



ulm university universität
uulm

Angewandte Stochastik I

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

Ulm

Sommersemester 2013

Vorwort

Das vorliegende Skript der Vorlesung Angewandte Stochastik gibt eine Einführung in die Problemstellungen der Wahrscheinlichkeitstheorie und der Statistik für Studierende der nicht mathematischen (jedoch mathematisch arbeitenden) Studiengänge wie Elektrotechnik, Informatik, Physik, usw. Es entstand in den Jahren 2005–2013, in denen ich diesen Vorlesungskurs an der Universität Ulm mehrmals gehalten habe.

Ich bedanke mich bei Herrn Tobias Scheinert und Herrn Michael Wiedler für die Umsetzung meiner Vorlesungsnotizen in \LaTeX .

Ulm, den 5. Juli 2013
Evgeny Spodarev

Inhaltsverzeichnis

Inhaltsverzeichnis	i
1 Einführung	1
1.1 Über den Begriff “Stochastik”	1
1.2 Geschichtliche Entwicklung der Stochastik	2
1.3 Typische Problemstellungen der Stochastik	5
2 Wahrscheinlichkeiten	6
2.1 Ereignisse	7
2.2 Wahrscheinlichkeitsräume	10
2.3 Beispiele	13
2.3.1 Klassische Definition der Wahrscheinlichkeiten	13
2.3.2 Geometrische Wahrscheinlichkeiten	19
2.3.3 Bedingte Wahrscheinlichkeiten	20
3 Zufallsvariablen	26
3.1 Definition und Beispiele	26
3.2 Verteilungsfunktion	27
3.3 Grundlegende Klassen von Verteilungen	30
3.3.1 Diskrete Verteilungen	30
3.3.2 Absolut stetige Verteilungen	35
3.3.3 Mischungen von Verteilungen	39
3.4 Verteilungen von Zufallsvektoren	40
3.5 Stochastische Unabhängigkeit	45
3.5.1 Unabhängige Zufallsvariablen	45
3.6 Funktionen von Zufallsvektoren	47
4 Momente von Zufallsvariablen	52
4.1 Erwartungswert	53
4.2 Varianz	56
4.3 Kovarianz und Korrelationskoeffizient	59
4.4 Höhere und gemischte Momente	60
4.5 Ungleichungen	62

5 Grenzwertsätze	64
5.1 Gesetze der großen Zahlen	64
5.1.1 Schwaches Gesetz der großen Zahlen	65
5.1.2 Anwendung der Gesetze der großen Zahlen	67
5.2 Zentraler Grenzwertsatz	68
5.2.1 Klassischer zentraler Grenzwertsatz	68
5.2.2 Grenzwertsatz von Lindeberg	70
5.2.3 Konvergenzgeschwindigkeit im zentralen Grenzwertsatz	72
 Literatur	 74
 Index	 75

Kapitel 1

Einführung

1.1 Über den Begriff “Stochastik”

Wahrscheinlichkeitsrechnung ist eine Teildisziplin von Stochastik. Dabei kommt das Wort “Stochastik” aus dem Griechischen $\sigma\tau\omega\chi\alpha\sigma\tau\iota\kappa\eta$ - “die Kunst des Vermutens” (von $\sigma\tau\omega\chi\omega\xi$ - “Vermutung, Ahnung, Ziel”).

Dieser Begriff wurde von Jacob Bernoulli in seinem Buch “Ars conjectandi” geprägt (1713), in dem das erste Gesetz der großen Zahlen bewiesen wurde.

Stochastik beschäftigt sich mit den Ausprägungen und quantitativen Merkmalen von Zufall. Aber was ist Zufall? Gibt es Zufälle überhaupt? Das ist eine philosophische Frage, auf die jeder seine eigene Antwort suchen muss. Für die moderne Mathematik ist der Zufall eher eine Arbeitshypothese, die viele Vorgänge in der Natur und in der Technik ausreichend gut zu beschreiben scheint. Insbesondere kann der Zufall als eine Zusammenwirkung mehrerer Ursachen aufgefasst werden, die sich dem menschlichen Verstand entziehen (z.B. Brownsche Bewegung). Andererseits gibt es Studienbereiche (wie z.B. in der Quantenmechanik), in denen der Zufall eine essentielle Rolle zu spielen scheint (die Unbestimmtheitsrelation von Heisenberg). Wir werden die Existenz des Zufalls als eine wirkungsvolle Hypothese annehmen, die für viele Bereiche des Lebens zufriedenstellende Antworten liefert.



Abbildung 1.1: Jacob Bernoulli (1654-1705)

Stochastik kann man in folgende Gebiete unterteilen:

- Wahrscheinlichkeitsrechnung oder Wahrscheinlichkeitstheorie (Grundlagen)
- Statistik (Umgang mit den Daten)
- Stochastische Prozesse (Theorie zufälliger Zeitreihen und Felder)

- Simulation

Diese Vorlesung ist nur dem ersten Teil gewidmet.

1.2 Geschichtliche Entwicklung der Stochastik

1. *Vorgeschichte:*

Die Ursprünge der Wahrscheinlichkeitstheorie liegen im Dunklen der alten Zeiten. Ihre Entwicklung ist in der ersten Phase den Glücksspielen zu verdanken. Die ersten Würfelspiele konnte man in Altägypten, I. Dynastie (ca. 3500 v. Chr.) nachweisen. Auch später im klassischen Griechenland und im römischen Reich waren solche Spiele Mode (Kaiser August (63 v. Chr. -14 n. Chr.) und Claudius (10 v.Chr. - 54 n. Chr.).

Gleichzeitig gab es erste wahrscheinlichkeitstheoretische Überlegungen in der Versicherung und im Handel. Die älteste uns bekannte Form der Versicherungsverträge stammt aus dem Babylon (ca. 4-3 T. J. v.Chr., Verträge über die Seetransporte von Gütern). Die ersten Sterbetafeln in der Lebensversicherung stammen von dem römischen Juristen Ulpian (220 v.Chr.). Die erste genau datierte Lebensversicherungspolice stammt aus dem Jahre 1347, Genua.

Der erste Wissenschaftler, der sich mit diesen Aufgabenstellungen aus der Sicht der Mathematik befasst hat, war G. Cardano, der Erfinder der Cardan-Welle. In seinem Buch "Liber de ludo alea" sind zum ersten Mal Kombinationen von Ereignissen beim Würfeln beschrieben, die vorteilhaft für den Spieler sind. Er hat auch als erster die



Abbildung 1.2:
Gerolamo Cardano
(1501-1576)

$$\frac{\text{Anzahl vorteilhafter Ereignisse}}{\text{Anzahl aller Ereignisse}}$$

als Maß für Wahrscheinlichkeit entdeckt.

2. *Klassische Wahrscheinlichkeiten (XVII-XVIII Jh.):*

Diese Entwicklungsperiode beginnt mit dem Briefwechsel zwischen Blaise Pascal und Pierre de Fermat. Sie diskutierten Probleme, die von Chevalier de Méré (Antoine Gombaud (1607-1684)) gestellt wurden. Anbei ist eines seiner Probleme:

Was ist wahrscheinlicher: mindestens eine 6 bei 4 Würfeln eines Würfels oder mindestens ein Paar (6,6) bei 24 Würfeln von 2 Würfeln zu bekommen?

$$A = \{\text{mind. eine 6 bei 4 Würfeln eines Würfels}\}$$

$$B = \{\text{mind. (6,6) bei 24 Würfeln von zwei Würfeln}\}$$

Die Antwort:

$$P(\text{mind. eine 6 in 4 Würfeln})$$

$$= P(A) = 1 - P(\bar{A}) = 1 - P(\text{keine 6}) = 1 - \left(\frac{5}{6}\right)^4 = 0,516 > 0,491 = 1 - \left(\frac{35}{36}\right)^{24}$$

$$= 1 - P(\bar{B}) = P(B) = P(\text{mind. 1 (6,6) in 24 Würfeln von 2 Würfeln})$$



Abbildung 1.3: Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665) und Christian Huygens (1629-1695)

Weitere Entwicklung:

1657	Christian Huygens “De Ratiociniis in Ludo Alea” (Operationen mit Wahrscheinlichkeiten)
1713	Jacob Bernoulli “Ars Conjectandi” (Wahrscheinlichkeit eines Ereignisses und Häufigkeit seines Eintretens)

3. *Entwicklung analytischer Methoden (XVIII-XIX Jh.)* von Abraham de Moivre, Thomas Bayes (1702-1761), Pierre Simon de Laplace, Carl Friedrich Gauß, Simeon Denis Poisson (vgl. Abb. 1.4).

Entwicklung der Theorie bezüglich Beobachtungsfehlern und der Theorie des Schießens (Artilleriefeuer). Erste nicht-klassische Verteilungen wie Binomial- und Normalverteilung ($f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$), Poisson-Verteilung, zentraler Grenzwertsatz von De Moivre.

St.-Petersburger Schule von Wahrscheinlichkeiten:

(P.L. Tschebyschew, A.A. Markow, A.M. Ljapunow)

– Einführung von Zufallsvariablen, Erwartungswerten, Wahrscheinlichkeitsfunktionen, Markow-Ketten, abhängigen Zufallsvariablen.

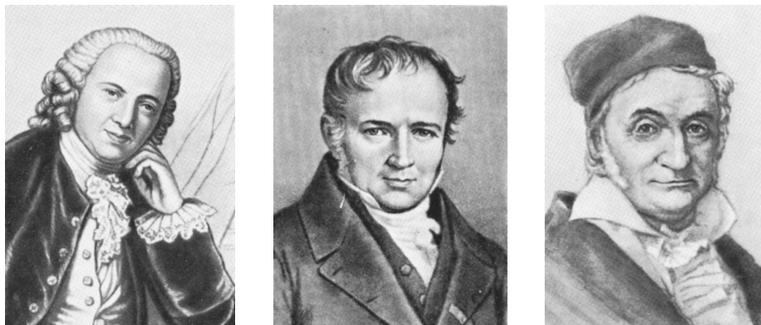


Abbildung 1.4: Abraham de Moivre (1667-1754), Pierre Simon de Laplace (1749-1827) und Karl Friedrich Gauß (1777-1855)



Abbildung 1.5: Simeon Denis Poisson (1781-1840), P. L. Tschebyschew (1821-1894) und A. A. Markow (1856-1922)

4. *Moderne Wahrscheinlichkeitstheorie (XX Jh.) David Hilbert, 8.8.1900, II. Mathematischer Kongress in Paris, Problem Nr. 6:*

Axiomatisierung von physikalischen Disziplinen, wie z.B. Wahrscheinlichkeitstheorie.

R. v. Mises: frequentistischer Zugang: $P(A) = \lim_{n \rightarrow \infty} \frac{\#A \text{ in } n \text{ Versionen}}{n}$

Antwort darauf: A.N. Kolmogorow führt Axiome der Wahrscheinlichkeitstheorie basierend auf der Maß- und Integrationstheorie von Borel und Lebesgue (1933) ein.

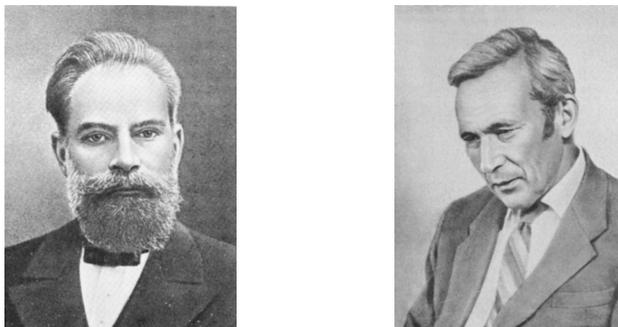
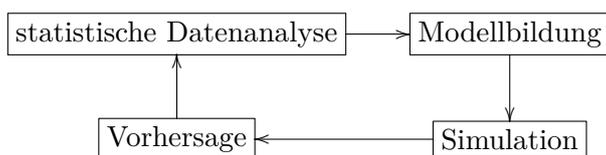


Abbildung 1.6: A. M. Ljapunow (1857-1918) und A. H. Kolmogorow (1903-1987)

1.3 Typische Problemstellungen der Stochastik



1. Modellierung von Zufallsexperimenten, d.h. deren adäquate theoretische Beschreibung.
2. Bestimmung von
 - Wahrscheinlichkeiten von Ereignissen
 - Mittelwerten und Varianzen von Zufallsvariablen
 - Verteilungsgesetzen von Zufallsvariablen
3. Näherungsformel und Lösungen mit Hilfe von Grenzwertsätzen
4. Schätzung von Modellparametern in der Statistik, Prüfung statistischer Hypothesen

Kapitel 2

Wahrscheinlichkeiten

Wahrscheinlichkeitstheorie befasst sich mit (im Prinzip unendlich oft) wiederholbaren Experimenten, in Folge derer ein Ereignis auftreten kann (oder nicht). Solche Ereignisse werden “zufällige Ereignisse” genannt. Sei A ein solches Ereignis. Wenn $n(A)$ die Häufigkeit des Auftretens von A in n Experimenten ist, so hat man bemerkt, dass $\frac{n(A)}{n} \rightarrow c$ für große n ($n \rightarrow \infty$). Diese Konstante c nennt man “Wahrscheinlichkeit von A ” und bezeichnet sie mit $P(A)$.

Beispiel: n -maliger Münzwurf: (siehe Abbildung 2.1) faire Münze, d.h. $n(A) \approx n(\bar{A})$, $A = \{\text{Kopf}\}$, $\bar{A} = \{\text{Zahl}\}$

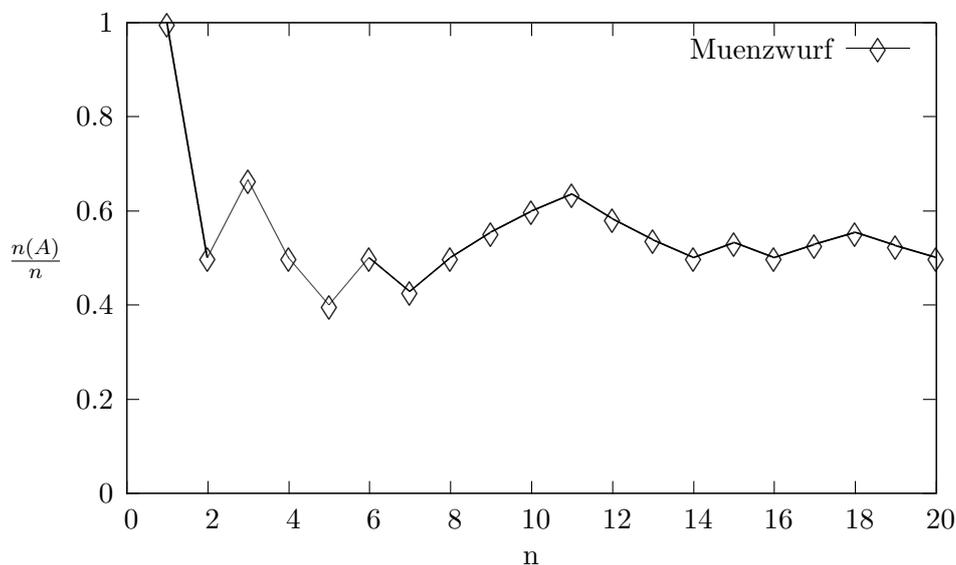


Abbildung 2.1: Relative Häufigkeit $\frac{n(A)}{n}$ des Ereignisses “Kopf” beim n -maligen Münzwurf

Man kann leicht feststellen, dass $\frac{n(A)}{n} \approx \frac{1}{2}$ für große n . $\implies P(A) = \frac{1}{2}$. Um dies zu verifizieren, hat Buffon in XVIII Jh. 4040 mal eine faire Münze geworfen, davon war 2048 mal Kopf, so dass $\frac{n(A)}{n} = 0,508$. Pierson hat es 24000 mal gemacht: es ergab $n(A) = 12012$ und somit $\frac{n(A)}{n} \approx 0.5005$.

In den Definitionen, die wir bald geben werden, soll diese empirische Begriffsbildung $P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$ ihren Ausdruck finden. Zunächst definieren wir, für welche Ereignisse A die Wahrscheinlichkeit $P(A)$ überhaupt eingeführt werden kann.

offene Fragen:

1. Was ist $P(A)$?
2. Für welche A ist $P(A)$ definiert?

2.1 Ereignisse

Sei E ein Grundraum und $\Omega \subseteq E$ sei die Menge von Elementarereignissen ω (Grundmenge).

Ω kann als Menge der möglichen Versuchsergebnisse interpretiert werden. Man nennt Ω manchmal auch *Grundgesamtheit* oder *Stichprobenraum*.

Definition 2.1.1 Eine Teilmenge A von Ω ($A \subset \Omega$) wird *Ereignis* genannt. Dabei ist $\{\omega\} \subset \Omega$ ein *Elementarereignis*, das das Versuchsergebnis ω darstellt. Falls bei einem Versuch das Ergebnis $\omega \in A$ erzielt wurde, so sagen wir, dass A eintritt.

Beispiel 2.1.1

1. *Einmaliges Würfeln*: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $E = \mathbb{N}$
2. *n-maliger Münzwurf*: $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i \in \{0, 1\}\}$
 $E = \mathbb{N}^n$, $\omega_i = \begin{cases} 1, & \text{falls ein "Kopf" im } i\text{-ten Wurf} \\ 0, & \text{sonst} \end{cases}$

Weiter werden wir E nicht mehr spezifizieren.

Tabelle 2.1: Wahrscheinlichkeitstheoretische Bedeutung von Mengenoperationen

$A = \emptyset$	unmögliches Ereignis
$A = \Omega$	wahres Ereignis
$A \subset B$	aus dem Eintreten von A folgt auch, dass B eintritt.
$A \cap B = \emptyset$	(disjunkte, <i>unvereinbare</i> Ereignisse): A und B können nicht gleichzeitig eintreten.
$A = B \cup C$	Mindestens eines der Ereignisse B und C tritt ein.
$A = \cup_{i=1}^n A_i$	Ereignis $A =$ "Es tritt mindestens ein Ereignis A_i ein"
$A = B \cap C$	Ereignis $A =$ "Es treten B und C gleichzeitig ein."
$A = \cap_{i=1}^n A_i$	Ereignis $A =$ "Es treten alle Ereignisse A_1, \dots, A_n ein"
$\bar{A} = A^c$	Das Ereignis A tritt nicht ein.
$A = B \setminus C$	Ereignis A tritt genau dann ein, wenn B eintritt, aber nicht C
$A = B \Delta C$	Ereignis A tritt genau dann ein, wenn B oder C eintreten (<i>nicht gleichzeitig!</i>)

Anmerkung: $A = B \Delta C = (B \setminus C) \cup (C \setminus B)$ (symmetrische Differenz)

Definition 2.1.2 Ereignisse A_1, A_2, A_3, \dots heißen *paarweise disjunkt* oder *unvereinbar*, wenn $A_i \cap A_j = \emptyset \quad \forall i \neq j$

Beispiel 2.1.2 *Zweimaliges Würfeln:* $\Omega = \{(\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}, i = 1, 2\}$,
 $A =$ "die Summe der Augenzahlen ist 6" = $\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$.
Die Ereignisse $B =$ "Die Summe der Augenzahlen ist ungerade" und
 $C = \{(3, 5)\}$ sind unvereinbar.

Oft sind nicht alle Teilmengen von Ω als Ereignisse sinnvoll. Deswegen beschränkt man sich auf ein Teilsystem von Ereignissen mit bestimmten Eigenschaften; und zwar soll dieses Teilsystem abgeschlossen bezüglich Mengenoperationen sein.

Definition 2.1.3 Eine nicht leere Familie \mathcal{F} von Ereignissen aus Ω heißt *Algebra*, falls

1. $A \in \mathcal{F} \implies \bar{A} \in \mathcal{F}$
2. $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$

Beispiel 2.1.3 1. Die Potenzmenge $\mathcal{P}(\Omega)$ = (die Menge aller Teilmengen von Ω) ist eine Algebra.

2. Im Beispiel 2.1.2 ist $\mathcal{F} = \{\emptyset, A, \bar{A}, \Omega\}$ eine Algebra. Dagegen ist $\mathcal{F} = \{\emptyset, A, B, C, \Omega\}$ keine Algebra: z.B. $A \cup C \notin \mathcal{F}$.

Lemma 2.1.1 (*Eigenschaften einer Algebra:*) Sei \mathcal{F} eine Algebra von Ereignissen aus Ω . Es gelten folgende Eigenschaften:

1. $\emptyset, \Omega \in \mathcal{F}$
2. $A, B \in \mathcal{F} \implies A \setminus B \in \mathcal{F}$
3. $A_1, \dots, A_n \in \mathcal{F} \implies \bigcup_{i=1}^n A_i \in \mathcal{F}$ und $\bigcap_{i=1}^n A_i \in \mathcal{F}$

Beweis

1. $\mathcal{F} \neq \emptyset \implies \exists A \in \mathcal{F} \implies \bar{A} \in \mathcal{F}$ nach Definition $\implies A \cup \bar{A} = \Omega \in \mathcal{F}$;
 $\emptyset = \bar{\Omega} \in \mathcal{F}$.
2. $A, B \in \mathcal{F}, \quad A \setminus B = A \cap \bar{B} = \overline{(\bar{A} \cup B)} \in \mathcal{F}$.
3. *Induktiver Beweis:*
 $n = 2$: $A, B \in \mathcal{F} \implies A \cap B = \overline{(\bar{A} \cup \bar{B})} \in \mathcal{F}$
 $n = k \mapsto n = k + 1$: $\bigcap_{i=1}^{k+1} A_i = (\bigcap_{i=1}^k A_i) \cap A_{k+1} \in \mathcal{F}$.

□

Für die Entwicklung einer gehaltvollen Theorie sind aber Algebren noch zu allgemein. Manchmal ist es auch notwendig, unendliche Vereinigungen $\bigcup_{i=1}^{\infty} A_i$ oder unendliche Schnitte $\bigcap_{i=1}^{\infty} A_i$ zu betrachten, um z.B. Grenzwerte von Folgen von Ereignissen definieren zu können. Dazu führt man Ereignissysteme ein, die σ -Algebren genannt werden:

Definition 2.1.4

1. Eine Algebra \mathcal{F} heißt σ -Algebra, falls aus $A_1, A_2, \dots, \in \mathcal{F}$ folgt, dass $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
2. Das Paar (Ω, \mathcal{F}) heißt *Messraum*, falls \mathcal{F} eine σ -Algebra der Teilmengen von Ω ist.

Beispiel 2.1.4 1. $\mathcal{F} = \mathcal{P}(\Omega)$ ist eine σ -Algebra.

2. *Beispiel einer Algebra \mathcal{F} , die keine σ -Algebra ist:*

Sei \mathcal{F} die Klasse von Teilmengen aus $\Omega = \mathbb{R}$, die aus endlichen Vereinigungen von disjunkten Intervallen der Form $(-\infty, a], (b, c]$ und (d, ∞)

$a, b, c, d \in \mathbb{R}$, besteht. Offensichtlich ist \mathcal{F} eine Algebra. Dennoch ist \mathcal{F} keine σ -Algebra, denn $[b, c] = \underbrace{\bigcap_{n=1}^{\infty} (b - \frac{1}{n}, c]}_{\in \mathcal{F}} \notin \mathcal{F}$.

2.2 Wahrscheinlichkeitsräume

Auf einem Messraum (Ω, \mathcal{F}) wird ein *Wahrscheinlichkeitsmaß* durch folgende *Axiome von Kolmogorow* eingeführt:

Definition 2.2.1

1. Die Mengenfunktion $P : \mathcal{F} \rightarrow [0, 1]$ heißt *Wahrscheinlichkeitsmaß* auf \mathcal{F} , falls
 - (a) $P(\Omega) = 1$ (*Normiertheit*)
 - (b) $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$, A_n paarweise disjunkt $\implies P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ (*σ -Additivität*)
2. Das Tripel (Ω, \mathcal{F}, P) heißt *Wahrscheinlichkeitsraum*.
3. $\forall A \in \mathcal{F}$ heißt $P(A)$ *Wahrscheinlichkeit* des Ereignisses A .

Bemerkung 2.2.1 1. Nachfolgend werden nur solche $A \subset \Omega$ *Ereignisse* genannt, die zu der ausgewählten σ -Algebra \mathcal{F} von Ω gehören. Alle anderen Teilmengen $A \subset \Omega$ sind demnach *keine* Ereignisse.

2. \mathcal{F} kann nicht immer als $\mathcal{P}(\Omega)$ gewählt werden. Falls Ω endlich oder abzählbar ist, ist dies jedoch möglich. Dann kann $P(A)$ auf $(\Omega, \mathcal{P}(\Omega))$ als $P(A) = \sum_{\omega \in A} P(\{\omega\})$ definiert werden (klassische Definition der Wahrscheinlichkeiten).

Falls z.B. $\Omega = \mathbb{R}$ ist, dann kann \mathcal{F} nicht mehr als $\mathcal{P}(\mathbb{R})$ gewählt werden, weil ansonsten \mathcal{F} eine große Anzahl von *pathologischen* Ereignissen enthalten würde, für die z.B. der Begriff der Länge nicht definiert ist.

Definition 2.2.2 Sei \mathcal{U} eine beliebige Klasse von Teilmengen aus Ω . Dann ist durch

$$\sigma(\mathcal{U}) = \bigcap_{\mathcal{U} \subset \mathcal{F}, \mathcal{F} \text{ } \sigma\text{-Alg. von } \Omega} \mathcal{F}$$

eine σ -Algebra gegeben, die *minimale σ -Algebra, die \mathcal{U} enthält*, genannt wird.

Übungsaufgabe 2.2.1 Zeigen Sie bitte, dass die in Definition 2.2.2 definierte Klasse $\sigma(\mathcal{U})$ tatsächlich eine σ -Algebra darstellt.

Definition 2.2.3 Sei $\Omega = \mathbb{R}^d$. Sei \mathcal{U} = Klasse aller offenen Teilmengen von \mathbb{R}^d . Dann heißt $\sigma(\mathcal{U})$ die *Borel σ -Algebra* auf \mathbb{R}^d und wird mit $\mathfrak{B}_{\mathbb{R}^d}$ bezeichnet. Elemente von $\mathfrak{B}_{\mathbb{R}^d}$ heißen *Borel-Mengen*. Diese Definition kann auch für einen beliebigen topologischen Raum Ω (nicht unbedingt \mathbb{R}^d) gegeben werden.

Übungsaufgabe 2.2.2 Zeigen Sie, dass $\mathfrak{B}_{\mathbb{R}^d}$ alle $\{xy, x \in \mathbb{R}^d\}$ alle geschlossenen und insbesondere kompakten Teilmengen von \mathbb{R}^d enthält, z.B. $[a, b]^d \in \mathfrak{B}_{\mathbb{R}^d}$, $a \subset b$.

Satz 2.2.1 Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n, A, B \subset \mathcal{F}$. Dann gilt:

1. $P(\emptyset) = 0$.
2. $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, falls A_i paarweise disjunkt sind.
3. Falls $A \subset B$, dann ist $P(B \setminus A) = P(B) - P(A)$.

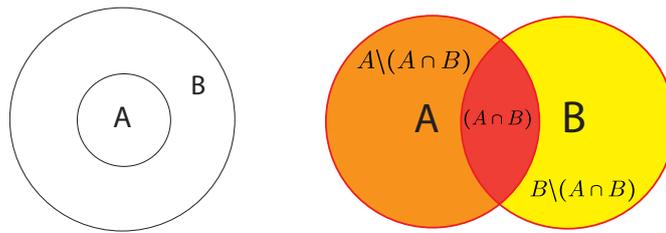


Abbildung 2.2: Illustration zu Satz 2.2.1, 3) und 4)

4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, $\forall A, B \in \mathcal{F}$.
5. $P(\overline{B}) = 1 - P(B)$

Beweis

1. $\emptyset = \bigcup_{i=1}^{\infty} \emptyset \implies P(\emptyset) = \sum_{i=1}^{\infty} P(\emptyset) \leq 1 \implies P(\emptyset) = 0$
2. $P(\bigcup_{i=1}^n A_i) = P(\bigcup_{i=1}^n A_i \cup \bigcup_{k=n+1}^{\infty} \emptyset) = \sum_{i=1}^n P(A_i) + \sum_{i=n+1}^{\infty} 0 = \sum_{i=1}^n P(A_i)$
3. $P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \implies$ geht.
4. Benutze 2), 3) und $A \cup B = [A \setminus (A \cap B)] \cup [A \cap B] \cup [B \setminus (A \cap B)]$

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

(vgl. Abb. 2.2).

$$5. \overline{B} = \Omega \setminus B \Rightarrow P(\overline{B}) = \underbrace{P(\Omega)}_{=1} - P(B) = 1 - P(B).$$

□

Folgerung 2.2.1

Es gelten folgende Eigenschaften von P für $A_1, \dots, A_n, A, B \in \mathcal{F}$:

1. $P(A^C) = 1 - P(A)$
2. $A \subseteq B \implies P(A) \leq P(B)$ (Monotonie)
3. $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$, $\forall A_1, \dots, A_n \in \mathcal{F}$ (Subadditivität)
4. *Siebformel*:

$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} P(A_{k_1} \cap \dots \cap A_{k_i})$$

Beweis

1. $1 = P(\Omega) = P(A \cup A^C) = P(A) + P(A^C) \implies P(A^C) = 1 - P(A)$
2. (Folgt aus Satz 2.2.1) $P(B) = P(A) + \underbrace{P(B \setminus A)}_{\geq 0} \geq P(A)$
3. Induktion nach n : $n=2$: $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$
4. Induktion nach n : Der Fall $n = 2$ folgt aus Satz 2.2.1, 4). Der Rest ist klar.

□

Übungsaufgabe 2.2.3 Führen Sie die Induktion bis zum Ende durch.

Die Eigenschaft $P(\bigcup_{n=1}^k A_n) \leq \sum_{n=1}^k P(A_n)$ heißt *Subadditivität des Wahrscheinlichkeitsmaßes P* . Diese Eigenschaft gilt jedoch auch für unendlich viele A_n , wie folgendes Korollar zeigt:

Folgerung 2.2.2 *σ -Subadditivität*: Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsmaß und $\{A_n\}_{n \in \mathbb{N}}$ eine Folge von Ereignissen. Dann gilt $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$, wobei die rechte Seite nicht unbedingt endlich sein soll (dann ist die Aussage trivial).

Definition 2.2.4 1. Ereignisse A und B heißen (*stochastisch*) *unabhängig*, falls $P(A \cap B) = P(A) \cdot P(B)$.

2. Eine Folge von Ereignissen $\{A_n\}_{n \in \mathbb{N}}$ (diese Folge kann auch endlich viele Ereignisse enthalten!) heißt (*stochastisch*) *unabhängig in ihrer Gesamtheit*, falls $\forall n \forall i_1 < i_2 < \dots < i_n$

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = \prod_{k=1}^n P(A_{i_k}).$$

2.3 Beispiele

In diesem Abschnitt betrachten wir die wichtigsten Beispiele für Wahrscheinlichkeitsräume (Ω, \mathcal{F}, P) . Wir beginnen mit:

2.3.1 Klassische Definition der Wahrscheinlichkeiten

Hier wird ein Grundraum Ω mit $|\Omega| < \infty$ ($|A| = \#A$ – die Anzahl von Elementen in A) betrachtet. Dann kann \mathcal{F} als $\mathcal{P}(\Omega)$ gewählt werden. Die klassische Definition von Wahrscheinlichkeiten geht von der Annahme aus, dass alle Elementarereignisse ω gleich wahrscheinlich sind:

Definition 2.3.1

1. Ein Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit $|\Omega| < \infty$ ist *endlicher Wahrscheinlichkeitsraum*.
2. Ein endlicher Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit $\mathcal{F} = \mathcal{P}(\Omega)$ und

$$\forall \omega \in \Omega \quad P(\{\omega\}) = \frac{1}{|\Omega|}$$

heißt *Laplacescher Wahrscheinlichkeitsraum*. Das eingeführte Maß heißt *klassisches* oder *laplacesches Wahrscheinlichkeitsmaß*.

Bemerkung 2.3.1 Für die klassische Definition der Wahrscheinlichkeit sind alle Elementarereignisse $\{\omega\}$ gleich wahrscheinlich:

$\forall \omega \in \Omega \quad P(\{\omega\}) = \frac{1}{|\Omega|}$. Nach der Additivität von Wahrscheinlichkeitsmaßen gilt:

$$P(A) = \frac{|A|}{|\Omega|} \quad \forall A \subset \Omega.$$

(Beweis: $P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}$). Dabei heißt

$$P(A) = \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl aller Fälle}}.$$

Beispiel 2.3.11. *Problem von Galilei:*

Ein Landsknecht hat Galilei (manche sagen, es sei Huygens passiert) folgende Frage gestellt: Es werden 3 Würfel gleichzeitig geworfen. Was ist wahrscheinlicher: Die Summe der Augenzahlen ist 11 oder 12? Nach Beobachtung sollte 11 öfter vorkommen als 12. Doch ist es tatsächlich so?

- Definieren wir den Wahrscheinlichkeitsraum $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_i \in \{1, \dots, 6\}\}$,
 $|\Omega| = 6^3 = 216 < \infty$, $\mathcal{F} = \mathcal{P}(\Omega)$; sei

$$\begin{aligned} B &:= \{\text{Summe der Augenzahlen } 11\} \\ &= \{\omega \in \Omega : \omega_1 + \omega_2 + \omega_3 = 11\} \\ C &:= \{\text{Summe der Augenzahlen } 12\} \\ &= \{\omega \in \Omega : \omega_1 + \omega_2 + \omega_3 = 12\}. \end{aligned}$$

- *Lösung des Landknechtes:* 11 und 12 können folgendermaßen in die Summe von 3 Summanden zerlegt werden:
 $11 = 1 + 5 + 5 = 1 + 4 + 6 = 2 + 3 + 6 = 2 + 4 + 5 = 3 + 3 + 5 = 3 + 4 + 4 \implies |B| = 6 \implies P(B) = \frac{6}{6^3} = \frac{1}{36}$
 $12 = 1 + 5 + 6 = 2 + 4 + 6 = 2 + 5 + 5 = 3 + 4 + 5 = 3 + 3 + 6 = 4 + 4 + 4 \implies P(C) = \frac{6}{6^3} = \frac{1}{36}$.
 Dies entspricht jedoch nicht der Erfahrung.

Die Antwort von Galilei war, dass der Landsknecht mit nicht unterscheidbaren Würfeln gearbeitet hat, somit waren Kombinationen wie (1,5,5), (5,1,5) und (5,5,1) identisch und wurden nur einmal gezählt. In der Tat ist es anders: Jeder Würfel hat eine Nummer, ist also von den anderen Zwei zu unterscheiden. Daher gilt $|B| = 27$ und $|C| = 25$, was daran liegt, das (4,4,4) nur einmal gezählt wird. Also

$$P(B) = \frac{|B|}{|\Omega|} = \frac{27}{216} > P(C) = \frac{|C|}{|\Omega|} = \frac{25}{216}$$

2. *Geburtstagsproblem:* Es gibt n Studenten in einem Jahrgang an der Uni, die dieselbe Vorlesung Angewandte Stochastik I besuchen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 Studenten den Geburtstag am selben Tag feiern? Sei $M = 365 =$ Die Anzahl der Tage im Jahr. Dann gilt

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, M\}\}, |\Omega| = M^n < \infty.$$

n	4	16	22	23	40	64
$P(A_n)$	0,016	0,284	0,476	0,507	0,891	0,997

Tabelle 2.2: Geburtstagsproblem

Sei

$$\begin{aligned} A_n &= \{\text{min. 2 Studenten haben am gleichen Tag Geb.}\} \subset \Omega, \\ A_n &= \{\omega \in \Omega : \exists i, j \in \{1, \dots, n\}, i \neq j : \omega_i = \omega_j\}, \\ P(A_n) &= ? \end{aligned}$$

Ansatz: $P(A_n) = 1 - P(\bar{A}_n)$, wobei

$$\bar{A}_n = \{\omega \in \Omega : \omega_i \neq \omega_j \quad \forall i \neq j \text{ in } \omega = (\omega_1, \dots, \omega_n)\}$$

$|\bar{A}_n| = M(M-1)(M-2)\dots(M-n+1)$. Somit gilt

$$\begin{aligned} P(\bar{A}_n) &= \frac{M(M-1)\dots(M-n+1)}{M^n} \\ &= \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \dots \left(1 - \frac{n-1}{M}\right) \end{aligned}$$

und

$$P(A_n) = 1 - \left(1 - \frac{1}{M}\right) \dots \left(1 - \frac{n-1}{M}\right)$$

Für manche n gibt Tabelle 2.2 die numerischen Wahrscheinlichkeiten von $P(A_n)$ an.

Es gilt offensichtlich $P(A_n) \approx 1$ für $n \rightarrow M$. Interessanterweise ist $P(A_n) \approx 0,5$ für $n = 23$. Dieses Beispiel ist ein Spezialfall eines so genannten *Urnenmodells*: In einer Urne liegen M durchnummerierte Bälle. Aus dieser Urne werden n Stück auf gut Glück mit Zurücklegen entnommen. Wie groß ist die Wahrscheinlichkeit, dass in der Stichprobe mindestens 2 Gleiche vorkommen?

3. *Urnenmodelle*: In einer Urne gibt es M durchnummerierte Bälle. Es werden n Stück "zufällig" entnommen. Das Ergebnis dieses Experimentes ist eine Stichprobe (j_1, \dots, j_n) , wobei j_m die Nummer des Balls in der m -ten Ziehung ist. Es werden folgende Arten der Ziehung betrachtet:

- mit Zurücklegen
- ohne Zurücklegen

- mit Reihenfolge
- ohne Reihenfolge

Das Geburtstagsproblem ist somit ein Urnenproblem mit Zurücklegen und mit Reihenfolge.

Demnach werden auch folgende Grundräume betrachtet:

(a) Ziehen mit Reihenfolge und mit Zurücklegen:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \underbrace{\{1, \dots, M\}}_{=K}\} = K^n, \quad |\Omega| = M^n$$

(b) Ziehen mit Reihenfolge und ohne Zurücklegen:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in K, \omega_i \neq \omega_j, i \neq j\},$$

$$|\Omega| = M(M-1) \dots (M-n+1) = \frac{M!}{(M-n)!}$$

Spezialfall: $M=n$ (Permutationen): $\implies |\Omega| = M!$

(c) Ziehen ohne Reihenfolge und mit Zurücklegen:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in K, \omega_1 \leq \omega_2 \leq \dots \leq \omega_n\}$$

Dies ist äquivalent zu der Verteilung von n Teilchen auf M Zellen ohne Reihenfolge \iff das Verteilen von $M-1$ Trennwänden der Zellen unter n Teilchen (vgl. Abb. 2.3). Daher ist

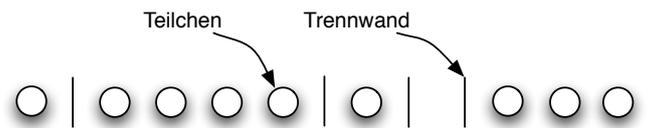


Abbildung 2.3: Ziehen ohne Reihenfolge und mit Zurücklegen

$$|\Omega| = \frac{(M+n-1)!}{n!(M-1)!} = \binom{M+n-1}{n}$$

(d) Ziehen ohne Reihenfolge und ohne Zurücklegen:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n); \omega_i \in K, \omega_1 < \omega_2 < \dots < \omega_n\}$$

$$|\Omega| = \frac{M!}{(M-n)!n!} = \binom{M}{n}$$

Auswahl von n aus M Kugeln in einer Urne	mit Zurücklegen	ohne Zurücklegen	
mit Reihenfolge	M^n (<i>Maxwell-Boltzmann-Statistik</i>)	$\frac{M!}{(M-n)!}$	unterscheidbare Teilchen
ohne Reihenfolge	$\binom{M+n-1}{n}$ (<i>Bose-Einstein-Statistik</i>)	$\binom{M}{n}$ (<i>Fermi-Dirac-Statistik</i>)	nicht unterscheidbare Teilchen
	mit Mehrfachbelegung	ohne Mehrfachbelegung	Verteilung von n Teilchen auf M Zellen.

Tabelle 2.3: Die Potenz $|\Omega|$ der Grundgesamtheit Ω in Urnenmodellen.

Ein Experiment der Mehrfachziehung aus einer Urne entspricht der Verteilung von n (unterschiedlichen oder nicht unterscheidbaren) Teilchen (wie z.B. Elektronen, Protonen, usw.) auf M Energieebenen oder Zellen (mit oder ohne Mehrfachbelegung dieser Ebenen) in der statistischen Physik. Die entsprechenden Namen der Modelle sind in Tabelle 2.3 zusammengeführt. So folgen z.B. Elektronen, Protonen und Neutronen der so genannten *Fermi-Dirac-Statistik* (nicht unterscheidbare Teilchen ohne Mehrfachbelegung). Photonen und Prionen folgen der *Bose-Einstein-Statistik* (nicht unterscheidbare Teilchen mit Mehrfachbelegung). Unterscheidbare Teilchen, die dem *Exklusionsprinzip von Pauli* folgen (d.h. ohne Mehrfachbelegung), kommen in der Physik nicht vor.

4. *Lotterie-Beispiele*: ein Urnenmodell ohne Reihenfolge und ohne Zurücklegen; In einer Lotterie gibt es M Lose (durchnummeriert von 1 bis M), davon n Gewinne ($M \geq 2n$). Man kauft n Lose. Mit welcher Wahrscheinlichkeit gewinnt man mindestens einen Preis?

Laut Tabelle 2.3 ist $|\Omega| = \binom{M}{n}$,

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \neq \omega_j, i \neq j, \omega_i \in \{1 \dots M\}\}.$$

Sei $A = \{\text{es gibt mind. 1 Preis}\}.$

$$\begin{aligned}
P(A) &= 1 - P(\bar{A}) = 1 - P(\text{es werden keine Preise gewonnen}) \\
n &= 1 - \frac{\binom{M-n}{n}}{|\Omega|} = 1 - \frac{\frac{(M-n)!}{n!(M-2n)!}}{\frac{M!}{n!(M-n)!}} \\
&= 1 - \frac{(M-n)(M-n-1)\dots(M-2n+1)}{M(M-1)\dots(M-n+1)} \\
&= 1 - \left(1 - \frac{n}{M}\right) \left(1 - \frac{n}{M-1}\right) \dots \left(1 - \frac{n}{M-n+1}\right)
\end{aligned}$$

Um ein Beispiel zu geben, sei $M = n^2$. Dann gilt:

$P(A) \xrightarrow{n \rightarrow \infty} 1 - e^{-1} \approx 0,632$, denn $e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$. Die Konvergenz ist schnell, $P(A) = 0,670$ schon für $n = 10$.

5. *Hypergeometrische Verteilung*: Nehmen wir jetzt an, dass M Kugeln in der Urne zwei Farben tragen können: schwarz und weiß. Seien S schwarze und W weiße Kugeln gegeben ($M = S + W$). Wie groß ist die Wahrscheinlichkeit, dass aus n zufällig entnommenen Kugeln (ohne Reihenfolge und ohne Zurücklegen) s schwarz sind?
Sei $A = \{\text{unter } n \text{ entnommenen Kugeln } s \text{ schwarze}\}$. Dann ist

$$P(A) = \frac{\binom{S}{s} \binom{W}{n-s}}{\binom{M}{n}}.$$

Diese Wahrscheinlichkeiten bilden die so genannte *hypergeometrische Verteilung*.

Um ein numerisches Beispiel zu geben, seien 36 Spielkarten gegeben. Sie werden zufällig in zwei gleiche Teile aufgeteilt. Wie groß ist die Wahrscheinlichkeit, dass die Anzahl von roten und schwarzen Karten in diesen beiden Teilen gleich ist?

Lösung: hypergeometrische Wahrscheinlichkeiten mit $M = 36$, $S = W = n = 18$, $s = \frac{18}{2} = 9$, $w = s = 9$. Dann ist

$$P(A) = \frac{\binom{18}{9} \binom{18}{9}}{\binom{36}{18}} = \frac{(18!)^4}{36!(9!)^4}.$$

Wenn man die Formel von Stirling

$$n! \approx \sqrt{2\pi n} n^n e^{-n}$$

benutzt, so kommt man auf

$$P(A) \approx \frac{(\sqrt{2\pi 18} \cdot 18^{18} e^{-18})^4}{\sqrt{2\pi 36} \cdot 36^{36} e^{-36} (\sqrt{2\pi 9} \cdot 9^9 e^{-9})^4} \approx \frac{2}{\sqrt{18\pi}} \approx \frac{4}{15} \approx 0.26$$

2.3.2 Geometrische Wahrscheinlichkeiten

Hier sei ein Punkt π zufällig auf eine beschränkte Teilmenge Ω von \mathbb{R}^d geworfen. Wie groß ist die Wahrscheinlichkeit, dass π die Teilmenge $A \subset \Omega$ trifft? Um dieses Experiment formalisieren zu können, dürfen wir nur solche Ω und A zulassen, für die der Begriff des d -dimensionalen Volumens (Lebesgue-Maß) wohl definiert ist. Daher werden wir nur Borelsche Teilmengen von \mathbb{R}^d betrachten. Also sei $\Omega \in \mathcal{B}_{\mathbb{R}^d}$ und $|\cdot|$ das Lebesgue-Maß auf \mathbb{R}^d , $|\Omega| < \infty$. Sei $\mathcal{F} = \mathcal{B}_{\mathbb{R}^d} \cap \Omega$ (vgl. Abb. 2.4).

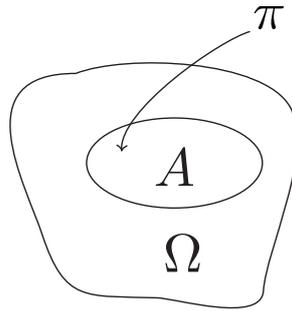


Abbildung 2.4: Zufälliger Punkt π auf Ω .

Definition 2.3.2

1. Das Wahrscheinlichkeitsmaß auf (Ω, \mathcal{F}) gegeben durch

$$P(A) = \frac{|A|}{|\Omega|}, \quad A \in \mathcal{F}$$

heißt *geometrische Wahrscheinlichkeit* auf Ω .

2. Das Tripel (Ω, \mathcal{F}, P) heißt *geometrischer Wahrscheinlichkeitsraum*.

Im Folgenden betrachten wir ein paar berühmte Probleme, die mit der geometrischen Wahrscheinlichkeit zu tun haben:

1. Die Koeffizienten p und q einer quadratischen Gleichung $x^2 + px + q = 0$ werden zufällig im Intervall $(0, 1)$ gewählt. Wie groß ist die Wahrscheinlichkeit, dass die Lösungen x_1, x_2 dieser Gleichung reelle Zahlen sind?

Hier ist $\Omega = \{(p, q) : p, q \in (0, 1)\} = (0, 1)^2$, $\mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap \Omega$.

$$A = \{x_1, x_2 \in \mathbb{R}\} = \{(p, q) \in \Omega : p^2 \geq 4q\},$$

denn $x_1, x_2 \in \mathbb{R}$ genau dann, wenn die Diskriminante $D = p^2 - 4q \geq 0$. Also gilt $A = \{(p, q) \in [0, 1]^2 : q \leq \frac{1}{4}p^2\}$ und

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\int_0^1 \frac{1}{4}p^2 dp}{1} = \frac{1}{12},$$

vgl. Abb. 2.5.

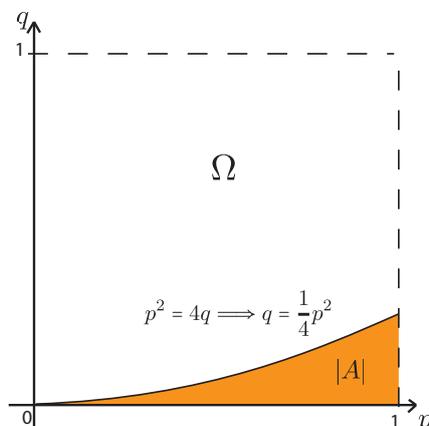


Abbildung 2.5: Wahrscheinlichkeit für reelle Lösungen einer quadratischen Gleichung

2.3.3 Bedingte Wahrscheinlichkeiten

Um den Begriff der bedingten Wahrscheinlichkeit intuitiv einführen zu können, betrachten wir zunächst das Beispiel der klassischen Wahrscheinlichkeiten: Sei (Ω, \mathcal{F}, P) ein Laplacescher Wahrscheinlichkeitsraum mit $|\Omega| = N$. Seien A und B Ereignisse aus \mathcal{F} . Dann gilt

$$P(A) = \frac{|A|}{N}, \quad P(A \cap B) = \frac{|A \cap B|}{N}.$$

Wie groß ist die Wahrscheinlichkeit $P(A|B)$ von A unter der Bedingung, dass B eintritt?

Da B eingetreten ist, ist die Gesamtanzahl aller Elementarereignisse hier gleich $|B|$. Die Elementarereignisse, die zu A beim Eintreten von B führen, liegen alle in $A \cap B$. Somit ist die Anzahl der "günstigen" Fälle hier $|A \cap B|$ und wir bekommen

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

Dies führt zu folgender Definition:

Definition 2.3.3

Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum, $A, B \in \mathcal{F}$, $P(B) > 0$. Dann ist die *bedingte Wahrscheinlichkeit* von A unter der Bedingung B

gegeben durch

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

Diese Definition kann in Form des sogenannten Multiplikationssatzes gegeben werden:

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

Übungsaufgabe 2.3.1 Zeigen Sie, dass $P(\cdot|B)$ für $B \in \mathcal{F}$, $P(B) > 0$ ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{F}) ist.

Satz 2.3.1 Seien $A_1, \dots, A_n \in \mathcal{F}$ Ereignisse mit $P(A_1 \cap \dots \cap A_{n-1}) > 0$, dann gilt $P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1})$.

Übungsaufgabe 2.3.2 Beweisen Sie den Satz 2.3.1.

Beweisidee: Induktion bezüglich n .

An dieser Stelle sollte man zu den unabhängigen Ereignissen zurückkehren. A und B sind nach Definition 2.2.4 unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$. Dies ist äquivalent zu $P(A|B) = P(A)$, falls $P(B) > 0$. Es sei allerdings an dieser Stelle angemerkt, dass die Definition 2.2.4 allgemeiner ist, weil sie auch den Fall $P(B) = 0$ zulässt.

Übungsaufgabe 2.3.3 Zeigen Sie folgendes:

1. Seien $A, B \in \mathcal{F}$. A und B sind (stochastisch) unabhängig genau dann, wenn A und \bar{B} oder $(\bar{A}$ und $\bar{B})$ unabhängig sind.
2. Seien $A_1, \dots, A_n \in \mathcal{F}$. Ereignisse A_1, \dots, A_n sind stochastisch unabhängig in ihrer Gesamtheit genau dann, wenn B_1, \dots, B_n unabhängig in ihrer Gesamtheit sind, wobei $B_i = A_i$ oder $B_i = \bar{A}_i$ für $i = 1, \dots, n$.
3. Seien $A, B_1, B_2 \in \mathcal{F}$ mit $B_1 \cap B_2 = \emptyset$. Sei A und B_1 , A und B_2 unabhängig. Zeigen Sie, dass A und $B_1 \cup B_2$ ebenfalls unabhängig sind.

Bemerkung 2.3.2 Der in Definition 2.2.4 gegebene Begriff der stochastischen Unabhängigkeit ist viel allgemeiner als die sogenannte Unabhängigkeit im Sinne des Gesetzes von Ursache und Wirkung. In den folgenden Beispielen wird man sehen, dass zwei Ereignisse stochastisch unabhängig sein können, obwohl ein kausaler Zusammenhang zwischen ihnen besteht. Somit ist die stochastische Unabhängigkeit allgemeiner, und nicht an das Gesetz von Ursache und Wirkung gebunden. In der Praxis allerdings ist man gut beraten, Ereignisse, die keinen kausalen Zusammenhang haben als stochastisch unabhängig zu deklarieren.

Beispiel 2.3.2 1. *Abhängige und unabhängige Ereignisse:*

Es werde ein Punkt $\pi = (X, Y)$ zufällig auf $[0, 1]^2$ geworfen. $\Omega =$

$[0, 1]^2$, $\mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap [0, 1]$. Betrachten wir $A = \{X \geq a\}$ und $B = \{Y \geq b\}$. Dann gilt

$$\begin{aligned} P(A \cap B) &= P(X \geq a, Y \geq b) \\ &= \frac{(1-a)(1-b)}{1} \\ &= P(A) \cdot P(B), \end{aligned}$$

insofern sind A und B stochastisch unabhängig. Allerdings kann für $B = \{\pi \in \Delta CDE\}$ leicht gezeigt werden, dass A und B voneinander abhängig sind.

2. Es können $n + 1$ Ereignisse konstruiert werden, die abhängig sind, wobei beliebige n von ihnen unabhängig sind, $\forall n \in \mathbb{N}$.
3. *Kausale und stochastische Unabhängigkeit:*

Auf das Intervall $[0, 1]$ wird auf gut Glück ein Punkt π geworfen. Sei x die Koordinate von π in $[0, 1]$. Betrachten wir die binäre Zerlegung der Zahl x :

$$x = \sum_{k=1}^{\infty} \frac{a_k}{2^k}, \quad a_n \in \{0, 1\}.$$

Dann ist klar, dass es einen starken kausalen Zusammenhang zwischen $\{a_n\}_{n=1}^{\infty}$ gibt, weil sie alle durch x verbunden sind. Man kann jedoch zeigen, dass die Ereignisse $\{a_k = B_k\}, k \in \mathbb{N}$ für alle $j = 0, 1$ unabhängig in ihrer Gesamtheit sind, und dass $P(a_k = j) = 1/2 \forall k \in \mathbb{N}, j = 0, 1$.

Definition 2.3.4 Sei $\{B_n\}$ eine endliche oder abzählbare Folge von Ereignissen aus \mathcal{F} . Sie heißt eine *messbare Zerlegung* von Ω , falls

1. B_n paarweise disjunkt sind: $B_i \cap B_j = \emptyset, \quad i \neq j$
2. $\bigcup_n B_n = \Omega$
3. $P(B_n) > 0 \quad \forall n$.

Satz 2.3.2 (*Formel der totalen Wahrscheinlichkeit, Bayes'sche Formel*):

Sei $\{B_n\} \subset \mathcal{F}$ eine messbare Zerlegung von Ω und $A \in \mathcal{F}$ ein beliebiges Ereignis, dann gilt

1. *Die Formel der totalen Wahrscheinlichkeit:*

$$P(A) = \sum_n P(A|B_n) \cdot P(B_n)$$

2. *Bayes'sche Formel:*

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_n P(A|B_n) \cdot P(B_n)} \quad \forall i$$

falls $P(A) > 0$. Die Summen in 1) und 2) können endlich oder unendlich sein, je nach Anzahl der B_n .

Beweis 1. Da $\Omega = \bigcup_n B_n$, ist $A = A \cap \Omega = A \cap (\bigcup_n B_n) = \bigcup_n (A \cap B_n)$ eine disjunkte Vereinigung von Ereignissen $A \cap B_n$, und es gilt

$$P(A) = P\left(\bigcup_n (A \cap B_n)\right) \stackrel{\text{\(\sigma\)-Add. v. } P}{=} \sum_n P(A \cap B_n) \stackrel{\text{S. 2.3.1}}{=} \sum_n P(A|B_n)P(B_n)$$

2.

$$P(B_i|A) \stackrel{\text{Def. 2.3.3}}{=} \frac{P(B_i \cap A)}{P(A)} \stackrel{\text{S. 2.3.1 u. 2.3.2 1)}}{=} \frac{P(B_i)P(A|B_i)}{\sum_n P(A|B_n)P(B_n)}$$

□

Bemerkung 2.3.3 Die Ereignisse B_n heißen oft ‘Hypothesen’. Dann ist $P(B_n)$ die so genannte *a-priori-Wahrscheinlichkeit von B_n* , also vor dem ‘Experiment’ A . Die Wahrscheinlichkeiten $P(B_n|A)$ werden als Wahrscheinlichkeiten des Auftretens von B_n ‘nach dem Experiment A ’ interpretiert. Daher heißen sie auch oft ‘*a-posteriori-Wahrscheinlichkeiten von B_n* ’. Die Formel von Bayes verbindet also die a-posteriori-Wahrscheinlichkeiten mit den a-priori-Wahrscheinlichkeiten.

Beispiel 2.3.3

1. *Routing-Problem:*

Im Internet muss ein Paket von Rechner S (Sender) auf den Rechner E (Empfänger) übertragen werden. In Abb. 2.6 ist die Geometrie des Computernetzes zwischen S und E schematisch dargestellt, wobei R_1, R_2, R_3 und R_4 (und andere Knoten des Graphen) jeweils andere Rechner sind, die sich an der Übertragung beteiligen können. Wir gehen davon aus, dass die Richtung der weiteren Übertragung des Paketes in den Knoten zufällig aus allen möglichen Knoten gewählt wird (mit gleicher Wahrscheinlichkeit). So ist z.B.

$$P(\underbrace{\text{von } S \text{ wird Router } R_i \text{ gewählt}}_{=A_i}) = \frac{1}{4} \quad i = 1, \dots, n.$$

Offensichtlich stellen die Ereignisse A_1, A_2, A_3, A_4 eine messbare Zerlegung von Ω dar. Nach Satz 2.3.2, 1) gilt also für $A = \{\text{das Paket erreicht } E \text{ aus } S\}$

$$P(A) = \sum_{i=1}^4 P(A|A_i) \cdot P(A_i) = \frac{1}{4} \sum_{i=1}^4 P(A|A_i).$$

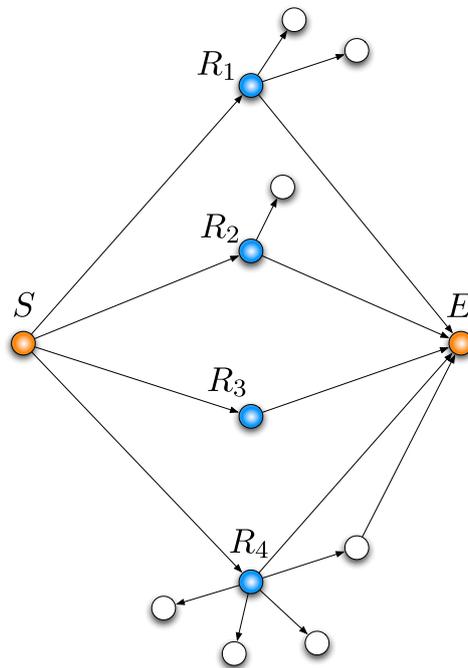


Abbildung 2.6: Routing-Problem: Computernetzwerk

Dabei können $P(A|A_i)$ aus dem Graphen eindeutig bestimmt werden:

$$P(A|A_1) = \frac{1}{3}, \quad P(A|A_2) = \frac{1}{2},$$

$$P(A|A_3) = 1, \quad P(A|A_4) = \frac{2}{5}.$$

Es gilt also

$$P(A) = \frac{1}{4} \left(\frac{1}{3} + \frac{1}{2} + 1 + \frac{2}{5} \right) = \frac{67}{120} \approx 0,5.$$

2. In einer Urne gibt es zwei Münzen. Die erste ist fair (Wahrscheinlichkeit des Kopfes und der Zahl = $\frac{1}{2}$), die zweite ist allerdings nicht fair mit $P(\text{Kopf}) = \frac{1}{3}$. Aus der Urne wird eine Münze zufällig genommen und geworfen. In diesem Wurf ist das Ereignis Kopf. Wie groß ist die Wahrscheinlichkeit, dass die Münze fair war?

Sei

$$A_1 = \{\text{Faire Münze ausgewählt}\}$$

$$A_2 = \{\text{Nicht faire Münze ausgewählt}\}$$

$$A = \{\text{Es kommt Kopf im Münzwurf}\}$$

$$P(A_1|A) = ?$$

Dann gilt $P(A_1) = P(A_2) = \frac{1}{2}$, $P(A|A_1) = \frac{1}{2}$, $P(A|A_2) = \frac{1}{3}$, daher gilt nach der Bayesschen Formel

$$P(A_1|A) = \frac{P(A_1) \cdot P(A|A_1)}{P(A_1) \cdot P(A|A_1) + P(A_2) \cdot P(A|A_2)} = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2} \cdot (\frac{1}{2} + \frac{1}{3})} = \frac{3}{5}.$$

Kapitel 3

Zufallsvariablen

3.1 Definition und Beispiele

Definition 3.1.1 1. Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ heißt *Zufallsvariable*, falls sie $\mathcal{B}_{\mathbb{R}}$ -messbar ist, mit anderen Worten,

$$\forall B \in \mathcal{B}_{\mathbb{R}} \quad X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

2. Eine Abbildung $X : \Omega \rightarrow \mathbb{R}^n$, $n \geq 1$ heißt *Zufallsvektor*, falls sie $\mathcal{B}_{\mathbb{R}^n}$ -messbar ist, mit anderen Worten,

$$\forall B \in \mathcal{B}_{\mathbb{R}^n} \quad X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Offensichtlich bekommt man aus Definition 3.1.1, 2) auch 3.1.1, 1) für $n = 1$.

Beispiel 3.1.1 1. *Indikator-Funktion eines Ereignisses:*

Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum und A ein Ereignis aus \mathcal{F} . Betrachten wir

$$X(\omega) = I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}.$$

Diese Funktion von ω nennt man *Indikator-Funktion des Ereignisses* A . Sie ist offensichtlich messbar und somit eine Zufallsvariable:

$$X^{-1}(B) = \begin{cases} A & \text{falls } 1 \in B, 0 \notin B \\ \bar{A} & \text{falls } 1 \notin B, 0 \in B \\ \Omega & \text{falls } 0, 1 \in B \\ \emptyset & \text{falls } 0, 1 \notin B \end{cases} \in \mathcal{F} \quad \forall B \in \mathcal{B}_{\mathbb{R}}$$

2. *n-maliger Münzwurf:*

Sei $\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{0, 1\}\} = \{0, 1\}^n$ mit

$$\omega_i = \begin{cases} 1, & \text{falls Kopf im } i\text{-ten Münzwurf} \\ 0, & \text{sonst} \end{cases}$$

für $i = 1, \dots, n$. Sei $\mathcal{F} = \mathcal{P}(\Omega)$. Definieren wir

$$X(\omega) = X((\omega_1, \dots, \omega_n)) = \sum_{i=1}^n \omega_i$$

als die Anzahl der Köpfe im n -maligen Münzwurf, so kann man zeigen, dass X \mathcal{F} -messbar ist und somit eine Zufallsvariable ist. X ist stetig als lineare Abbildung, somit ist für alle B offen oder geschlossen nach continuous mapping theorem $X^{-1}(B)$ offen oder geschlossen. Weiter per Definition einer σ -Algebra.

Satz 3.1.1 Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ ist genau dann eine Zufallsvariable, wenn $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$.

3.2 Verteilungsfunktion

Definition 3.2.1 Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße.

1. Die Funktion $F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$, $x \in \mathbb{R}$ heißt *Verteilungsfunktion* von X . Offensichtlich ist $F_X : \mathbb{R} \rightarrow [0, 1]$.
2. Die Mengenfunktion $P_X : \mathcal{B}_{\mathbb{R}} \rightarrow [0, 1]$ gegeben durch

$$P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}), B \in \mathcal{B}_{\mathbb{R}}$$

heißt *Verteilung* von X .

Bemerkung 3.2.1 Folgende gekürzte Schreibweise wird benutzt:

$$F_X(x) = P(X \leq x), \quad P_X(B) = P(X \in B).$$

Beispiel 3.2.1

Hier geben wir Verteilungsfunktionen für Zufallsvariablen aus dem Beispiel 3.1.1 an.

1. *Indikator-Funktion:*
Sei $X(\omega) = I_A(\omega)$. Dann ist

$$F_X(x) = P(I_A \leq x) = \begin{cases} 1, & x \geq 1, \\ P(\bar{A}), & x \in [0, 1), \\ 0, & x < 0 \end{cases}$$

vgl. Abb. 3.1.

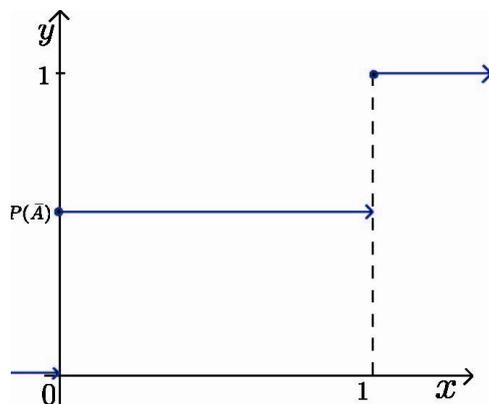


Abbildung 3.1: Verteilungsfunktion von I_A

2. n -maliger Münzwurf:

Sei $X =$ Anzahl Kopf in n Münzwürfen. $P(\text{Kopf in einem Wurf}) = p$, $p \in (0, 1)$. Dann gilt

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

und somit

$$F_X(x) = P(X \leq x) = \sum_{0 \leq k \leq [x]} P(x = k) = \sum_{k=0}^{[x]} \binom{n}{k} p^k (1 - p)^{n-k},$$

$\forall x \in [0, n]$, vgl. Abb. 3.2. Es gilt

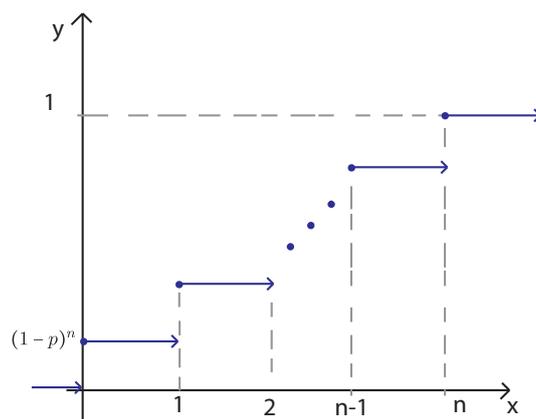


Abbildung 3.2: Verteilungsfunktion einer $Bin(n, p)$ -Verteilung

$$\begin{aligned}
 F_X(0) &= P(X \leq 0) = P(X = 0) = (1 - p)^n, \\
 F_X(x) &= P(X \leq x) = 0 \quad \text{für } x < 0, \\
 F_X(n) &= P(X \leq n) = 1.
 \end{aligned}$$

Diese Verteilung wird später *Binomial-Verteilung* mit Parametern n, p genannt: $\text{Bin}(n, p)$

Satz 3.2.1 Sei X eine beliebige Zufallsvariable und $F_X : \mathbb{R} \rightarrow [0, 1]$ ihre Verteilungsfunktion. F_X besitzt folgende Eigenschaften:

1. *Asymptotik*: $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
2. *Monotonie*: $F_X(x) \leq F_X(x + h)$, $\forall x \in \mathbb{R}, h \geq 0$.
3. *Rechtsseitige Stetigkeit*: $\lim_{x \rightarrow x_0+0} F_X(x) = F_X(x_0) \quad \forall x_0 \in \mathbb{R}$.

Bemerkung 3.2.2

1. Im Satz 3.2.1 wurde gezeigt, dass eine Verteilungsfunktion F_X monoton nicht-fallend, rechtsseitig stetig und beschränkt auf $[0, 1]$ ist. Diese Eigenschaften garantieren, dass F_X höchstens abzählbar viele Sprungstellen haben kann. In der Tat kann F_X wegen $F_X \uparrow$ und $0 \leq F_X \leq 1$ nur eine endliche Anzahl von Sprungstellen mit Sprunghöhe $> \varepsilon$ besitzen, $\forall \varepsilon > 0$. Falls ε_n die Menge \mathbb{Q} aller rationaler Zahlen durchläuft, wird somit gezeigt, dass die Anzahl aller möglichen Sprungstellen höchstens abzählbar sein kann. Die Grafik einer typischen Verteilungsfunktion ist in Abb. 3.3 dargestellt.

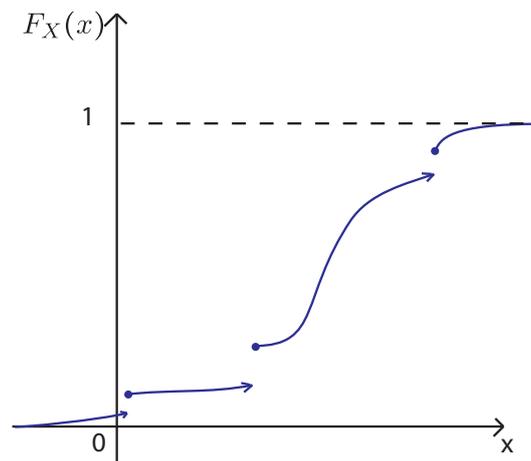


Abbildung 3.3: Typische Verteilungsfunktion

2. Mit Hilfe von F_X können folgende Wahrscheinlichkeiten leicht berechnet werden: $\forall a, b$ gilt: $-\infty \leq a < b \leq +\infty$

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

$$P(a \leq X \leq b) = F_X(b) - \lim_{x \rightarrow a-0} F_X(x),$$

denn

$$P(a < X \leq b) = P(\{X \leq b\} \setminus \{X \leq a\}) = P(X \leq b) - P(X \leq a)$$

$$= F_X(b) - F_X(a),$$

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F_X(b) - \lim_{x \rightarrow a-0} F_X(x)$$

mit $P(X < a) = P(X \leq a) - P(X = a) = \lim_{x \rightarrow a-0} F_X(x)$ nach Stetigkeit von P_X .

Da $P(X < a) = F(X \leq a) - P(X = a)$ gilt, ist somit

$$\lim_{x \rightarrow a-0} F_X(x) \neq F_X(a)$$

und F_X im Allgemeinen nicht linksseitig stetig.

Übungsaufgabe 3.2.1 Drücken Sie die Wahrscheinlichkeiten $P(a < X < b)$ und $P(a \leq X < b)$ mit Hilfe von F_X aus.

Satz 3.2.2 Falls eine Funktion $F(x)$ die Eigenschaften 1) bis 3) des Satzes 3.2.1 erfüllt, dann existiert ein Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) und eine Zufallsvariable X , definiert auf diesem Wahrscheinlichkeitsraum, derart, dass $F_X(x) = F(x)$, $\forall x \in \mathbb{R}$.

Satz 3.2.3 Die Verteilung P_X einer Zufallsvariable X wird eindeutig durch die Verteilungsfunktion F_X von X bestimmt.

3.3 Grundlegende Klassen von Verteilungen

In diesem Abschnitt werden wir Grundtypen von Verteilungen betrachten, die dem Schema aus Abbildung 3.4 zu entnehmen sind.

3.3.1 Diskrete Verteilungen

Definition 3.3.1 1. Die Verteilung einer Zufallsvariablen X heißt *diskret*, falls eine höchstens abzählbare Teilmenge $C \subset \mathbb{R}$ (Wertebereich von X) mit $P(X \in C) = 1$ existiert. Manchmal wird auch die Zufallsvariable X selbst als diskret bezeichnet.

2. Falls X eine diskrete Zufallsvariable mit Wertebereich $C = \{x_1, x_2, x_3, \dots\}$ ist, dann heißt $\{p_k\}$ mit $p_k = P(X = x_k)$, $k = 1, 2, \dots$ *Wahrscheinlichkeitsfunktion* oder *Zähldichte* von X .

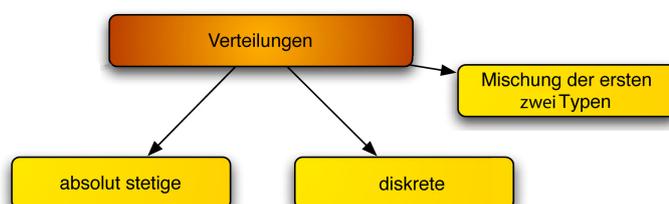


Abbildung 3.4: Verteilungstypen

Bemerkung 3.3.1

1. Beispiele für diskrete Wertebereiche C sind $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \{0, 1, \dots, n\}$, $n \in \mathbb{N}$.
2. Für die Zähldichte $\{p_k\}$ einer diskreten Zufallsvariable X gilt offenbar $0 \leq p_k \leq 1 \quad \forall n$ und $\sum_k p_k = 1$. Diese Eigenschaften sind für eine Zähldichte charakteristisch.
3. Die Verteilung P_X einer diskreten Zufallsvariable X wird eindeutig durch ihre Zähldichte $\{p_k\}$ festgelegt:

$$\begin{aligned}
 P_X(B) &= P_X(B \cap C) = P_X\left(\bigcup_{x_i \in B} \{x_i\}\right) = \sum_{x_i \in B} P(X = x_i) \\
 &= \sum_{x_i \in B} p_i, \quad B \in \mathcal{B}_{\mathbb{R}}.
 \end{aligned}$$

Insbesondere gilt $F_X(x) = \sum_{x_k \leq x} p_k \implies P_X$ festgelegt nach Satz 3.2.3.

Wichtige diskrete Verteilungen:

Die Beispiele 3.1.1 und 3.2.1 liefern uns zwei wichtige diskrete Verteilungen mit Wertebereichen $\{0, 1\}$ und $\{0, 1, \dots, n\}$. Das sind

1. *Bernoulli-Verteilung:*
 $X \sim \text{Ber}(p)$, $p \in [0, 1]$ (abkürzende Schreibweise für “Zufallsvariable X ist Bernoulli-verteilt mit Parameter p ”), falls

$$X = \begin{cases} 1, & \text{mit Wahrscheinlichkeit } p, \\ 0, & \text{mit Wahrscheinlichkeit } 1-p. \end{cases}$$

Dann gilt $C = \{0, 1\}$ und $p_0 = 1 - p$, $p_1 = p$ (vgl. Beispiel 3.1.1, 1) mit $X = I_A$).

2. *Binomialverteilung:*

$X \sim \text{Bin}(n, p)$, $p \in [0, 1]$, $n \in \mathbb{N}$, falls $C = \{0, \dots, n\}$ und

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

Interpretation:

$X = \#\{\text{Erfolge in einem } n \text{ mal unabhängig wiederholten Versuch}\}$, wobei $p = \text{Erfolgswahrscheinlichkeit in einem Versuch}$ (vgl. Beispiel 3.2.1, 2) mit $X = \#\{\text{Kopf}\}$).

3. *Geometrische Verteilung:*

$X \sim \text{Geo}(p)$, $p \in [0, 1]$, falls $C = \mathbb{N}$, und

$$P(X = k) = p_k = (1-p)p^{k-1}, \quad k \in \mathbb{N}.$$

Interpretation: $X = \#\{\text{unabhängige Versuche bis zum ersten Erfolg}\}$, wobei $1-p = \text{Erfolgswahrscheinlichkeit in einem Versuch}$.

4. *Hypergeometrische Verteilung:*

$X \sim \text{HG}(M, S, n)$, $M, S, n \in \mathbb{N}$, $S, n \leq M$, falls

$$X : \Omega \rightarrow \{0, 1, 2, \dots, \min\{n, S\}\}$$

und

$$P(X = k) = p_k = \frac{\binom{S}{k} \binom{M-S}{n-k}}{\binom{M}{n}}, \quad k = 0, 1, \dots, \min\{n, S\}.$$

Interpretation: Urnenmodell aus Beispiel 2.3.1, 5) mit

$$X = \#\{\text{schwarze Kugeln bei } n \text{ Entnahmen aus einer Urne}\}$$

mit insgesamt S schwarzen und $M - S$ weißen Kugeln.

5. *Gleichverteilung:*

$X \sim U\{x_1, \dots, x_n\}$, $n \in \mathbb{N}$, falls $X : \Omega \rightarrow \{x_1, \dots, x_n\}$ mit

$$p_k = P(X = x_k) = \frac{1}{n}, \quad k = 1, \dots, n$$

(wobei U in der Bezeichnung von Englischen “uniform” kommt).

Interpretation: $\{p_k\}$ ist eine Laplacesche Verteilung (klassische Definition von Wahrscheinlichkeiten, vgl. Abschnitt 2.3.1).

6. *Poisson-Verteilung:*

$X \sim \text{Poisson}(\lambda)$, $\lambda > 0$, falls $X : \Omega \rightarrow \{0, 1, 2, \dots\} = \mathbb{N} \cup \{0\}$ mit

$$p_k = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N} \cup \{0\}.$$

Interpretation: $X = \#\{\text{Ereignisse im Zeitraum } [0, 1]\}$, λ ist die Rate (Häufigkeit), mit der Ereignisse passieren können, wobei

$$P(1 \text{ Ereignis tritt während } \Delta t \text{ ein}) = \lambda|\Delta t| + o(|\Delta t|),$$

$$P(> 1 \text{ Ereignis tritt während } \Delta t \text{ ein}) = o(|\Delta t|), \quad |\Delta t| \rightarrow 0$$

und $\#\{\text{Ereignisse in Zeitintervall } \Delta t_i\}$, $i = 1, \dots, n$ sind unabhängig, falls Δt_i , $i = 1, \dots, n$ disjunkte Intervalle aus \mathbb{R} sind. Hier $|\Delta t|$ ist die Länge des Intervalls Δt .

z. B.

$$X = \#\{\text{Schäden eines Versicherers in einem Geschäftsjahr}\}$$

$$X = \#\{\text{Kundenanrufe eines Festnetzanbieters an einem Tag}\}$$

$$X = \#\{\text{Elementarteilchen in einem Geiger-Zähler in einer Sekunde}\}.$$

Satz 3.3.1 (Approximationssatz)

1. *Binomiale Approximation:* Die hypergeometrische Verteilung $HG(M, S, n)$ kann für $M, S \rightarrow \infty$, $\frac{S}{M} \rightarrow p$ durch eine $\text{Bin}(n, p)$ -Verteilung approximiert werden: Für $X \sim HG(M, S, n)$ gilt

$$p_k = P(X = k) = \frac{\binom{S}{k} \binom{M-S}{n-k}}{\binom{M}{n}} \underset{M, S \rightarrow \infty, \frac{S}{M} \rightarrow p}{\rightsquigarrow} \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n$$

2. *Poissonsche Approximation oder Gesetz der seltenen Ereignisse:* Die Binomialverteilung $\text{Bin}(n, p)$ kann für $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \lambda$ durch eine Poisson-Verteilung $\text{Poisson}(\lambda)$ approximiert werden:

$$X \sim \text{Bin}(n, p), \quad p_k = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \underset{n \rightarrow \infty, p \rightarrow 0, np \rightarrow \lambda}{\rightsquigarrow} e^{-\lambda} \frac{\lambda^k}{k!}$$

mit $k = 0, 1, 2, \dots$

Beweis 1. Falls $M, S \rightarrow \infty, \frac{S}{M} \rightarrow p \in (0, 1)$, dann gilt

$$\begin{aligned} \frac{\binom{S}{k} \binom{M-S}{n-k}}{\binom{M}{n}} &= \frac{S!}{k!(S-k)!} \cdot \frac{(M-S)!}{(M-S-n+k)!(n-k)!} \\ &= \frac{n!}{(n-k)!k!} \cdot \underbrace{\frac{S}{M}}_{\rightarrow p} \underbrace{\frac{(S-1)}{(M-1)}}_{\rightarrow p} \dots \underbrace{\frac{(S-k+1)}{(M-k+1)}}_{\rightarrow p} \\ &\quad \cdot \underbrace{\frac{(M-S)}{(M-k)}}_{\rightarrow 1-p} \underbrace{\frac{(M-S-1)}{\dots}}_{\rightarrow 1-p} \dots \underbrace{\frac{(M-S-n+k+1)}{(M-n+1)}}_{\rightarrow 1-p} \\ &\xrightarrow[M, S \rightarrow \infty]{\frac{S}{M} \rightarrow p} \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

2. Falls $n \rightarrow \infty, p \rightarrow 0, np \rightarrow \lambda > 0$, dann gilt

$$\begin{aligned} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{1}{k!} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\rightarrow 1} \cdot \underbrace{(np)^k}_{\rightarrow \lambda^k} \underbrace{\frac{(1-p)^n}{(1-p)^k}}_{\rightarrow e^{-\lambda}} \\ &\rightarrow e^{-\lambda} \frac{\lambda^k}{k!} \text{ f\"ur } n \rightarrow \infty, p \rightarrow 0, np \rightarrow \lambda, \end{aligned}$$

weil

$$\frac{(1-p)^n}{(1-p)^k} \underset{n \rightarrow \infty}{\sim} \frac{(1-\frac{\lambda}{n})^n}{1} \xrightarrow[n \rightarrow \infty]{} e^{-\lambda}, \text{ da } p \sim \frac{\lambda}{n} (n \rightarrow \infty).$$

□

Bemerkung 3.3.2 1. Die Aussage 1) aus Satz 3.3.1 wird dann verwendet, wenn M und S in $HG(M, S, n)$ -Verteilung groß werden ($n < 0, 1 \cdot M$). Dabei wird die direkte Berechnung von hypergeometrischen Wahrscheinlichkeiten umständlich.

2. Genauso wird die Poisson-Approximation verwendet, falls n groß und p entweder bei 0 oder bei 1 liegt. Dann können binomiale Wahrscheinlichkeiten nur schwer berechnet werden.

3. Bei allen diskreten Verteilungen ist die zugehörige Verteilungsfunktion eine stückweise konstante Treppenfunktion (vgl. Bsp. 1, 2 im Abschnitt 3.2.1).

3.3.2 Absolut stetige Verteilungen

Im Gegensatz zu diskreten Zufallsvariablen ist der Wertebereich einer absolut stetigen Zufallsvariablen überabzählbar.

Definition 3.3.2 Die Verteilung einer Zufallsvariablen X heißt *absolut stetig*, falls die Verteilungsfunktion von F_X folgende Darstellung besitzt:

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad x \in \mathbb{R}, \quad (3.1)$$

wobei $f_X : \mathbb{R} \rightarrow \mathbb{R}_+ = [0, \infty)$ eine Lebesgue-integrierbare Funktion auf \mathbb{R} ist, die *Dichte* der Verteilung von X heißt und das Integral in (3.1) als Lebesgue-Integral zu verstehen ist.

Daher wird oft abkürzend gesagt, dass die Zufallsvariable X absolut stetig (verteilt) mit Dichte f_X ist.

Im folgenden Satz zeigen wir, dass die Verteilung P_X einer absolut stetigen Zufallsvariablen eindeutig durch ihre Dichte f_X bestimmt wird:

Satz 3.3.2 Sei X eine Zufallsvariable mit Verteilung P_X .

1. X ist absolut stetig verteilt genau dann, wenn

$$P_X(B) = \int_B f_X(y) dy, \quad B \in \mathcal{B}_{\mathbb{R}}. \quad (3.2)$$

2. Seien X und Y absolut stetige Zufallsvariablen mit Dichten f_X, f_Y und Verteilungen P_X und P_Y . Es gilt $P_X = P_Y$ genau dann, wenn $f_X(x) = f_Y(x)$ für fast alle $x \in \mathbb{R}$, d.h. für alle $x \in \mathbb{R} \setminus A$, wobei $A \in \mathcal{B}_{\mathbb{R}}$ und $\int_A dy = 0$ (das Lebesgue-Maß von A ist Null).

Bemerkung 3.3.3 (*Eigenschaften der absolut stetigen Verteilungen*): Sei X absolut stetig verteilt mit Verteilungsfunktion F_X und Dichte f_X .

1. Für die Dichte f_X gilt: $f_X(x) \geq 0 \quad \forall x$ und $\int_{-\infty}^{\infty} f_X(x) dx = 1$ (vgl. Abb. 3.5).

Diese Eigenschaften sind charakteristisch für eine Dichte, d.h. eine beliebige Funktion f , die diese Eigenschaften erfüllt, ist die Dichte einer absolut stetigen Verteilung.

2. Es folgt aus (3.2), dass

- (a) $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(y) dy, \quad \forall a < b, a, b \in \mathbb{R}$

- (b) $P(X = x) = \int_{\{x\}} f_X(y) dy = 0, \quad \forall x \in \mathbb{R},$

- (c) $f_X(x)\Delta x$ als Wahrscheinlichkeit $P(X \in [x, x + \Delta x])$ interpretiert werden kann, falls f_X stetig in der Umgebung von x und Δx klein

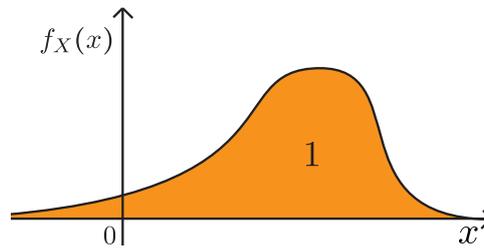


Abbildung 3.5: Die Fläche unter dem Graphen einer Dichtefunktion ist gleich eins.

ist.

In der Tat, mit Hilfe des Mittelwertsatzes bekommt man

$$\begin{aligned}
 P(X \in [x, x + \Delta x]) &= \int_x^{x+\Delta x} f_X(y) dy \\
 &= f_X(\xi) \cdot \Delta x, \quad \xi \in (x, x + \Delta x) \\
 &\stackrel{\Delta x \rightarrow 0}{\approx} (f_X(x) + o(1))\Delta x \\
 &= f_X(x) \cdot \Delta x + o(\Delta x),
 \end{aligned}$$

weil $\xi \rightarrow x$ für $\Delta x \rightarrow 0$ und f_X stetig in der Umgebung von x ist.

3. Es folgt aus [2b](#), dass die Verteilungsfunktion F_X von X eine stetige Funktion ist. F_X kann keine Sprünge haben, weil die Höhe eines Sprunges von F_X in x genau $P(X = x) = 0$ darstellt (vgl. [Abb. 3.6](#)).

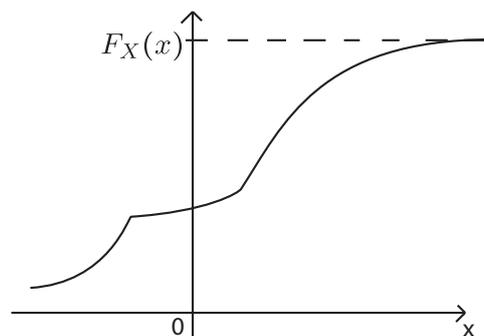


Abbildung 3.6: Eine absolut stetige Verteilungsfunktion

4. Sehr oft wird f_X als (stückweise) stetig angenommen. Dann ist das Integral in [Definition 3.3.2](#) das (uneigentliche) Riemann-Integral. F_X ist im Allgemeinen nur an jeder Stetigkeitsstelle von ihrer Dichte f_X differenzierbar.

5. In den Anwendungen sind Wertebereiche aller Zufallsvariablen endlich. Somit könnte man meinen, dass für Modellierungszwecke nur diskrete Zufallsvariablen genügen. Falls der Wertebereich einer Zufallsvariable X jedoch sehr viele Elemente x enthält, ist die Beschreibung dieser Zufallsvariable mit einer absolut stetigen Verteilung günstiger, denn man braucht nur eine Funktion f_X (Dichte) anzugeben, statt sehr viele Einzelwahrscheinlichkeiten $p_k = P(X = x_k)$ aus den Daten zu schätzen.

Wichtige absolut stetige Verteilungen

1. *Normalverteilung (Gauß-Verteilung):*
 $X \sim N(\mu, \sigma^2)$ für $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, falls

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

(vgl. Abb. 3.7).

μ heißt der *Mittelwert* von X und σ die *Standardabweichung* bzw.

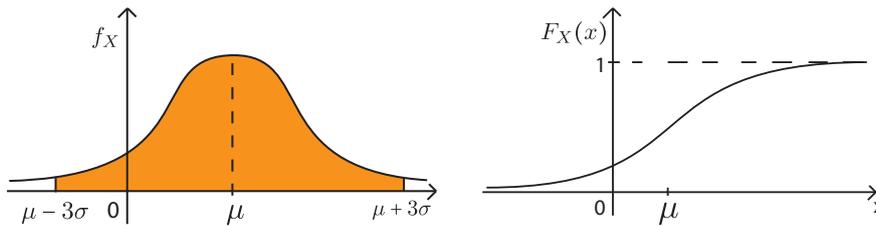


Abbildung 3.7: Dichte und Verteilungsfunktion der $N(\mu, \sigma^2)$ -Verteilung

Streuung, denn es gilt die sogenannte “ 3σ -Regel” (Gauß, 1821):

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \geq 0,9973$$

Spezialfall $N(0, 1)$: In diesem Fall sieht die Dichte f_X folgendermaßen aus:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Interpretation:

X = Messfehler einer physikalischen Größe μ , σ = Streuung des Messfehlers. Die Verteilungsfunktion $F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$ kann nicht analytisch berechnet werden (vgl. Abb. 3.7).

2. Gleichverteilung auf $[a, b]$:
 $X \sim U[a, b]$, $a < b$, $a, b \in \mathbb{R}$, falls

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{sonst} \end{cases} \quad (\text{vgl. Abb. 3.8}).$$

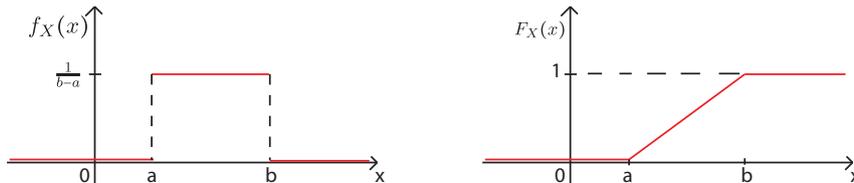


Abbildung 3.8: Dichte und Verteilungsfunktion der Gleichverteilung $U[a, b]$.

Interpretation:

X = Koordinate eines zufällig auf $[a, b]$ geworfenen Punktes (geometrische Wahrscheinlichkeit). Für $F_X(x)$ gilt:

$$F_X(x) = \begin{cases} 1, & x \geq b, \\ \frac{x-a}{b-a}, & x \in [a, b], \\ 0, & x < a \end{cases} \quad (\text{vgl. Abb. 3.8}).$$

3. Exponentialverteilung:
 $X \sim \text{Exp}(\lambda)$ für $\lambda > 0$, falls

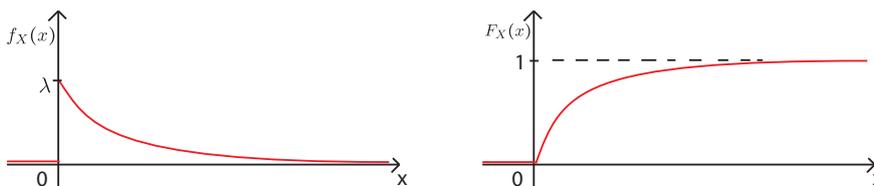


Abbildung 3.9: Dichte und Verteilungsfunktion der Exponentialverteilung $\text{Exp}(\lambda)$.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{sonst} \end{cases} \quad (\text{vgl. Abb. 3.9}).$$

Interpretation:

X = Zeitspanne der fehlerfreien Arbeit eines Geräts, z.B. eines Netzservers oder einer Glühbirne, λ = Geschwindigkeit, mit der das Gerät

kaputt geht. $F_X(x)$ hat folgende Gestalt:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases} \quad (\text{vgl. Abb. 3.9}).$$

4. Cauchy-Verteilung:

$X \sim \text{Cauchy}(\alpha, \lambda)$, falls für $\lambda > 0$, $\alpha \in \mathbb{R}$

$$f_X(x) = \frac{\lambda}{\pi(\lambda^2 + (x - \alpha)^2)}, \quad x \in \mathbb{R}, \quad \text{vgl. Abb. 3.10}$$

Die Verteilungsfunktion

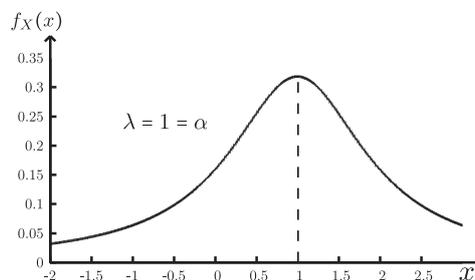


Abbildung 3.10: Dichte der $\text{Cauchy}(\alpha, \lambda)$ -Verteilung

$$F_X(x) = \frac{\lambda}{\pi} \int_{-\infty}^x \frac{dy}{\lambda^2 + (y - \alpha)^2} \quad x \in \mathbb{R}$$

kann nicht analytisch berechnet werden.

Die praktische Interpretation dieser Verteilung ist schwierig, weil sie keinen Mittelwert besitzt. Sie erscheint jedoch als Grenzwert anderer Verteilungen.

3.3.3 Mischungen von Verteilungen

Durch eine lineare Kombination der Verteilungen aus dem Abschnitt 3.3.1-3.3.2 kann eine neue Verteilung konstruiert werden, die Mischung genannt wird. Ihre Verteilungsfunktion ist

$$F(x) = p_1 F_1(x) + p_2 F_2(x), \quad x \in \mathbb{R},$$

wobei $0 \leq p_i \leq 1$, $i = 1, 2$, $p_1 + p_2 = 1$,

F_1 ist die Verteilungsfunktion einer diskreten Zufallsvariable X_1 ,

F_2 ist die Verteilungsfunktion einer absolut stetigen Zufallsvariable X_2 .

Bemerkung 3.3.4 Diese Mischung von 2 Verteilungen P_{X_i} kann folgendermaßen realisiert werden. Definieren wir eine diskrete Zufallsvariable N , die unabhängig von X_i ist, durch

$$N = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p_1, \\ 2 & \text{mit Wahrscheinlichkeit } p_2. \end{cases}$$

Dann gilt $X \stackrel{d}{=} X_N$ für $X \sim F_X$, wobei “ $\stackrel{d}{=}$ ” die *Gleichheit in Verteilung* ist (aus dem Englischen “d” für “distribution”): $X \stackrel{d}{=} Y$, falls $P_X = P_Y$.

3.4 Verteilungen von Zufallsvektoren

In der Definition 3.1.1, 2) wurden Zufallsvektoren bereits eingeführt. Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}^n$ ein n -dimensionaler Zufallsvektor. Bezeichnen wir seine Koordinaten als (X_1, \dots, X_n) . Dann folgt aus Definition 3.1.1, 2), dass X_i , $i = 1, \dots, n$ Zufallsvariablen sind. Umgekehrt kann man einen beliebigen Zufallsvektor X definieren, indem man seine Koordinaten $X_1 \dots X_n$ als Zufallsvariablen auf demselben Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) einführt (Übungsaufgabe).

Definition 3.4.1 Sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor auf (Ω, \mathcal{F}, P) .

1. Die *Verteilung* von X ist die Mengenfunktion $P_X : \mathcal{B}_{\mathbb{R}^n} \rightarrow [0, 1]$ mit $P_X(B) = P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})$, $B \in \mathcal{B}_{\mathbb{R}^n}$.
2. Die *Verteilungsfunktion* $F_X : \mathbb{R}^n \rightarrow [0, 1]$ von X ist gegeben durch $F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$ $x_1, \dots, x_n \in \mathbb{R}$. Sie heißt manchmal auch die *gemeinsame* oder die *multivariate Verteilungsfunktion* von X , um sie von folgenden *marginalen Verteilungsfunktionen* zu unterscheiden.
3. Sei $\{i_1, \dots, i_k\}$ ein Teilvektor von $\{1, \dots, n\}$. Die multivariate Verteilungsfunktion F_{i_1, \dots, i_k} des Zufallsvektors $(X_{i_1}, \dots, X_{i_k})$ heißt *marginale Verteilungsfunktion* von X . Insbesondere für $k = 1$ und $i_1 = i$ spricht man von den so genannten *Randverteilungen*:

$$F_{X_i}(x) = P(X_i \leq x), \quad i = 1, \dots, n.$$

Satz 3.4.1 (*Eigenschaften multivariater Verteilungsfunktionen*):

Sei $F_X : \mathbb{R}^n \rightarrow [0, 1]$ die Verteilungsfunktion eines Zufallsvektors $X = (X_1, \dots, X_n)$. Dann gelten folgende Eigenschaften:

1. *Asymptotik*:

$$\lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_n) = 0, \quad \forall i = 1, \dots, n \quad \forall x_1, \dots, x_n \in \mathbb{R},$$

$$\lim_{x_1, \dots, x_n \rightarrow +\infty} F_X(x_1, \dots, x_n) = 1,$$

$$\lim_{x_j \rightarrow +\infty \forall j \notin \{i_1, \dots, i_k\}} F_X(x_1, \dots, x_n) = F_{(X_{i_1}, \dots, X_{i_k})}(x_{i_1}, \dots, x_{i_k}),$$

wobei $F_{(X_{i_1}, \dots, X_{i_k})}(x_{i_1}, \dots, x_{i_k})$ die Verteilungsfunktion der marginalen Verteilung von

$$(X_{i_1} \dots X_{i_k}) \text{ ist, } \{i_1, \dots, i_k\} \subset \{1, \dots, n\}.$$

Insbesondere gilt

$$\lim_{x_j \rightarrow +\infty, j \neq i} F_X(x_1, \dots, x_n) = F_{X_i}(x_i), \quad \forall i = 1, \dots, n$$

(Randverteilungsfunktion).

2. *Monotonie:* $\forall (x_1, \dots, x_n) \in \mathbb{R}^n \quad \forall h_1, \dots, h_n \geq 0$

$$F_X(x_1 + h_1, \dots, x_n + h_n) \geq F_X(x_1, \dots, x_n)$$

3. *Rechtsseitige Stetigkeit:*

$$F_X(x_1, \dots, x_n) = \lim_{y_i \rightarrow x_i + 0, i=1, \dots, n} F_X(y_1, \dots, y_n) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

Beweis Analog zum Satz 3.2.1. □

Definition 3.4.2 Die Verteilung eines Zufallsvektors $X = (X_1, \dots, X_n)$ heißt

1. *diskret*, falls eine höchstens abzählbare Menge $C \subseteq \mathbb{R}^n$ existiert, für die $P(X \in C) = 1$ gilt. Die Familie von Wahrscheinlichkeiten

$$\{P(X = x), x \in C\}$$

heißt dann *Wahrscheinlichkeitsfunktion* oder *Zähldichte* von X .

2. *absolut stetig*, falls eine Funktion $f_X : \mathbb{R}^n \rightarrow [0, 1]$ existiert, die Lebesgue-integrierbar auf \mathbb{R}^n ist und für die gilt

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(y_1, \dots, y_n) dy_n \dots dy_1,$$

$\forall (x_1, \dots, x_n) \in \mathbb{R}^n$. f_X heißt *Dichte* der gemeinsamen Verteilung von X .

Lemma 3.4.1 Sei $X = (X_1, \dots, X_n)$ ein diskreter (bzw. absolut stetiger) Zufallsvektor mit Zähldichte $P(X = x)$ (bzw. Dichte $f_X(x)$). Dann gilt:

1. Die Verteilung P_X von X ist gegeben durch

$$P_X(B) = \sum_{x \in B} P(X = x) \text{ bzw. } P_X(B) = \int_B f_X(x) dx, \quad B \in \mathcal{B}_{\mathbb{R}^n}.$$

2. Die Koordinaten $X_i, i = 1, \dots, n$ sind ebenfalls diskrete bzw. absolut stetige Zufallsvariablen mit der Randzähldichte

$$\begin{aligned} P(X_i = x) &= \\ &= \sum_{(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n) \in C} P(X_1 = y_1, \dots, X_{i-1} = y_{i-1}, X_i = x, X_{i+1} = y_{i+1}, \dots, X_n = y_n) \end{aligned}$$

bzw. Randdichte

$$f_{X_i}(x) = \int_{\mathbb{R}^{n-1}} f_X(y_1, \dots, y_{i-1}, x, y_{i+1}, \dots, y_n) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n$$

$\forall x \in \mathbb{R}$.

Beweis

1. Folgt aus dem eindeutigen Zusammenhang zwischen einer Verteilung und ihrer Verteilungsfunktion.
2. Die Aussage für diskrete Zufallsvektoren ist trivial. Sei nun $X = (X_1, \dots, X_n)$ absolut stetig. Dann folgt aus Satz 3.4.1

$$\begin{aligned} F_{X_i}(x) &= \lim_{y_j \rightarrow +\infty, j \neq i} F_X(x_1 \dots x_{i-1}, x, x_{i+1}, \dots, x_n) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^x \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(y_1, \dots, y_n) dy_n \dots dy_1 \\ &\stackrel{\text{S. v. Fubini}}{=} \int_{-\infty}^x \underbrace{\left(\int_{\mathbb{R}^{n-1}} f_X(y_1, \dots, y_n) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n \right)}_{f_{X_i}(y_i)} dy_i \end{aligned}$$

Somit ist X_i absolut stetig verteilt mit Dichte f_{X_i} .

□

Beispiel 3.4.1 *Verschiedene Zufallsvektoren:*

1. *Polynomiale Verteilung:*

$X = (X_1, \dots, X_k) \sim \text{Polynom}(n, p_1, \dots, p_k), \quad n \in \mathbb{N}, p_i \in [0, 1],$
 $i = 1, \dots, k, \quad \sum_{i=1}^n p_i = 1,$ falls X diskret verteilt ist mit Zähldichte

$$P(X = x) = P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

$\forall x = (x_1, \dots, x_k)$ mit $x_i \in \mathbb{N} \cup \{0\}$ und $\sum_{i=1}^k x_i = n$. Die polynomiale Verteilung ist das k -dimensionale Analogon der Binomialverteilung. So

sind die Randverteilungen von $X_i \sim \text{Bin}(n, p_i)$, $i = 1, \dots, k$. (Bitte prüfen Sie dies als Übungsaufgabe!). Es gilt $P(\sum_{i=1}^k X_i = n) = 1$.

Interpretation:

Es werden n Versuche durchgeführt. In jedem Versuch kann eines aus insgesamt k Merkmalen auftreten. Sei p_i die Wahrscheinlichkeit des Auftretens von Merkmal i in einem Versuch. Sei

$$X_i = \#\{\text{Auftretens von Merkmal } i \text{ in } n \text{ Versuchen}\}, \quad i = 1, \dots, k.$$

Dann ist $X = (X_1, \dots, X_k) \sim \text{Polynom}(n, p_1, \dots, p_k)$.

2. *Gleichverteilung:*

$X \sim \mathcal{U}(A)$, wobei $A \subset \mathbb{R}^n$ eine beschränkte Borel-Teilmenge von \mathbb{R}^n ist, falls $X = (X_1, \dots, X_n)$ absolut stetig verteilt mit der Dichte

$$f_X(x_1, \dots, x_n) = \begin{cases} \frac{1}{|A|}, & (x_1, \dots, x_n) \in A, \\ 0, & \text{sonst,} \end{cases}$$

ist (vgl. Abb. 3.11). Im Spezialfall $A = \prod_{i=1}^n [a_i, b_i]$ (Parallelepiped)

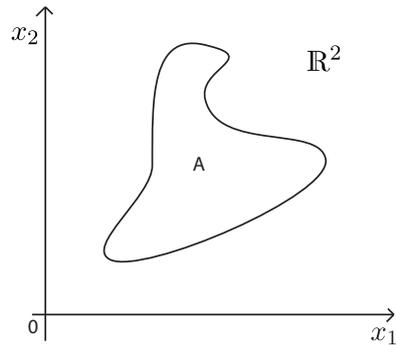


Abbildung 3.11: Wertebereich A einer zweidimensionalen Gleichverteilung

sind alle Randverteilungen von X_i ebenso Gleichverteilungen:

$$X_i \sim U[a_i, b_i], \quad i = 1, \dots, n.$$

Interpretation:

$X = (X_1, \dots, X_n)$ sind Koordinaten eines zufälligen Punktes, der gleichwahrscheinlich auf A geworfen wird. Dies ist die geometrische Wahrscheinlichkeit, denn $P(X \in B) = \int_B f_X(y) dy = \frac{|B|}{|A|}$ für $B \in \mathcal{B}_{\mathbb{R}^n} \cap A$.

3. *Multivariate Normalverteilung:*

$X = (X_1, \dots, X_n) \sim N(\mu, K)$, $\mu \in \mathbb{R}^n$, K eine positiv definite $(n \times n)$ -Matrix, falls X absolut stetig verteilt mit Dichte

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \cdot \det K}} \exp\left\{-\frac{1}{2}(X - \mu)^T K^{-1}(X - \mu)\right\}, \quad x \in \mathbb{R}^n$$

ist.

Spezialfall zweidimensionale Normalverteilung:

Falls $n = 2$ und

$$\mu = (\mu_1, \mu_2), \quad K = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

dann gilt $\det K = \sigma_1^2\sigma_2^2(1 - \rho^2)$ und

$$f_X(x_1, x_2) = \frac{1}{\sqrt{1 - \rho^2} \cdot 2\pi\sigma_1\sigma_2} \times \\ \times \exp\left\{-\frac{1}{2(1 - \rho^2)} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right)\right\},$$

$(x_1, x_2) \in \mathbb{R}^2$, weil

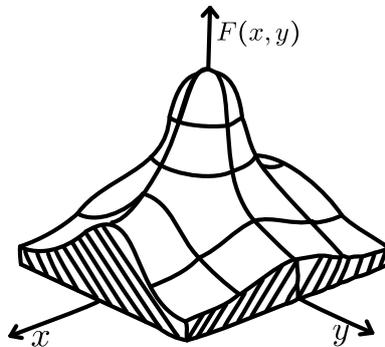


Abbildung 3.12: Grafik der Dichte einer zweidimensionalen Normalverteilung

$$K^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}, \quad (\text{vgl. Abb. 3.12}).$$

Übungsaufgabe 3.4.1 Zeigen Sie, dass $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$. Diese Eigenschaft der Randverteilungen gilt für alle $n \geq 2$. Somit ist die multivariate Normalverteilung ein mehrdimensionales Analogon der eindimensionalen $N(\mu, \sigma^2)$ -Verteilung.

Interpretation:

Man feuert eine Kanone auf das Ziel mit Koordinaten (μ_1, μ_2) . Dann sind $X = (X_1, X_2)$ die Koordinaten des Treffers. Durch die Streuung wird $(X_1, X_2) = (\mu_1, \mu_2)$ nur im Mittel. σ_1^2 und σ_2^2 sind Maße für die Genauigkeit des Feuers.

3.5 Stochastische Unabhängigkeit

3.5.1 Unabhängige Zufallsvariablen

Definition 3.5.1

1. Seien X_1, \dots, X_n Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Sie heißen *unabhängig*, falls

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n), \quad x_1, \dots, x_n \in \mathbb{R}$$

oder äquivalent dazu,

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n).$$

2. Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Diese Folge besteht aus *unabhängigen Zufallsvariablen*, falls $\forall k \in \mathbb{N} \forall i_1 < i_2 < \dots < i_k X_{i_1}, X_{i_2}, \dots, X_{i_k}$ unabhängige Zufallsvariablen (im Sinne der Definition 1) sind.

Lemma 3.5.1 Die Zufallsvariablen X_1, \dots, X_n sind genau dann unabhängig, wenn für alle $B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$ gilt

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n) \quad (3.3)$$

Satz 3.5.1 (*Charakterisierung der Unabhängigkeit von Zufallsvariablen*)

1. Sei (X_1, \dots, X_n) ein diskret verteilter Zufallsvektor mit dem Wertebereich C . Seine Koordinaten X_1, \dots, X_n sind genau dann unabhängig, wenn

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$$

$\forall (x_1, \dots, x_n) \in C$.

2. Sei (X_1, \dots, X_n) ein absolut stetiger Zufallsvektor mit der Dichte $f_{(X_1, \dots, X_n)}$ und Randdichten f_{X_i} . Seine Koordinaten X_1, \dots, X_n sind genau dann unabhängig, wenn

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

für fast alle $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Beispiel 3.5.1

1. *Multivariate Normalverteilung:*

Mit Hilfe des Satzes 3.5.1 kann gezeigt werden, dass die Komponenten X_1, \dots, X_n von

$$X = (X_1, \dots, X_n) \sim N(\mu, K)$$

genau dann unabhängig sind, wenn $k_{ij} = 0$, $i \neq j$, wobei $K = (k_{ij})_{i,j=1}^n$. Insbesondere gilt im zweidimensionalen Fall (vgl. Bsp. 3 Seite 44), dass X_1 und X_2 unabhängig sind, falls $\rho = 0$.

Übungsaufgabe 3.5.1 Zeigen Sie es!

2. *Multivariate Gleichverteilung:*

Die Komponenten des Vektors $X = (X_1, \dots, X_n) \sim \mathcal{U}(A)$ sind genau dann unabhängig, falls $A = \prod_{i=1}^n [a_i, b_i]$ ist. In der Tat gilt dann

$$f_X(x_1, \dots, x_n) = \begin{cases} \frac{1}{|A|} = \frac{1}{\prod_{i=1}^n (b_i - a_i)} = \prod_{i=1}^n \frac{1}{b_i - a_i} = \prod_{i=1}^n f_{X_i}(x_i), & x \in A \\ 0, & \text{sonst} \end{cases}$$

mit $x = (x_1, \dots, x_n)$, wobei

$$f_X(x_i) = \begin{cases} 0, & \text{falls } x_i \notin [a_i, b_i] \\ \frac{1}{b_i - a_i}, & \text{sonst.} \end{cases}$$

Implizit haben wir an dieser Stelle benutzt, dass

$$X_i \sim \mathcal{U}[a_i, b_i], \quad i = 1, \dots, n$$

(Herleitung: $\int_{\mathbb{R}^{n-1}} f_X(x) dx_n \dots dx_{i+1} dx_{i-1} \dots dx_1 = \frac{1}{b_i - a_i}$, $x_i \in [a_i, b_i]$).

Übungsaufgabe 3.5.2

Zeigen Sie die Notwendigkeit der Bedingung $A = \prod_{i=1}^n [a_i, b_i]$!

3. Es gibt Beispiele von Zufallsvariablen X_1 und X_2 , die stochastisch abhängig von einander sind, so dass X_1^2 stochastisch unabhängig von X_2^2 ist.

Unterschied: Kausale bzw. stochastische Abhängigkeit!

Definition 3.5.2 Eine Funktion $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $n, m \in \mathbb{N}$ heißt *Borelsche Funktion*, falls sie $\mathcal{B}_{\mathbb{R}^n}$ -messbar ist, d.h. $\forall B \in \mathcal{B}_{\mathbb{R}^m}$ ist $\varphi^{-1}(B) \in \mathcal{B}_{\mathbb{R}^n}$.

Bemerkung 3.5.1

1. Sei $X = (X_1, \dots, X_n)^T$ ein Zufallsvektor, und $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sei Borelsch. Dann $Y = (Y_1, \dots, Y_m)^T = \phi(X)$ ebenfalls ein Zufallsvektor.
2. Seien X_1, X_2 Zufallsvariablen auf (Ω, \mathcal{F}, P) und $\varphi_1, \varphi_2 : \mathbb{R} \rightarrow \mathbb{R}$ Borelsche Funktionen. Falls X_1 und X_2 unabhängig sind, dann sind auch $\varphi_1(X_1)$ und $\varphi_2(X_2)$ unabhängig.

3.6 Funktionen von Zufallsvektoren

Lemma 3.6.1 Jede stetige Funktion $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist Borel-messbar. Insbesondere sind Polynome $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ der Form

$$\varphi(x_1, \dots, x_n) = \sum_{i=0}^k a_i x_1^{m_{1i}} \dots x_n^{m_{ni}},$$

$k \in \mathbb{N}$, $a_0, a_1, \dots, a_n \in \mathbb{R}$, $m_{1i}, \dots, m_{ni} \in \mathbb{N} \cup \{0\}$, $i = 1, \dots, k$ Borel-messbar.

Satz 3.6.1 (*Transformationssatz*)

Sei X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) .

1. Falls die Abbildung $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ stetig und streng monoton wachsend ist, dann gilt $F_{\varphi(X)}(x) = F_X(\varphi^{-1}(x)) \quad \forall x \in \mathbb{R}$, wobei φ^{-1} die Umkehrfunktion von φ ist. Falls $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ stetig und streng monoton fallend ist, dann gilt $F_{\varphi(X)}(x) = 1 - F_X(\varphi^{-1}(x)) + P(X = \varphi^{-1}(x))$, $x \in \mathbb{R}$.
2. Falls X absolut stetig mit Dichte f_X ist und $C \subset \mathbb{R}$ eine offene Menge mit $P(X \in C) = 1$ ist, dann ist $\varphi(X)$ absolut stetig mit Dichte $f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y)) \cdot |(\varphi^{-1})'(y)|$, $y \in \varphi(C)$, falls φ eine auf C stetig differenzierbare Funktion mit $\varphi'(x) \neq 0$, $x \in C$ ist.

Beweis

1. Falls φ monoton wachsend ist, gilt für $x \in \mathbb{R}$, dass $F_{\varphi(X)}(x) = P(\varphi(X) \leq x) = P(X \leq \varphi^{-1}(x)) = F_X(\varphi^{-1}(x))$.

Für φ monoton fallend gilt für $x \in \mathbb{R}$

$$\begin{aligned} F_{\varphi(X)}(x) &= P(\varphi(X) \leq x) = P(X \geq \varphi^{-1}(x)) \\ &= 1 - P(X < \varphi^{-1}(x)) = 1 - F_X(\varphi^{-1}(x)) + P(X = \varphi^{-1}(x)). \end{aligned}$$

2. Nehmen wir o.B.d.A. an, dass $C = \mathbb{R}$ und $\varphi'(x) > 0 \forall x \in \mathbb{R}$.
 Für $\varphi'(x) < 0$ verläuft der Beweis analog. Aus Punkt 1) folgt

$$\begin{aligned} F_{\varphi(X)}(x) &= F_X(\varphi^{-1}(x)) \\ &= \int_{-\infty}^{\varphi^{-1}(x)} f_X(y) dy \\ &\stackrel{\varphi^{-1}(t)=y}{=} \int_{-\infty}^x f_X(\varphi^{-1}(t)) \cdot |(\varphi^{-1})'(t)| dt, \quad x \in \mathbb{R}. \end{aligned}$$

Hieraus folgt $f_{\varphi(X)}(t) = f_X(\varphi^{-1}(t)) \cdot |(\varphi^{-1})'(t)|$, $t \in \mathbb{R}$.

□

Satz 3.6.2 (*lineare Transformation*)

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable und $a, b \in \mathbb{R}$, $a \neq 0$. Dann gilt Folgendes:

1. Die Verteilungsfunktion der Zufallsvariable $aX + b$ ist gegeben durch

$$F_{aX+b}(x) = \begin{cases} F_X\left(\frac{x-b}{a}\right), & a > 0 \\ 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right), & a < 0. \end{cases}$$

2. Falls X absolut stetig mit Dichte f_X ist, dann ist $aX + b$ ebenfalls absolut stetig mit Dichte

$$f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right).$$

Beweis 1. Der Fall $a > 0$ ($a < 0$) folgt aus dem Satz 3.6.1, 1), weil $\varphi(x) = aX + b$ stetig und monoton wachsend bzw. fallend ist.

2. Folgt aus dem Satz 3.6.1, 2), weil $\varphi(x) = aX + b$ stetig differenzierbar auf $C = \mathbb{R}$ (offen) mit $\varphi'(x) = a \neq 0$ ist.

□

Satz 3.6.3 (*Quadrierung*)

Sei X eine Zufallsvariable auf (Ω, \mathcal{F}, P) .

1. Die Verteilungsfunktion von X^2 ist gegeben durch

$$F_{X^2}(x) = \begin{cases} F_X(\sqrt{x}) - F_X(-\sqrt{x}) + P(X = -\sqrt{x}), & \text{falls } x \geq 0 \\ 0, & \text{sonst.} \end{cases}$$

2. Falls X absolut stetig mit Dichte f_X ist, dann ist auch X^2 absolut stetig mit Dichte

$$f_{X^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}} (f_X(\sqrt{x}) + f_X(-\sqrt{x})), & x > 0 \\ 0, & \text{sonst.} \end{cases}$$

Beweis 1. Für $x < 0$ gilt $F_{X^2}(x) = P(X^2 \leq x) = 0$.

Für $x \geq 0$ gilt

$$\begin{aligned} F_{X^2}(x) &= P(X^2 \leq x) = P(|X| \leq \sqrt{x}) \\ &= P(-\sqrt{x} \leq X \leq \sqrt{x}) = P(X \leq \sqrt{x}) - P(X < -\sqrt{x}) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x}) + P(X = -\sqrt{x}). \end{aligned}$$

2. Wegen 1) gilt $F_{X^2}(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x})$, da im absolut stetigen Fall $P(X = -\sqrt{x}) = 0 \quad \forall x \geq 0$. Daher gilt

$$\begin{aligned} F_{X^2}(x) &= \int_{-\sqrt{x}}^{\sqrt{x}} f_X(y) dy = \\ &= \int_0^{\sqrt{x}} f_X(y) dy + \int_{-\sqrt{x}}^0 f_X(y) dy \\ &\stackrel{\substack{=} \\ y=\sqrt{t} \text{ oder } y=-\sqrt{t}}}{=} \int_0^x \frac{1}{2\sqrt{t}} f_X(\sqrt{t}) dt + \int_0^x \frac{1}{2\sqrt{t}} f_X(-\sqrt{t}) dt \\ &= \int_0^x \frac{1}{2\sqrt{t}} (f_X(\sqrt{t}) + f_X(-\sqrt{t})) dt, \quad \forall x \geq 0. \end{aligned}$$

Daher gilt die Aussage 2) des Satzes. □

Beispiel 3.6.1

1. Falls $X \sim N(0, 1)$, dann ist $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$.
2. Falls $X \sim N(\mu, \sigma^2)$, dann heißt die Zufallsvariable $Y = e^X$ *lognormalverteilt mit Parametern μ und σ^2* . Diese Verteilung wird sehr oft in ökonomischen Anwendungen benutzt.

Zeigen Sie, dass die Dichte von Y durch

$$f_Y(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{\log(x-\mu)^2}{2\sigma^2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

gegeben ist.

3. Falls $X \sim N(0, 1)$, dann heißt $Y = X^2$ χ_1^2 -verteilt (*Chi-Quadrat-Verteilung*) mit einem Freiheitsgrad.

Zeigen Sie, dass die Dichte von Y durch

$$f_Y(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

gegeben ist.

Die χ^2 -Verteilung wird in der Statistik sehr oft als die sogenannte *Prüfverteilung* bei der Konstruktion von statistischen Tests und Konfidenzintervallen verwendet.

Satz 3.6.4 (*Summe von unabhängigen Zufallsvariablen*)

Sei $X = (X_1, X_2)$ ein absolut stetiger Zufallsvektor mit Dichte f_X . Dann gilt Folgendes:

1. Die Zufallsvariable $Y = X_1 + X_2$ ist absolut stetig mit Dichte

$$f_Y(x) = \int_{\mathbb{R}} f_X(y, x - y) dy, \quad \forall x \in \mathbb{R}. \quad (3.4)$$

2. Falls X_1 und X_2 unabhängig sind, dann heißt der Spezialfall

$$f_Y(x) = \int_{\mathbb{R}} f_{X_1}(y) f_{X_2}(x - y) dy, \quad x \in \mathbb{R}$$

von (3.4) *Faltungsformel*.

Beweis

2) ergibt sich aus 1) für $f_X(x, y) = f_{X_1}(x) \cdot f_{X_2}(y) \quad \forall x, y \in \mathbb{R}$.

Beweisen wir also 1):

$$\begin{aligned} P(Y \leq t) &= P(X_1 + X_2 \leq t) = \int_{(x,y) \in \mathbb{R}^2: x+y \leq t} f_X(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{t-x} f_X(x, y) dy dx \\ &\stackrel{y \rightarrow z = x+y}{=} \int_{-\infty}^{\infty} \int_{-\infty}^t f_X(x, z - x) dz dx \\ &\stackrel{\text{Fubini}}{=} \int_{-\infty}^t \underbrace{\int_{\mathbb{R}} f_X(x, z - x) dx}_{=f_Y(z)} dz, t \in \mathbb{R}. \end{aligned}$$

Somit ist $f_Y(z) = \int_{\mathbb{R}} f_X(x, z - x) dx$ die Dichte von $Y = X_1 + X_2$. □

Folgerung 3.6.1 (*Faltungsstabilität der Normalverteilung*):

Falls die Zufallsvariablen X_1, \dots, X_n unabhängig mit

$$X_i \sim N(\mu_i, \sigma_i^2) \quad i = 1, \dots, n$$

sind, dann gilt

$$X_1 + \dots + X_n \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Beweis Induktion bzgl. n . Den Fall $n = 2$ sollten Sie als Übungsaufgabe lösen. Der Rest des Beweises ist trivial. \square

Satz 3.6.5 (*Produkt und Quotient von Zufallsvariablen*):

Sei $X = (X_1, X_2)$ ein absolut stetiger Zufallsvektor mit Dichte f_X . Dann gilt Folgendes:

1. Die Zufallsvariable $Y = X_1 \cdot X_2$ und $Z = \frac{X_1}{X_2}$ sind absolut stetig verteilt mit Dichten

$$f_Y(x) = \int_{\mathbb{R}} \frac{1}{|t|} f_X(x/t, t) dt,$$

bzw.

$$f_Z(x) = \int_{\mathbb{R}} |t| f_X(x \cdot t, t) dt, x \in \mathbb{R}.$$

2. Falls X_1 und X_2 unabhängig sind, dann gilt der Spezialfall der obigen Formeln

$$f_Y(x) = \int_{\mathbb{R}} \frac{1}{|t|} f_{X_1}(x/t) f_{X_2}(t) dt,$$

bzw.

$$f_Z(x) = \int_{\mathbb{R}} |t| f_{X_1}(x \cdot t) f_{X_2}(t) dt, x \in \mathbb{R}.$$

Beweis Analog zu dem Beweis des Satzes 3.6.4. \square

Beispiel 3.6.2 Zeigen Sie, dass $X_1/X_2 \sim \text{Cauchy}(0,1)$, falls $X_1, X_2 \sim N(0,1)$ und unabhängig sind:

$$f_{X_1/X_2}(x) = \frac{1}{\pi(x^2 + 1)}, \quad x \in \mathbb{R}.$$

Bemerkung 3.6.1 Da X_1 und X_2 absolut stetig verteilt sind, tritt das Ereignis $\{X_2 = 0\}$ mit Wahrscheinlichkeit Null ein. Daher ist $X_1(\omega)/X_2(\omega)$ wohl definiert für fast alle $\omega \in \Omega$. Für $\omega \in \Omega : X_2(\omega) = 0$ kann $X_1(\omega)/X_2(\omega)$ z.B. als 1 definiert werden. Dies ändert den Ausdruck der Dichte von X_1/X_2 nicht.

Kapitel 4

Momente von Zufallsvariablen

Weitere wichtige Charakteristiken von Zufallsvariablen sind ihre so genannten *Momente*, darunter der Erwartungswert und die Varianz. Zusätzlich wird in diesem Kapitel die Kovarianz von zwei Zufallsvariablen als Maß ihrer Abhängigkeit diskutiert. Um diese Charakteristiken einführen zu können, brauchen wir die Definition des Lebesgue-Integrals auf beliebigen messbaren Räumen.

Beispiel 4.0.3 Sei X eine diskrete Zufallsvariable mit dem endlichen Wertebereich $C = \{x_1, \dots, x_n\}$ und Zähldichte $\{p_i\}_{i=1}^n$, wobei $p_i = P(X = x_i)$, $i = 1, \dots, n$. Wie soll der Mittelwert von X berechnet werden? Aus der Antike sind drei Ansätze zur Berechnung des Mittels von n Zahlen bekannt:

- das arithmetische Mittel: $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$
- das geometrische Mittel: $\bar{g}_n = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$
- das harmonische Mittel: $\bar{h}_n = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$

Um \bar{g}_n und \bar{h}_n berechnen zu können, ist die Bedingung $x_i > 0$ bzw. $x_i \neq 0$ $i = 1, \dots, n$ eine wichtige Voraussetzung. Wir wollen jedoch diese Mittel für beliebige Wertebereiche einführen. Somit fallen diese zwei Möglichkeiten schon weg. Beim arithmetischen Mittel werden beliebige x_i zugelassen, jedoch alle gleich gewichtet, unabhängig davon, ob der Wert x_{i_0} wahrscheinlicher als alle anderen Werte ist und somit häufiger in den Experimenten vorkommt.

Deshalb ist es naheliegend, das gewichtete Mittel $\sum_{i=1}^n x_i \omega_i$ zu betrachten, $\forall i \omega_i \geq 0$, $\sum_{i=1}^n \omega_i = 1$, wobei das Gewicht ω_i die relative Häufigkeit des Vorkommens des Wertes x_i in den Experimenten ausdrückt. Somit ist es natürlich, $\omega_i = p_i$ zu setzen, $i = 1, \dots, n$, und schreiben $EX = \sum_{i=1}^n x_i p_i$.

Dieses Mittel wird “Erwartungswert der Zufallsvariable X ” genannt. Der Buchstabe “ \mathbb{E} ” kommt aus dem Englischen: “Expectation”. Für die Gleichverteilung auf $\{x_1, \dots, x_n\}$, d.h. $p_i = \frac{1}{n}$, stimmt $\mathbb{E}X$ mit dem arithmetischen Mittel \bar{x}_n überein. Wie wir bald sehen werden, kann

$$\mathbb{E}X = \sum_{i=1}^n x_i P(X = x_i)$$

geschrieben werden.

4.1 Erwartungswert

Somit können wir folgende Definition angeben:

Definition 4.1.1

1. Sei X eine diskret verteilte Zufallsvariable mit Wertebereich C und Zähldichte $P_X(x)$. Der Erwartungswert von X ist definiert als

$$\mathbb{E}X = \sum_{x \in C} x P_X(x),$$

falls $\sum_{x \in C} |x| P_X(x) < \infty$.

2. Sei X absolut stetig verteilt mit Dichte f_X . Der Erwartungswert von X ist definiert als

$$\mathbb{E}X = \int_{\mathbb{R}} x f_X(x) dx,$$

falls $\underbrace{\int_{\mathbb{R}} |x| f_X(x) dx}_{\mathbb{E}|X|} < \infty$

Satz 4.1.1 (Eigenschaften des Erwartungswertes)

Seien X, Y integrierbare Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Dann gilt Folgendes:

1. Falls $X(\omega) = I_A(\omega)$ für ein $A \in \mathcal{F}$, dann gilt $\mathbb{E}X = P(A)$.
2. Falls $X(\omega) \geq 0$, für fast alle $\omega \in \Omega$, dann ist $\mathbb{E}X \geq 0$.
3. *Additivität*: für beliebige $a, b \in \mathbb{R}$ gilt $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$.
4. *Monotonie*: Falls $X \geq Y$ für fast alle $\omega \in \Omega$ (man sagt dazu “fast sicher” und schreibt “f.s.”), dann gilt $\mathbb{E}X \geq \mathbb{E}Y$.
Falls $0 \leq X \leq Y$ fast sicher und lediglich vorausgesetzt wird, dass Y integrierbar ist, dann ist auch X integrierbar.

5. $|\mathbb{E}X| \leq \mathbb{E}|X|$.
6. Falls X fast sicher auf Ω beschränkt ist, dann ist X integrierbar.
7. Falls X und Y unabhängig sind, dann gilt $\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$.
8. Falls $X \geq 0$ fast sicher und $\mathbb{E}X = 0$, dann gilt $X = 0$ fast sicher.

Bemerkung 4.1.1

1. Aus dem Satz 4.1.1, 3) und 7) folgt per Induktion, dass

- (a) Falls X_1, \dots, X_n integrierbare Zufallsvariablen sind und $a_1, \dots, a_n \in \mathbb{R}$, dann ist $\sum_{i=1}^n a_i X_i$ eine integrierbare Zufallsvariable und es gilt

$$\mathbb{E} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i \mathbb{E}X_i.$$

- (b) Falls X_1, \dots, X_n zusätzlich unabhängig sind und $\mathbb{E}|X_1 \dots X_n| < \infty$, dann gilt

$$\mathbb{E} \left(\prod_{i=1}^n X_i \right) = \prod_{i=1}^n \mathbb{E}X_i.$$

2. Die Aussage 7) des Satzes 4.1.1 gilt nicht in umgekehrter Richtung: aus $\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$ folgt im Allgemeinen nicht die Unabhängigkeit von Zufallsvariablen X und Y . Als Illustration betrachten wir folgendes Beispiel:

- (a) Seien X_1, X_2 unabhängige Zufallsvariablen mit $\mathbb{E}X_1 = \mathbb{E}X_2 = 0$. Setzen wir $X = X_1$ und $Y = X_1 \cdot X_2$. X und Y sind abhängig und dennoch

$$\mathbb{E}(XY) = \mathbb{E}(X_1^2 X_2) = \mathbb{E}X_1^2 \cdot \mathbb{E}X_2 = 0 = \mathbb{E}X \cdot \mathbb{E}Y = 0 \cdot \mathbb{E}Y = 0.$$

- (b) In der Vorlesung Stochastik III wird folgendes bewiesen: falls der Zufallsvektor (X, Y) normalverteilt ist, dann sind X und Y unabhängig genau dann, wenn $\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$.

Folgerung 4.1.1

1. Falls X absolut stetig verteilt mit Dichte f_X ist, dann gilt

$$\mathbb{E}g(X) = \int_{\mathbb{R}^n} g(x) f_X(x) dx.$$

2. Falls X diskret verteilt mit dem Wertebereich $C = \{x_1, x_2, \dots\} \subset \mathbb{R}^n$ ist, dann gilt

$$\mathbb{E}g(X) = \sum_i g(x_i) P(X = x_i) = \sum_{x \in C} g(x) P(X = x).$$

Beispiele für die Berechnung des Erwartungswertes:

1. *Poisson-Verteilung:* Sei $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Dann gilt

$$\begin{aligned} EX &= \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1} \cdot \lambda}{(k-1)!} \\ &\stackrel{k-1=n}{=} e^{-\lambda} \cdot \lambda \cdot \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} \lambda e^{\lambda} = \lambda \implies EX = \lambda. \end{aligned}$$

2. *Normalverteilung:* Sei $X \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Zeigen wir, dass $EX = \mu$.

$$\begin{aligned} EX &= \int_{\mathbb{R}} xf_X(x) dx = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{\infty} xe^{-\left(\frac{x-\mu}{\sigma}\right)^2 \cdot \frac{1}{2}} dx \\ &\stackrel{y=\frac{x-\mu}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-\frac{y^2}{2}} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{\mathbb{R}} ye^{-\frac{y^2}{2}} dy}_{=0} + \frac{\mu}{\sqrt{2\pi}} \underbrace{\int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy}_{=\sqrt{2\pi}} = \mu, \end{aligned}$$

weil

$$\int_{\mathbb{R}} ye^{-\frac{y^2}{2}} dy \stackrel{t=\frac{y^2}{2}}{=} \left(\int_{-\infty}^0 + \int_0^{+\infty} \right) e^{-t} dt = - \left(\int_0^{+\infty} - \int_0^{+\infty} e^{-t} \right) = 0;$$

$$\begin{aligned} \left(\int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy \right)^2 &= \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \cdot \int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} dx dy \\ &\stackrel{(x,y) \mapsto \text{Polarkoord. } (r,\varphi)}{=} \int_0^{2\pi} \int_0^{+\infty} e^{-\frac{r^2}{2}} r dr d\varphi \\ &= 2\pi \cdot \int_0^{+\infty} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) \\ &\stackrel{\frac{r^2}{2}=t}{=} 2\pi(-1) \int_0^{+\infty} d(e^{-t}) = 2\pi \end{aligned}$$

$$\implies \int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi} \implies EX = \mu.$$

4.2 Varianz

Neben dem ‘‘Mittelwert’’ einer Zufallsvariablen, den der Erwartungswert reprasentiert, gibt es weitere Charakteristiken, die fur die praktische Beschreibung der zufalligen Vorgange in der Natur und Technik sehr wichtig sind. Die Varianz z.B. beschreibt die Streuung der Zufallsvariablen um ihren Mittelwert. Sie wird als mittlere quadratische Abweichung vom Erwartungswert eingefuhrt:

Definition 4.2.1 Sei X eine Zufallsvariable mit $EX^2 < \infty$.

1. Die *Varianz* der Zufallsvariablen X wird als $\text{Var } X = E(X - EX)^2$ definiert.
2. $\sqrt{\text{Var } X}$ heit *Standardabweichung* von X .
3. Seien X und Y zwei Zufallsvariablen mit $E|XY| < \infty$. Die Groe $\text{Cov}(X, Y) = E(X - EX)(Y - EY)$ heit *Kovarianz* der Zufallsvariablen X und Y .
4. Falls $\text{Cov}(X, Y) = 0$, dann heien die Zufallsvariablen X und Y *unkorreliert*.

Satz 4.2.1 (*Eigenschaften der Varianz und der Kovarianz*)

Seien X, Y zwei Zufallsvariablen mit $E(X^2) < \infty, E(Y^2) < \infty$. Dann gelten folgende Eigenschaften:

1. $\text{Cov}(X, Y) = E(XY) - EX \cdot EY, \quad \text{Var } X = E(X^2) - (EX)^2. \quad (4.1)$
2. $\text{Cov}(aX + b, cY + d) = ac \cdot \text{Cov}(X, Y), \quad \forall a, b, c, d \in \mathbb{R}, \quad (4.2)$
 $\text{Var}(aX + b) = a^2 \text{Var}(X), \quad \forall a, b \in \mathbb{R}. \quad (4.3)$
3. $\text{Var } X \geq 0$. Es gilt $\text{Var } X = 0$ genau dann, wenn $X = EX$ fast sicher.
4. $\text{Var}(X + Y) = \text{Var } X + \text{Var } Y + 2\text{Cov}(X, Y)$.
5. Falls X und Y unabhangig sind, dann sind sie unkorreliert, also gilt $\text{Cov}(X, Y) = 0$.

Beweis 1. Beweisen wir die Formel $\text{Cov}(X, Y) = E(XY) - EX \cdot EY$. Die Darstellung (4.1) fur die Varianz ergibt sich aus dieser Formel fur $X = Y$.

$$\begin{aligned} \text{Cov}(X, Y) &= E(X - EX)(Y - EY) \\ &= E(XY - XEY - YEX + EX \cdot EY) \\ &= E(XY) - EX \cdot EY - EY \cdot EX + EX \cdot EY \\ &= E(XY) - EX \cdot EY. \end{aligned}$$

2. Beweisen wir die Formel (4.2). Die Formel (4.3) folgt aus (4.2) für $X = Y$, $a = c$ und $b = d$. Es gilt

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= \text{E}((aX + b - aEX - b)(cY + d - cEY - d)) \\ &= \text{E}(ac(X - EX)(Y - EY)) \\ &= ac \text{Cov}(X, Y), \quad \forall a, b, c, d \in \mathbb{R}. \end{aligned}$$

3. Es gilt offensichtlich

$$\text{Var } X = \text{E}(X - EX)^2 \geq 0, \text{ da } (X - EX)^2 \geq 0 \quad \forall \omega \in \Omega.$$

Falls $X = EX$ fast sicher, dann gilt $(X - EX)^2 = 0$ fast sicher und somit $\text{E}(X - EX)^2 = 0 \implies \text{Var } X = 0$.

Falls umgekehrt $\text{Var } X = 0$, dann bedeutet es $\text{E}(X - EX)^2 = 0$ für $(X - EX)^2 \geq 0$. Damit folgt nach Satz 4.1.1, 8) $(X - EX)^2 = 0$ fast sicher $\implies X = EX$ fast sicher.

4.
$$\begin{aligned} \text{Var}(X + Y) &= \text{E}(X + Y)^2 - (\text{E}(X + Y))^2 \\ &= \text{E}(X^2 + 2XY + Y^2) - (\text{E}X + \text{E}Y)^2 \\ &= \text{E}(X^2) + 2\text{E}(XY) + \text{E}Y^2 - (\text{E}X)^2 - 2\text{E}X \cdot \text{E}Y - (\text{E}Y)^2 \\ &= \underbrace{\text{E}(X^2) - (\text{E}X)^2}_{\text{Var } X} + \underbrace{\text{E}(Y^2) - (\text{E}Y)^2}_{\text{Var } Y} + \underbrace{2(\text{E}(XY) - \text{E}X \cdot \text{E}Y)}_{\text{Cov}(X, Y)} \\ &= \text{Var } X + \text{Var } Y + 2\text{Cov}(X, Y). \end{aligned}$$

5. Falls X und Y unabhängig sind, dann gilt nach dem Satz 4.1.1, 7) $\text{E}(XY) = \text{E}X \cdot \text{E}Y$ und somit $\text{Cov}(X, Y) = \text{E}(XY) - \text{E}X \cdot \text{E}Y = 0$. □

Folgerung 4.2.1 1. Es gilt $\text{Var } a = 0 \quad \forall a \in \mathbb{R}$.

2. Falls $\text{Var } X = 0$, dann ist $X = \text{const}$ fast sicher.
 3. Für Zufallsvariablen X_1, \dots, X_n mit $\text{E}X_i^2 < \infty$, $i = 1, \dots, n$ gilt

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var } X_i + 2 \sum_{i < j} \text{Cov}(X_i, X_j).$$

4. Falls X_1, \dots, X_n paarweise unkorreliert sind, dann gilt

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var } X_i.$$

Dies gilt insbesondere dann, wenn die Zufallsvariablen X_1, \dots, X_n paarweise unabhängig sind.

Beispiel 4.2.11. *Poisson-Verteilung:*

Sei $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Zeigen wir, dass $\text{Var } X = \lambda$.

Es ist uns bereits bekannt, dass $\text{E}X = \lambda$. Somit gilt

$$\begin{aligned}
 \text{Var } X &= \text{E}(X^2) - \lambda^2 = \sum_{k=0}^{\infty} k^2 \underbrace{e^{-\lambda} \frac{\lambda^k}{k!}}_{=P(X=k)} - \lambda^2 \\
 &= e^{-\lambda} \sum_{k=1}^{\infty} (k(k-1) + k) \frac{\lambda^k}{k!} - \lambda^2 \\
 &= e^{-\lambda} \sum_{k=1}^{\infty} k(k-1) \frac{\lambda^k}{k!} + \underbrace{\sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!}}_{=\text{E}X=\lambda} - \lambda^2 \\
 &= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2} \cdot \lambda^2}{(k-2)!} + \lambda - \lambda^2 \\
 &\stackrel{m=k-2}{=} \lambda^2 \underbrace{\sum_{m=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^m}{m!}}_{=1} + \lambda - \lambda^2 = \lambda.
 \end{aligned}$$

$$\boxed{\implies \text{Var } X = \lambda.}$$

2. *Normalverteilung:*

Sei $X \sim N(\mu, \sigma^2)$. Zeigen wir, dass $\text{Var } X = \sigma^2$. Wie wir wissen, gilt $\text{E}X = \mu$ und somit

$$\begin{aligned}
 \text{Var } X &= \text{E}(X - \mu)^2 = \int_{\mathbb{R}} (x - \mu)^2 \underbrace{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{=f_X(x)} dx \\
 &\stackrel{y=\frac{x-\mu}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \sigma^2 \int_{\mathbb{R}} y^2 e^{-\frac{y^2}{2}} dy \\
 &= \frac{1}{\sqrt{2\pi}} \sigma^2 \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2}} d\left(\frac{y^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \sigma^2 \int_{-\infty}^{\infty} y d\left(e^{-\frac{y^2}{2}}\right) \\
 &= \frac{1}{\sqrt{2\pi}} \sigma^2 \left(-y e^{-\frac{y^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) \\
 &= \frac{1}{\sqrt{2\pi}} \sigma^2 (-0 + \sqrt{2\pi}) = \sigma^2.
 \end{aligned}$$

4.3 Kovarianz und Korrelationskoeffizient

Definition 4.3.1 Seien X und Y zwei Zufallsvariablen mit $0 < \text{Var } X$, $\text{Var } Y < \infty$. Die Größe

$$\varrho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X} \sqrt{\text{Var } Y}}$$

heißt *Korrelationskoeffizient* von X und Y .

$\varrho(X, Y)$ ist ein Maß für die lineare Abhängigkeit der Zufallsvariablen X und Y .

Satz 4.3.1 (*Eigenschaften des Korrelationskoeffizienten*):

Seien X und Y zwei Zufallsvariablen mit $0 < \text{Var } X$, $\text{Var } Y < \infty$. Dann gilt

1. $|\varrho(X, Y)| \leq 1$.
2. $\varrho(X, Y) = 0$ genau dann, wenn X und Y unkorreliert sind. Eine hinreichende Bedingung dafür ist die Unabhängigkeit von X und Y .
3. $|\varrho(X, Y)| = 1$ genau dann, wenn X und Y fast sicher linear abhängig sind, d.h., $\exists a \neq 0, b \in \mathbb{R} : P(Y = aX + b) = 1$.

Beweis Setzen wir

$$\bar{X} = \frac{X - \text{E}X}{\sqrt{\text{Var } X}}, \quad \bar{Y} = \frac{Y - \text{E}Y}{\sqrt{\text{Var } Y}}.$$

Diese Konstruktion führt zu den sogenannten *standardisierten Zufallsvariablen* \bar{X} und \bar{Y} , für die

$$\begin{aligned} \text{E}\bar{X} &= 0, & \text{Var } \bar{X} &= 1, & \text{Cov}(\bar{X}, \bar{Y}) &= \text{E}(\bar{X}\bar{Y}) = \varrho(X, Y) \\ \text{E}\bar{Y} &= 0, & \text{Var } \bar{Y} &= 1. \end{aligned}$$

1. Es gilt

$$\begin{aligned} 0 &\leq \text{Var}(\bar{X} \pm \bar{Y}) = \text{E}(\bar{X} \pm \bar{Y})^2 = \underbrace{\text{E}(\bar{X})^2}_{\text{Var } \bar{X}=1} \pm 2\text{E}(\bar{X} \cdot \bar{Y}) + \underbrace{\text{E}(\bar{Y})^2}_{\text{Var } \bar{Y}=1} \\ &= 2 \pm 2\varrho(X, Y) \implies 1 \pm \varrho(X, Y) \geq 0 \implies |\varrho(X, Y)| \leq 1. \end{aligned}$$

2. Folgt aus der Definition 4.3.1 und dem Satz 4.2.1, 5).

3. “ \Leftarrow ” Falls $Y = aX + b$ fast sicher, $a \neq 0, b \in \mathbb{R}$, dann gilt Folgendes: Bezeichnen wir $\text{E}X = \mu$ und $\text{Var } X = \sigma^2$. Dann ist $\text{E}Y = a\mu + b$, $\text{Var } Y = a^2 \cdot \sigma^2$ und somit

$$\begin{aligned} \varrho(X, Y) &= \text{E}(\bar{X}\bar{Y}) = \text{E}\left(\frac{X - \mu}{\sigma} \cdot \frac{aX + b - a\mu - b}{|a| \cdot \sigma}\right) \\ &= \text{E}\left(\underbrace{\left(\frac{X - \mu}{\sigma}\right)^2}_{\bar{X}^2} \cdot \text{sgn } a\right) = \text{sgn } a \cdot \underbrace{\text{E}(\bar{X}^2)}_{\text{Var } \bar{X}=1} = \text{sgn } a = \pm 1. \end{aligned}$$

“ \Rightarrow ” Sei $|\rho(X, Y)| = 1$. O.B.d.A. betrachten wir den Fall $\rho(X, Y) = 1$. Aus 1) gilt $\text{Var}(\bar{X} - \bar{Y}) = 2 - 2\rho(X, Y) = 0 \implies \bar{X} - \bar{Y} = \text{const}$ fast sicher aus dem Satz 4.2.1, 3). Somit sind X und Y linear abhängig.

Für den Fall $\rho(X, Y) = -1$ betrachten wir analog

$$\text{Var}(\bar{X} + \bar{Y}) = 2 + 2\rho(X, Y) = 0.$$

□

4.4 Höhere und gemischte Momente

Außer des Erwartungswertes, der Varianz und der Kovarianz gibt es eine Reihe von weiteren Charakteristiken von Zufallsvariablen, die für uns von Interesse sind.

Definition 4.4.1 Seien X, X_1, \dots, X_n Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) .

1. Der Ausdruck $\mu_k = E(X^k)$, $k \in \mathbb{N}$ heißt *k-tes Moment* der Zufallsvariablen X .
2. $\bar{\mu}_k = E(X - EX)^k$, $k \in \mathbb{N}$ heißt *k-tes zentriertes Moment* der Zufallsvariablen X .
3. $E(X_1^{k_1} \cdot \dots \cdot X_n^{k_n})$, $k_1, \dots, k_n \in \mathbb{N}$ heißt *gemischtes Moment* der Zufallsvariablen X_1, \dots, X_n der Ordnung $k = k_1 + \dots + k_n$.
4. $E[(X_1 - EX_1)^{k_1} \cdot \dots \cdot (X_n - EX_n)^{k_n}]$ heißt *zentriertes gemischtes Moment* der Zufallsvariablen X_1, \dots, X_n der Ordnung $k = k_1 + \dots + k_n$.

Anmerkung:

- (a) Die angegebenen Momente müssen nicht unbedingt existieren, beispielsweise existiert EX^k , $k \in \mathbb{N}$ für $X \sim \text{Cauchy}(0, 1)$ nicht.
- (b) Offensichtlich ist $\text{Var} X$ das zweite zentrierte Moment von X , genauso wie $\text{Cov}(X, Y)$ das zweite zentrierte gemischte Moment von X und Y ist. Dabei haben Momente dritter und vierter Ordnung eine besondere Bedeutung:

Definition 4.4.2 1. Der Quotient

$$\gamma_1 = \text{Sch}(X) = \frac{\bar{\mu}_3}{\sqrt{(\bar{\mu}_2)^3}} = \frac{E(X - EX)^3}{\sqrt{(\text{Var } X)^3}} = E(\bar{X}^3)$$

heißt *Schiefte* oder *Symmetriekoeffizient* der Verteilung von X . Falls $\gamma_1 > 0$, dann ist die Verteilung von X rechtsschief bzw. linkssteil (für

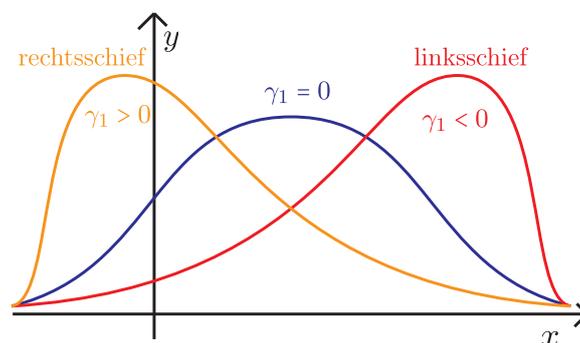


Abbildung 4.1: Veranschaulichung der Schiefe einer Verteilung an Hand der Grafik ihrer Dichte

$\gamma_1 < 0$ linksschief bzw. rechtssteil) vgl. hierzu Abbildung 4.1. Es ist ein Maß für die Symmetrie der Verteilung.

2. Der Ausdruck

$$\gamma_2 = \frac{\bar{\mu}_4}{\bar{\mu}_2^2} - 3 = \frac{E(X - EX)^4}{(\text{Var } X)^2} - 3 = E(\bar{X}^4) - 3$$

heißt *Wölbung (Exzess)* der Verteilung von X . Es ist ein Maß für die “Spitzigkeit” der Verteilung:

- $\gamma_2 > 0$ – Verteilung steilgipflig
- $\gamma_2 < 0$ – Verteilung flachgipflig, vgl. Abb. 4.2.

Die beiden Kerngrößen messen Abweichungen der Verteilung von X

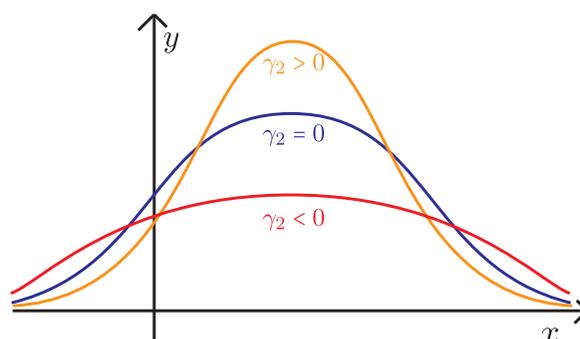


Abbildung 4.2: Veranschaulichung der Wölbung einer Verteilung an Hand der Grafik ihrer Dichte

von einer Gaußschen $N(\mu, \sigma^2)$ -Verteilung, für die $\gamma_1 = \gamma_2 = 0$.

Übungsaufgabe 4.4.1 Beweisen Sie, dass $\gamma_1 = \gamma_2 = 0$ für $X \sim N(\mu, \sigma^2)$.

4.5 Ungleichungen

Satz 4.5.1 (*Ungleichung von Markow*):

Sei X eine Zufallsvariable mit $E|X|^r < \infty$ für ein $r \geq 1$. Dann gilt

$$P(|X| \geq \varepsilon) \leq \frac{E|X|^r}{\varepsilon^r} \quad \forall \varepsilon > 0.$$

Beweis Es gilt

$$\begin{aligned} E|X|^r &= \underbrace{E(|X|^r \cdot I(|X| \leq \varepsilon))}_{\geq 0} + E(|X|^r \cdot I(|X| > \varepsilon)) \\ &\geq E(\varepsilon^r \cdot I(|X| > \varepsilon)) = \varepsilon^r \cdot P(|X| > \varepsilon), \end{aligned}$$

daher $P(|X| > \varepsilon) \leq \frac{E|X|^r}{\varepsilon^r}$. □

Folgerung 4.5.1 (*Ungleichung von Tschebyschew*).

1. Sei X eine Zufallsvariable mit $EX^2 < \infty$ und $\varepsilon > 0$. Dann gilt

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

2. Sei $Ee^{sX} < \infty$ für ein $s > 0$. Dann gilt

$$P(X \geq \varepsilon) \leq \frac{Ee^{\lambda X}}{e^{\lambda \varepsilon}}, \quad \forall \varepsilon > 0 \quad \forall 0 \leq \lambda \leq s.$$

Beweis Benutze die Markow-Ungleichung für die Zufallsvariable

1. $Y = X - EX$ und $r = 2$ und
2. $Y = e^{\lambda X} \geq 0$, $\varepsilon = e^{\lambda \varepsilon_0}$, $r = 1$.

□

Beispiel 4.5.1 Der Durchmesser der Mondscheibe wird aus den Bildern der Astrophotographie folgendermaßen bestimmt: bei jeder Aufnahme der Mondscheibe wird ihr Durchmesser X_i gemessen. Nach n Messungen wird der Durchmesser als $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ aus allen Beobachtungen geschätzt. Sei $\mu = EX_i$ der wahre (unbekannte) Wert des Monddurchmessers. Wie viele Beobachtungen n müssen durchgeführt werden, damit die Schätzung \bar{X}_n weniger als um 0,1 vom Wert μ mit Wahrscheinlichkeit von mindestens 0,99 abweicht? Mit anderen Worten, finde n : $P(|\bar{X}_n - \mu| \leq 0,1) \geq 0,99$. Diese Bedingung ist äquivalent zu $P(|\bar{X}_n - \mu| > 0,1) \leq 1 - 0,99 = 0,01$. Sei $\text{Var } X_i = \sigma^2 > 0$. Dann gilt

$$\text{Var } \bar{X}_n = \frac{1}{n^2} \cdot \sum_{i=1}^n \text{Var } X_i = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n},$$

wobei vorausgesetzt wird, dass alle Messungen X_i unabhängig voneinander durchgeführt werden. Somit gilt nach der Ungleichung von Tschebyschew

$$P(|\bar{X}_n - \mu| > 0,1) \leq \frac{\text{Var } \bar{X}_n}{0,1^2} = \frac{\sigma^2}{n \cdot 0,01},$$

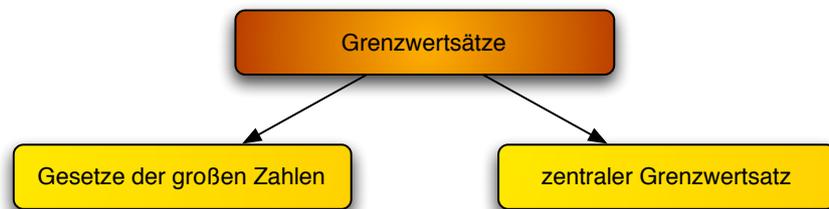
woraus folgt, dass

$$n \geq \frac{\sigma^2}{(0,01)^2} = 10^4 \cdot \sigma^2.$$

Für $\sigma = 1$ braucht man z.B. mindestens 10000 Messungen! Diese Zahl zeigt, wie ungenau die Schranke in der Ungleichung von Tschebyschew ist. Eine viel genauere Antwort ($n \geq 670$) kann man mit Hilfe des zentralen Grenzwertsatzes bekommen. Dies wird allerdings erst im Kapitel 5 behandelt.

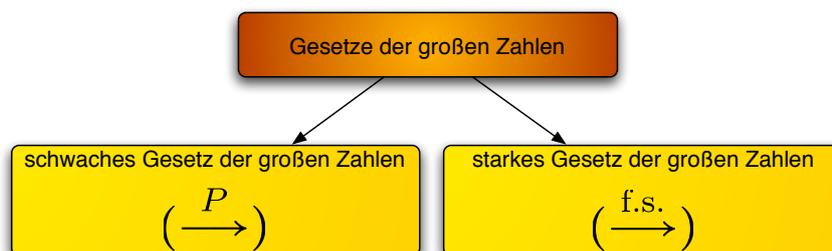
Kapitel 5

Grenzwertsätze



In diesem Kapitel betrachten wir Aussagen der Wahrscheinlichkeitsrechnung, die Näherungsformeln von großer anwendungsbezogener Bedeutung liefern. Dies wird an mehreren Beispielen erläutert.

5.1 Gesetze der großen Zahlen



Ein typisches Gesetz der großen Zahlen besitzt die Form

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} EX_0, \quad (5.1)$$

wobei $\{X_n\}_{n \in \mathbb{N}}$ unabhängige identisch verteilte Zufallsvariablen mit $X_n \stackrel{d}{=} X_0$, $E|X_0| < \infty$ sind.

Die Konvergenz in (5.1) wird entweder in Wahrscheinlichkeit oder fast sicher verstanden.

Definition 5.1.1 Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Sei X eine weitere Zufallsvariable auf (Ω, \mathcal{F}, P) . Man sagt, die Folge $\{X_n\}$ konvergiert gegen X für $n \rightarrow \infty$

1. *fast sicher oder mit Wahrscheinlichkeit 1* ($X_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} X$), falls

$$P(\{\omega \in \Omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)\}) = 1.$$

2. *in Wahrscheinlichkeit oder stochastisch* ($X_n \xrightarrow[n \rightarrow \infty]{P} X$), falls

$$\forall \varepsilon > 0 \quad P(|X_n - X| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Wenn \xrightarrow{P} gemeint ist, spricht man von dem *schwachen Gesetz der großen Zahlen*. Falls $\xrightarrow{\text{f.s.}}$ gemeint ist, heißt die Aussage (5.1) *starkes Gesetz der großen Zahlen*.

Im Folgenden verwenden wir die Bezeichnungen $S_n = \sum_{i=1}^n X_i$, $\bar{X}_n = \frac{S_n}{n}$, $\forall n \in \mathbb{N}$ für eine Folge $\{X_n\}_{n \in \mathbb{N}}$ von Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) .

5.1.1 Schwaches Gesetz der großen Zahlen

Satz 5.1.1 (*Markow*)

Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit $EX_i^2 < \infty \forall i$. Falls

$$\text{Var } \bar{X}_n \xrightarrow[n \rightarrow \infty]{} 0, \tag{5.2}$$

dann gilt

$$\bar{X}_n - \frac{1}{n} \sum_{i=1}^n EX_i \xrightarrow[n \rightarrow \infty]{P} 0.$$

Beweis Da $X_i \in L^2(\Omega, \mathcal{F}, P) \quad \forall i \in \mathbb{N}$, gilt durch die wiederholte Anwendung der Minkowski-Ungleichung für $p = 2$

$$\sqrt{E\bar{X}_n^2} \leq \frac{1}{n} \sum_{i=1}^n \sqrt{EX_i^2} < \infty$$

und somit $\bar{X}_n \in L^2(\Omega, \mathcal{F}, P)$. Aus der Ungleichung von Tschebyschew folgt für alle $\varepsilon > 0$

$$P\left(\left|\bar{X}_n - \frac{1}{n} \sum_{i=1}^n EX_i\right| > \varepsilon\right) \leq \frac{E(\bar{X}_n - E\bar{X}_n)^2}{\varepsilon^2} = \frac{\text{Var } \bar{X}_n}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Somit gilt

$$\bar{X}_n - E\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} 0.$$

□

Folgerung 5.1.1 Seien die Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ im Satz 5.1.1 unabhängig. Dann gilt Folgendes:

1. Die Bedingung $\text{Var } \bar{X}_n \xrightarrow{n \rightarrow \infty} 0$ bekommt die Form

$$\frac{1}{n^2} \sum_{i=1}^n \text{Var } X_i \xrightarrow{n \rightarrow \infty} 0.$$

2. Falls $\text{Var } X_n \leq c = \text{const} \quad \forall n \in \mathbb{N}$, dann gilt die Bedingung (5.2) und somit die Aussage des Satzes 5.1.1 (Satz von Tschebyschew).
3. Insbesondere ist die Bedingung $\text{Var } X_n \leq c = \text{const} \quad \forall n \in \mathbb{N}$ erfüllt, falls $\{X_n\}_{n \in \mathbb{N}}$ unabhängige identisch verteilte Zufallsvariablen mit

$$EX_n = \mu, \text{Var } X_n = \sigma^2 < \infty$$

sind. Dann nimmt das schwache Gesetz der großen Zahlen die klassische Form

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} \mu$$

an.

Die Existenz der zweiten Momente ist für das schwache Gesetz der großen Zahlen nicht entscheidend. So kann man mit Hilfe der charakteristischen Funktionen folgenden Satz beweisen:

Satz 5.1.2 (*Schwaches Gesetz der großen Zahlen von Kchintschin*)

Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen, $n \in \mathbb{N}$, $X_n \in L^1$, mit demselben Erwartungswert $EX_n = \mu < \infty$. Dann gilt

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} \mu.$$

Satz 5.1.3 (*Starkes Gesetz der großen Zahlen von Kolmogorow*)

Seien $\{X_n\}_{n \in \mathbb{N}}$ unabhängige identisch verteilte Zufallsvariablen. Es gilt $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mu$ genau dann, wenn $\exists EX_n = \mu < \infty$.

5.1.2 Anwendung der Gesetze der großen Zahlen

1. Monte-Carlo-Methoden zur numerischen Integration

Sei $g : [0, 1]^d \rightarrow \mathbb{R}$ eine beliebige stetige Funktion. Wie kann man mit Hilfe der Gesetze der großen Zahlen

$$\int_{[0,1]^d} g(x) dx = \int_0^1 \dots \int_0^1 g(x_1, \dots, x_d) dx_1 \dots dx_d$$

numerisch berechnen?

Der Algorithmus ist wie folgt:

- Generiere eine Folge von Realisierungen von unabhängigen identisch verteilten Zufallsvariablen

$$X_1, \dots, X_n \text{ mit } X_i \sim [0, 1]^d, i = 1, \dots, n.$$

- Setze

$$\int_{[0,1]^d} g(x) dx \approx \frac{1}{n} \sum_{i=1}^n g(X_i) \quad (5.3)$$

für große n . Dieser Vorgang ist berechtigt, denn nach dem Satz 5.1.3 gilt

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}g(X_1) = \int_{[0,1]^d} g(x) dx$$

und somit gilt (5.3) für ausreichend große n .

Bemerkung:

Dieselbe Methode kann durch geeignete Transformation vom Integrationsgebiet $G \subset \mathbb{R}^d$ und andere Wahl von Zufallsvariablen X_i auf die Berechnung von $\int_G g(x) dx$ erweitert werden, G kompakte Teilmenge von \mathbb{R}^d . So genügt es nur $X_i \sim U(G)$, $i = 1, \dots, n$ zu betrachten.

2. Numerische Berechnung der Zahl π :

Wie kann π mit Hilfe eines Rechners beliebig genau berechnet werden? Dazu wird das starke Gesetz der großen Zahlen wie folgt verwendet:

- Generiere Realisierungen von unabhängig und identisch verteilten Zufallvektoren $X_1, \dots, X_n \in \mathbb{R}^2$ mit $X_i \sim U[-1, 1]^2$, $i = 1, \dots, n$.
- Es gilt

$$\pi \approx \frac{4}{n} \sum_{i=1}^n I(|X_i| \leq 1) \quad (5.4)$$

für große n .

In der Tat, nach dem Satz 5.1.3 gilt

$$\frac{1}{n} \sum_{i=1}^n I(|X_i| \leq 1) \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbf{E}I(|X_1| \leq 1) = P(|X_1| \leq 1) = \frac{|B_1(0)|}{|[-1, 1]^2|} = \frac{\pi}{2^2} = \frac{\pi}{4}.$$

Somit ist die Verwendung der Berechnungsformel (5.4) berechtigt für große n .

5.2 Zentraler Grenzwertsatz

Für die Gesetze der großen Zahlen wurde die Normierung $\frac{1}{n}$ der Summe $S_n = \sum_{i=1}^n X_i$ gewählt, um $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{} \mathbf{E}X_1$ zu beweisen. Falls jedoch eine andere Normierung gewählt wird, so sind andere Grenzwertaussagen möglich. Im Fall der Normierung $\frac{1}{\sqrt{n}}$ spricht man von zentralen Grenzwertsätzen: unter gewissen Voraussetzungen gilt also

$$\frac{S_n - n\mathbf{E}X_1}{\sqrt{n\text{Var } X_1}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1).$$

5.2.1 Klassischer zentraler Grenzwertsatz

Satz 5.2.1 Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit $\mathbf{E}X_i = \mu$, $\text{Var } X_i = \sigma^2$, wobei $0 < \sigma^2 < \infty$. Dann gilt

$$P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) \xrightarrow[n \rightarrow \infty]{} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \quad \forall x \in \mathbb{R},$$

wobei $\Phi(x)$ die Verteilungsfunktion einer $N(0, 1)$ -verteilten Zufallsvariablen ist.

Folgerung 5.2.1 Unter den Voraussetzungen des Satzes 5.2.1 gilt zusätzlich

1.

$$P\left(\frac{S_n - n\mu}{\sqrt{n} \cdot \sigma} < x\right) \xrightarrow[n \rightarrow \infty]{} \Phi(x) \quad \forall x \in \mathbb{R},$$

2.

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n} \cdot \sigma} \leq b\right) \xrightarrow[n \rightarrow \infty]{} \Phi(b) - \Phi(a) \quad \forall a, b \in \mathbb{R}, a \leq b.$$

Beispiel 5.2.1 1. *Satz von de Moivre–Laplace*

Falls $X_n \sim \text{Bernoulli}(p)$, $n \in \mathbb{N}$ unabhängig sind und $p \in (0, 1)$, dann

genügt die Folge $\{X_n\}_{n \in \mathbb{N}}$ mit $EX_n = p$, $\text{Var } X_n = p(1-p)$ den Voraussetzungen des Satzes 5.2.1. Das Ergebnis

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1)$$

wurde mit einfachen Mitteln als erster zentraler Grenzwertsatz von Abraham de Moivre (1667-1754) bewiesen und trägt daher seinen Namen. Es kann folgendermaßen interpretiert werden:

Falls die Anzahl n der Experimente groß wird, so wird die Binomialverteilung von $S_n \sim \text{Bin}(n, p)$, das die Anzahl der Erfolge in n Experimenten darstellt, approximiert durch

$$\begin{aligned} P(a \leq S_n \leq b) &= P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right), \end{aligned}$$

$\forall a, b \in \mathbb{R}$, $a \leq b$, wobei $p \in (0, 1)$ als die Erfolgswahrscheinlichkeit in einem Experiment interpretiert wird. So kann z.B. S_n als die Anzahl von Kopf in n Würfeln einer fairen Münze ($p = \frac{1}{2}$) betrachtet werden. Hier gilt also

$$P(a \leq S_n \leq b) \underset{n \text{-groß}}{\approx} \Phi\left(\frac{2b - n}{\sqrt{n}}\right) - \Phi\left(\frac{2a - n}{\sqrt{n}}\right), \quad a < b.$$

2. Berechnen wir die Anzahl der notwendigen Messungen des Monddurchmessers im Beispiel 4.5.1 mit Hilfe des zentralen Grenzwertsatzes: finde $n \in \mathbb{N}$ mit

$$P(|\bar{X}_n - \mu| \leq 0,1) > 0,99,$$

oder äquivalent dazu

$$P(|\bar{X}_n - \mu| > 0,1) \leq 0,01,$$

wobei $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ist. Es gilt

$$\begin{aligned} P(|\bar{X}_n - \mu| \leq 0,01) &= P\left(-0,1 \leq \frac{S_n - n\mu}{n} \leq 0,1\right) \\ &= P\left(-0,1 \frac{\sqrt{n}}{\sigma} \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq 0,1 \frac{\sqrt{n}}{\sigma}\right) \\ &\underset{n \text{ groß}}{\approx} \Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{0,1\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) - 1, \end{aligned}$$

weil $N(0, 1)$ eine symmetrische Verteilung ist und somit

$$\Phi(x) = 1 - \Phi(-x) \quad \forall x \in \mathbb{R}$$

gilt.

Es muss also $2\Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) - 1 > 0,99$ erfüllt sein. Dies ist äquivalent zu

$$\Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) > \frac{1,99}{2} = 0,995 \iff \frac{0,1\sqrt{n}}{\sigma} > \Phi^{-1}(0,995)$$

oder

$$\begin{aligned} n &> \frac{\sigma^2}{(0,1)^2} \left(\Phi^{-1}(0,995)\right)^2 = \frac{\sigma^2 \left(\Phi^{-1}(0,995)\right)^2}{0,01} \\ &= \frac{\sigma^2 (2,58)^2}{0,01} = \sigma^2 \cdot 665,64. \end{aligned}$$

Für $\sigma^2 = 1$ ergibt sich die Antwort

$$\boxed{n \geq 666}$$

5.2.2 Grenzwertsatz von Lindeberg

Im klassischen zentralen Grenzwertsatz wurden Folgen von unabhängigen und identisch verteilten Zufallsvariablen betrachtet. In diesem Abschnitt lassen wir die Voraussetzung der identischen Verteiltheit fallen und formulieren einen allgemeineren Grenzwertsatz in der Form von Lindeberg.

Satz 5.2.2 (Lindeberg)

Sei $\{X_{nk}, 1 \leq k \leq n, n \in \mathbb{N}\}$, eine Folge von unabhängigen Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit den Eigenschaften

$$EX_{nk} = 0, 0 < \sigma_{nk}^2 = \text{Var } X_{nk} < \infty, \sum_{k=1}^n \sigma_{nk}^2 = 1, \quad k \in \{1, \dots, n\}, n \in \mathbb{N}.$$

Falls $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} \left(X_{nk}^2 \cdot I(|X_{nk}| > \varepsilon) \right) = 0, \quad (5.5)$$

dann gilt

$$\sum_{k=1}^n X_{nk} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1).$$

Bemerkung 5.2.1

1. Der klassische zentrale Grenzwertsatz stellt einen Spezialfall des Satzes 5.2.2 dar: falls $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathbb{E}X_n = \mu$, $0 < \sigma^2 = \text{Var } X_n < \infty$ $\forall n \in \mathbb{N}$ ist, setzen wir $X_{nk} = \frac{X_k - \mu}{\sqrt{n}\sigma}$, $k = 1, \dots, n$, $n \in \mathbb{N}$. Es gilt $\mathbb{E}X_{nk} = 0$, $\text{Var } X_{nk} = \frac{1}{n}$, $k = 1, \dots, n$, $\sum_{k=1}^n \frac{1}{n} = 1$ und

$$\begin{aligned} & \sum_{k=1}^n \mathbb{E} \left(X_{nk}^2 I(|X_{nk}| > \varepsilon) \right) \\ &= \frac{1}{n\sigma^2} \sum_{k=1}^n \mathbb{E} \left((X_k - \mu)^2 \cdot I(|X_k - \mu| > \sqrt{n}\sigma) \right) \\ &\stackrel{X_k\text{-id. vert.}}{=} \frac{1}{\sigma^2} \mathbb{E} \left(\underbrace{(X_1 - \mu)^2}_Z \cdot I(|X_1 - \mu| > \sigma\sqrt{n}) \right) \\ &= \frac{1}{\sigma^2} \int_{|x| > \sigma\sqrt{n}} x^2 dF_Z(x) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Somit sind alle Voraussetzungen des Satzes 5.2.2 erfüllt.

2. Die Lindeberg-Bedingung (5.5) nennt man auch *gleichmäßige asymptotische Kleinheit von X_{nk}* aus folgendem Grund: aus (5.5) folgt

$$\begin{aligned} & \max_{1 \leq k \leq n} P(|X_{nk}| > \varepsilon) \stackrel{\text{Folg. 4.5.1}}{\leq} \max_{1 \leq k \leq n} \frac{\mathbb{E}X_{nk}^2}{\varepsilon^2} \\ &\stackrel{\forall \delta > 0}{=} \frac{1}{\varepsilon^2} \left(\max_{1 \leq k \leq n} \mathbb{E} \left(X_{nk}^2 \cdot I(|X_{nk}| \leq \delta) \right) + \max_{1 \leq k \leq n} \mathbb{E} \left(X_{nk}^2 \cdot I(|X_{nk}| > \delta) \right) \right) \\ &\leq \frac{\delta^2}{\varepsilon^2} + \frac{1}{\varepsilon^2} \sum_{k=1}^n \mathbb{E} \left(X_{nk}^2 \cdot I(|X_{nk}| > \delta) \right) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

wegen (5.5) und weil $\delta > 0$ beliebig klein gewählt werden kann.

Für den Beweis des Satzes 5.2.2 brauchen wir eine Reihe von Hilfsergebnissen auf die wir zum Teil im folgenden eingehen.

Bemerkung 5.2.2 Statt die Zufallsvariable X_{nk} in der Formulierung des Satzes 5.2.2 zu verwenden, können auch die Zufallsvariablen $\{X_n\}$ (wie im klassischen zentralen Grenzwertsatz) verwendet werden, wie folgende äquivalente Formulierung zeigt:

Satz 5.2.3 (Lindeberg)

Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit $\mathbb{E}X_n = \mu_n$, $0 < \sigma_n^2 = \text{Var } X_n < \infty \forall n$. Sei $S_n = \sum_{i=1}^n X_i$, $D_n^2 = \sum_{i=1}^n \sigma_i^2$. Falls

$$\frac{1}{D_n^2} \sum_{k=1}^n \mathbb{E} \left((X_k - \mu_k)^2 \cdot I(|X_k - \mu_k| > \varepsilon D_n) \right) \xrightarrow{n \rightarrow \infty} 0,$$

dann gilt

$$\frac{S_n - \sum_{i=1}^n \mu_i}{D_n} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1).$$

Beweis Setzen wir $X_{nk} = \frac{X_k - \mu_k}{D_n}$ und wenden wir den Satz 5.2.2 auf diese Folge von $\{X_{nk}\}_{k=1}^{n \in \mathbb{N}}$ an, die den Voraussetzungen des Satzes genügt. \square

Folgerung 5.2.2 Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit $|X_n| \leq c < \infty \forall n \in \mathbb{N}$ und $D_n \xrightarrow[n \rightarrow \infty]{} \infty$. Dann gilt der zentrale Grenzwertsatz in der Form des Satzes 5.2.3.

Beweis Wir müssen die Gültigkeit der Lindeberg-Bedingung (5.5) prüfen: aus der Ungleichung von Tschebyschew folgt

$$\begin{aligned} \mathbb{E} \left((X_k - \mu_k)^2 \cdot I(|X_k - \mu_k| > \varepsilon D_n) \right) &\leq_{|X_k| \leq c, |\mathbb{E}X_k| \leq c} (2c)^2 P(|X_k - \mu_k| > \varepsilon D_n) \\ &\leq 4c^2 \frac{\sigma_k^2}{\varepsilon^2 D_n^2} \xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

für $1 \leq k \leq n$ und somit

$$\frac{1}{D_n^2} \sum_{k=1}^n \mathbb{E} \left((X_k - \mu_k)^2 \cdot I(|X_k - \mu_k| > \varepsilon D_n) \right) \leq 4c^2 \frac{\overbrace{\sum_{k=1}^n \sigma_k^2}^{=D_n^2}}{\varepsilon^2 D_n^4} = \frac{4c^2}{\varepsilon^2 D_n^2} \xrightarrow[n \rightarrow \infty]{} 0,$$

weil $D_n^2 \xrightarrow[n \rightarrow \infty]{} \infty$. \square

5.2.3 Konvergenzgeschwindigkeit im zentralen Grenzwertsatz

In diesem Abschnitt möchten wir die *Schnelligkeit der Konvergenz im zentralen Grenzwertsatz* untersuchen. Damit aber diese Fragestellung überhaupt sinnvoll erscheint, muss die Konvergenz im zentralen Grenzwertsatz gleichmäßig sein:

$$\sup_{x \in \mathbb{R}} \left| P \left(\sum_{k=1}^n X_{nk} \leq x \right) - \Phi(x) \right| \xrightarrow[n \rightarrow \infty]{} 0.$$

Das ist tatsächlich der Fall, wie aus der Stetigkeit von $\Phi(x)$ und dem folgenden Lemma 5.2.1 hervorgeht.

Lemma 5.2.1 Sei $X_n \xrightarrow[n \rightarrow \infty]{d} X$, wobei $F_n(x) = P(X_n \leq X)$ und $F(x) = P(X \leq x)$. Falls $F(x)$ stetig ist $\forall x \in \mathbb{R}$, dann ist die Konvergenz von F_n zu F gleichmäßig:

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{} 0.$$

Satz 5.2.4 (*Berry-Esséen*)

Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathbb{E}X_n = \mu$, $\text{Var } X_n = \sigma^2 > 0$, $\mathbb{E}|X_n|^3 < \infty$. Sei

$$F_n(x) = P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right), \quad x \in \mathbb{R}, n \in \mathbb{N}.$$

Dann gilt

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{c \cdot \mathbb{E}|X_1 - \mu|^3}{\sigma^3 \sqrt{n}},$$

wobei c eine Konstante ist, $\frac{1}{\sqrt{2\pi}} \leq c < 0,4785$, $\frac{1}{\sqrt{2\pi}} \approx 0,39894$.

Literaturverzeichnis

- [1] H. Dehling, B. Haupt. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Springer, Berlin, 2003.
- [2] H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter, Berlin, 1991.
- [3] A. A. Borovkov. *Wahrscheinlichkeitstheorie: eine Einführung*. Birkhäuser, Basel, 1976.
- [4] W. Feller. *An introduction to probability theory and its applications. Vol I/II*. J. Wiley & Sons, New York, 1970/71.
- [5] H. O. Georgii. *Stochastik*. de Gruyter, Berlin, 2002.
- [6] B. V. Gnedenko. *Einführung in die Wahrscheinlichkeitstheorie*. Akademie, Berlin, 1991.
- [7] C. Hesse. *Angewandte Wahrscheinlichkeitstheorie*. Vieweg, Braunschweig, 2003.
- [8] A. F. Karr. *Probability*. Springer, New York, 1993.
- [9] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie*. Vieweg, Braunschweig, 2002.
- [10] J. Jacod, P. Protter. *Probability essentials*. Springer, Berlin, 2003.
- [11] L. Sachs. *Angewandte Statistik*. Springer, 2004.
- [12] A. N. Shiryaev. *Probability*. Springer, New York, 1996.
- [13] J. M. Stoyanov. *Counterexamples in probability*. Wiley & Sons, 1987.
- [14] H. Tijms. *Understanding probability. Chance rules in everyday life*. Cambridge University Press, 2004.
- [15] P. Gänßler, W. Stute. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 1977.

Index

- a-posteriori-Wahrscheinlichkeiten, 23
- a-priori-Wahrscheinlichkeit, 23
- absolut stetige Verteilung, 35
- Algebra, 8
 - σ -Algebra, 9
 - Borelsche σ -Algebra, 11
 - Eigenschaften, 9
 - erzeugende Algebra, 10
 - minimale σ -Algebra, 10
- Approximation
 - Binomiale, 33
 - Poissonsche, 33
- Approximationssatz, 33
- arithmetisches Mittel, *siehe* Mittel
- Axiome von Kolmogorow, 10

- Bayesche Formel, 22
- bedingte Wahrscheinlichkeit, 20
- Bernoulli-Verteilung, 31
- Berry
 - Satz von Berry-Esséen, 73
- Binomiale Approximation, 33
- Binomialverteilung, 29, 32
- Borel-Mengen, 11
- Bose-Einstein-Statistik, 17

- Cauchy-Verteilung, 39

- Dichte, 35, 41
- disjunkt, 8
- diskrete Verteilung, *siehe* Verteilung
- 3σ -Regel, 37

- endlicher Wahrscheinlichkeitsraum, 13
- Ereignis, 7
- Erwartungswert, *siehe* Momente
- Esséen
 - Satz von Berry-Esséen, 73
- Exklusionsprinzip von Pauli, 17

- Exponentialverteilung, 38
- Exzess, 61

- Faltungsformel, 50
- Faltungsstabilität
 - Normalverteilung, 50
- Fermi-Dirac-Statistik, 17
- flachgipflig, *siehe* Exzess

- Gaußsche-Verteilung, *siehe* Normalverteilung
- Geburtstagsproblem, 14
- gemischte Momente, *siehe* Momente
- Geometrische Verteilung, 32
- geometrische Wahrscheinlichkeit, 19
- geometrisches Mittel, *siehe* Mittel
- gewichtetes Mittel, *siehe* Mittel
- Gleichheit in Verteilung, 40
- Gleichverteilung, 32, 43
- Gleichverteilung auf $[a, b]$, 38
- Grenzwertsätze, 64
 - Gesetz der großen Zahlen, 64–68
 - Anwendungen, 67–68
 - schwaches Gesetz der großen Zahlen, 65
 - zentraler Grenzwertsatz, 68
 - Grenzwertsatz von Lindeberg, 70
 - klassischer, 68–70
 - Konvergenzgeschwindigkeit, 72
- Grundgesamtheit, 7

- höhere Momente, *siehe* Momente
- harmonisches Mittel, *siehe* Mittel
- hypergeometrische Verteilung, 18, 32

- Indikator-Funktion, 26, 27

- k -tes Moment, *siehe* Momente
- kausale und stochastische Unabhängigkeit, 22

Kchintschin, schwaches Gesetz der großen Zahlen von, *siehe* Grenzwertsätze
 klassisches Wahrscheinlichkeitsmaß, 13
 Kolmogorow
 Gesetz der großen Zahlen, 66
 Konvergenz
 fast sicher, 65
 in Wahrscheinlichkeit, 65
 mit Wahrscheinlichkeit 1, 65
 stochastisch, 65
 Korrelationskoeffizient, 59
 Kovarianz, *siehe* Momente

 laplacesches Wahrscheinlichkeitsmaß, 13
 Lindeberg
 Grenzwertsatz von, *siehe* Grenzwertsätze
 Satz von, 71
 lineare Abhängigkeit, Maß für, *siehe* Korrelationskoeffizient
 linksschief, *siehe* Schiefe
 linkssteil, *siehe* Schiefe

 Markow
 schwaches Gesetz der großen Zahlen, *siehe* Grenzwertsätze
 Ungleichung von, *siehe* Ungleichungen
 Maxwell–Boltzmann–Statistik, 17
 messbare Zerlegung, 22
 Messraum, 9
 Mischungen von Verteilungen, 39
 Mittel
 arithmetisches, 52
 geometrisches, 52
 gewichtetes, 52
 harmonisches, 52
 Mittelwert, 37
 mittlere quadratische Abw. des EW, *siehe* Varianz
 Momente, 52
 Erwartungswert, 53
 absolut stetiger ZV, 54
 Additivität, 53
 diskreter ZV, 54
 Monotonie, 53
 Normalverteilung, 55
 Poisson–Verteilung, 55
 gemischte, 60
 zentrales gemischtes Moment, 60
 höhere, 60
 k–tes Moment, 60
 k–tes zentriertes Moment, 60
 Kovarianz, 56
 Varianz, 56
 Addition, 57
 Normalverteilung, 58
 Poisson–Verteilung, 58
 Mondscheibe – Messung des Diameters, 62
 Monte–Carlo, Methode zur numerischen Integration, 67
 Multiplikationssatz, 21
 multivariate Verteilungsfunktion, 40

 Normalverteilung, *siehe* Verteilung
 Erwartungswert, *siehe* Momente
 Varianz, *siehe* Momente

 paarweise disjunkt, 8
 Pauli, Exklusionsprinzip von, 17
 π , Berechnung von, 67
 Poisson–Verteilung, 33
 Erwartungswert, *siehe* Momente
 Varianz, *siehe* Momente
 Poissonsche Approximation, 33
 Polynomiale Verteilung, 42
 Potenzmenge, 9
 Prüfverteilung, 50
 Problem von Galilei, 14

 rechtsschief, *siehe* Schiefe
 rechtssteil, *siehe* Schiefe
 Routing–Problem, 23

 Schiefe, 60
 linksschief, 61
 linkssteil, 60
 rechtsschief, 60
 rechtssteil, 61
 Siebformel, 12
 σ –Algebra, *siehe* Algebra
 σ –Subadditivität, 12
 Spitzigkeit, *siehe* Exzess
 Standardabweichung, 37, 56
 Standardnormalverteilung, *siehe* Verteilung
 steilgipflig, *siehe* Exzess
 Stichprobenraum, 7

- (stochastisch) unabhängig, 12
- stochastische Unabhängigkeit, 12
- Streuung, 37
- Subadditivität des Wahrscheinlichkeitsmaßes P , 12
- Symmetriekoeffizient, 60
- symmetrische Differenz, 8

- totale Wahrscheinlichkeit, *siehe* Bayesche Formel, 22
- Transformationssatz, 47
 - linear, 48
- Tschebyschew, Ungleichung von, *siehe* Ungleichungen

- Unabhängigkeit, 45
 - Charakterisierung, 45
- Ungleichungen
 - Markow, 62
 - Tschebyschew, 62
- unkorreliert, 56
 - falls unabhängig, 56
- unvereinbare Ereignisse, 8
- Urnenmodell, 15

- Varianz, *siehe* Momente
- Verteilung, 27
 - absolut stetig, 35
 - Cauchy-, 39
 - Eigenschaften, 35
 - Exponential-, 38
 - Gleich-, 38
 - Normal-, 37, 50
 - Standardnormal-, 37
 - diskret, 30
 - Beispiele, 31
 - Bernoulli-, 31
 - Binomial-, 32
 - geometrische, 32
 - Gleich-, 32
 - hypergeometrische, 32
 - Poisson-, 33
 - quadrierte ZV, 48
 - von Zufallsvektoren, 40
 - absolut stetig, 41
 - diskret, 41
 - Gleichverteilung, 43
 - multivariate Gleichverteilung, 46
 - multivariate Normalverteilung, 44, 46
 - Polynomiale Verteilung, 42
 - Verteilungsfunktion, 40
- Verteilungsfunktion, 27
 - Asymptotik, 29
 - Eigenschaften, 29
 - Monotonie, 29
 - rechtsseitige Stetigkeit, 29

- Wölbung, *siehe* Exzess
- Wahrscheinlichkeit, 10
- Wahrscheinlichkeitsfunktion, 30, 41
- Wahrscheinlichkeitsmaß, 10
- Wahrscheinlichkeitsraum, 10

- Zähldichte, 30, 41
- zentrales gemischtes Moment, *siehe* Momente
- zentriertes Moment, *siehe* Momente
- Zufallsvariable, 26
 - Momente, 52
 - Produkt, 51
 - Quotient, 51
 - Summe von, 50
- Zufallsvektor, 26