dichte gilt

  $$t_{[p;n]} = -t_{[1-p;n]}.$$

  Für $n > 50$ kann von der Näherung

  $$t_{[p;n]} \approx z_{[p]}$$

  Gebrauch gemacht werden, wobei $z_{[p]}$ das p-Quantil der Standardnormalverteilung ist.

D–1.33 *F*-Verteilung oder FISHER-Verteilung

Die Zufallsvariable χ_1^2 sei $\chi^2(m)$-verteilt, die Zufallsvariable χ_2^2 sei $\chi^2(n)$-verteilt, und beide Zufallsvariablen seien unabhängig voneinander. Die Verteilung der Zufallsvariable

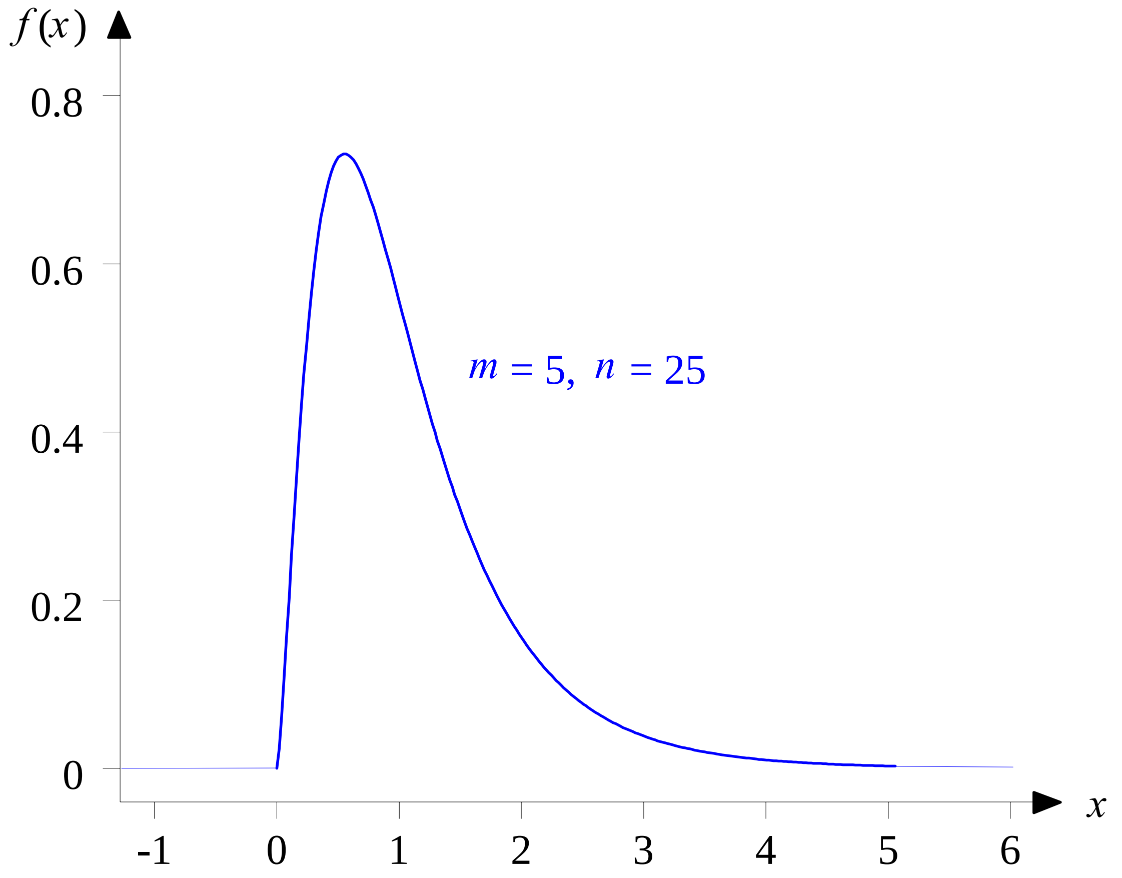
$$F = \frac{\chi_1^2 / m}{\chi_2^2 / n}$$

heißt *F-Verteilung mit m und n Freiheitsgraden* oder kurz $F(m,n)$*-Verteilung*.

Hierbei sind m und n wieder positive ganze Zahlen. Sie sind die Parameter der Verteilung.

Eigenschaften der F-Verteilung

- Die Zufallsvariable F nimmt stets nicht-negative Werte an. Die Dichte der $F(5, 25)$-Verteilung zeigt folgende Abbildung.



- Die p-Quantile der $F(m,n)$-Verteilung bezeichnen wir mit $f_{[p;m,n]}$. Sie sind für ausgewählte p und ausgewählte Freiheitsgrade in der Formelsammlung vertafelt. Wegen

$$F \sim F(m,n) \quad \Leftrightarrow \quad \frac{1}{F} \sim F(n,m)$$

(vgl. D–1.33) gilt

$$\boxed{f_{[p;m,n]} = \frac{1}{f_{[1-p;n,m]}}} \ .$$

Deswegen reicht die Vertafelung der linken Quantile.

- Für sehr große Freiheitsgrade m gilt

$$f_{[p;n,m]} \approx \frac{\chi^2_{[p;n]}}{n} \qquad \text{(Gleichheitszeichen für } m \to \infty\text{);}$$

für sehr große n gilt

$$f_{[p;n,m]} \approx \frac{m}{\chi^2_{[1-p;m]}} \qquad \text{(Gleichheitszeichen für } n \to \infty\text{).}$$