

Einführung in die Stochastik

Bergische Universität Wuppertal WiSe 2022/23

Hanno Gottschalk

G.16.01

Tel: 0202-439-2516

e-mail: hanno.gottschalk@uni-wuppertal.de

Sprechstunde: Nach Vereinbarung

Inhaltsverzeichnis

1. Kombinatorik	5
2. Wahrscheinlichkeitsräume	21
3. Bedingte Wahrscheinlichkeiten	35
4. Markov-Ketten	51
5. Unabhängigkeit	63
6. Diskrete Zufallsvariablen	71
7. Kontinuierliche Zufallsvariablen – Grundlagen	87
8. Kontinuierliche Verteilungen – Beispiele	101
9. Erwartungswert, Varianz und Momente	115
10. Erzeugende Funktionen	129
11. Multivariate Verteilungen	139
12. Unabhängige und unkorrelierte Zufallsvariablen	151
13. Faltung von Verteilungen	165
14. Gesetz der großen Zahlen	171
15. Der zentrale Grenzwertsatz	179
16. Statistische Modelle	189
17. Konfidenzintervalle	201
18. Asymptotische Konfidenzintervalle	211
19. Maximum Likelihood Schätzer	217
A. Tabellen	227

1. Kombinatorik

Am Anfang der Wahrscheinlichkeitstheorie im europäischen Kulturkreis im 16. und 17. Jahrhundert standen zum einen die Theorie des Glücksspiels und andererseits kaufmännische Fragen, bei denen es um Geschäftsrisiken geht. Besonders für die Glücksspieltheorie hat sich folgender Ansatz bewährt, der mit der symmetrischen Beschaffenheit von Würfeln oder anderen Gerätschaften zur “Produktion von Zufall” einhergeht.

Prinzip 1.1 (Prinzip des unzureichenden Grundes). Wenn man bei einem Experiment mit zufälligem Ausgang, bei dem die endlich vielen möglichen Ausgänge vorab bekannt sind, keinen Grund hat, einen Ausgang für bevorzugt oder weniger bevorzugt anzusehen, dann misst man jedem möglichen Ausgang dieselbe Wahrscheinlichkeit zu, nämlich

$$p = \frac{1}{\text{Anzahl der möglichen Ausgänge}} . \quad (1.1)$$

Bemerkung 1.2. Beim Prinzip des unzureichenden Grundes handelt es sich nicht um eine mathematische Definition, sondern um eine empirische Vorschrift wie man gewisse Sachverhalte modellieren kann. Dies ist ein eher philosophisches Prinzip und man könnte sich je nach Sachverhalt ebenfalls andere “Prinzipien” ausdenken.

Das folgende Beispiel zeigt einige Anwendungen dieses Prinzips. Beachten Sie, dass die Pfeile ‘ \Rightarrow ’ nicht für eine mathematisch-logische Schlußfolgerung stehen, sondern für die Anwendung des Prinzip des unzureichenden Grundes.

Beispiel 1.3.

- (i) Münzwurf mit einer mir unbekannten Münze mit Ausgängen $\{K, Z\}$ (bei “Kante” wird noch einmal geworfen) $\Rightarrow p = \frac{1}{2}$ sowohl für K als auch für Z
- (ii) Würfeln mit einem mir unbekannten Würfel mit Seiten $\{1, 2, 3, 4, 5, 6\} \Rightarrow p = \frac{1}{6}$ für jede Augenzahl $i \in \{1, 2, 3, 4, 5, 6\}$
- (iii) Die Frage, ob der DAX morgen steigt oder fällt, die ich mir als ökonomischer Laie stelle $\Rightarrow p = \frac{1}{2}$ sowohl für steigen als auch für fallen

Das Prinzip des unzureichenden Grundes ist die Grundlage für das folgende Modell eines Wahrscheinlichkeitsraumes.

1. Kombinatorik

Definition 1.4.

- (i) Ein **Laplacescher Wahrscheinlichkeitsraum** besteht aus einer endlichen Menge $\Omega = \{\omega_1, \dots, \omega_n\}$ und der Zuweisung einer uniformen Eintretens-Wahrscheinlichkeit $p = \frac{1}{\#\Omega} = \frac{1}{n}$ für jedes sogenannte **Elementarereignis** $\omega_i, i = 1, \dots, n$.
- (ii) Ein (möglicherweise zusammengesetztes) **Ereignis** ist eine Untermenge $A \subseteq \Omega$. Jedem solchen Ereignis wird die folgende Wahrscheinlichkeit zugewiesen:

$$P(A) = \frac{\#A}{\#\Omega} . \quad (1.2)$$

Hier steht $\#A$ für die Anzahl der Elemente in A . Beachten Sie auch, dass unsere Definition von ‘Teilmenge’ stets $\Omega \subseteq \Omega$ und $\emptyset \subseteq \Omega$ mit einschließt.

Im Folgenden identifizieren wir Elementarereignisse $\omega \in \Omega$ mit den zugehörigen einelementigen Teilmengen $\{\omega\} \subseteq \Omega$.

Beispiel 1.5.

- (i) Würfeln mit einem Würfel: Bestimme Wahrscheinlichkeit für Ereignis $A = \{2, 4, 6\} =$ ‘Augenzahl gerade’

$$P(A) = \frac{\#A}{\#\Omega} = \frac{3}{6} = \frac{1}{2} . \quad (1.3)$$

- (ii) Würfeln mit zwei unterscheidbaren Würfeln: Gefragt ist nach der Wahrscheinlichkeit, dass beide Würfel zusammen genau 7 Augen zeigen.

Lösung. Es gibt $\#\Omega = 36$ Möglichkeiten für die angezeigten Augen zweier (unterscheidbarer) Würfel

$$\Omega = \begin{array}{cccccc} \{ & (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ & (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ & (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ & (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ & (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ & (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \} \quad (1.4)$$

Genau die Elemente auf der von links unten aufsteigenden Diagonalen des oben abgebildeten Zahlenquadrates gehören zum Ereignis $A_7 = \{(a,b) \in \Omega : a+b=7\}$ für $q=7$. Offensichtlich gibt es genau 6 solcher Elemente, also

$$P(A_7) = \frac{6}{36} = \frac{1}{6} . \quad (1.5)$$

Wenn wir weiter nach den Wahrscheinlichkeiten fragen, dass die Gesamtaugenanzahl $q = 2, \dots, 12$ erreicht wird, so müssen wir entsprechend die Elemente auf den von

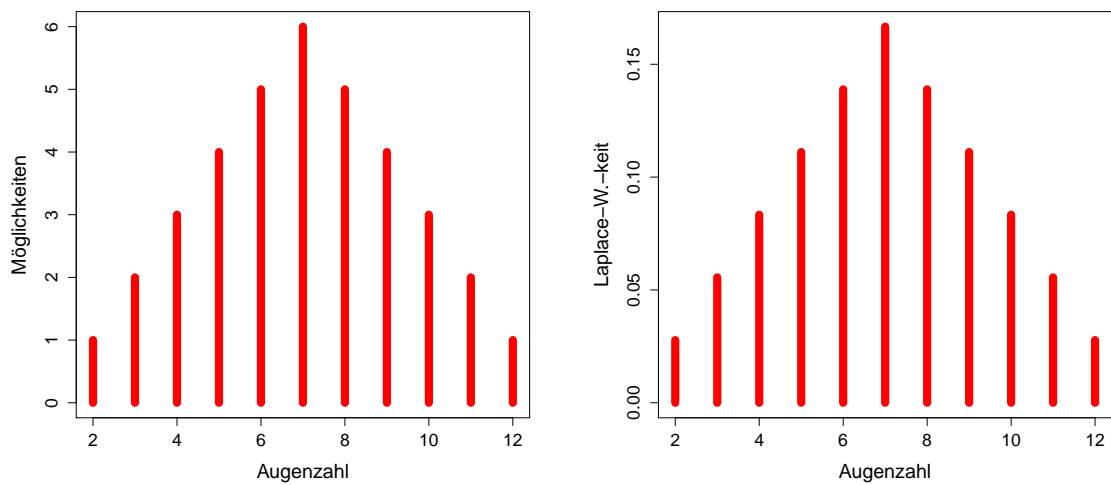


Abbildung 1.1: Anzahl der Möglichkeiten und Laplace-Wahrscheinlichkeit für die Augenzahl beim Wurf mit zwei fairen Würfeln.

links nach rechts aufsteigenden Nebendiagonalen abzählen. Da jede Nebendiagonale jeweils ein Element weniger hat, als die darunter/darüber liegende und die Anzahl der Elemente auf der ‘roten’ Hauptdiagonalen bekannt ist, kann man sich leicht überzeugen, dass

$$P(A_q) = \frac{6 - |7 - q|}{36}, \quad q = 2, \dots, 12. \quad (1.6)$$

Dies ist in Abbildung 1.1 visualisiert. \diamond

Aus Beispiel 1.5 ersieht man, dass das Abzählen von Möglichkeiten oft ein gewisses Geschick erfordert. Eine Sammlung von geschickten Abzählmethoden stellt das mathematische Feld der **Kombinatorik** zur Verfügung. Hierbei werden besonders häufig rekursive Abzählverfahren verwendet. Rekursion ist eng verbunden mit der mathematischen Methode der **vollständigen Induktion**.

Prinzip 1.6 (Prinzip der vollständigen Induktion). Es sei C_n , $n \in \mathbb{N}$, eine Folge von logischen Aussagen. Es gelte

- (I1) C_1 ist wahr (Induktionsanfang)
- (I2) Unter der Voraussetzung C_n (Induktionshypothese) gilt auch C_{n+1} (Induktionsschritt), symbolisch: $C_n \Rightarrow C_{n+1}$.

Dann gilt C_n für alle $n \in \mathbb{N}$.

Bemerkung 1.7. Das oben genannte ‘Prinzip’ ist aus den Peano-Axiomen für die natürlichen Zahlen ableitbar bzw. es ist ein Teil des Peano-Axiomensystems (\rightarrow Induktionsaxiom).

1. Kombinatorik

Satz 1.8. Es seien $\Omega_1, \dots, \Omega_k$ Mengen mit jeweils $n_1, \dots, n_k \in \mathbb{N}$ Elementen, wobei $k \in \mathbb{N}$. Dann gilt für

$$\Omega := \times_{j=1}^k \Omega_j := \{(\omega_1, \dots, \omega_k) : \omega_j \in \Omega_j, j = 1, \dots, k\} \quad (1.7)$$

die Aussage

$$\#\Omega = n_1 \cdots n_k = \prod_{j=1}^k n_j . \quad (1.8)$$

Bemerkung 1.9. In der Sprache der Kombinatorik bedeutet dieser Satz das Folgende: Es gibt $\prod_{j=1}^k n_j$ Möglichkeiten aus den Mengen $\Omega_1, \dots, \Omega_k$ jeweils ein Element auszuwählen.

Beweis von Satz 1.8. Zunächst gilt für zwei Mengen A und B mit n bzw. m Elementen, dass $\#(A \times B) = nm$. Wenn man streng vorgeht, müßte man auch dies mittels vollständiger Induktion beweisen. Wir überlassen dies der Leserin/dem Leser.

Wir beweisen nun den Satz mittels vollständiger Induktion: Für $k = 1$ (Induktionsanfang) ist wegen $\Omega = \Omega_1$ offensichtlich $\#\Omega = n_1$, wie gefordert. Als Induktionshypothese setzen wir voraus, dass

$$\# \times_{j=1}^k \Omega_j = \prod_{j=1}^k n_j . \quad (1.9)$$

Es sei nun $\Omega = \times_{j=1}^{k+1} \Omega_j = (\times_{j=1}^k \Omega_j) \times \Omega_{k+1}$. Es gilt der Induktionsschritt, da

$$\#\Omega = (\# \times_{j=1}^k \Omega_j) n_{k+1} = \left(\prod_{j=1}^k n_j \right) n_{k+1} = \prod_{j=1}^{k+1} n_j , \quad (1.10)$$

wobei wir im zweiten Schritt die Induktionshypothese verwendet haben. \square

Satz 1.10 (Auswahl ‘ k aus n ’). Es seien $k, n \in \mathbb{N}_0$ mit $k \leq n$.

(i) Die Anzahl von Möglichkeiten, k Elemente aus einer n -elementigen Menge unter Berücksichtigung der Reihenfolge auszuwählen, beträgt

$$n(n-1)(n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!} . \quad (1.11)$$

Die Fakultät ist definiert durch $n! := \prod_{j=1}^n j$, für $n \in \mathbb{N}$, und $0! = 1$.

(ii) Die Anzahl von Möglichkeiten, k Elemente aus einer n -elementigen Menge ohne Berücksichtigung der Reihenfolge zu wählen, beträgt

$$\frac{n(n-1)(n-2) \cdots (n-k+1)}{k(k-1)(k-2) \cdots 1} = \frac{n!}{(n-k)! k!} . \quad (1.12)$$

Beweis. (i) Die Menge, aus der im j -ten Schritt ein Element ausgewählt werden kann, hat jeweils die folgende Größe:

1.	Auswahlschritt	$n_1 = n$
2.	Auswahlschritt	$n_2 = n - 1$
3.	Auswahlschritt	$n_3 = n - 2$
\vdots	\vdots	\vdots
$k - 1$.	Auswahlschritt	$n_{k-1} = n - k + 2$
k .	Auswahlschritt	$n_k = n - k + 1$

Nach Satz 1.8 ergeben sich somit

$$\prod_{j=1}^k n_j = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!} \quad (1.13)$$

Auswahlmöglichkeiten.

(ii) Wenn wir nach der Auswahl von k Elementen von der Reihenfolge absehen wollen, so müssen wir alle die Auswahlen, die sich nur in der Reihenfolge der gewählten Elemente unterscheiden, als nur eine einzige Auswahlmöglichkeit ansehen. Die Anzahl von Reihenfolgen für eine gegebene Auswahl von k Elementen kann wiederum als Auswahlprozess ‘ k aus k ’ unter Berücksichtigung der Reihenfolge verstanden werden. Es ergeben sich nach (i) also $k!$ Möglichkeiten, die Reihenfolge zu wählen. Dementsprechend ist die Anzahl ‘ k aus n ’ mit Berücksichtigung der Reihenfolge um einen Faktor $k!$ größer, als die entsprechende Anzahl ohne die Berücksichtigung der Reihenfolge. Indem wir wieder (i) verwenden, erhalten wir als Anzahl von Auswahlmöglichkeiten ohne Berücksichtigung der Reihenfolge $\frac{1}{k!} \frac{n!}{(n-k)!}$. \square

Definition 1.11. Der **Binomialkoeffizient** $\binom{n}{k}$ (sprich: ‘ n über k ’) ist definiert für $n, k \in \mathbb{N}_0$, $n \geq k$, als

$$\binom{n}{k} := \frac{n!}{(n-k)!k!} . \quad (1.14)$$

Beispiel 1.12. Sie verwenden bei *facebook* ein Passwort mit 6 Zeichen aus dem Alphabet [a-z] ohne Sonderzeichen, große Zeichen und doppelte Buchstaben. Affe Fips setzt sich an den Rechner und tippt blindlings 6 solche Buchstaben ein, ohne einen Buchstaben zu wiederholen. Wie groß ist die Wahrscheinlichkeit, dass Fips auf Anhieb Ihr Passwort knackt?

Lösung. Bei diesem Problem zählt die Reihenfolge. Also gibt es nach Satz 1.10 (i) $26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 = 165765600$ mögliche Passwörter. Da Fips blindlings tippt, können wir mit einer gewissen Berechtigung nach dem Prinzip des unzureichenden Grundes vorgehen und die Wahrscheinlichkeit als

$$p = \frac{1}{165765600} \quad (1.15)$$

1. Kombinatorik

bestimmen. Wäre Fips ein Hacker-Programm, dass um die 7000000 Möglichkeiten pro Sekunde austestet (das geht ohne Probleme mit einem Quad-Core aus dem Jahre 2013), so wäre Ihr Passwort allerdings spätestens nach wenigen Sekunden geknackt. \diamond

Der Philosoph und Mathematiker Blaise Pascal (1623–1662) fand folgendes Schema zur Darstellung der Binomialkoeffizienten $\binom{n}{k}$:

$$\begin{array}{c|cccccc}
 n & & & & & & \\
 0 & & & & & & 1 \\
 1 & & & 1 & & 1 & \\
 2 & & 1 & 2 & 1 & & \\
 3 & 1 & 3 & 3 & 1 & & \\
 4 & 1 & 4 & 6 & 4 & 1 & \\
 \dots & & & \text{usw.} & & &
 \end{array} \tag{1.16}$$

Hier wird k in jeder Zeile von links nach rechts durchgezählt, beginnend mit $k = 0$ und endend mit $k = n$. Die Außenseiten des **Pascalschen Dreiecks** werden jeweils durch 1 gebildet, die inneren Werte erhält man, indem man den Wert links oberhalb und rechts oberhalb der jeweiligen Zahl addiert. Das Pascalsche Dreieck können wir also als Visualisierung folgender Rekursionsformel verstehen.

Proposition 1.13. *Es gilt, für alle $k, n \in \mathbb{N}_0$ mit $n \geq 1$ und $1 \leq k \leq n - 1$,*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}. \tag{1.17}$$

Beweis. Der Beweis folgt unmittelbar aus folgender Rechnung,

$$\begin{aligned}
 \binom{n-1}{k-1} + \binom{n-1}{k} &= \frac{(n-1)!}{(n-1-(k-1))!(k-1)!} + \frac{(n-1)!}{(n-1-k)!k!} \\
 &= \frac{(n-1)!}{(n-1-k)!(k-1)!} \left(\frac{1}{n-k} + \frac{1}{k} \right) \\
 &= \frac{(n-1)!}{(n-1-k)!(k-1)!} \left(\frac{k+n-k}{(n-k)k} \right) \\
 &= \frac{(n-1)!n}{(n-1-k)!(n-k)(k-1)!k} = \frac{n!}{(n-k)!k!} = \binom{n}{k}. \quad \square
 \end{aligned} \tag{1.18}$$

Bemerkung 1.14. Ein kombinatorischer Beweis der Rekursionsformel aus Proposition 1.13 kann wie folgt gegeben werden: Eine Menge von k Elementen gewählt aus einer n -elementigen Menge $\{x_1, \dots, x_n\}$ kann dahingehend unterschieden werden, ob x_n in dieser Menge enthalten ist, oder nicht. Eine k -elementige Teilmenge von $\{x_1, \dots, x_n\}$, die x_n nicht enthält, muss man aus der Menge $\{x_1, \dots, x_{n-1}\}$ wählen. Von diesen Mengen gibt es $\binom{n-1}{k}$ verschiedene. Eine k -elementige Menge aus $\{x_1, \dots, x_n\}$, die x_n enthält, kann man durch freie Wahl von $k-1$ Elementen aus $\{x_1, \dots, x_{n-1}\}$ konstruieren. Von dieser Sorte gibt es

$\binom{n-1}{k-1}$ verschiedene. Da genau eine der beschriebenen Möglichkeiten zutreffen muss, ergibt sich die Rekursionsformel.

Man sieht also, dass die algebraische Rechnung und das kombinatorische Argument oft zwei Alternativen darstellen, von denen man eine zum Beweis eines Sachverhaltes auswählen kann. Kommt man z.B. mit einer algebraischen Identität mit ganzen Zahlen nicht weiter, könnte man nach einem äquivalenten kombinatorischen Problem Ausschau halten. Dieses Vorgehen – Beweis durch innermathematische Übersetzung in ein äquivalentes Problem, das der Mathematikerin/dem Mathematiker eine andere Intuition erlaubt – ist sehr typisch für die Mathematik.

Wir betrachten nun eine wichtige Anwendung der Binomialkoeffizienten.

Proposition 1.15 (Binomialformel). *Es seien $x, y \in \mathbb{K}$ (mit $\mathbb{K} = \mathbb{R}$ oder \mathbb{C}). Dann gilt, für alle $n \in \mathbb{N}$,*

$$(x + y)^n = \sum_{j=0}^n \binom{n}{j} x^j y^{n-j} . \quad (1.19)$$

Beweis. Wir geben einen Beweis durch vollständige Induktion. Der Induktionsanfang ($n = 1$) ist erfüllt, da

$$(x + y)^1 = x + y = \underbrace{\binom{1}{0}}_{=1} x^0 y^1 + \underbrace{\binom{1}{1}}_{=1} x^1 y^0 = \sum_{j=0}^1 \binom{1}{j} x^j y^{1-j} . \quad (1.20)$$

Aus der Induktionshypothese (1.19) folgt der Induktionsschritt, da

$$\begin{aligned} (x + y)^{n+1} &= (x + y) (x + y)^n = x (x + y)^n + y (x + y)^n \\ &= \sum_{j=0}^n \binom{n}{j} x^{j+1} y^{n-j} + \sum_{j=0}^n \binom{n}{j} x^j y^{n-j+1} \\ &= \sum_{j=1}^{n+1} \binom{n}{j-1} x^j y^{n+1-j} + \sum_{j=0}^n \binom{n}{j} x^j y^{n+1-j} \\ &= \underbrace{\binom{n}{n}}_{=1=\binom{n+1}{n+1}} x^{n+1} y^0 + \sum_{j=1}^n \underbrace{\left[\binom{n}{j-1} + \binom{n}{j} \right]}_{=\binom{n+1}{j}} x^j y^{n+1-j} + \underbrace{\binom{n}{0}}_{=1=\binom{n+1}{0}} x^0 y^{n+1} \\ &= \sum_{j=0}^{n+1} \binom{n+1}{j} x^j y^{n+1-j} . \end{aligned} \quad (1.21)$$

Im Schritt $*$ haben wir die Rekursionsformel aus Proposition 1.13 verwendet. \square

Bemerkung 1.16. Auch hier gibt es wieder eine alternative, kombinatorische Beweisidee: Wir schreiben

$$(x + y)^n = \underbrace{(x + y) \cdots (x + y)}_{n \text{ Mal}} . \quad (1.22)$$

1. Kombinatorik

Wollen wir nun nach dem Ausmultiplizieren die Summanden zählen, in denen die Potenz x^j auftritt, so müssen wir aus den obigen n Faktoren genau j auswählen, von denen jeweils ein x her stammt. Für diese Wahl gibt es nach Satz 1.10 (ii) $\binom{n}{j}$ Möglichkeiten. Also muss dies der Vorfaktor der Potenz $x^j y^{n-j}$ sein.

Der Binomialkoeffizient hat die folgende Verallgemeinerung.

Definition 1.17. Es seien $n_1, \dots, n_k \in \mathbb{N}_0$ und $n = \sum_{j=1}^k n_j$, wobei $k \in \mathbb{N}$. Der **Multinomialkoeffizient** (sprich: ‘ n über n_1, \dots, n_k ’) ist definiert als

$$\binom{n}{n_1, \dots, n_k} := \frac{n!}{n_1! \dots n_k!} . \quad (1.23)$$

Der Multinomialkoeffizient findet Anwendung in folgendem kombinatorischen Problem.

Satz 1.18. Es seien $n_1, \dots, n_k \in \mathbb{N}_0$ und $n = \sum_{j=1}^k n_j$, wobei $k \in \mathbb{N}$. Dann gibt es $\binom{n}{n_1, \dots, n_k}$ Möglichkeiten n Elemente in k Gruppen ohne Berücksichtigung der Reihenfolge innerhalb der Gruppen aufzuteilen, wobei die j -te Gruppe n_j Elemente hat, für alle $j = 1, \dots, k$.

Beweis. Wir wählen sukzessive die Elemente für die erste, zweite, \dots , k -te Gruppe aus und bestimmen die Anzahl der Möglichkeiten:

1.	Auswahlschritt	$\binom{n}{n_1}$	Möglichkeiten
2.	Auswahlschritt	$\binom{n-n_1}{n_2}$	Möglichkeiten
3.	Auswahlschritt	$\binom{n-n_1-n_2}{n_3}$	Möglichkeiten
\vdots	\vdots	\vdots	\vdots
$k-1$.	Auswahlschritt	$\binom{n-n_1-n_2-\dots-n_{k-2}}{n_{k-1}}$	Möglichkeiten
k .	Auswahlschritt	$\binom{n-n_1-n_2-\dots-n_{k-1}}{n_k}$	Möglichkeiten

Es ergibt sich für die Anzahl der Gesamtmöglichkeiten

$$\begin{aligned} & \binom{n}{n_1} \binom{n-n_1}{n_2} \dots \binom{n-n_1-n_2-\dots-n_{k-1}}{n_k} \\ &= \frac{n!}{n_1! (n-n_1)!} \frac{(n-n_1)!}{n_2! (n-n_1-n_2)!} \dots \frac{(n-n_1-\dots-n_{k-1})!}{n_k! \underbrace{(n-n_1-n_2-\dots-n_k)!}_{=0!=1}} \\ &= \frac{n!}{n_1! n_2! \dots n_k!} = \binom{n}{n_1, \dots, n_k} . \end{aligned} \quad (1.24)$$

In der mittleren Zeile haben wir verwendet, dass sich die rot markierten Faktoren $(n-n_1-\dots-n_j)!$ für $j = 1, \dots, k-1$ durch Teleskopkürzung heraus heben. \square

Es gibt noch sehr viele weitere interessante kombinatorische Zahlen. Stellvertretend für vieles Andere möchten wir hier auf die Kombinatorik der Partitionen eingehen, bei denen zwar die Anzahl k der Gruppen, nicht aber deren Größe vorgeschrieben ist. Dafür sollen alle n Elemente der ursprünglichen Menge verteilt werden.

Definition 1.19.

- (i) Die Anzahl der Möglichkeiten, eine n -elementige Menge in k nicht leere Gruppen aufzuteilen, wobei die Reihenfolge der k Gruppen keine Rolle spielt, heißt $B_{n,k}$, die **Stirlingsche Zahl** zu n und k . Es gilt

$$B_{n,k} := \sum_{\substack{n_1, \dots, n_k \in \mathbb{N} \\ n_1 + \dots + n_k = n}} \frac{1}{k!} \binom{n}{n_1, \dots, n_k}. \quad (1.25)$$

- (ii) Die Anzahl der Möglichkeiten, eine n -elementige Menge in eine beliebige Anzahl von nicht leeren Gruppen einzuteilen, heißt **Bellsche Zahl** B_n . Es gilt

$$B_n := \sum_{k=1}^n B_{n,k}. \quad (1.26)$$

Das Aufteilen einer Menge in k nicht leere Gruppen wird durch den Begriff der Partition formalisiert.

Definition 1.20. Eine Menge $I = \{I_1, \dots, I_k\}$ von Teilmengen von Ω , $I_j \subseteq \Omega$, $j = 1, \dots, k$, $k \in \mathbb{N}$, heißt **k -Partition** von Ω , falls

- (i) $I_j \neq \emptyset$, für alle $j = 1, \dots, k$,
- (ii) $I_i \cap I_j = \emptyset$, für alle $i, j = 1, \dots, k$, $i \neq j$,
- (iii) $\bigcup_{j=1}^k I_j = \Omega$.

Es sei $\mathcal{P}^{(k)}(\Omega)$ die Menge aller k -Partitionen und $\mathcal{P}(\Omega) := \bigcup_{k=1}^n \mathcal{P}^{(k)}(\Omega)$ die Menge aller Partitionen.

Lemma 1.21. Es gilt $\#\mathcal{P}^{(k)}(\Omega) = B_{n,k}$ und $\#\mathcal{P}(\Omega) = B_n$ für $\#\Omega = n$, $k \in \mathbb{N}$, $k \leq n$.

Beweis. Dies folgt direkt aus der Definition der Stirling- bzw. Bell-Zahlen. □

Partitionen (und damit die Bellschen Zahlen) treten an allen Ecken und Enden der Mathematik und ihrer Anwendung in den Naturwissenschaften auf. Ersteres gilt etwa für die Zahlentheorie, Hopf-Algebren oder auch für den zentralen Grenzwertsatz (darauf kommen wir noch zurück). In den Naturwissenschaften sind es z.B. die statistische Mechanik oder die Quantenfeldtheorie, in welchen Partitionen eine wichtige Rolle spielen.

Ein wichtiges und relativ einfaches Beispiel für das Auftreten von Partitionen ist die verallgemeinerte Kettenregel, auch bekannt als die Formel von Faà di Bruno. Diese geht auf den italienischen Mathematiker und Geistlichen Francesco Faà di Bruno (1825–1888) zurück.

1. Kombinatorik

Satz 1.22 (Verallgemeinerte Kettenregel). *Es seien $f: \mathbb{R} \rightarrow \mathbb{R}$ und $g: \mathbb{R}^d \rightarrow \mathbb{R}$ beliebig oft differenzierbare Funktionen und $f \circ g: \mathbb{R}^d \rightarrow \mathbb{R}$ die Verknüpfung von f und g . $f^{(k)}$ bezeichne die k -te Ableitung von f . Für $(\alpha_1, \dots, \alpha_n) \in \{1, \dots, d\}^{\times n}$ und $I_j = \{\beta_1, \dots, \beta_q\} \subseteq \{1, \dots, n\}$ schreibe*

$$\frac{\partial^{I_j} g}{\partial x^{I_j}} = \frac{\partial^q g}{\partial x^{\alpha_{\beta_1}} \dots \partial x^{\alpha_{\beta_q}}} . \quad (1.27)$$

Dann gilt

$$\frac{\partial^n f \circ g}{\partial x^{\alpha_1} \dots \partial x^{\alpha_n}} = \sum_{k=1}^n f^{(k)} \circ g \sum_{I \in \mathcal{P}^{(k)}(\{1, \dots, n\})} \prod_{j=1}^k \frac{\partial^{I_j} g}{\partial x^{I_j}} . \quad (1.28)$$

Bemerkung 1.23. Bevor wir Satz 1.22 beweisen, möchten wir einige Spezialfälle betrachten: Ist z.B. $f(t) = \exp(t)$, so ist $f^{(k)}(t) = f(t)$ und die Formel (1.28) vereinfacht sich zu

$$\frac{\partial^n \exp(g)}{\partial x^{\alpha_1} \dots \partial x^{\alpha_n}} = \exp(g) \sum_{I \in \mathcal{P}(\{1, \dots, n\})} \prod_{j=1}^k \frac{\partial^{I_j} g}{\partial x^{I_j}} . \quad (1.29)$$

Ist nun g eine quadratische Funktion, d.h. $g(x) = \frac{1}{2} x' A x$ mit A einer symmetrischen $d \times d$ -Matrix, so ist $\frac{\partial^{I_j} g}{\partial x^{I_j}} \upharpoonright_{x=0} = 0$ falls $\#I_j \neq 2$ und $\frac{\partial^{I_j} g}{\partial x^{I_j}} \upharpoonright_{x=0} = A_{i,k}$ falls $I_j = \{i, k\}$. Dementsprechend vereinfacht sich der obige Ausdruck für $x = 0$ für ungerade n zu 0 und für gerade n zu

$$\frac{\partial^n \exp(\frac{1}{2} x' A x)}{\partial x^{\alpha_1} \dots \partial x^{\alpha_n}} \upharpoonright_{x=0} = \sum_{\substack{I \in \mathcal{P}_{\text{Paar}}(\{1, \dots, n\}) \\ I = \{\{i_1, j_1\}, \dots, \{i_{\frac{n}{2}}, j_{\frac{n}{2}}\}\}}} \prod_{q=1}^{\frac{n}{2}} A_{i_q, j_q} . \quad (1.30)$$

$\mathcal{P}_{\text{Paar}}(\Omega)$ sind die Einteilungen von Ω in Paare. Die Formel (1.30) wird uns zur Berechnung der sogenannten Momente für Gaussverteilungen nützen. In der Elementarteilchenphysik ist (1.30) als Satz von Wick bekannt und ist ein wesentlicher Baustein für die Berechnung der Streuquerschnitte in Reaktionen von Elementarteilchen.

Beweis von Satz 1.22. Der Induktionsanfang für $n = 1$ ist einfach die Leibnizregel. Unter Verwendung der Induktionshypothese (1.28) kann der Induktionsschritt wie folgt vollzogen werden:

$$\begin{aligned} \frac{\partial^{n+1} f \circ g}{\partial x^{\alpha_1} \dots \partial x^{\alpha_{n+1}}} &= \frac{\partial}{\partial x^{\alpha_{n+1}}} \frac{\partial^n f \circ g}{\partial x^{\alpha_1} \dots \partial x^{\alpha_n}} \\ &= \frac{\partial}{\partial x^{\alpha_{n+1}}} \left(\sum_{k=1}^n f^{(k)} \circ g \sum_{I \in \mathcal{P}^{(k)}(\{1, \dots, n\})} \prod_{j=1}^k \frac{\partial^{I_j} g}{\partial x^{I_j}} \right) \\ &= \sum_{k=1}^n f^{(k+1)} \circ g \sum_{I \in \mathcal{P}^{(k)}(\{1, \dots, n\})} \frac{\partial g}{\partial x^{\alpha_{n+1}}} \prod_{j=1}^k \frac{\partial^{I_j} g}{\partial x^{I_j}} \\ &\quad + \sum_{k=1}^n f^{(k)} \circ g \sum_{I \in \mathcal{P}^{(k)}(\{1, \dots, n\})} \sum_{q=1}^k \frac{\partial^{I_q \cup \{n+1\}} g}{\partial x^{I_q \cup \{n+1\}}} \prod_{\substack{j=1 \\ j \neq q}}^k \frac{\partial^{I_j} g}{\partial x^{I_j}} . \end{aligned} \quad (1.31)$$

Der erste Term auf der rechten Seite kann aufgefaßt werden als induziert durch eine $k + 1$ -Partition, bei der $I_{k+1} = \{n + 1\}$. Somit enthält er alle Terme, die von Partitionen induziert werden, in denen $n + 1$ alleine eine Gruppe ausmacht. Ist jedoch $n + 1$ nicht alleine in einer Gruppe, so muss es zwangsläufig in eine Gruppe einer Partition von $\{1, \dots, n\}$ hinzugefügt werden. Gibt es k Gruppen, so können wir für $q = 1, \dots, k$ eine Partition I_q auswählen, zu der wir $n + 1$ hinzufügen möchte. Genau diese Möglichkeiten treten jedoch im zweiten Term auf der rechten Seite auf. Also finden wir

$$\frac{\partial^{n+1} f \circ g}{\partial x^{\alpha_1} \dots \partial x^{\alpha_{n+1}}} = \sum_{k=1}^{n+1} f^{(k)} \circ g \sum_{I \in \mathcal{P}^{(k)}(\{1, \dots, n+1\})} \prod_{j=1}^k \frac{\partial^{I_j} g}{\partial x^{I_j}}. \quad (1.32)$$

□

Korollar 1.24. Ist in Satz 1.22 $d = 1$, d.h. $g : \mathbb{R} \rightarrow \mathbb{R}$, dann gilt

$$(f \circ g)^{(n)} = \sum_{k=1}^n f^{(k)} \circ g \sum_{\substack{n_1, \dots, n_k \in \mathbb{N} \\ n_1 + \dots + n_k = n}} \frac{1}{k!} \binom{n}{n_1, \dots, n_k} \prod_{j=1}^k g^{(n_j)}. \quad (1.33)$$

Beweis. Für eine Partition $I = \{I_1, \dots, I_k\}$ mit $\#I_j = n_j$ gilt $\frac{\partial^{I_j} g}{\partial x^{I_j}} = g^{(n_j)}$ unabhängig von den Werten in I_j . Wir können also in (1.28) alle Partitionen $I \in \mathcal{P}^{(k)}(\{1, \dots, n\})$ mit gleicher Gruppengröße n_1, \dots, n_k zusammenfassen. Von diesen gibt es nach Satz 1.18 gerade $\frac{1}{k!} \binom{n}{n_1, \dots, n_k}$ viele. Somit folgt (1.33) aus (1.28). □

Es ist der interessierten Leserin/dem interessierten Leser überlassen, die Formel (1.33) für den Fall $g(x) = \exp(\lambda x)$ weiter zu vereinfachen.

1.1. Übungsaufgaben

1. Selbsttest

1. Denken Sie sich weitere Beispiele für Laplacesche Wahrscheinlichkeitsräume aus! Spezifizieren Sie hierzu die Menge Ω der Elementarereignisse und geben Sie eine Begründung warum die Anwendung des Prinzips des unzureichenden Grundes für Ihre Beispiele sinnvoll ist.
2. Ein Laplacescher Wahrscheinlichkeitsraum Ω hat 10 Elemente und eine Teilmenge $A \subseteq \Omega$ hat 3 Elemente. Wie groß ist die Wahrscheinlichkeit $P(A)$?
3. In der Antike wurden oft *vierseitige* Würfel verwendet. In einem Spiel wird mit zwei fairen vierseitigen Würfeln gewürfelt und das Ergebnis addiert. Welche Zahl tritt am Häufigsten auf?
4. Wie viele Möglichkeiten gibt es 12 SpielerInnen in zwei gleich große Teams – genannt Team A und Team B – einzuteilen? Beide Teams sollen unterscheidbar sein, d.h. wenn dieselben SpielerInnen in Team A bzw. Team B spielen, zählt das als zwei verschiedene Möglichkeiten.
5. Nehmen wir an, dass Sie sich 3 Stochastikbücher gekauft haben. Zuhause haben Sie noch weitere 2 Mathematikbücher aus einer anderen Vorlesung.
 - a) Wieviele Möglichkeiten haben Sie die 5 Bücher in Ihrem Regal anzuordnen? Geben Sie eine kurze Erklärung!
 - b) Um beim Lernen nicht durcheinander zu kommen, möchten Sie nun die Stochastikbücher im linken Teil des Regals anordnen und die anderen Bücher im rechten Teil. Wieviele Möglichkeiten haben Sie nun?

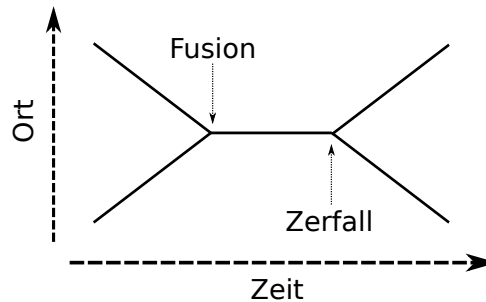
Nun das ganze etwas abstrakter: Sie besuchen $N \in \mathbb{N}$ Vorlesungen. Für die Vorlesung $i \in \{1, 2, \dots, N\}$ haben Sie $n_i \in \mathbb{N}_0$ Bücher.

- c) Wieviele Möglichkeiten haben Sie alle Ihre Bücher im Regal anzuordnen, wenn Sie wie in Aufgabenteil a) vorgehen?
- d) Wieviele Möglichkeiten sind es, wenn Sie die Bücher für die unterschiedlichen Fächer in separate Teile des Regals einordnen möchten?

2. Transfer

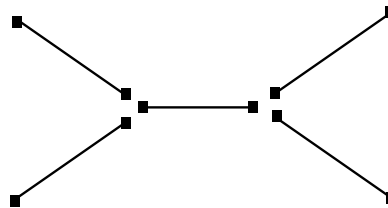
Das Higgs-Boson ist seit seiner Entdeckung im Juli 2012 am Large Hadron Collider (LHC) am CERN in Genf in aller Munde. Hierbei handelt es sich um ein Elementarteilchen, welches durch seine besonderen Eigenschaften den subatomaren Teilchen, z.B. den Elektronen, Myonen und Quarks, ihre Masse verleiht. Für ihre theoretischen Beiträge zum Higgs-Boson wurden im Jahr 2013 François Englert und Peter Higgs mit dem Nobelpreis für Physik ausgezeichnet.

Für die Berechnung der Reaktionsraten in den Streuexperimenten am LHC verwenden theoretische Physiker sogenannte Feynman-Graphen und Feynman-Regeln. Die Feynman-Graphen visualisieren eine Teilchenreaktion, z.B. beschreibt der folgende Graph die Fusion zweier Teilchen in ein Teilchen, welches danach wieder in 2 Teilchen zerfällt:



Die Feynman-Regeln ordnen solchen Graphen mathematische Formeln zu, mit welchen sich die Reaktionsraten berechnen lassen. Dies werden wir im folgenden nicht näher betrachten.

Wir möchten nun ein kombinatorisches Problem betrachten in dem es darum geht solche Feynman-Graphen abzuzählen. Dazu beobachten wir zuerst, dass wenn man die Reaktionspunkte im obigen Graphen etwas “auseinanderschiebt” man folgendes Bild erhält:



Somit erhält man diesen Feynman-Graphen indem man 10 Punkte in gewisse Paare einteilt und diese mit Linien verbindet.

- Wie viele Möglichkeiten gibt es $2n$ Punkte in Paare einzuteilen, wobei $n \in \mathbb{N}$?
- Wie viele Möglichkeiten sind es, wenn wir die Reihenfolge in der die Paarung durchgeführt wird nicht berücksichtigen?
- Zeichnen Sie ähnlich wie im zweiten Bild alle möglichen Paarungen für $n = 2$, i.e. für $2n = 4$ Punkte.
- Unter Annahme der Abzählweise aus Teil b), wie viele verschiedene Paarungen (Feynman-Graphen) muss sich eine Physikerin/ein Physiker anschauen, wenn sie/er eine Reaktion mit 6 äußeren Punkten und zwei 4-valenten inneren Vertices berechnen möchte?

Hinweis: In unserem Abzählverfahren kommt es nur auf die Gesamtanzahl der Punkte an, also auf die Summe der äußeren Punkte und der inneren Punkte in den Vertices. Im obigen Fall sind dies $6 + 4 + 4 = 14$.

3. Theorie

1. Sei $\Omega_0 := \emptyset$ die leere Menge und, für $n \in \mathbb{N}$, $\Omega_n := \{1, 2, \dots, n\}$ die n -elementige Menge. Die Potenzmenge $\mathbb{P}(\Omega)$ einer Menge Ω sei definiert als “die Menge aller Teilmengen von Ω ”, d.h. $\mathcal{P}(\Omega) := \{A : A \subseteq \Omega\}$. Es gilt zum Beispiel $\mathbb{P}(\Omega_0) = \{\emptyset\}$, $\mathbb{P}(\Omega_1) = \{\emptyset, \{1\}\}$ und $\mathbb{P}(\Omega_2) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. Die Mächtigkeit $|\Omega| \in \mathbb{N}_0$ einer endlichen Menge Ω ist definiert als die Anzahl der enthaltenen Elemente. Beweisen Sie mittels vollständiger Induktion, dass für alle $n \in \mathbb{N}_0$ gilt $|\mathbb{P}(\Omega_n)| = 2^n$.
2. In Satz 1.10 wurde die Anzahl der Möglichkeiten für die Auswahl ‘ k aus n ’ *ohne* Zurücklegen bestimmt. Mit Berücksichtigung der Reihenfolge gibt es $\frac{n!}{(n-k)!}$ Möglichkeiten und ohne Berücksichtigung der Reihenfolge sind es $\binom{n}{k}$. Ziel dieser Aufgabe ist es den Fall ‘ k aus n ’ *mit* Zurücklegen zu studieren.
 - a) Betrachten Sie zuerst den Fall $n=3$ und $k=2$. Wieviele Möglichkeiten gibt es, 2 Elemente aus einer 3-elementigen Menge mit Zurücklegen und Berücksichtigung der Reihenfolge auszuwählen? Wieviele Möglichkeiten sind es, wenn wir die Reihenfolge nicht berücksichtigen? Zählen Sie nach!
 - b) Versuchen Sie nun eine allgemeine Formel für die Auswahl ‘ k aus n ’ mit Zurücklegen und Berücksichtigung der Reihenfolge abzuleiten. Geben Sie nicht nur die Formel, sondern auch eine Erklärung an!
 - c) Lösen Sie auch das allgemeine Problem im Falle, dass die Reihenfolge nicht berücksichtigt wird.
 - d) Überprüfen Sie die Ihre Formeln aus b) und c) mit den in a) bestimmten Möglichkeiten für $n=3$ und $k=2$!
3. In Definition 1.19 wurden die Bellschen Zahlen B_n , $n \in \mathbb{N}$, als die Anzahl der möglichen Partitionen einer n -elementigen Menge definiert. Wir definieren weiter $B_0 = 1$. Im folgenden möchten wir in mehreren Schritten die Rekursionsformel

$$B_{n+1} = \sum_{k=0}^n \binom{n}{k} B_k$$

beweisen. Wir wählen hierzu einen kombinatorischen Beweis. Wenn Sie lieber einen algebraischen Beweis angeben möchten, dann können Sie das auch tun.

- 1.) Zerlegen Sie eine n -elementige Menge in disjunkte Teilmengen mit k und $n - k$ Elementen. Wie viele Möglichkeiten gibt es dafür?
- 2.) Sie fügen nun das $n + 1$ -te Element zu der $n - k$ -elementigen Teilmenge hinzu. Wie viele Partitionen der $n + 1$ -elementigen Menge gibt es, die genau diese $n - k + 1$ -elementige Teilmenge enthalten?
- 3.) Geben Sie ein Argument, wie man aus 1.) und 2.) die Rekursionsformel beweisen kann!

4. Aktivität

Eine empirische Tatsache, welche unter dem Namen Benfordsches Gesetz bekannt ist, zeigt auf, dass die Anwendung des Prinzips des unzureichenden Grundes nicht immer zu der korrekten Modellierung führt. Betrachtet man aus empirischen Datensätzen die erste Ziffer der darin enthaltenen Zahlen, so könnte man vermuten, dass aufgrund der vermeintlichen Symmetrie der Zahlen $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ jede Zahl mit Wahrscheinlichkeit $p = \frac{1}{9}$ auftritt. Dies ist jedoch im Allgemeinen nicht der Fall und es stellt sich heraus, dass die 1 häufiger auftritt als z.B. die 9. Informieren Sie sich über die Gründe für dieses Verhalten und testen Sie das Benfordsche Gesetz experimentell, z.B. mit wahllos der Tageszeitung entnommenen Zahlen.

1. Kombinatorik

1.2. Literatur

Die Kombinatorik als Abzählen von gleich wahrscheinlichen Möglichkeiten ist eng mit dem Briefwechsel zwischen Blaise Pascal und Pierre Fermat verbunden, siehe z.B. [Dev09]. Hier findet man auch das Pascalsche Dreieck. Die früheste bekannte Darstellung im Kontext der Arithmetik und der Binomischen Formel finden sich in Indien bereits im 2. Jhdt. a.D. . Weitere bekannte Nennungen entstammen der arabisch-persischen Mathematik etwa durch Al Karaji (953-1029) und Omar Kayyam (1084-1131), sowie in China im 13. Jhdt. durch Yang Huis. Im Europäischen Raum waren die Binomialkoeffizienten und ihre Berechnung z.B. dem deutschen Mathematiker Peter Albian schon bekannt, der das 'Pascalsche' Dreieck gegen 1531 auf dem Titel eines Wirtschaftsmathematischen Buches veröffentlichte. Die Zuschreibung an den Wiederentdecker Pascal erfolgte ein knappes Jahrhundert nach der Veröffentlichung von Pascal im Jahre 1555 durch die französischen Mathematiker P.-R. de Monfort und A. de Moivre.

Die explizite Formulierung der Laplace-Wahrscheinlichkeit als gleiche Wahrscheinlichkeit für alle in Frage kommenden Möglichkeiten war schon G. F. Leibnitz geläufig und ist konzeptionelle Grundlage in J. Bernoullis Buch *ars conjugandi* [Ber99], vgl. [Hac07, Kapitel 14] für weitere historische Kommentare.

Die verallgemeinerte Kettenregel, bekannt als Faà di Brunos Formel, ist benannt nach dem italienischen Priester und Mathematiker F. Faà di Bruno, der sie 1855 publizierte [Fra55], jedoch findet sich diese Formel auch schon in dem Lehrbuch von L. F. A. Arbogast aus dem Jahr 1800, [Arb00].

Die heute ist die Kombinatorik ein eigenständiger Zweig der Mathematik, bei dem in Folge des Einflusses von G.-C. Rota (1932-1999) insbesondere algebraische Methoden eine große Rolle spielen [KRY09]. Eine aktuelle Einführung in das Gebiet der Kombinatorik wird z.B. in [JJ04] gegeben.

2. Wahrscheinlichkeitsräume

Die Theorie der Laplaceschen Wahrscheinlichkeitsräume ist unzureichend, wenn wir auch Zufallsexperimente mit kontinuierlichem Ausgang oder mit unendlich vielen, aber abzählbaren, Alternativen modellieren möchten.

Beispiel 2.1. Folgende Zufallsexperimente lassen sich nicht mit Laplaceschen Wahrscheinlichkeitsräumen modellieren:

- (i) Anzahl von Würfeln mit einem Würfel, bis eine 6 kommt (diskret, aber unendlich viele Alternativen)
- (ii) Eintrittszeitpunkt des nächsten Erdbebens in einer bestimmten Region (Zeit vergeht kontinuierlich)

Bei abzählbar unendlich vielen möglichen Ausgängen ist offensichtlich die Laplace-Hypothese der gleichen Wahrscheinlichkeit für jeden denkbaren Ausgang nicht sinnvoll, da für jedes Elementarereignis die Wahrscheinlichkeit $p = 0$ herauskäme, was keine Information über die Wahrscheinlichkeiten allgemeiner Ereignisse beinhaltet. Dies kann jedoch repariert werden, indem man für die Wahrscheinlichkeit der unterschiedlichen Elementarereignisse unterschiedliche Werte zulässt. Bei kontinuierlichen Ausgängen kann man darüber hinaus die Wahrscheinlichkeit von zusammengesetzten Ereignissen – etwa dem Eintritt des nächsten Erdbebens im Laufe des nächsten Jahres – nicht auf die Wahrscheinlichkeiten der Elementarereignisse zurückführen, da die Wahrscheinlichkeit, dass z.B. ein Erdbeben genau zu einem beliebig genau spezifizierten Zeitpunkt $t \in \mathbb{R}$ eintritt, stets Null ist. Somit ist klar, dass wir einen flexibleren Rahmen für die Wahrscheinlichkeitstheorie benötigen, wenn wir nicht jeden Einzelfall gesondert behandeln möchten. Dieses Problem wurde von David Hilbert (1862–1943) in einer berühmten Rede im Jahr 1900 zu einem Teilproblem der 23 'Jahrhundertprobleme' des 20. Jahrhunderts erhoben (6. Problem).

In der Mathematik wird allgemein anerkannt, dass dieser Rahmen 1933 durch die Axiome des russischen Mathematikers A. N. Kolmogorov (1903–1987) gesteckt wurde. Kolmogorov begriff die Wahrscheinlichkeitstheorie als Spezialfall der Maß- und Integrationstheorie, die in unserem Rahmen leider nicht ausführlich abgehandelt werden kann.

Definition 2.2. Ein **messbarer Raum** (Ω, \mathcal{A}) besteht aus einer Menge Ω und einer σ -**Algebra** \mathcal{A} , d.h. einem Mengensystem \mathcal{A} von Teilmengen von Ω mit den folgenden Eigenschaften:

- (i) $\Omega \in \mathcal{A}$

2. Wahrscheinlichkeitsräume

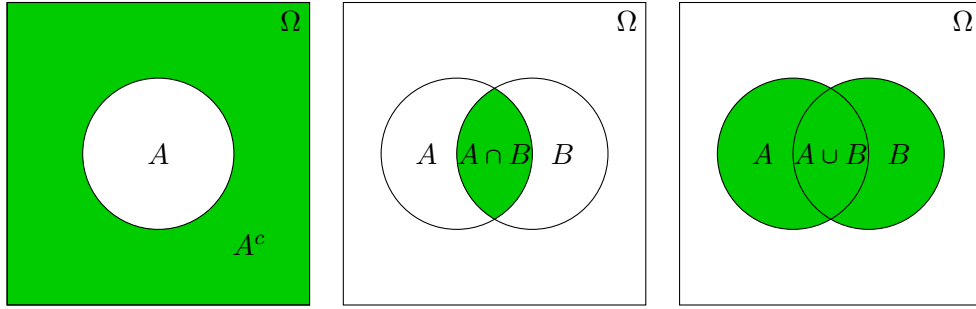


Abbildung 2.1: Venn Diagramme für Komplement A^c , Schnitt $A \cap B$ und Vereinigung $A \cup B$.

- (ii) $A \in \mathcal{A} \Rightarrow A^c = \Omega \setminus A \in \mathcal{A}$
- (iii) $A_j \in \mathcal{A}, j \in \mathbb{N} \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$

Beispiel 2.3.

- (i) Die kleinstmögliche σ -Algebra über einer Menge Ω ist $\{\emptyset, \Omega\}$.
- (ii) Die größte denkbare σ -Algebra ist die Potenzmenge von Ω , also $\mathbb{P}(\Omega) = \{A : A \subseteq \Omega\}$ die Mengen aller Teilmengen von Ω inklusive der unechten Teilmengen Ω und \emptyset .

Die Interpretation der Ω -Teilmengen in \mathcal{A} ist die von zufallsabhängigen Ereignissen, die eintreten können, oder auch nicht. Ist A ein Ereignis, so steht das Komplement A^c für das Gegenereignis 'A tritt nicht ein'. Ω ist das **sichere Ereignis** und $\emptyset = \Omega^c$ das **unmögliche Ereignis**.

Ein Ereignis in Beispiel 2.1 (ii) ist z.B. $A =$ 'nächstes Erdbeben tritt im nächsten Jahr ein' sowie $A^c =$ 'nächstes Erdbeben tritt im nächsten Jahr nicht ein'. $\Omega = \mathbb{R}_+ = [0, \infty)$ wäre die Ansammlung aller möglichen Eintrittszeitpunkte von Erdbeben, gemessen ab der in einem bestimmten Moment gegebenen Zeit.

Innerhalb der σ -Algebra \mathcal{A} können wir neue Ereignisse aus zuvor gegebenen Ereignissen konstruieren. Dies erfolgt – wie schon oben bei der Verneinung – durch Mengenoperationen. Die Vereinigung von zwei Ereignissen $A \cup B$, wobei $A, B \in \mathcal{A}$, steht z.B. für das Ereignis, das wenigstens eines der beiden Ereignisse A oder B eintritt. Der Schnitt $A \cap B$ steht für das gleichzeitige Eintreten von A und B . 'Oder' und 'und' sind hierbei im Sinne der entsprechenden logischen Verknüpfung zu verstehen.

Die Mengenoperationen in der σ -Algebra stellen wir bequem durch das **Venn Diagramm** dar. In diesem Diagramm veranschaulichen wir Mengen und Mengenoperationen als Flächen, siehe Abbildung 2.1 für einige Beispiele.

Proposition 2.4. *Es gelten die folgenden Regeln für Mengenoperationen:*

$$(i) (A \cap B)^c = A^c \cup B^c \text{ und } \left(\bigcap_{j=1}^{\infty} A_j\right)^c = \bigcup_{j=1}^{\infty} A_j^c$$

$$(ii) \quad (A \cup B)^c = A^c \cap B^c \text{ und } \left(\bigcup_{j=1}^{\infty} A_j\right)^c = \bigcap_{j=1}^{\infty} A_j^c$$

Beweis. Indem wir die Mengen A, B, A_j als logische Aussagen über Elemente der Menge Ω auffassen (A ist die Menge aller $\omega \in \Omega$ für die die logische Aussage $\omega \in A$ WAHR ist), können wir den Durchschnitt von Mengen \cap als logisches UND verstehen, die Vereinigung als logisches ODER und die Komplementbildung als logische Verneinung auffassen. Die Aussagen folgen nun aus den entsprechenden logischen Sachverhalten. \square

Diese Proposition zeigt insbesondere, dass neben abzählbaren Vereinigungen auch abzählbare Durchschnitte in \mathcal{A} liegen. Bei genauem Hinsehen ergibt sich nun die Frage, ob die σ -Algebra auch unter endlichen, d.h. nicht abzählbar-unendlich vielen, Mengenoperationen geschlossen ist. Für die Vereinigung ist dies klar, indem wir zu endlich vielen Mengen $A_1, \dots, A_n \in \mathcal{A}$ weitere, leere Mengen $A_{n+1} = \emptyset, A_{n+2} = \emptyset, \dots \in \mathcal{A}$ hinzufügen, weswegen $\bigcup_{j=1}^n A_j = \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$ folgt. Für den Durchschnitt folgt dies wieder durch Komplementbildung.

Wir kommen nun darauf zurück, wie man Ereignissen Wahrscheinlichkeiten zuordnen kann. Zur Beantwortung dieser Frage steht bisher nur das Symmetrieprinzip für den Laplaceschen Wahrscheinlichkeitsraum zur Verfügung, das in vielen Situationen nicht anwendbar ist. Ein von Richard von Mises (1883–1953) vertretener Ansatz besteht darin, Zufallsexperimente in sehr großer Anzahl zu wiederholen, und die Wahrscheinlichkeiten als Grenzwert der relativen Häufigkeit, mit der ein Ereignis gemessen an den Wiederholungen des Zufallsexperimentes auftritt, zu definieren. Das Problem mit diesem Ansatz ist die Notwendigkeit beliebig oft wiederholbarer Zufallsexperimente, die in vielen relevanten Fällen nicht gegeben ist. Auch wird der Begriff der Wahrscheinlichkeit auf dem Begriff des Zufallsexperimentes errichtet, was das Problem der Mathematisierung eher verschiebt als löst.

Hier folgen wir dem axiomatischen Ansatz, bei dem die Rechenregeln für Wahrscheinlichkeiten festgelegt werden, die Frage der Anwendbarkeit auf die Praxis und die philosophische Interpretation jedoch offen gehalten werden.

Definition 2.5 (Kolmogorov Axiome für Wahrscheinlichkeiten). Es sei (Ω, \mathcal{A}) ein messbarer Raum. Ein **Wahrscheinlichkeitsmaß** P auf (Ω, \mathcal{A}) ist eine Abbildung $P: \mathcal{A} \rightarrow [0, 1]$, die jedem Ereignis $A \in \mathcal{A}$ die Wahrscheinlichkeit $P(A)$ zuweist. Hierbei erfüllt P folgende zwei Bedingungen:

$$(i) \quad P(\Omega) = 1 \quad (\text{Normierung})$$

$$(ii) \quad \text{Seien } A_j \in \mathcal{A}, j \in \mathbb{N}, \text{ paarweise disjunkt, d.h. } A_j \cap A_k = \emptyset, \forall k \neq j, \text{ so gilt}$$

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \quad (\sigma\text{-Additivität}) \quad (2.1)$$

Das Tripel (Ω, \mathcal{A}, P) heißt **Wahrscheinlichkeitsraum**.

2. Wahrscheinlichkeitsräume

Bemerkung 2.6. Zunächst einmal scheinen die σ -Additivität und der Begriff der σ -Algebra unnötig kompliziert zu sein – täte es nicht auch die endliche Additivität? Leider nein! Denn später wollen wir mit der Wahrscheinlichkeitsrechnung auch Erwartungswerte von Zufallszahlen (genauer: Zufallsvariablen) berechnen. Dafür brauchen wir Integrale, die wiederum durch Grenzwerte definiert werden müssen. Um diese Grenzwerte gut kontrollieren zu können, ist die σ -Additivität genau das richtige Werkzeug.

Beispiel 2.7.

- (i) Es sei Ω eine endliche Menge und $\mathcal{A} = \mathbb{P}(\Omega)$ die maximale σ -Algebra von Ω (Potenzmenge). Dann ist das **Laplace-Wahrscheinlichkeitsmaß** $P(A) = \frac{\#A}{\#\Omega}$ ein Wahrscheinlichkeitsmaß im Sinne von Definition 2.5. Wir bezeichnen das so konstruierte Tripel (Ω, \mathcal{A}, P) auch als Laplaceschen Wahrscheinlichkeitsraum.
- (ii) Es sei Ω eine beliebige, nicht notwendiger Weise endliche, Menge und \mathcal{A} eine beliebige σ -Algebra über Ω . Für $x \in \Omega$ und $A \in \mathcal{A}$ definiere das **Dirac-Maß**

$$\delta_x(A) = \begin{cases} 1 & , \text{ falls } x \in A , \\ 0 & , \text{ sonst .} \end{cases} \quad (2.2)$$

Das Dirac-Maß steht somit aus Sicht der Wahrscheinlichkeitstheorie für das deterministische Extrem, dass stets das Elementarereignis x eintritt.

Proposition 2.8 (Rechenregeln für Wahrscheinlichkeitsmaße). *Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gelten für P folgende Regeln:*

- (i) $P(\emptyset) = 0$
- (ii) Seien $A_j \in \mathcal{A}$, $j = 1, \dots, n$, paarweise disjunkt, so gilt

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j) \quad (\text{endliche Additivität}) \quad (2.3)$$

- (iii) Für $A, B \in \mathcal{A}$, so dass $A \subseteq B$, gilt $P(A) \leq P(B)$ (Isotonie)
- (iv) $P(A \cup B) \leq P(A) + P(B)$, $\forall A, B \in \mathcal{A}$ (Sub-Additivität)
- (v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, $\forall A, B \in \mathcal{A}$
- (vi) $P(A^c) = 1 - P(A)$, $\forall A \in \mathcal{A}$ (Gegenwahrscheinlichkeit)

Beweis. (i) Da $A_j = \emptyset \in \mathcal{A}$, $j \in \mathbb{N}$, eine Folge von paarweise disjunkten Ereignissen definiert ($\emptyset \cap \emptyset = \emptyset$) folgt mit der σ -Additivität

$$P(\emptyset) = P\left(\bigcup_{j=1}^{\infty} \emptyset\right) = \sum_{j=1}^{\infty} P(\emptyset) \in [0, 1] , \quad (2.4)$$

was sofort $P(\emptyset) = 0$ nach sich zieht.

(ii) Für endlich viele paarweise disjunkte Ereignisse A_j , $j = 1, \dots, n$ ergänze $A_j = \emptyset$, für $j = n+1, n+2, \dots$. Dann ist A_j , $j \in \mathbb{N}$, immer noch paarweise disjunkt und $\bigcup_{j=1}^n A_j = \bigcup_{j=1}^{\infty} A_j$. Deshalb erhalten wir nun aus der σ -Additivität (2.1)

$$P\left(\bigcup_{j=1}^n A_j\right) = P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) = \sum_{j=1}^n P(A_j) + \sum_{j=n+1}^{\infty} P(\emptyset) = \sum_{j=1}^n P(A_j) . \quad (2.5)$$

(iii) Für $A, B \in \mathcal{A}$ mit $A \subseteq B$ gilt $B \setminus A = B \cap A^c \in \mathcal{A}$. Mittels (ii) folgt nun

$$P(B) = P(A \cup (B \setminus A)) = P(A) + \underbrace{P(B \setminus A)}_{\geq 0} \geq P(A) . \quad (2.6)$$

(iv) & (v) Für $A, B \in \mathcal{A}$ beliebig gilt $A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$ als disjunkte Vereinigung, vergleiche Abbildung 2.1. Also wieder mit (ii) erhalten wir

$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= P((A \setminus B) \cup (A \cap B)) + P((B \setminus A) \cup (A \cap B)) - P(A \cap B) \\ &= P(A) + P(B) - \underbrace{P(A \cap B)}_{\geq 0} \leq P(A) + P(B) . \end{aligned} \quad (2.7)$$

(vi) Aus der Normiertheit und (ii) folgt $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ und somit $P(A^c) = 1 - P(A)$. \square

Beispiel 2.9. Sie bewerben sich auf zwei Stellen. Sie sind sich zu 80% sicher, dass Sie Stelle A bekommen und mit einer Zusage bei Stelle B rechnen sie zu 50%. Die Wahrscheinlichkeit, dass Sie zwischen A und B aussuchen können, veranschlagen Sie mit 45%. Wie groß ist dann Ihre (subjektive) Wahrscheinlichkeit, dass Sie am Ende ohne eine Stelle da stehen?

Lösung. Wir tragen zusammen, was an Angaben gemacht wurde: $P(A) = 0.8$, $P(B) = 0.5$, $P(A \cap B) = 0.45$. Gefragt ist jedoch die Wahrscheinlichkeit von dem Ereignis $A^c \cap B^c =$ 'Nicht Job A und nicht Job B '. Nach Proposition 2.8 (v) gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.5 - 0.45 = 0.85 . \quad (2.8)$$

Da aber $A^c \cap B^c = (A \cup B)^c$, siehe Proposition 2.4 (ii), haben wir nun nach Proposition 2.8 (vi)

$$P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0.85 = 0.15 . \quad (2.9)$$

\diamond

An dieser Stelle möchten wir einen kurzen historischen Kommentar einbringen: Ursprünglich wurden die Kolmogorov-Axiome über endliche Additivität und Stetigkeit des Wahrscheinlichkeitsmaßes formuliert. Ein natürlicher Stetigkeitsbegriff ist der folgende:

2. Wahrscheinlichkeitsräume

Definition 2.10. Es sei (Ω, \mathcal{A}) ein messbarer Raum und $P : \mathcal{A} \rightarrow [0, 1]$ eine endlich additive Abbildung mit $P(\Omega) = 1$.

- (i) Eine Folge von Mengen $A_j \in \mathcal{A}$, $j \in \mathbb{N}$, heißt **isoton** steigend (bzw. fallend) gegen eine Menge $A \in \mathcal{A}$, falls $A_j \subseteq A_{j+1}$ ($A_j \supseteq A_{j+1}$) $\forall j \in \mathbb{N}$ sowie $\bigcup_{j \in \mathbb{N}} A_j = A$ ($\bigcap_{j \in \mathbb{N}} A_j = A$). Wir schreiben $A_j \uparrow A$ ($A_j \downarrow A$).
- (ii) P heißt **unterhalbstetig** (**oberhalbstetig**), falls $A_j \uparrow A$ ($A_j \downarrow A$) impliziert dass $P(A_j) \rightarrow P(A)$ für $j \rightarrow \infty$.

Wir werden nun zeigen, dass die historische Version der Kolmogorov-Axiome zu unserer Version aus Definition 2.5 äquivalent ist.

Proposition 2.11. Es seien (Ω, \mathcal{A}) und $P : \mathcal{A} \rightarrow [0, 1]$ wie in Definition 2.10. Die Abbildung P ist genau dann ein Wahrscheinlichkeitsmaß, wenn P unterhalbstetig oder oberhalbstetig ist.

Beweis. Wir geben den Beweis nur für die Unterhalbstetigkeit. Oberhalbstetigkeit folgt aus Unterhalbstetigkeit durch Komplementbildung.

‘ \Rightarrow ’: Es gelte $A_j \uparrow A$ in \mathcal{A} . Setze $B_1 = A_1$ und $B_j = A_j \setminus A_{j-1}$ für $j \geq 2$. Dann sind die B_j paarweise disjunkt und es gilt $A_n = \bigcup_{j=1}^n B_j$ und $A = \bigcup_{j=1}^{\infty} B_j$ als disjunkte Vereinigung. Aus der σ -Additivität von P erhalten wir

$$\begin{aligned} P(A) &= P\left(\bigcup_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} P(B_j) \\ &= \lim_{n \rightarrow \infty} \sum_{j=1}^n P(B_j) = \lim_{n \rightarrow \infty} P\left(\bigcup_{j=1}^n B_j\right) = \lim_{n \rightarrow \infty} P(A_n) . \end{aligned} \quad (2.10)$$

‘ \Leftarrow ’: Seien B_j , $j \in \mathbb{N}$, paarweise disjunkte Mengen in \mathcal{A} . Setze $A_n := \bigcup_{j=1}^n B_j$ und $A := \bigcup_{j=1}^{\infty} B_j$. Dann gilt wegen der Definition von unendlichen Reihen, der endlichen Additivität, der Unterhalbstetigkeit von P und $A_n \uparrow A$,

$$\sum_{j=1}^{\infty} P(B_j) = \lim_{n \rightarrow \infty} \sum_{j=1}^n P(B_j) = \lim_{n \rightarrow \infty} P\left(\bigcup_{j=1}^n B_j\right) = \lim_{n \rightarrow \infty} P(A_n) = P(A) = P\left(\bigcup_{j=1}^{\infty} B_j\right) . \quad (2.11)$$

Dies ist genau die σ -Additivität von P . □

Die σ -Additivität gibt uns eine Vorschrift wie man Wahrscheinlichkeiten von disjunkten Vereinigungen von Ereignissen bestimmen kann. Es stellt sich somit sofort die folgende Frage: Was kann man über die Berechnung der Wahrscheinlichkeit einer beliebigen (endlichen) Vereinigung von Ereignissen sagen, die nicht notwendiger Weise disjunkt sind? Erste Hinweise geben Proposition 2.8 (v) & (vi) für den Fall von zwei Ereignissen A und B . Doch zunächst ein motivierendes

Beispiel 2.12. Ein Maschine besteht aus n Teilen, die alle funktionieren müssen, damit sie läuft. Das Ereignis ‘Teil j funktioniert’ bezeichnen wir mit A_j . Dementsprechend haben wir mit $\epsilon_j = 1 - P(A_j) = P(A_j^c)$ die Versagenswahrscheinlichkeit von Teil j . Können wir nun auf einfache Weise eine konservative Abschätzung für die Wahrscheinlichkeit abgeben, dass die ganze Maschine läuft? Konservativ heißt hierbei, dass wir die Zuverlässigkeit lieber unter- als überschätzen.

Eine mögliche Antwort liefert die folgende

Proposition 2.13 (Bonferroni-Ungleichung). *Es sei $A_j \in \mathcal{A}$ und $P(A_j) = 1 - \epsilon_j$, für $j = 1, \dots, n$. Dann gilt*

$$P\left(\bigcap_{j=1}^n A_j\right) \geq 1 - \sum_{j=1}^n \epsilon_j. \quad (2.12)$$

Beweis. Durch induktive Anwendung von Proposition 2.8 (v) und der Nichtnegativität von Wahrscheinlichkeiten erhält man zunächst $P(\bigcup_{j=1}^n A_j^c) \leq \sum_{j=1}^n \epsilon_j$. Nun wenden wir Proposition 2.4 (i) und Proposition 2.8 (vi) an und erhalten

$$P\left(\bigcap_{j=1}^n A_j\right) = 1 - P\left(\left(\bigcap_{j=1}^n A_j\right)^c\right) = 1 - P\left(\bigcup_{j=1}^n A_j^c\right) \geq 1 - \sum_{j=1}^n \epsilon_j. \quad \square \quad (2.13)$$

Lösung von Beispiel 2.12. Die Maschine funktioniert wenn alle Einzelteile funktionieren, also wenn das Ereignis $A = \bigcap_{j=1}^n A_j$ eintritt. Mittels der Bonferroni-Ungleichung (2.12) können wir die Wahrscheinlichkeit von A konservativ abschätzen als $P(A) \geq 1 - \sum_{j=1}^n \epsilon_j$. \diamond

Bemerkung 2.14. Wir weisen auf ein mögliches Problem mit der Bonferroni-Ungleichung hin. Zur Illustration betrachten wir den Spezialfall, dass $\epsilon_j = \epsilon > 0$ für alle $j = 1, \dots, n$. Dann erhalten wir aus (2.12) die Abschätzung $P(A) \geq 1 - n\epsilon$. Für hinreichend großes n (genauer: $n \geq \frac{1}{\epsilon}$) gibt diese Ungleichung keinerlei Information, da sie einfach nur besagt, dass $P(A)$ größer als die in diesem Falle nichtpositive Zahl $1 - n\epsilon$ sein muss, was sowieso per Definition von P gilt. Analoges gilt für die Abschätzung $P(\bigcup_{j=1}^n A_j^c) \leq \sum_{j=1}^n \epsilon_j$, welche aus der Verneinung der Ereignisse in (2.12) hervorgeht.

Um das oben genannte Problem zu beheben, möchten wir nun eine Formel für die Wahrscheinlichkeit einer Vereinigung von n Ereignissen A_1, \dots, A_n ableiten, wobei wir voraussetzen müssen, dass wir die Wahrscheinlichkeiten beliebiger Durchschnitte $A_I = \bigcap_{j \in I} A_j$, $I \subseteq \{1, \dots, n\}$ kennen. In manchen Situationen wird dies der Fall sein, in anderen könnten zumindest Abschätzungen vorliegen.

Satz 2.15 (Siebformel von Poincaré und Sylvester). *Es seien $A_j \in \mathcal{A}$ Ereignisse, $j = 1, \dots, n$. Dann gilt*

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{j=1}^n P(A_j) - \sum_{1 \leq j_1 < j_2 \leq n} P(A_{j_1} \cap A_{j_2}) \\ &\quad + \sum_{1 \leq j_1 < j_2 < j_3 \leq n} P(A_{j_1} \cap A_{j_2} \cap A_{j_3}) - \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n) \\ &= \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{\#I+1} P(A_I). \end{aligned} \quad (2.14)$$

2. Wahrscheinlichkeitsräume

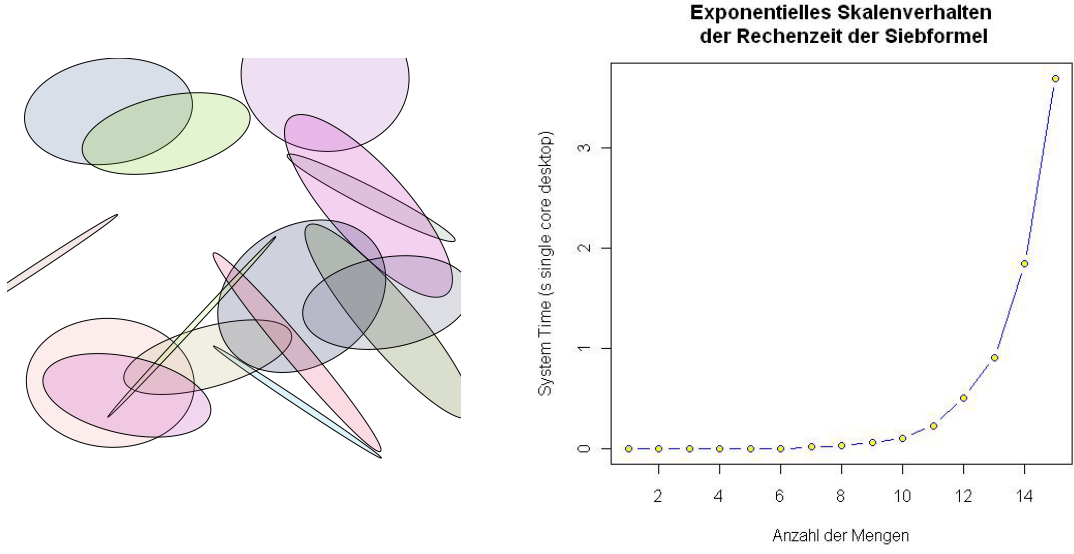


Abbildung 2.2: Links: Eine große Menge sich teilweise überschneidender Mengen - mit R erzeugt. Rechts: Anstieg der Auswertungszeit mit n bei der Implementierung der Siebformel durch den Pseudocode.

Bemerkung 2.16. Der kombinatorische Charakter der obigen Formel wird sehr deutlich. Hierbei ist die letzte Zeile offensichtlich nur eine kompakte (aber auch ein wenig abstrakte) Schreibweise für die Formel im mittleren Teil der Gleichung.

Beweis von Satz 2.15. Wir beweisen per Induktion über n . Der Fall $n = 2$ ist Proposition 2.8 (v) und kann als Induktionsanfang dienen. Wir setzen nun $B = \bigcup_{i=1}^{n+1} A_i$, $B' = \bigcup_{i=1}^n A_i$ und verwenden zunächst den Induktionsanfang und dann zwei Mal die Induktionshypothese. Schließlich schreiben wir alle auftretenden Ausdrücke als Terme in der Siebformel mit $n + 1$, wobei wir Fallunterscheidungen anwenden, ob $\{n + 1\}$ gleich der Menge I , nicht enthalten in I oder echt enthalten in der Menge I ist. Da eine der drei Alternativen zutreffen muss, erhalten wir das Resultat. In Formeln lautet dieser Beweis wie folgt

$$\begin{aligned}
 P(B) &= P(A_{n+1}) + P(B') - P(A_{n+1} \cap B') \\
 &= P(A_{n+1}) + P\left(\bigcup_{i=1}^n A_i\right) - P\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) \\
 &= P(A_{n+1}) + \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{\#I+1} P(A_I) - \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{\#I+1} P(A_I \cap A_{n+1}) \\
 &= \sum_{\substack{\emptyset \neq I \subseteq \{1, \dots, n+1\} \\ I = \{n+1\}}} (-1)^{\#I+1} P(A_I) + \sum_{\substack{\emptyset \neq I \subseteq \{1, \dots, n+1\} \\ \{n+1\} \notin I}} (-1)^{\#I+1} P(A_I) \\
 &\quad + \sum_{\substack{\emptyset \neq I \subseteq \{1, \dots, n+1\} \\ \{n+1\} \in I, I \neq \{n+1\}}} (-1)^{\#I+1} P(A_I) = \sum_{\emptyset \neq I \subseteq \{1, \dots, n+1\}} (-1)^{\#I+1} P(A_I) . \quad \square \quad (2.15)
 \end{aligned}$$

Bemerkung 2.17. Induktionsbeweise geben oft einen Hinweis, wie man eine Formel wie die von Poincaré und Sylvester implementieren kann. Der folgende Pseudocode wendet den ersten Schritt im Beweis von Satz 2.15 rekursiv an:

```
    L=liste(A1,...,An)
|sieb(L) procedure
|  falls länge(L)=1
|    return P(A1)
|  sonst
|    L1=liste(A1,...,An-1)
|    L2=liste(A1∩An,...,An-1 ∩An)
|    return P(An)+sieb(L1)-sieb(L2)
|  ende falls
|ende procedure
```

Leider steigt die Auswertungszeit für diesen Algorithmus exponentiell mit n an, siehe Abbildung 2.2.

2.1. Übungsaufgaben

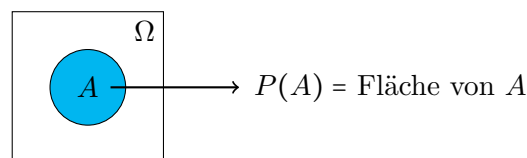
1. Selbsttest

1. Welche der drei Axiome für σ -Algebren erfüllt das Mengensystem

$$\mathcal{A} = \{\{1, 2, 3\}, \{1\}, \{2, 3\}\}$$

über der Menge $\Omega = \{1, 2, 3\}$? Nennen Sie zuerst die drei Axiome und überprüfen Sie dann deren Gültigkeit!

2. Begründen Sie wieso $(\Omega, \mathbb{P}(\Omega))$ für jede beliebige Menge Ω ein messbarer Raum ist!
3. Es seien A, B, C drei Ereignisse. Geben Sie die Mengen an, welche die folgenden Ereignisse charakterisieren:
- (i) keines der drei Ereignisse treten ein,
 - (ii) genau zwei der drei Ereignisse treten ein,
 - (iii) höchstens zwei der drei Ereignisse treten ein.
4. Ziel dieser Aufgabe ist es ein bildliches Verständnis von Wahrscheinlichkeitsmaßen und ihren Eigenschaften zu erhalten. Dazu stellen wir uns die Menge Ω und ihre messbaren Teilmengen $A \in \mathcal{A}$ mittels Venn Diagrammen vor. Das Wahrscheinlichkeitsmaß $P: \mathcal{A} \rightarrow [0, 1]$, $A \mapsto P(A)$ soll der Teilmenge $A \subseteq \Omega$ den Flächeninhalt $P(A)$ zuweisen, also



- a) Interpretieren Sie die Axiome für Wahrscheinlichkeitsmaße aus Definition 2.5 mittels Venn Diagrammen und diesem Bild!
 - b) Geben Sie bildliche Argumente warum die Rechenregeln für Wahrscheinlichkeitsmaße aus Proposition 2.8 gelten müssen!
 - c) Geben Sie ebenfalls eine bildliche Erklärung für die Siebformel für $P(A \cup B \cup C)$ indem Sie geeignete Venn Diagramme studieren.
5. Sie möchten einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) zu einer gegebenen endlichen Menge $\Omega = \{1, \dots, n\}$ konstruieren. Weiter möchten Sie, dass das Wahrscheinlichkeitsmaß auf allen Elementarereignissen $\{i\} \subseteq \Omega$ definiert ist, d.h. alle $\{i\}$ sollen in der σ -Algebra \mathcal{A} enthalten sein. Welche Möglichkeiten haben Sie \mathcal{A} zu wählen?
6. Können Sie alle Wahrscheinlichkeitsmaße P auf dem messbaren Raum $(\Omega, \mathbb{P}(\Omega))$ mit $\Omega = \{1, 2\}$ charakterisieren? Gibt es neben den Dirac-Maßen und dem Laplaceschen Wahrscheinlichkeitsmaß noch weitere Möglichkeiten P zu wählen?

7. Das Ereignis A trete mit Wahrscheinlichkeit 60% ein und das Ereignis B mit Wahrscheinlichkeit 50%.
- a) Schätzen Sie durch eine obere Schranke die Wahrscheinlichkeit ab, dass A oder B eintritt!
 - b) Ist die Abschätzung in a) sinnvoll?
 - c) Können die Ereignisse A und B disjunkt sein?
 - d) Nun haben Sie ebenfalls die Zusatzinformation, dass die Wahrscheinlichkeit das A nicht oder B nicht eintritt 60% beträgt. Berechnen Sie hiermit $P(A \cup B)$ exakt!

2. Transfer

1. Geben Sie für die folgenden Zufallsexperimente einen geeigneten Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) an! Spezifizieren Sie hierzu die Menge Ω , die σ -Algebra \mathcal{A} und das Wahrscheinlichkeitsmaß P .
 - a) Einmaliges Würfeln mit einem Würfel mit Augenzahlen 1 bis 6.
 - b) Einmaliges Würfeln mit n solchen Würfeln.
 - c) Lotto 6 aus 49, d.h. Ziehen ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge von 6 Kugeln aus einer Menge von 49 Kugeln.
2. In der Finanzwelt geht es darum geschickte Entscheidungen zu treffen. Eine Finanzexpertin mit langjähriger Erfahrung macht nur bei jeder 200sten Entscheidung für eine Transaktion einen Fehler.
 - a) Geben Sie eine konservative Abschätzung für die Wahrscheinlichkeit, dass die Expertin die nächsten 90 Entscheidungen korrekt trifft!
 - b) Ab welcher Anzahl von Entscheidungen n ist eine Abschätzung wie unter a) nicht mehr sinnvoll?

3. Theorie

1. Es sei $\Omega = \{1, \dots, n\}$ eine endliche Menge und $\mathbb{P}(\Omega)$ die Potenzmenge. Wir betrachten das Wahrscheinlichkeitsmaß P auf dem messbaren Raum $(\Omega, \mathbb{P}(\Omega))$, welches durch $P(\{i\}) = p_i \geq 0$, für alle $i = 1, \dots, n$, definiert ist. Aufgrund der Normierung $P(\Omega) = 1$ gilt natürlich $\sum_{i=1}^n p_i = 1$. Dem Wahrscheinlichkeitsraum $(\Omega, \mathbb{P}(\Omega), P)$ weisen wir eine charakteristische Zahl zu, die sogenannte *Shannon Entropie*, welche den Informationsgehalt bzw. die Unordnung misst. Diese ist definiert durch $H = -\sum_{i=1}^n p_i \log(p_i)$.
 - a) Informieren Sie sich im Internet über Anwendungsbereiche der Shannon Entropie und erklären Sie diese kurz!

2. Wahrscheinlichkeitsräume

b) In vielen Anwendungen geht es darum die Entropie zu extremieren. Bestimmen Sie für $n = 2$ und $n = 3$ die Extrema von H ! Handelt es sich hierbei um Minima oder Maxima?

c) Bestimmen Sie für allgemeines n das Maximum von H !

Hinweis: Sie können entweder das Extremwertproblem unter der Nebenbedingung $\sum_{i=1}^n p_i = 1$ mit Lagrange-Multiplikatoren lösen oder das erwartete Resultat per Widerspruchsbeweis zeigen.

2. Sie haben sich vielleicht gewundert, warum man in der Definition eines Wahrscheinlichkeitsraumes (Ω, \mathcal{A}, P) die σ -Algebra \mathcal{A} nicht immer durch die maximale σ -Algebra, die Potenzmenge $\mathbb{P}(\Omega)$ von Ω , fixiert hat. Für abzählbare Mengen Ω ist dies in der Tat die übliche Wahl, jedoch für überabzählbare Ω zeigen sich Probleme mit $\mathcal{A} = \mathbb{P}(\Omega)$, wie Sie anhand des folgenden Beispiels sehen sollen:

Sei $\Omega = S^1 := \{z \in \mathbb{C} : |z| = 1\}$ der Einheitskreis. Dann existiert auf der Potenzmenge $\mathcal{A} = \mathbb{P}(\Omega)$ **kein** Wahrscheinlichkeitsmaß P das unter Phasentransformationen invariant ist, d.h. für das gilt

$$P(A_\tau) = P(A) \text{ , für alle } A \subseteq \Omega \text{ , } \tau \in \mathbb{R} \text{ , } A_\tau := \{e^{i\tau} z : z \in A\} \text{ .}$$

Beweisen Sie diese Aussage!

Hinweis: Informieren Sie sich zuvor über *Vitali-Mengen*. Da dieser Beweis recht schwierig ist, sollten Sie sich Inspiration durch das Internet oder durch fortgeschrittenere Fachliteratur einholen.

2.2. Literatur

Die Grundlagen der Maß- und Integrationstheorie in Form der Ausschöpfungsmethode und des doppelten Widerspruchsbeweises reichen zurück bis in die Antike und sind verbunden mit den Namen Eudoxos von Knidos (c.a. 390-340 a.D.) und Archimedes von Syrakus (c.a. 287-212 a.D.). In der mittelalterlichen und neuzeitlichen Mathematik blieb die Maßtheorie stets ein aktives Feld [Pau02].

Die Moderne Maß- und Integrationstheorie ist mit den Namen O. Stolz (1842-1905), G. Cantor (1845-1918), Giuseppe Peano (1858-1932) und Émile Borel (1871-1956) verbunden. Das Lebesgue-Maß wurde von H. Lebesgue (1875-1941) in seiner Dissertation eingeführt. J. Radon (1887-1956) löste die Verbindung von der Integrationstheorie zu Volumen- oder Oberflächenbestimmungen und begründete die abstrakte Maßtheorie, die im Weiteren von O. Nikodym (1887-1974), C. Carathéodory (1873-1950), Eugene Dynkin (*1924) und vielen anderen erforscht wurde [Pau02].

Die Axiomatisierung der Wahrscheinlichkeitstheorie gelang im Jahr 1933 dem Russischen Mathematiker A. N. Kolmogorov [Kol33]. Auf die Notwendigkeit einer mathematisch-exakten Fassung des Wahrscheinlichkeitsbegriffs hatte D. Hilbert (1862-1943) auf seiner berühmten Rede auf dem Internationalen Mathematikerkongress in Berlin im Jahre 1900 hingewiesen (23. Jahrhundertproblem). Dem allgemein akzeptierten Ansatz von Kolmogorov waren verschiedene andere Axiomatisierungsversuche, u.a. durch R. von Mises (1883-1953) [vM72] und G. Bohlmann (1869-1928), vorangegangen, die sich in der Folge aus unterschiedlichen Gründen nicht durchsetzten: Bohlmanns Axiomensystem verzichtete auf die σ -Additivität und konnte von daher den Begriff des Erwartungswertes nicht angemessen erklären. Von Mises Begriff der maximal ungeordneten Folge von Ereignissen blieb mathematisch solange mathematisch unklar, bis Kolmogorov ihn mit unabhängig, identisch verteilten Zufallsvariablen zusammen brachte.

Die nach H. Poincaré (1854-1912) und J. Sylvester (1814-1897) benannte Siebformel wurde von dem letztgenannten im Jahre 1883 veröffentlicht. Sie wird aber schon A. de Moivre zugeschrieben. Von D. Da Silva (1814-1878) wurde sie bereits im Jahre 1854 veröffentlicht.

Moderne Abhandlungen über die Maßtheorie findet man in [Bau90] und [Hal74].

3. Bedingte Wahrscheinlichkeiten

Unsere Einschätzung über die Wahrscheinlichkeit eines bestimmten Ereignisses B kann sich je nach dem Stand unserer Vorabinformation über andere, mit B zusammenhängende Ereignisse ändern. Diesen Sachverhalt formalisiert die folgende

Definition 3.1. Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ zwei Ereignisse mit $P(A) > 0$. Die **bedingte Wahrscheinlichkeit** $P(B|A)$ von B , gegeben dass A eintritt, ist definiert als

$$P(B|A) := \frac{P(A \cap B)}{P(A)} . \quad (3.1)$$

Man spricht das verkürzend als ‘Wahrscheinlichkeit von B gegeben A ’ aus. Vergleichen Sie dies auch mit Abbildung 3.1.

Beispiel 3.2. Damit ein Versicherungsunternehmen eine Lebensversicherung kalkulieren kann, braucht es Informationen über den nicht absehbaren Todeszeitpunkt der Versicherungsnehmer/innen. Mit bestimmten statistischen Verfahren¹ können Wahrscheinlichkeiten für den Tod im n -ten Lebensjahr geschätzt werden - Abbildung 3.2 (Links) zeigt ein historisches Beispiel. Die Daten sind vom Max Planck Institut für demografische Forschung, siehe www.lifetable.org.

Hätte eine Versicherung in dieser Zeit mit diesen Daten die Sterblichkeit ihrer Kunden kalkuliert, hätte das Unternehmen diese erheblich unterschätzt. Nehmen wir als Beispiel den Versicherungsabschluss mit einem dreißigjährigen Mann. Dann interessiert weniger die Wahrscheinlichkeit $P(\text{‘Mann stirbt mit } 30 + k \text{ Jahren’})$, als vielmehr die bedingte Wahrscheinlichkeit (Tote schließen im Allgemeinen keine Lebensversicherung ab!)

$$P(\text{‘Mann stirbt mit } 30 + k \text{ Jahren’} | \text{‘Mann lebt am 30-ten Geburtstag’}) . \quad (3.2)$$

Eine entsprechende Grafik finden Sie in Abbildung 3.2 (Mitte). Da diese Darstellung jedoch vom Eintrittsalter abhängt, arbeiten Versicherungen mit so genannten Sterbetafeln, in denen die (geschätzten) einjährigen Sterbewahrscheinlichkeiten tabellarisiert sind:

$$h(a) = P(\text{‘Mann stirbt mit } a \text{ Jahren’} | \text{‘Mann lebt am } a\text{-ten Geburtstag’}) . \quad (3.3)$$

Eine entsprechende Darstellung der Daten findet sich in Abbildung 3.2 (Rechts). Gut zu erkennen ist die durch die hohe Kindersterblichkeit bedingte “Badewannenform”.

Die folgende Proposition stellt wichtige Eigenschaften der bedingten Wahrscheinlichkeit zusammen.

¹Wir ignorieren hier die Probleme, die sich aus dem demografischen Trend hin zur längeren Lebenszeit ergeben.

3. Bedingte Wahrscheinlichkeiten

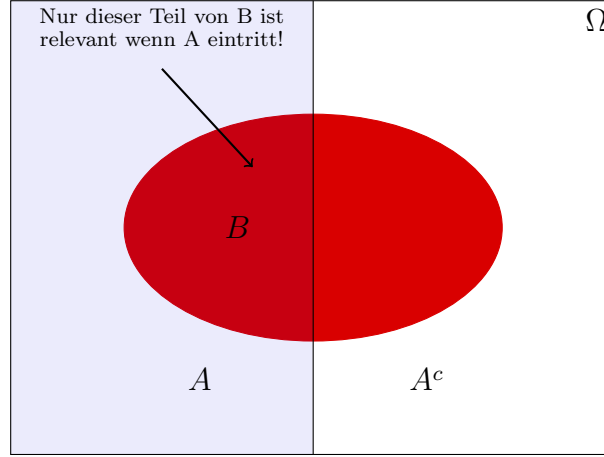


Abbildung 3.1: Venn Diagramm für $P(B|A)$: Nur der Teil A des Ereignisraumes Ω ist relevant für die Berechnung von $P(B|A)$.

Proposition 3.3. *Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A \in \mathcal{A}$ mit $P(A) > 0$. Dann gilt:*

- (i) *Die Abbildung $P(\cdot|A) : \mathcal{A} \rightarrow [0, 1]$, $B \mapsto P(B|A)$ ist ein Wahrscheinlichkeitsmaß auf dem messbaren Raum (Ω, \mathcal{A}) .*
- (ii) *Die Abbildung $P(\cdot|A) : \mathcal{A}_A \rightarrow [0, 1]$, $B \mapsto P(B|A)$ ist ein Wahrscheinlichkeitsmaß auf dem messbaren Raum (A, \mathcal{A}_A) , wobei $\mathcal{A}_A := \{C \cap A : C \in \mathcal{A}\}$.*

Beweis. Zuerst müssen wir zeigen, dass $P(B|A) \in [0, 1]$. Aus (3.1) und $P(C) \geq 0$, für alle $C \in \mathcal{A}$, folgt, dass $P(B|A) \geq 0$. Die Ungleichung $P(B|A) \leq 1$ folgt aus (3.1) und der Isotonie von P , d.h. $P(A) \geq P(A \cap B)$. Es bleibt somit nur noch die Normiertheit und die σ -Additivität zu zeigen. Für (i) ist die Normiertheit eine Konsequenz der folgenden Rechnung $P(\Omega|A) = \frac{P(\Omega \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$. Für (ii) findet man analog $P(A|A) = \frac{P(A \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$. Es seien nun $A_j \in \mathcal{A}$, $j \in \mathbb{N}$, paarweise disjunkt, so gilt die σ -Additivität für (i)

$$\begin{aligned} P\left(\bigcup_{j=1}^{\infty} A_j \middle| A\right) &= \frac{P\left(\left(\bigcup_{j=1}^{\infty} A_j\right) \cap A\right)}{P(A)} = \frac{P\left(\bigcup_{j=1}^{\infty} (A_j \cap A)\right)}{P(A)} \\ &= \frac{\sum_{j=1}^{\infty} P(A_j \cap A)}{P(A)} = \sum_{j=1}^{\infty} P(A_j|A) . \end{aligned} \quad (3.4)$$

Wählt man $A_j \in \mathcal{A}_A$ so folgt aus dieser Rechnung auch die σ -Additivität für (ii). \square

Der folgende Satz über bedingte Wahrscheinlichkeiten sorgte historisch für Aufruhr, weil man die Änderung der Wahrscheinlichkeit für B gegeben A als kausale Abhängigkeit von B von A missverstand. Wäre dies richtig, so würde der Satz von Bayes zeigen, dass man dies umkehren kann und von der Wirkung ‘kausal’ auf die Ursache schließen könnte.

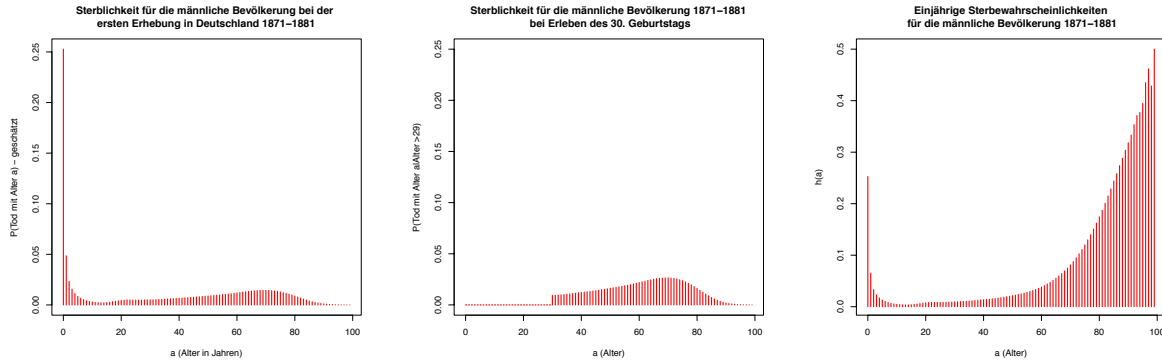


Abbildung 3.2: Links: Historische Sterbewahrscheinlichkeiten. Mitte: Sterbewahrscheinlichkeiten bedingt auf das Erleben des 30. Geburtstages. Rechts: Einjährige Sterbewahrscheinlichkeiten.

Satz 3.4 (Fallunterscheidungsformel und Satz von Bayes (1763)). *Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A_j \in \mathcal{A}$, $j = 1, \dots, n$, paarweise disjunkte Ereignisse mit $P(A_j) > 0$ und $\bigcup_{j=1}^n A_j = \Omega$.*

(i) *Für alle $B \in \mathcal{A}$ gilt*

$$P(B) = \sum_{j=1}^n P(B|A_j) P(A_j) . \quad (3.5)$$

(ii) *Für alle $B \in \mathcal{A}$ mit $P(B) > 0$ gilt*

$$P(A_k|B) = \frac{P(B|A_k) P(A_k)}{\sum_{j=1}^n P(B|A_j) P(A_j)} . \quad (3.6)$$

Beweis. (i) Mit der endlichen Additivität von P und (3.1) gilt

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P\left(B \cap \left(\bigcup_{j=1}^n A_j\right)\right) = P\left(\bigcup_{j=1}^n (B \cap A_j)\right) \\ &= \sum_{j=1}^n P(B \cap A_j) = \sum_{j=1}^n P(B|A_j) P(A_j) . \end{aligned} \quad (3.7)$$

(ii) Nach (3.1) und (i) gilt

$$P(A_k|B) = \frac{P(B \cap A_k)}{P(B)} = \frac{P(B|A_k) P(A_k)}{P(B)} = \frac{P(B|A_k) P(A_k)}{\sum_{j=1}^n P(B|A_j) P(A_j)} . \quad \square \quad (3.8)$$

3. Bedingte Wahrscheinlichkeiten

Beispiel 3.5. Wir betrachten das Dopingproblem in der olympischen Wintersportart Curling. Obwohl es im Curling bereits Dopingfälle gegeben hat, nehmen wir an, dass höchstens 1% der Curling-Athlet/innen gedopt sind, also $P(D) = 0.01$ und $P(D^c) = 0.99$.

Da Curling eine olympische Sportart ist, wird jede/r Athlet/in auf Doping getestet. Das Ereignis T stehe für einen positiven Dopingtest und das Ereignis T^c für einen negativen Testausgang. Die Charakteristik des Testes sei gegeben wie folgt:

	D	D^c			D	D^c
T	$P(T D)$	$P(T D^c)$	$=$	T	0.99	0.01
T^c	$P(T^c D)$	$P(T^c D^c)$		T^c	0.01	0.99
Σ	1	1		Σ	1	1

Der Test hat also eine Wahrscheinlichkeit von 99% gedopte Sportler/innen positiv anzuzeigen und nicht gedopte Sportler/innen negativ. Ist ein/e Sportler/in nicht gedopt, so wird sie zu 1% zu Unrecht durch den positiven Test beschuldigt werden. Dies nennt man eine falsch positive Anzeige oder auch den α -Fehler. Ein/e gedopete/r Sportler/in kommt zu 1% ungeschoren davon. Dies nennt man eine falsch negative Anzeige, oder den β -Fehler des Tests.

Nun wird ein/e Curler/in positiv getestet. Wie groß ist die Wahrscheinlichkeit, dass diese Person tatsächlich gedopt hat?

Lösung. Zunächst einmal ist die naive Antwort 99% falsch, denn sie berücksichtigt nicht, dass viel mehr der getesteten Curler/innen ehrlich als unehrlich sind. Deshalb wächst die Chance auf einen falsch positiven Test mit der Anzahl der Gelegenheiten falsch positiv zu testen.

Die richtige Antwort liefert die bedingte Wahrscheinlichkeit $P(D|T)$, die wir nach dem Satz von Bayes berechnen:

$$P(D|T) = \frac{P(T|D)P(D)}{P(D)P(T|D) + P(D^c)P(T|D^c)} = \frac{0.99 \times 0.01}{0.01 \times 0.99 + 0.99 \times 0.01} = \frac{1}{2}. \quad (3.9)$$

Hier scheint eine B-Probe angebracht. ◇

Die folgende Formel zeigt, wie man die Wahrscheinlichkeit dafür, dass mehrere Ereignisse alle eintreten, durch bedingte Wahrscheinlichkeiten ausdrücken kann.

Proposition 3.6 (Produktformel). *Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A_j \in \mathcal{A}$, $j = 1, \dots, n$, Ereignisse mit $P(\cap_{j=1}^{n-1} A_j) > 0$. Dann gilt*

$$\begin{aligned} P\left(\bigcap_{j=1}^n A_j\right) &= P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}) \\ &= \prod_{j=1}^n P\left(A_j \middle| \bigcap_{l=1}^{j-1} A_l\right), \end{aligned} \quad (3.10)$$

wobei wir die Konvention $P(A_1) = P(A_1 | \cap_{l=1}^{1-1} A_l)$ verwendet haben.

Beweis. Wir beweisen diese Proposition durch vollständige Induktion über n . Der Induktionsanfang $n = 1$ ist trivialerweise erfüllt und im Falle $n = 2$ ist (3.10) lediglich die Definition der bedingten Wahrscheinlichkeit.

Für den Induktionsschritt $n \Rightarrow n + 1$ setzen wir $B = \bigcap_{j=1}^n A_j$. Es gilt nach der Induktionshypothese $P(B) = \prod_{j=1}^n P(A_j | \bigcap_{l=1}^{j-1} A_l)$. Des Weiteren ist aufgrund der Hypothese der Proposition $P(\bigcap_{j=1}^{n-1} A_j) \geq P(\bigcap_{j=1}^n A_j) > 0$. Unter Verwendung der Induktionshypothese finden wir

$$\begin{aligned} P\left(\bigcap_{j=1}^{n+1} A_j\right) &= P(B \cap A_{n+1}) = P(B) P(A_{n+1}|B) \\ &= \left(\prod_{j=1}^n P\left(A_j \middle| \bigcap_{l=1}^{j-1} A_l\right)\right) P(A_{n+1}|B) \\ &= \left(\prod_{j=1}^n P\left(A_j \middle| \bigcap_{l=1}^{j-1} A_l\right)\right) P\left(A_{n+1} \middle| \bigcap_{l=1}^n A_l\right) \\ &= \prod_{j=1}^{n+1} P\left(A_j \middle| \bigcap_{l=1}^{j-1} A_l\right) . \quad \square \end{aligned} \tag{3.11}$$

Oft ist es in mehrschrittigen Zufallsexperimenten einfacher, Szenarien Schritt für Schritt zu modellieren. So wird z.B. in der industriellen Risikoanalyse oft mit Fehlerbäumen gearbeitet, in welchen man die Entstehungsbedingungen eines katastrophalen Unfalls betrachtet und, soweit möglich, die einzelnen Entstehungsschritte mit Wahrscheinlichkeiten unterlegt.

Beispiel 3.7. Nach der internationalen Norm ISO 21789 darf ein Gaskraftwerk nur dann bedenkenlos betrieben werden, wenn die Wahrscheinlichkeit eines Personenunfalls in einem betreffenden Jahr für die am stärksten gefährdete Person Eins zu eine Millionen beträgt (grüner Bereich). Beträgt die Wahrscheinlichkeit Eins zu Tausend, so darf das Kraftwerk zwar vorerst weiter betrieben werden, es wird jedoch empfohlen, risikosenkende Maßnahmen einzuleiten (gelber Bereich). Übersteigt die Wahrscheinlichkeit diese Marke, so muss das Kraftwerk solange heruntergefahren werden, bis die Risikofaktoren beseitigt sind (roter Bereich).

In einer etwas älteren Anlage, in der die Arbeiter/innen im gefährdeten Bereich im drei Schichten Betrieb an 220 Tagen im Jahr arbeiten, hat das verantwortliche Risk-Management-Team im Rahmen einer *Failure Modes and Effects Analysis* (FMEA) der Berteiberfirma die folgenden Risikofaktoren identifiziert:

- 1) Ereignis R = ‘Bruch einer Rotorscheibe durch Materialversprödung’, Wahrscheinlichkeit in einem Jahr $P(R) = 1/1000$. Wahrscheinlichkeit, dass eine Person in Folge dessen zu Schaden kommen, $P(S|R) = 0.5$.
- 2) Ereignis B = ‘Brand in der Ölzufuhr’, Wahrscheinlichkeit in einem Jahr $P(B) = 1/50$. Wahrscheinlichkeit, dass Feuer in der Ölzufuhr auf Dieseltank übergreift $P(D|B) = 1/10$. Wahrscheinlichkeit, dass in Folge eines Brandes der Dieseltanks eine Person zu Schaden kommt $P(S|D) = 1/2$.

3. Bedingte Wahrscheinlichkeiten

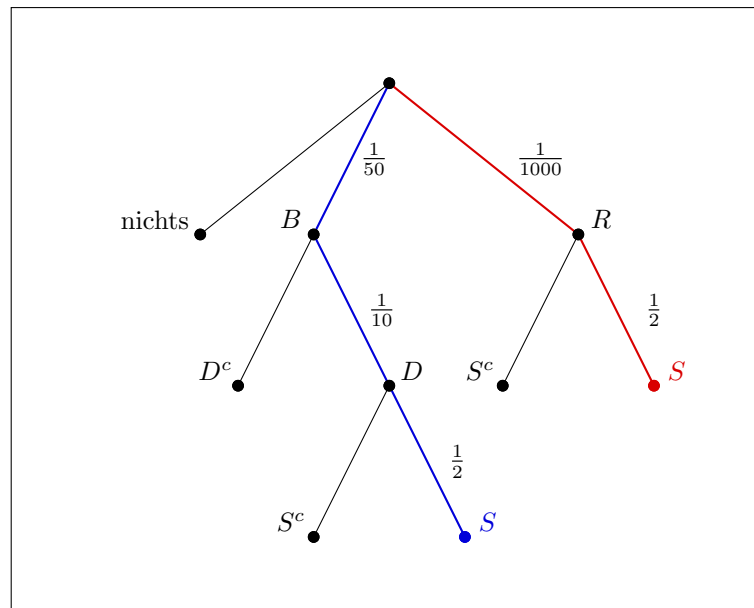


Abbildung 3.3: Wahrscheinlichkeitsbaum bzw. Fehlerbaum zu Beispiel 3.7.

Das Team schließt aus, dass in einem Jahr Ereignis R und Ereignis B gleichzeitig auftreten können, da nach einem Unfall das Kraftwerk auf jeden Fall erstmal stillgelegt und überholt wird. Der Eintritt eines solchen Unfalls hängt nicht von der Schicht ab. Die am stärksten gefährdeten Personen sind beiden Risiken während der gesamten Schicht ausgesetzt.

Wenn wir die ISO 21789 zu Grunde legen – welche Konsequenz muss das Team daraus ziehen?

Lösung. Zunächst stellen wir einen so genannten **Wahrscheinlichkeitsbaum** auf (in der Mathematik stehen diese oft auf dem Kopf), der die Szenarien abbildet, vergleiche Abbildung 3.3.

Mit der Produktformel berechnet man nun ‘entlang der Äste’ die Wahrscheinlichkeit, dass eine bestimmte anwesende Person in einem Szenario zu Schaden kommt. Anschließend summiert man die Szenarien, für die das zutrifft, auf. Das Ergebnis ist

$$P(S) = \frac{1}{1000} \times \frac{1}{2} + \frac{1}{50} \times \frac{1}{10} \times \frac{1}{2} = \frac{1}{2000} + \frac{1}{1000} = \frac{3}{2000} \quad (3.12)$$

Dies wäre die Risikozahl für eine gefährdete Person, die ‘immer da ist’. Dieser Wert liegt im roten Bereich.

Der Anteil der Anwesenheitszeit eines/einer Arbeiter/in im gefährdeten Personenkreis beträgt $p = 220/(365 \times 3) \approx 0.2$. Wir könnten die kritischen Szenarien also um eine weitere Verzweigung ‘Person da’ oder ‘Person nicht da’ erweitern. Also ist das Risiko auf die individuelle Person bezogen $p \times P(S) = \frac{3}{10000}$. Dies ist so grade mal im gelben Bereich.

Das Team braucht laut der ISO 21789 das Kraftwerk nicht sofort stillzulegen, sollte aber dringend Sicherheitsmaßnahmen zur Senkung des Risikos veranlassen. \diamond

Da Wahrscheinlichkeitsbäume offensichtlich nützlich sind, möchten wir dieses Konzept nun genauer fassen. Wir möchten mittels des folgenden Satzes zeigen, dass man aus der Modellierung von bedingten Wahrscheinlichkeiten ein Wahrscheinlichkeitsmaß konstruieren kann. Wir beschränken uns hier auf den Fall von endlichen Ereignisräumen. Mit mehr technischem Aufwand kann dieses Resultat erheblich erweitert werden.

Satz 3.8. *Es sei Ω_j eine endliche Menge und $\mathbb{P}(\Omega_j)$ die Potenzmenge (Menge aller Teilmengen) von Ω_j , für $j = 1, \dots, n$. Weiter sei*

$$\begin{array}{llll} P_1(\cdot) & \text{W.-Maß auf} & (\Omega_1, \mathbb{P}(\Omega_1)) & \\ P_2(\cdot|\omega_1) & \text{W.-Maß auf} & (\Omega_2, \mathbb{P}(\Omega_2)) & \forall \omega_1 \in \Omega_1 \\ P_3(\cdot|\omega_1, \omega_2) & \text{W.-Maß auf} & (\Omega_3, \mathbb{P}(\Omega_3)) & \forall \omega_1 \in \Omega_1, \omega_2 \in \Omega_2 \\ \vdots & \vdots & \vdots & \vdots \\ P_n(\cdot|\omega_1, \dots, \omega_{n-1}) & \text{W.-Maß auf} & (\Omega_n, \mathbb{P}(\Omega_n)) & \forall \omega_j \in \Omega_j, j = 1, \dots, n-1 \end{array}$$

Wir definieren $\Omega := \times_{j=1}^n \Omega_j = \{(\omega_1, \dots, \omega_n) : \omega_j \in \Omega_j, j = 1, \dots, n\}$ und betrachten den messbaren Raum $(\Omega, \mathbb{P}(\Omega))$. Wir setzen

$$P(\{\omega\}) := \prod_{j=1}^n P_j(\{\omega_j\}|\omega_1, \dots, \omega_{j-1}), \quad \forall \omega = (\omega_1, \dots, \omega_n) \in \Omega, \quad (3.13a)$$

$$P(A) := \sum_{\omega \in A} P(\{\omega\}), \quad \forall A \in \mathbb{P}(\Omega). \quad (3.13b)$$

Für Mengen $B \in \mathbb{P}(\Omega_j)$ bezeichnen mit $A_j(B) = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega : \omega_j \in B\}$ die Menge aller $\omega \in \Omega$ mit Wert im j -ten Eintrag aus B . Insbesondere schreiben wir $A_j(\omega_j)$ falls $B = \{\omega_j\}$ nur ein Element enthält.

Unter diesen Voraussetzungen gilt:

(i) P ist ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathbb{P}(\Omega))$.

(ii) Die bedingten Wahrscheinlichkeiten von P stimmen mit den vorgegebenen bedingten Wahrscheinlichkeiten überein, d.h. $\forall \omega = (\omega_1, \dots, \omega_n) \in \Omega$ und $A \in \mathbb{P}(\Omega_j)$ gilt

$$P(A_j(\omega_j)|A_1(\omega_1) \cap \dots \cap A_{j-1}(\omega_{j-1})) = P_j(A|\omega_1, \dots, \omega_{j-1}) \quad (3.14)$$

sofern $P(A_1(\omega_1) \cap \dots \cap A_{j-1}(\omega_{j-1})) > 0$. Für $j = 1$ ist die linke Seite als $P_1(\{\omega_1\})$ zu lesen.

(iii) Die Gleichung (3.14) legt P eindeutig fest.

Bemerkung 3.9. Satz 3.8 wirkt kompliziert, doch dahinter steht die sehr intuitive Idee des Wahrscheinlichkeitsbaums, vergleiche Abbildung 3.4. Dort ist $\Omega_1 = \{x_1, x_2, x_3\}$, $\Omega_2 = \{y_1, y_2, y_3\}$ und $\Omega_3 = \{z_1, z_2\}$.

3. Bedingte Wahrscheinlichkeiten

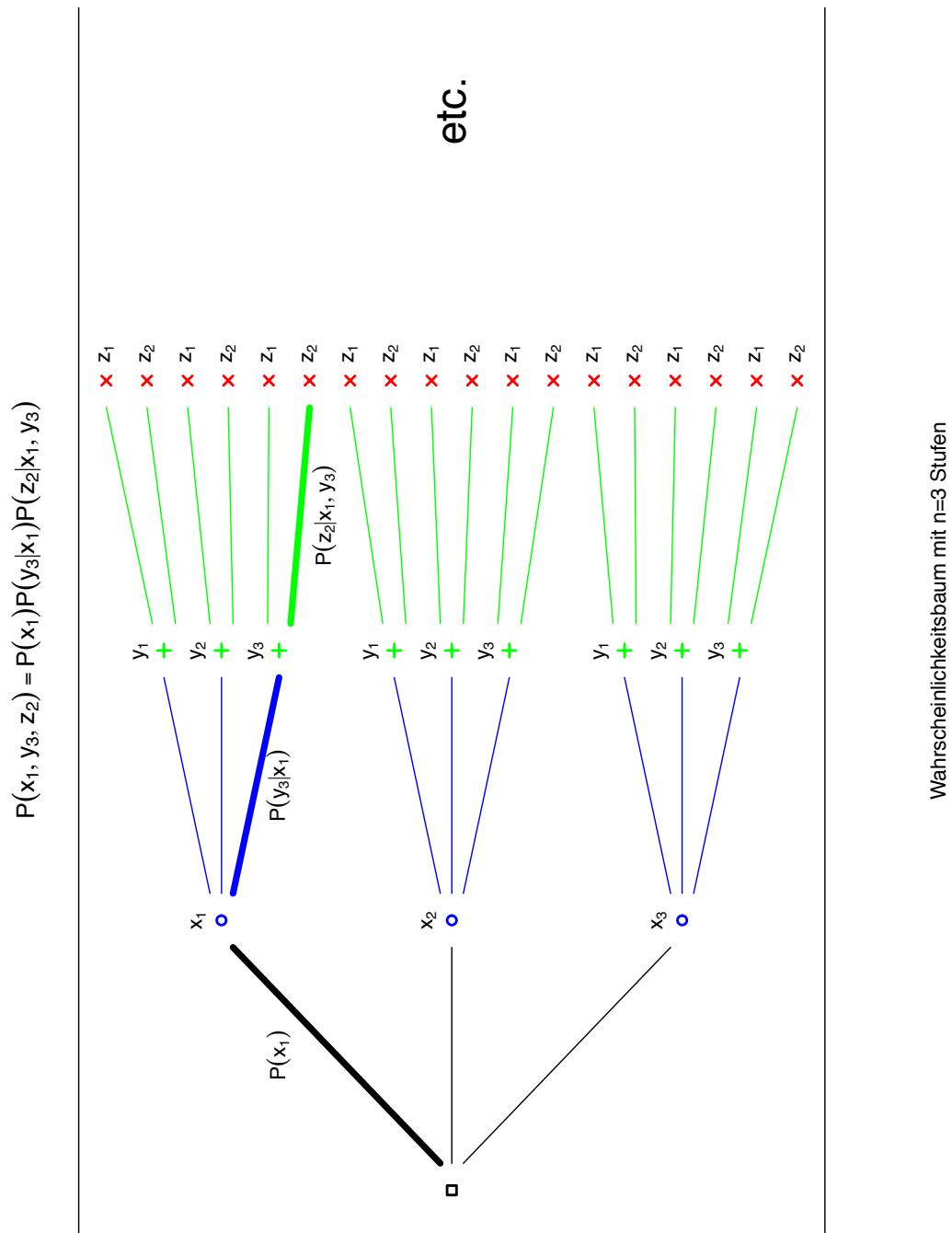


Abbildung 3.4: Wahrscheinlichkeitsbaum.

Beweis von Satz 3.8. (i) Die Additivität (entspricht σ -Additivität wegen $\#\mathbb{P}(\Omega) < \infty$) von P folgt unmittelbar aus (3.13b). Wir zeigen nun die Normiertheit. Für $n = 1$ folgt diese aus der Voraussetzung, dass P_1 ein Wahrscheinlichkeitsmaß ist. Wenn wir nun annehmen, dass die Normiertheit bis zu $n \in \mathbb{N}$ gilt (Induktionshypothese), so folgt mit (3.13a) und (3.13b)

$$\begin{aligned}
P(\Omega) &= \sum_{\omega \in \times_{j=1}^{n+1} \Omega_j} \prod_{j=1}^{n+1} P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) \\
&= \sum_{\omega \in \times_{j=1}^n \Omega_j} \left(\sum_{\omega_{n+1} \in \Omega_{n+1}} P_{n+1}(\{\omega_{n+1}\} | \omega_1, \dots, \omega_n) \right) \prod_{j=1}^n P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) \\
&= \sum_{\omega \in \times_{j=1}^n \Omega_j} \underbrace{P_{n+1}(\Omega_{n+1} | \omega_1, \dots, \omega_n)}_{=1} \prod_{j=1}^n P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) \\
&= 1 .
\end{aligned} \tag{3.15}$$

Im letzten Schritt wurde die Induktionshypothese angewendet. Es bleibt noch zu zeigen, dass $P(A) \in [0, 1]$ für alle $A \in \mathbb{P}(\Omega)$. Aufgrund von (3.13a) und der Nichtnegativität der P_j gilt $P(\{\omega\}) \geq 0$ und mittels (3.13b) ebenfalls $P(A) \geq 0$, für alle $A \in \mathbb{P}(\Omega)$. Verwendet man dies, so erhält man aus der Normiertheit und der σ -Additivität die Aussage $P(A) \leq 1$, also das gewünschte Resultat.

(ii) Mittels einer absteigenden vollständigen Induktion (beginnend mit $l = n$) beweisen wir zunächst, dass für $l = 1, \dots, n$ gilt

$$P(A_1(\omega_1) \cap \dots \cap A_l(\omega_l)) = \prod_{j=1}^l P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) . \tag{3.16}$$

Der Fall $l = n$ ist (3.13a). Ist dieser Sachverhalt bereits für l bewiesen, so gilt er auch für $l - 1$, denn

$$\begin{aligned}
P(A_1(\omega_1) \cap \dots \cap A_{l-1}(\omega_{l-1})) &= \sum_{\omega_l \in \Omega_l} P(A_1(\omega_1) \cap \dots \cap A_l(\omega_l)) \\
&= \sum_{\omega_l \in \Omega_l} \prod_{j=1}^l P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) \\
&= \sum_{\omega_l \in \Omega_l} \underbrace{P_l(\{\omega_l\} | \omega_1, \dots, \omega_{l-1})}_{=1} \times \prod_{j=1}^{l-1} P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) \\
&= \prod_{j=1}^{l-1} P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) .
\end{aligned} \tag{3.17}$$

Um nun (3.14) zu beweisen wählen wir ein $j \in \{1, \dots, n\}$ und betrachten zuerst $A = \{\omega_j\}$. Wir können mittels der Hypothese und der obigen Identität $\prod_{q=1}^{j-1} P_q(\{\omega_q\} | \omega_1, \dots, \omega_{q-1}) > 0$

3. Bedingte Wahrscheinlichkeiten

voraussetzen. Es gilt

$$\begin{aligned}
 P(A_j(\omega_j) | A_1(\omega_1) \cap \dots \cap A_{j-1}(\omega_{j-1})) &= \frac{P(A_1(\omega_1) \cap \dots \cap A_j(\omega_j))}{P(A_1(\omega_1) \cap \dots \cap A_{j-1}(\omega_{j-1}))} \\
 &= \frac{\prod_{q=1}^j P_q(\{\omega_q\} | \omega_1, \dots, \omega_{q-1})}{\prod_{q=1}^{j-1} P_q(\{\omega_q\} | \omega_1, \dots, \omega_{q-1})} \\
 &= P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) .
 \end{aligned} \tag{3.18}$$

Die Gleichung (3.14) für allgemeines $A \in \mathbb{P}(\Omega_j)$ folgt aus der Additivität beider Seiten.

(iii) Es sei P' ein zweites Wahrscheinlichkeitsmaß auf $(\Omega, \mathbb{P}(\Omega))$, welches (ii) erfüllt. Da sowohl P als auch P' die σ -Additivität erfüllen, gilt Gleichheit von P und P' dann und nur dann wenn $P(\{\omega\}) = P'(\{\omega\})$, für alle $\omega \in \Omega$. Wendet man (3.14) mehrmals an, so findet man für $P(\{\omega\})$ den Ausdruck

$$\begin{aligned}
 P(\{\omega\}) &= P\left(\bigcap_{j=1}^n A_j(\omega_j)\right) = P\left(\bigcap_{j=1}^{n-1} A_j(\omega_j)\right) P_n(\{\omega_n\} | \omega_1, \dots, \omega_{n-1}) \\
 &= \dots = \prod_{j=1}^n P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) .
 \end{aligned} \tag{3.19}$$

Da die gleiche Formel auch für $P'(\{\omega\})$ gilt und die rechte Seite nur von den P_j abhängt, so folgt die Aussage. \square

Die große Flexibilität des Wahrscheinlichkeitsbaum-Konzeptes wird durch die folgenden Beispiele illustriert.

Beispiel 3.10.

- (i) n -faches Wiederholen eines Glücksspiels: Es sei $\Omega_0 = \{x_1, \dots, x_q\}$ die Menge der möglichen Ausgänge eines Glücksspiels. Mit $P_0(\{x_j\}) = p_j$, wobei $\sum_{j=1}^q p_j = 1$, definieren wir ein Wahrscheinlichkeitsmaß auf $(\Omega_0, \mathbb{P}(\Omega_0))$. Wir setzen nun $\Omega_j = \Omega_0$ und

$$P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) = P_0(\{\omega_j\}) . \tag{3.20}$$

Die Wahrscheinlichkeiten beim j -ten Wurf hängen also nicht von dem Ausgang des Spielverlaufes für die ersten $j-1$ Spiele ab – ‘neues Spiel neues Glück’. Die Abbildung 3.5 zeigt eine typische Realisation mit $q = 6$, $P_0(\{x_j\}) = \frac{1}{6}$ und $n = 100$. Es dürfte klar werden, warum dieses Signal auch ‘Rauschen’ genannt wird.

- (ii) n -fache Irrfahrt: Wir setzen $\omega_0 = 0$ den Startpunkt der Irrfahrt und möchten nun in jedem Schritt mit der Wahrscheinlichkeit p einen Schritt in die positive Richtung machen, d.h. $\omega_j = \omega_{j-1} + 1$, oder mit der Wahrscheinlichkeit $(1-p)$ einen Schritt in die negative Richtung, d.h. $\omega_j = \omega_{j-1} - 1$. Wir realisieren dies im Rahmen der

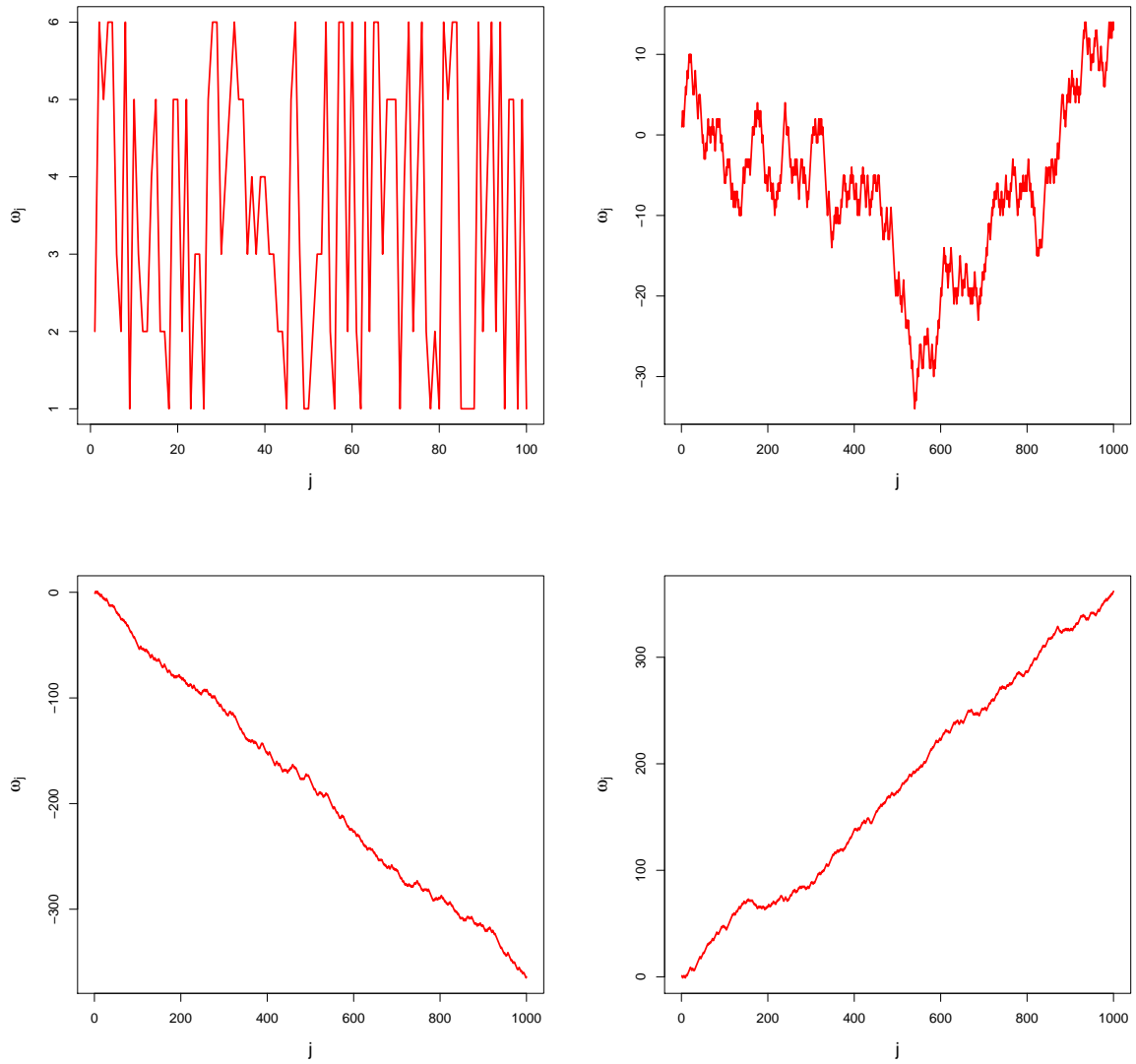


Abbildung 3.5: Links oben: Simulation von 100-fachem Würfeln. Rechts oben: Simulation fairer Irrfahrt ($p = 0.5$) mit 1000 Schritten. Unten: Simulation von 'unfairen' Irrfahrten ($p = 0.3$ links und $p = 0.7$ rechts).

3. Bedingte Wahrscheinlichkeiten

Wahrscheinlichkeitsbäume durch $\Omega_j = \{-j, -j+1, \dots, j-1, j\}$ (ausgestattet mit der Potenzmenge als σ -Algebra) und

$$P_j(\{\omega_j\}|\omega_1, \dots, \omega_{j-1}) = P_j(\{\omega_j\}|\omega_{j-1}) = \begin{cases} p & , \text{ für } \omega_j = \omega_{j-1} + 1 , \\ 1-p & , \text{ für } \omega_j = \omega_{j-1} - 1 , \\ 0 & , \text{ sonst .} \end{cases} \quad (3.21)$$

Die n -fache Irrfahrt können wir auch als die (zufällige) Entwicklung des Gesamtgewinns bei einem Spiel mit gleichen Einsätzen, aber nicht notwendiger Weise gleichen Chancen (falls $p \neq \frac{1}{2}$), verstehen. Siehe Abbildung 3.5 für einige simulierte, typische Spielverläufe für unterschiedliche p .

3.1. Übungsaufgaben

1. Selbsttest

1. Die Wahrscheinlichkeit, dass ein Ereignis A eintritt, beträgt $P(A) = 0.2$. Die Wahrscheinlichkeit, dass ein Ereignis B eintritt, beträgt $P(B) = 0.5$. Gegeben dass A eintritt, beträgt die Wahrscheinlichkeit von B 60%. Wie groß ist die Wahrscheinlichkeit von B , gegeben dass A *nicht* eintritt?

2. Es sei $\Omega = \{1, 2, \dots, n\}$ eine n -elementige Menge und (Ω, \mathcal{A}, P) der dazugehörige Laplacesche Wahrscheinlichkeitsraum. Weiter sei $\emptyset \neq A \subseteq \Omega$ eine nichtleere Teilmenge von Ω .
 - a) Zeigen Sie, dass $P(A) > 0$!

 - b) Betrachten Sie nun die bedingte Wahrscheinlichkeit $P(\cdot | A)$ als ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) , siehe Proposition 3.3. Ist $(\Omega, \mathcal{A}, P(\cdot | A))$ immer ein Laplacescher Wahrscheinlichkeitsraum? Falls nein, was sind die Bedingungen an A so dass dies gilt?

 - c) In Proposition 3.3 hatten wir den messbaren Raum (A, \mathcal{A}_A) definiert. Charakterisieren Sie in unserem Fall (d.h. (Ω, \mathcal{A}, P) Laplacescher Wahrscheinlichkeitsraum) die σ -Algebra \mathcal{A}_A !

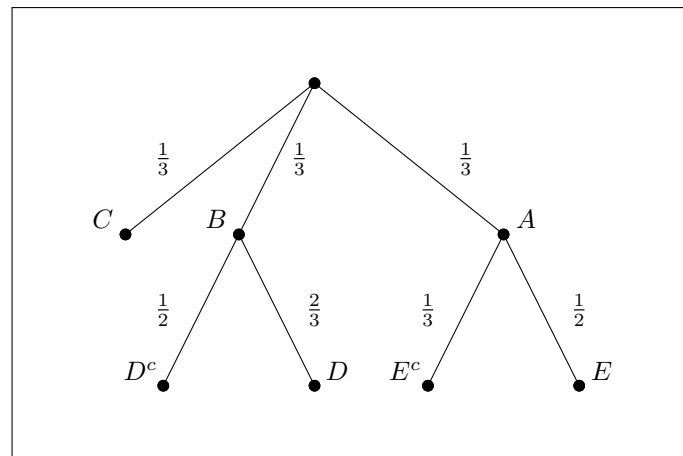
 - d) Ist $(A, \mathcal{A}_A, P(\cdot | A))$ immer ein Laplacescher Wahrscheinlichkeitsraum?

3. Hintereinander würfeln Sie mit einem fairen Würfel ein Mal und werfen eine faire Münze zwei Mal. Das Würfelspiel gewinnen Sie, wenn Sie eine 6 würfeln. Jedes der zwei Münzspiele gewinnen Sie, wenn ‘Kopf’ kommt. Sie sind Gesamtsieger/in, falls Sie von den drei Spielen mindestens zwei Mal gewonnen haben.

Zeichnen Sie einen Wahrscheinlichkeitsbaum und beantworten Sie die Frage, mit welcher Wahrscheinlichkeit Sie zum/zur Gesamtsieger/in werden!

4. Der unten stehende Wahrscheinlichkeitsbaum enthält logische Fehler. Finden und erklären Sie diese!

3. Bedingte Wahrscheinlichkeiten



2. Transfer

1. Als Anti-Dopingexpert/in im Radsport gehen Sie davon aus, dass 80% der Spitzenfahrer gedopt sind. Da die Dopingverfahren jedoch sehr fortgeschritten sind, wird ein gedopter Spitzenfahrer nur in 5% der Fälle durch den Test überführt. Bei einem unschuldigen Fahrer zeigt der Test in 1% der Fälle ein falsch positives Resultat. Beim Test des führenden Fahrers bei der Tour de France erhalten Sie ein negatives Testresultat. Wie groß ist die Wahrscheinlichkeit, dass der Fahrer tatsächlich sauber ist?
2. Manager Danny Dynamic sagt für das kommende Quartal der ABC AG einen Gewinn voraus. In den vorangegangenen Quartalen sagte der Manager allerdings auch für 70% der Quartale, in denen die Firma dann später keinen Gewinn machte, ebenfalls Gewinne voraus. Quartale, welche die Firma später mit Gewinn abschloss, prognostizierte der Manager zu 95% richtig. Die Firma machte in der Vergangenheit im Schnitt nur in einem Viertel aller Quartale einen Gewinn.

Sie sitzen dem Aufsichtsrat vor und lassen Dannys eloquente Powerpoint-Präsentation an sich vorbeirauschen. Unterdessen stellen Sie sich die folgenden Fragen:

- a) Wie hieß noch mal der mathematische Satz mit dem man herausfinden kann, ob Danny hier schon wieder Quatsch erzählt?
- b) Erstellen Sie eine 2×2 -Tabelle, in welche Sie die Wahrscheinlichkeiten für falsche und richtige Prognosen von Danny für Quartale mit und ohne Gewinn eintragen!
- c) Wie groß ist die Wahrscheinlichkeit, dass Danny Ihnen mal wieder Quatsch erzählt? Bei über 50% feuern Sie ihn auf der Stelle!

Sie kennen diese Firma schon lange, daher gehen Sie davon aus, dass sich hier nie etwas ändert und Ihre Erfahrungswerte aus der Vergangenheit auch für die Beurteilung der heutigen Präsentation aussagekräftig sind.

3. Die Panzerknacker starten eine Attacke auf Onkel Dagoberts Geldspeicher. Doch Onkel Dagobert hat vorgesorgt:
- 1) 60% der Eindringlinge werden durch seine Alarmanlagen, die den Vorplatz überwachen, entdeckt und festgenommen. Nach Personalienfeststellung werden sie dann wieder laufen gelassen.
 - 2) Von denen, die über den Vorplatz kommen, schaffen es nur 30% die vordere Stahltür des Geldspeichers zu knacken. Die anderen werden beim Versuch ertappt und landen im Gefängnis.
 - 3) Wenn die Eindringlinge erstmal im Geldspeicher sind, haben sie zwei Möglichkeiten weiterzukommen, 50% gehen über die Treppe und 50% kriechen durch den Lüftungsschacht.
 - 3a) Wenn sie über die Treppe gehen, fallen sie zu 80% in eine der geheimen Falltüren, die über eine Rutsche direkt ins Gefängnis führen.
 - 3b) Klettern sie durch den Lüftungsschacht, dann bleiben sie zu 50% darin hängen, wo sie Dagoberts Neffe Donald am nächsten Morgen beim Geldspeicher Putzen findet, befreit und einfach draußen laufen lässt.

Wer die Treppe oder den Lüftungsschacht überwindet, erbeutet das Geld von Onkel Dagobert.

Bitte lösen Sie die folgenden Aufgaben:

- a) Zeichnen Sie den Wahrscheinlichkeitsbaum für die Attacke der Panzerknacker auf den Geldspeicher.
- b) Wie groß ist die Wahrscheinlichkeit, dass die Panzerknacker das Geld erbeuten?
- c) Wie groß ist die Wahrscheinlichkeit, dass die Panzerknacker im Gefängnis landen?
- d) Die Panzerknacker haben das Geld geklaut. Mit welcher Wahrscheinlichkeit kamen sie über die Treppe?

3.2. Literatur

Mit die ersten Sterbetafeln wurden in London um 1660 aufgestellt, die Interpretation über bedingte Wahrscheinlichkeiten mißlang jedoch [[Hac07](#), Kapitel 12].

Das Bayes-Theorem wurde nach dem Tod von Thomas Bayes (1702-1761) im Jahre 1763 veröffentlicht. Bis heute beeinflußt dieser Essay stark den nach Bayes benannten Ansatz zur Interpretation von Wahrscheinlichkeiten, der starkes Gewicht auf den subjektiven Wahrscheinlichkeitsbegriff legt. Ein modernes, von Bayes inspiriertes Buch ist z.B. [[Rob07](#)].

Wahrscheinlichkeitsbäume sind heute in den Wirtschaftswissenschaften und der Risikotheorie sowie der Statistik sehr verbreiten, siehe z.B. [[FHT96](#), Kapitel 5.4].

4. Markov-Ketten

Markov-Ketten sind einfache Beispiele für stochastisch ablaufende Prozesse, d.h. mehrstufige und zufällige Ereignisse, die sich mit der Zeit entwickeln. Zum Glück kann man diese bereits mit den uns zur Verfügung stehenden Mitteln ganz gut beschreiben. Das Gebiet der stochastischen Prozesse im Allgemeinen ist riesig und auch zu Markov-Ketten gibt es viel Literatur. Deshalb werden wir hier als Einblick in diese Thematik nur eine Darstellung der wichtigsten Sachverhalte geben, mit einem Schwerpunkt auf der numerischen Simulation.

Definition 4.1 (Markov-Ketten). Es sei $\Omega_0 = \{x_1, \dots, x_q\}$ eine endliche Menge, genannt der **Zustandsraum**.

- (i) Eine **Anfangsverteilung** ist ein Vektor $\rho = (\rho_k)_{k=1, \dots, q}$ mit $\rho_k \in [0, 1]$ und $\sum_{k=1}^q \rho_k = 1$.
- (ii) Eine Matrix $p = (p_{kl})_{k, l=1, \dots, q}$ heißt **stochastische Matrix**, falls $p_{kl} \in [0, 1]$ und jede Zeilensumme den Wert 1 ergibt, d.h.

$$\sum_{l=1}^q p_{kl} = 1, \quad \forall k = 1, \dots, q. \quad (4.1)$$

- (iii) Eine **Markov-Kette** (der Schrittlänge n) ist ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , welcher aus einem Zustandsraum Ω_0 , einer Anfangsverteilung ρ und stochastischen Matrizen $p^{(j)}$, $j = 1, \dots, n$, wie folgt konstruiert ist: Wir setzen $\Omega = \times_{j=0}^n \Omega_j = \Omega_0^{\times(n+1)}$ und $\mathcal{A} = \mathbb{P}(\Omega)$. Mittels der Anfangsverteilung ρ und den stochastischen Matrizen $p^{(j)}$ definieren wir Wahrscheinlichkeitsmaße auf $(\Omega_j, \mathbb{P}(\Omega_j))$ durch $P_0(\{x_k\}) := \rho_k$ und

$$P_j(\{\omega_j\} | \omega_1, \dots, \omega_{j-1}) := P_j(\{\omega_j\} | \omega_{j-1}) := p_{kl}^{(j)}, \quad \text{für } \omega_j = x_l \text{ und } \omega_{j-1} = x_k, \quad j \geq 1. \quad (4.2)$$

Wir wählen auf (Ω, \mathcal{A}) das aus P_0 und (4.2) eindeutig bestimmte Wahrscheinlichkeitsmaß P , siehe Satz 3.8.

- (iv) Die Menge Ω einer Markov-Kette (Ω, \mathcal{A}, P) bezeichnet man auch als **Pfadraum** der Markov-Kette.
- (v) Falls die Übergangsmatrizen $p^{(j)} = p$ nicht von j abhängen, so spricht man von einer **stationären Markov-Kette**.

Bemerkung 4.2. Die Einträge der stochastischen Matrizen $p_{kl}^{(j)}$ kann man als Übergangswahrscheinlichkeiten (im Zeitschritt j) von Zustand x_k in Zustand x_l auffassen, vergleiche Abbildung 4.1.

4. Markov-Ketten

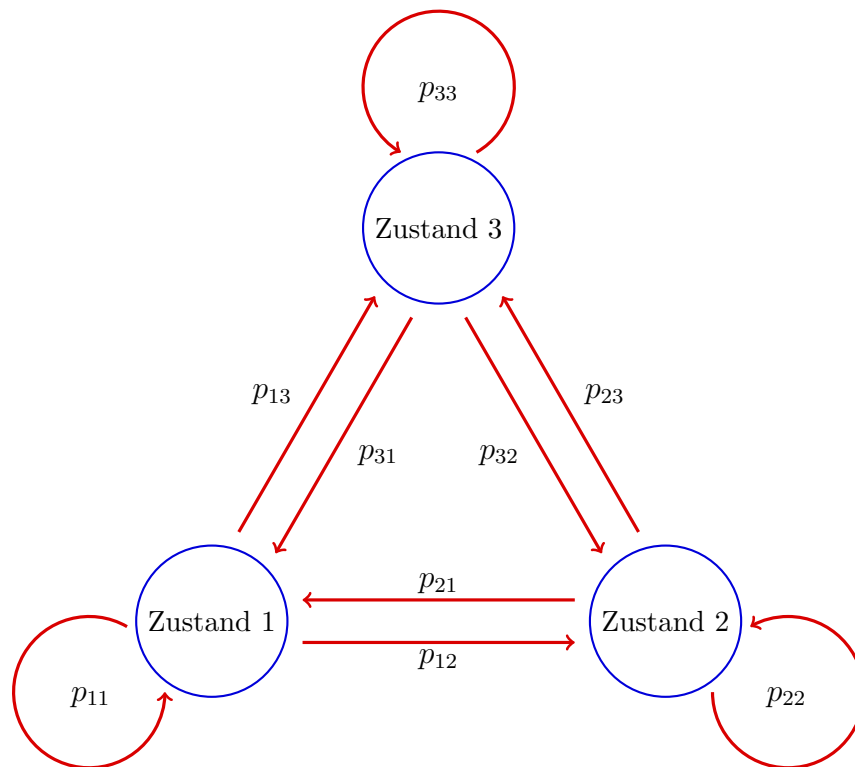
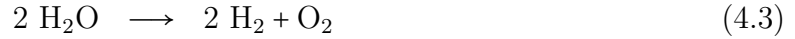


Abbildung 4.1: Schematische Darstellung einer Markov-Kette.

Beispiel 4.3 (Wasserstoffherstellung). Mittels Elektrolyse kann man Wasser (H_2O) in Wasserstoff (H_2) und Sauerstoff (O_2) gemäß der folgenden Reaktion umwandeln:



Die Gegenreaktion ist die sogenannte Knallgasreaktion:



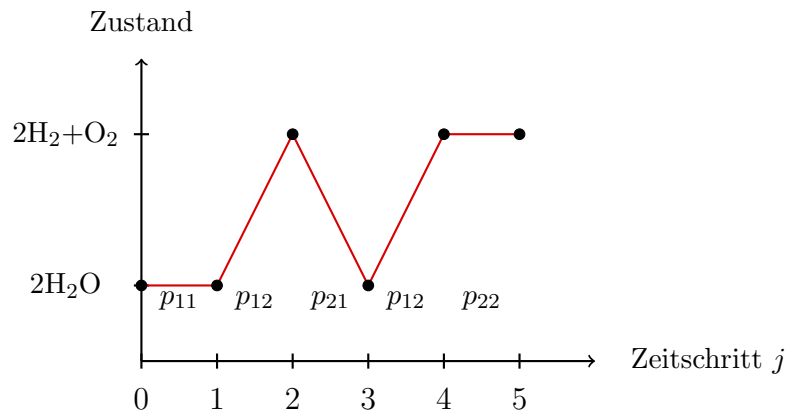
Wir betrachten den Zustandsraum $\Omega_0 = \{x_1, x_2\} = \{\text{“}2 \text{H}_2\text{O}\text{”}, \text{“}2 \text{H}_2 + \text{O}_2\text{”}\}$, also die zwei möglichen Zustände unserer chemischen Substanz. Durch die Reaktionen können nun Übergänge zwischen den einzelnen Zuständen auftreten (wir nehmen vereinfachend an, dass diese jeweils nur in diskreten Zeitschritten j statt finden). Diese Übergänge lassen sich durch eine stochastische Matrix beschreiben:

$$p = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \quad (4.5)$$

Hierbei ist p_{11} die Wahrscheinlichkeit, dass sich während eines Zeitschrittes “ $2 \text{H}_2\text{O}$ ” nicht umwandelt, und p_{12} die Wahrscheinlichkeit, dass sich während eines Zeitschrittes “ $2 \text{H}_2\text{O}$ ” in “ $2 \text{H}_2 + \text{O}_2$ ” umwandelt. Analog ist p_{21} die Wahrscheinlichkeit für den Übergang von “ $2 \text{H}_2 + \text{O}_2$ ” nach “ $2 \text{H}_2\text{O}$ ” und p_{22} die Wahrscheinlichkeit, dass sich “ $2 \text{H}_2 + \text{O}_2$ ” nicht umwandelt. Dementsprechend hängt die obere Zeile der stochastischen Matrix von der Reaktionsrate von (4.3) ab und die untere Zeile von der Reaktionsrate von (4.4).

Hat man nun ein bestimmtes Gemisch der beiden Substanzen gegeben, so bestimmt man eine Anfangsverteilung $\rho = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$ indem man für ρ_1 den relativen Anteil von “ $2 \text{H}_2\text{O}$ ” und für ρ_2 den relativen Anteil von “ $2 \text{H}_2 + \text{O}_2$ ” in dem Gemisch wählt. Es gilt offensichtlich $\rho_1 + \rho_2 = 1$.

Die stationäre Markov-Kette (Ω, \mathcal{A}, P) , welche sich aus den obigen Daten ergibt, modelliert nun die zeitliche, zufällige Entwicklung des relativen Anteils der einzelnen Substanzen im Gemisch. Ein typischer Pfad für diese Markov-Kette (mit Länge $n = 5$) sieht wie folgt aus



4. Markov-Ketten

Dieser Pfad ist ein zufälliges Ereignis, welchem man die Wahrscheinlichkeit $\rho_1 p_{11} p_{12} p_{21} p_{12} p_{22}$ zuweisen kann.

Wir fokussieren uns nun auf die Simulation von Markov-Ketten. Untenstehend ist der Simulationscode für stationäre Markov-Ketten mit der Stochastik-Programmiersprache R. Zuerst definieren wir die grundlegenden Bestandteile der Markov-Kette. Den Zustandsraum identifizieren wir hierbei mit den Zahlen $\{1, \dots, 6\}$. Da wir keine Lust haben, die stochastische Matrix und die Anfangsverteilung einzutippen, simulieren wir diese als Zufallszahlen:

```
# generate transition probabilities and initial distribution randomly

Omega0=1:6                                # define a state space -
                                           # must be coded as 'integer'

qState=length(Omega0)                     # determine number of states
rho0=runif(qState)                         # qState uniform [0, 1] random no's
rho0=rho0/sum(rho0)                       # normalize
p=matrix(runif(qState*qState),ncol=qState) # generate stoch matrix
for(i in 1:qState) p[i,]=p[i,]/sum(p[i,]) # normalize rows
n=100                                      # set length of markov chains
```

Nun definieren wir den eigentlichen Simulationsalgorithmus, führen die Simulation durch und visualisieren die Resultate (siehe Abbildung 4.2):

```
# simulate Markov chain

markovChain=function(rho0,p,n,Omega0){    # define function
  omega=integer(n)                        # initialize
  omega[1]=sample(Omega0,1,prob=rho0)     # init. value ~ rho0
  for(i in 2:n)
    omega[i]=sample(Omega0,1,prob=p[omega[i-1],]) # simulation loop
  return(omega)                           # return path of markov chain
}                                          # end of function body

omega=markovChain(rho0,p,n,Omega0)        # simulate one path

# visualize

plot(omega,type="l",lwd=2,col="red")
```

Beachten Sie, dass die erste Gleichung in (4.2) die Ursache von einer erheblichen Komplexitätsreduktion ist. Diese wichtige Eigenschaft hat deshalb einen eigenen Namen bekommen und wird die **Markov-Eigenschaft** eines mehrstufigen stochastischen Modells (bzw. einer Zeitreihe, was mathematisch fast dasselbe ist) genannt. Interpretieren wir die Stufe j

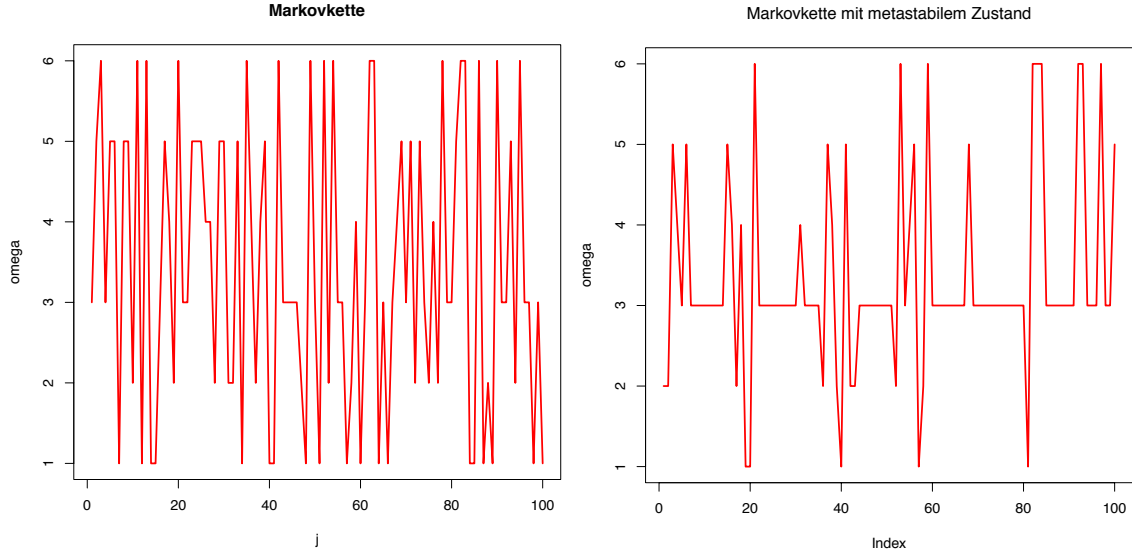


Abbildung 4.2: Zwei simulierte Pfade einer Markov-Kette mit Zustandsraum $\{1, \dots, 6\}$.

des mehrstufigen Modells als Zeit, so bedeutet die Markov-Eigenschaft, dass man über die Wahrscheinlichkeiten für die zukünftigen Zustände der Markov-Kette nicht mehr Information bekommen kann, als durch den ‘gegenwärtigen’ Zustand ω_{j-1} gegeben. Insbesondere spielt die Vergangenheit vor $j - 1$ keine Rolle, wenn wir ω_{j-1} erst einmal kennen.

Diese Eigenschaft der ‘Gedächtnislosigkeit’ spielt eine wichtige Rolle in der Modellierung physikalischer Prozesse. Doch auch in der *theoretischen* Behandlung von effektiven Finanz- und Aktienmärkten geht man davon aus, dass alle Information der Vergangenheit im gegebenen Preis ‘eingepreist’ ist und der Wertentwicklungsprozess eines Renten- oder Aktienpapiers deswegen die Markov-Eigenschaft haben sollte. Wie realistisch die Annahme effektiver Märkte ist, sollte natürlich im Einzelfall überprüft werden.

Betrachtet man mehrere Schritte auf einmal, so kann man praktischer Weise die Übergangswahrscheinlichkeiten für $m = 0, \dots, n - 1$ Schritte mit Hilfe der linearen Algebra berechnen.

Lemma 4.4. *Es sei $\Omega = \{x_1, \dots, x_q\}$ der Zustandsraum, $\rho = (\rho_k)_{k=1, \dots, q}$ die Anfangsverteilung und $p = (p_{kl})_{k,l=1, \dots, q}$ die stochastische Matrix einer stationären Markov-Kette (Ω, \mathcal{A}, P) . Mit $A_j(x)$ bezeichnen wir die ‘Pfade’ $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ der Markovkette, die im j -ten Eintrag den Wert x annimmt, $\omega_j = x$. Dann gilt für die Verteilung der Markov-Kette nach m Schritten $\rho^{(m)} := (\rho_k^{(m)})_{k=1, \dots, q}$, definiert durch $\rho_k^{(m)} := P(A_m(x_k))$, $k = 1, \dots, q$, die Gleichung*

$$\rho^{(m)} = \rho p^m. \quad (4.6)$$

Beweis. Einmal mehr erfolgt der Beweis durch vollständige Induktion. Der Induktionsanfang $\rho^{(0)} = \rho$ ist gemäß Satz 3.8 und Definition 4.1 trivial.

Es gelte (4.6) für m . Dann gilt nach der Fallunterscheidungsformel in Satz 3.4 (i) und

4. Markov-Ketten

der Induktionshypothese für $k = 1, \dots, q$

$$\begin{aligned}
 \rho_k^{(m+1)} &= P(A_{m+1}(x_k)) \\
 &= \sum_{l=1}^q P(A_{m+1}(x_k) | A_m(x_l)) P(A_m(x_l)) \\
 &= \sum_{l=1}^q p_{lk} \rho_l^{(m)} = \sum_{l=1}^q [\rho p^m]_l p_{lk} = [\rho p^{m+1}]_k \quad \square
 \end{aligned} \tag{4.7}$$

Wir setzen nun die Simulation fort¹, indem wir eine Markov-Kette in unterschiedlichen Anfangszuständen starten und dann zählen, wie oft die Markov-Kette in welchem Zustand gewesen ist. Indem wir dies durch die Anzahl der Simulationsschritte teilen bekommen wir die relative Aufenthaltshäufigkeit für die Zustände. Diese visualisieren wir in Abbildung 4.3 für große Schrittzahlen $n = 100000$.

```

# simulations for asymptotic behavior of markov chains
n=100000                                # set length of markov chain
rho0=c(1,0,0,0,0,0)                   # start in state 1
omega= markovChain(rho0,p,n)           # simulate markov chain
table(omega)/n                         # show rel. occupation freq.

# visualize

par(mfrow=c(1,2))                     # double display
barplot(table(omega)/n,main="Relative Häufigkeit
des Aufenthalts\n bei Start in Zustand 1",col="red") # plot rel. occupation freq.
rho0=c(0,0,0,0,0,1)                   # change initial conditions
omega= markovChain(rho0,p,n,Omega0)    # simulate markov chain again
barplot(table(omega)/n,main="Relative Häufigkeit
des Aufenthalts \n bei Start in Zustand 6",col="red") # plot rel. occupation freq.
par(mfrow=c(1,1))                     # reset display

```

Bemerkung 4.5. Die große Ähnlichkeit der beiden Panels in Abbildung 4.3 ist bemerkenswert in zweierlei Hinsicht: Zum Einen hängt die relative Aufenthaltshäufigkeit nicht vom gewählten Anfangspunkt, zum Anderen nicht von der (unterschiedlichen) Simulationshistorie ab. Beachten Sie, dass die Balkenhöhen von der zufällig generierten Matrix p abhängen.

Nun möchten wir die Simulation ein weiteres Mal fortsetzen. Anstelle der relativen Aufenthaltshäufigkeiten bei *einer* Simulation der Markov-Kette über ihren Verlauf betrachten wir nun die relative Häufigkeit des Aufenthalts nach einer gegebenen, großen Anzahl von Schritten, indem wir die Markov-Kette *sehr oft* simulieren und dann jeweils am Endpunkt auswerten. Dabei halten wir den Anfangspunkt fest:

¹Achten Sie darauf, dass die zuvor erzeugten Objekte noch nicht aus dem R-workspace gelöscht wurden!

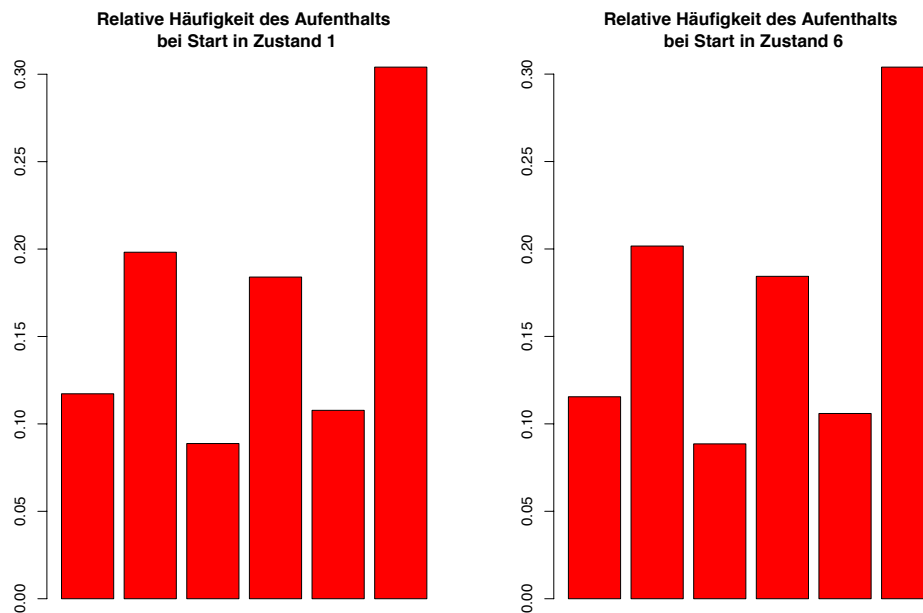


Abbildung 4.3: Die relativen Aufenthaltshäufigkeiten der simulierten Markov-Kette (mit Länge $n = 100000$) in den Zuständen $\{1, \dots, 6\}$ mit unterschiedlichen Startpunkten.

```
# simulation of ensemble average

nSim=10000                                # number of simulations
n=1000                                    # number of steps in MC
mcFinalState=integer(nSim)                # initialize

for(i in 1:nSim){                          # loop nSim simulations
  omega=markovChain(rho0,p,n,Omega0)      # one simulation
  mcFinalState[i]=tail(omega,1)           # read final state and store
}

# calculate probs according to (4.6)
rho0=c(0,0,0,0,0,1)                      # fix initial state to 1
rhoS=rho0                                 # initialize rhoS as rho0
for(i in 1:1000) rhoS=as.numeric(rhoS**p) # calculate rhoS by matrix mult.
```

4. Markov-Ketten

```
# visualization

par(mfrow=c(1,2))                # double display
barplot(tabulate(mcFinalState)/n,
main="Relative Häufigkeit
des Aufenthalts\n nach 1000 Schritten
\n in der Simulation",col="red")  # plot rel.  freq. of simulation
barplot(rhoS,
main="Aufenthaltswahrscheinlichkeit\n
nach Formel (4.6)",col="red")      # simulate results from matrix mult.
par(mfrow=c(1,1))                # reset display
```

Wieder ergibt sich (fast) dieselbe Verteilung auf die einzelnen Zustände wie in Abbildung 4.3. Werden solche Übereinstimmungen ‘entdeckt’, so lässt es vielen Mathematiker/innen oft keine Ruhe mehr, bis der dahinter steckende mathematische Sachverhalt aufgeklärt ist. Dabei gilt es zunächst, die richtigen Begrifflichkeiten zu finden.

Definition 4.6. Es sei $\Omega_0 = \{x_1, \dots, x_q\}$ ein Zustandsraum, ρ eine Anfangsverteilung und p eine stochastische Matrix.

- (i) Ein Wahrscheinlichkeitsmaß P^S auf $(\Omega_0, \mathbb{P}(\Omega_0))$ heißt **stationärer Zustand** der durch Ω_0 , ρ und p definierten stationären Markov-Kette, falls der zugehörige Vektor $\rho^S = (\rho_k^S)_{k=1, \dots, q}$, wobei $\rho_k^S = P^S(\{x_k\})$, $k = 1, \dots, q$, ein Linkseigenvektor von p zum Eigenwert 1 ist, d.h.

$$\rho^S = \rho^S p . \quad (4.8)$$

- (ii) Eine Markov-Kette heißt **stark mischend**, falls ein eindeutiger stationärer Zustand ρ^S existiert, so dass für alle Anfangsverteilungen ρ die Verteilungen nach n Schritten gegen diesen stationären Zustand konvergieren, also nach Lemma 4.4

$$\lim_{n \rightarrow \infty} \rho p^n = \rho^S \quad \forall \rho \in [0, 1]^{\times q} , \quad \sum_{k=1}^q \rho_k = 1 . \quad (4.9)$$

Bemerkung 4.7. Genau genommen müssten wir in der Aussage (ii) eine Familie von Markov-Ketten mit n Schritten, für alle $n \in \mathbb{N}$, betrachten. Alternativ kann man gleich eine Markov-Kette mit beliebiger Schrittzahl definieren. Für die letztere Lösung benötigt man jedoch die Fortsetzungssätze aus der Maßtheorie, die wir erst später erarbeiten werden.

Die in Abbildung 4.3 gezeigten Simulationsresultate lassen vermuten, dass relativ allgemeine Bedingungen hinreichend für die Mischeigenschaft einer Markov-Kette sind. Man kann zeigen, dass das Mischen stets auch zur **Ergodizität** der Markov-Kette führt, d.h. dass mit Wahrscheinlichkeit 1 die relativen Häufigkeiten des Aufenthalts der Markov-Kette in Zustand x_k gegen ρ_k^S konvergieren. Wir sehen hier von einem mathematischen Beweis ab und konzentrieren uns auf hinreichende Bedingungen für die Mischeigenschaft.

Satz 4.8. Es sei $p = (p_{kl})_{k,l=1,\dots,q}$ eine elementweise positive stochastische Matrix, d.h. $p_{kl} > 0$ für alle $k, l = 1, \dots, q$. Dann ist die zugehörige Markov-Kette stark mischend.

Beweis. Wir betrachten den Raum der Anfangsverteilungen

$$V := \left\{ \rho = (\rho_k)_{k=1,\dots,q} : \rho_k \in [0, 1] , \quad \sum_{k=1}^q \rho_k = 1 \right\} \quad (4.10)$$

mit der Metrik, die durch die Summennorm auf \mathbb{R}^q erzeugt ist, also

$$\|\rho - \rho'\| = \sum_{k=1}^q |\rho_k - \rho'_k| . \quad (4.11)$$

V ist bezüglich dieser Metrik ein vollständiger Raum.

Die Rechtsmultiplikation mit der stochastischen Matrix p induziert eine Abbildung $T : V \rightarrow V$, $\rho \mapsto T(\rho) = \rho p$. Wir wollen zeigen, dass T eine kontrahierende Abbildung ist, d.h. $\|T(\rho) - T(\rho')\| \leq C \|\rho - \rho'\|$ mit $C < 1$. Denn dann gilt nach dem Banachschen Fixpunktsatz, dass ein eindeutig bestimmter Fixpunkt $\rho^S \in V$ existiert, d.h. $T(\rho^S) = \rho^S p = \rho^S$, so dass $T^n(\rho) = \rho p^n \rightarrow \rho^S$ für $n \rightarrow \infty$.

Sei hierzu $\epsilon = \min\{p_{kl} : k, l = 1, \dots, q\} > 0$. Dann gilt

$$\begin{aligned} \|T(\rho) - T(\rho')\| &= \sum_{k=1}^q \left| \sum_{l=1}^q (\rho_l - \rho'_l) p_{lk} \right| \\ &= \sum_{k=1}^q \left| \sum_{l=1}^q (\rho_l - \rho'_l) [\epsilon + (p_{lk} - \epsilon)] \right| \\ &= \sum_{k=1}^q \left| \sum_{l=1}^q (\rho_l - \rho'_l) (p_{lk} - \epsilon) \right| \\ &\leq \sum_{k=1}^q \sum_{l=1}^q |\rho_l - \rho'_l| (p_{lk} - \epsilon) \\ &= \sum_{l=1}^q |\rho_l - \rho'_l| \sum_{k=1}^q (p_{lk} - \epsilon) \\ &= (1 - q\epsilon) \sum_{l=1}^q |\rho_l - \rho'_l| = (1 - \epsilon q) \|\rho - \rho'\| . \end{aligned} \quad (4.12)$$

Im dritten Schritt haben wir ausgenutzt, dass $\sum_{l=1}^q \epsilon(\rho_l - \rho'_l) = \epsilon(1 - 1) = 0$ und im vorletzten Schritt wurde verwendet, dass p stochastisch ist und somit insbesondere Zeilensumme 1 hat. \square

Bemerkung 4.9. Die Voraussetzungen von Satz 4.8 können dahingehend abgeschwächt werden, dass ein $k \in \mathbb{N}$ existiert, so dass p^k elementweise positiv ist.

4.1. Übungsaufgaben

1. Selbsttest

1. Eine Markov-Kette hat zwei Zustände x_1 und x_2 und startet in x_1 , d.h. der Anfangszustand ist $\rho = (1, 0)$. Die Übergangswahrscheinlichkeiten p_{kl} für den Übergang $k \rightarrow l$ sind gegeben durch die folgende stochastische Matrix

$$p = (p_{kl})_{k,l=1,2} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

Wie groß ist die Wahrscheinlichkeit, dass sich die Markov-Kette nach zwei Schritten in Zustand x_2 befindet?

2. Es sei $\Omega_0 = \{x_1, x_2\}$ ein 2-elementiger Zustandsraum, $\rho = (\rho_1, \rho_2)$ ein Vektor und

$$p = \begin{pmatrix} 1 - \epsilon_1 & \epsilon_2 \\ \epsilon_3 & 1 - \epsilon_4 \end{pmatrix}$$

eine Matrix, wobei $\rho_1, \rho_2, \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 \in \mathbb{R}$.

- a) Unter welchen Bedingungen spezifiziert der Vektor ρ einen Anfangszustand und die Matrix p eine stochastische Matrix? Nehmen Sie diese Bedingungen dann im Folgenden immer an!
- b) Hat die Markov-Kette, welche aus den obigen Daten bestimmt ist, einen stationären Zustand? Wenn ja, dann berechnen Sie diesen!
- c) Setzen Sie nun alle ϵ_i gleich, d.h. $\epsilon_i = \epsilon \ \forall i$. Starten Sie mit einem beliebigen Anfangszustand ρ und bestimmen Sie den Zustand nach n Zeitschritten! Konvergiert dieser Ausdruck für $n \rightarrow \infty$? Wenn ja, gegen welchen Zustand?

Hinweis: Diagonalisieren Sie die Matrix p .

2. Transfer

Ein Lebewesen kann sich in den folgenden drei Zuständen befinden: ‘gesund’, ‘krank’ und ‘tot’. Zwischen diesen Zuständen kann es Übergänge geben, z.B. kann ein krankes Lebewesen sich wieder heilen. Wir nehmen an, dass die Übergänge zufällig sind und nur zu diskreten Zeitschritten stattfinden können, und wir möchten diese Problem mittels einer stationären Markov-Kette beschreiben.

- a) Formulieren Sie den Zustandsraum für dieses Problem!
- b) Bestimmen Sie eine geeignete stochastische Matrix p , in welcher Sie den unmöglichen Übergängen schon die Übergangswahrscheinlichkeit 0 zuweisen!
- c) Mit welcher Anfangsverteilung ρ wird ein mit Sicherheit lebendiges Lebewesen beschrieben?

- d) Zeichnen Sie einen typischen Pfad für diese Markov-Kette! Gibt es offensichtliche stationäre Zustände in dieser Markov-Kette?
- e) Es seien nun die folgenden Übergangswahrscheinlichkeiten gegeben: Ein gesundes Lebewesen erkrankt mit Wahrscheinlichkeit 25% und stirbt eines plötzlichen Todes mit 5%. Ein kranken Lebewesen erfreut sich einer Genesung mit 45% und stirbt mit Wahrscheinlichkeit 25%. Wie groß ist die Wahrscheinlichkeit, dass in lebendiges Lebewesen noch zwei Zeitschritten tot ist?

Bemerkung: Dieses Spielzeug-Modell sollte Ihnen veranschaulichen, dass Markov-Ketten in der Tat in der Praxis von großen Nutzen sein können. Zum Beispiel ist es in der Versicherungsmathematik eine zentrale Aufgabe, Aussagen über die Invalidisierungs- und Sterbewahrscheinlichkeiten von Kunden zu treffen. Unser Spielzeug-Modell ist somit ein erster Schritt in diese Richtung, da es erlaubt die zeitliche Entwicklung von einigen Risiken ('krank' oder 'tot') zu beschreiben.

3. Theorie

Verallgemeinern Sie das Resultat aus Satz 4.8 dahingehend, dass Sie Bemerkung 4.9 mit einbeziehen. Explizit: Beweisen Sie, dass für eine stochastische Matrix p , so dass p^k elementweise positiv ist für ein $k \in \mathbb{N}$, die dazugehörige Markov-Kette stark mischen ist.

4. Aktivität

Spielen Sie mit dem R-code, indem Sie weitere Matrizen p und Anfangszustände ρ definieren². Diskutieren Sie in Kleingruppen, was gemeint ist, und finden Sie Beispiele für stochastische Matrizen und Anfangsverteilungen für die folgenden Fälle:

- a) Eine 'deterministische' Markov-Kette;
- b) Eine Markov-Kette mit 'Friedhof';
- c) Eine Markov-Kette mit 'metastabilem Zustand';
- d) Eine Markov-Kette 'im Gleichgewicht'.

Überprüfen Sie Ihre Lösung durch Simulation und Visualisierung! Bei dieser Aufgabe herrscht Internetverbot!

²Studieren Sie die folgenden Hilfeseiten `R>?matrix`, `R>?c`, insbesondere die Beispiele unten auf den Hilfeseiten.

4.2. Literatur

Die Markov-Ketten, A. A. Markov (1856-1922), wurden zunächst als Beispiele für Grenzwertsätze mit nicht-unabhängigen Zufallsvariablen verstanden [Mar06], vergleichen Sie z.B. mit Kapitel 14. und 15. in diesem Buch. Später wurden sie u.a. von L. Bachelier (1870-1946) und A. Einstein (1879-1955) zu und anderen zu Markov-Prozessen verallgemeinert, siehe z.B. [Bau91], [Geo09] oder [Kle06]. Die entgeltige

Für Markov-Ketten existieren unzählige Anwendungen, siehe z.B. [Pri13] für einen aktuellen Überblick.

5. Unabhängigkeit

Eine sehr einleuchtende Formulierung von Unabhängigkeit besagt, dass ein Ereignis B nicht von einem Ereignis A abhängt, wenn sich die Eintrittswahrscheinlichkeit von B , gegeben dass A eingetreten ist, nicht von der ursprünglichen Wahrscheinlichkeit für B unterscheidet. Dies würde für den Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) auf die folgende Definition hinauslaufen:

$$A, B \in \mathcal{A}, P(A) > 0, B \text{ unabhängig von } A \Leftrightarrow P(B|A) = P(B) . \quad (5.1)$$

Die rechte Seite von (5.1) kann äquivalent umgeformt werden zu

$$P(A \cap B) = P(A)P(B) . \quad (5.2)$$

Diese Formulierung macht auch dann noch Sinn, wenn $P(A) = 0$ ist, und bewahrt uns daher vor lästigen Fallunterscheidungen. Deshalb legen wir (5.2) unserer mathematischen Definition von Unabhängigkeit zugrunde.

Definition 5.1 (Unabhängigkeit). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

- (i) Zwei Ereignisse $A, B \in \mathcal{A}$ heißen **unabhängig**, falls $P(A \cap B) = P(A)P(B)$ gilt.
- (ii) Es sei $\{A_j\}_{j \in J} \subseteq \mathcal{A}$ eine Familie von Ereignissen mit J einer beliebigen Indexmenge. Dann heißt die Familie $\{A_j\}_{j \in J}$ eine **Familie unabhängiger Ereignisse**, wenn für alle $I \subseteq J$ endlich, d.h. $\#I < \infty$, gilt

$$P\left(\bigcap_{j \in I} A_j\right) = \prod_{j \in I} P(A_j) . \quad (5.3)$$

- (iii) Es seien $\{\mathcal{A}_j\}_{j \in J}$ Mengensysteme $\mathcal{A}_j \subseteq \mathcal{A}$. Dann sind *per definitionem* diese **Mengensysteme unabhängig** von einander, wenn für jede Wahl $A_j \in \mathcal{A}_j$, $j \in J$, die Familie von Ereignissen $\{A_j\}_{j \in J}$ eine Familie unabhängiger Ereignisse im Sinne von (ii) ist.

Bemerkung 5.2. Für die Gültigkeit von (ii) reicht im Allgemeinen nicht aus, dass (i) für je zwei unterschiedliche Ereignisse aus $\{A_j\}_{j \in J} \subseteq \mathcal{A}$ gilt. Sei z.B. J eine endliche Menge mit n Elementen, so beinhaltet (ii) eine Gleichung für jede der 2^n Teilmengen I von J . Die Gleichung für $I = \emptyset$ ist hierbei trivial. Also ergeben sich $2^n - 1$ nichttriviale Bedingungen. Paarweise Unabhängigkeit im Sinne von (i) für alle Paare A, B , $A \neq B$, ergibt jedoch nur $\binom{n}{2}$ Bedingungen, was im Allgemeinen eine viel kleinere Anzahl ist.

Lemma 5.3. Sei $\{A_j\}_{j \in J}$ eine Familie unabhängiger Ereignisse aus einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

5. Unabhängigkeit

- (i) Dann ist für $k \in J$ beliebig $\{A'_j\}_{j \in J}$ mit $A'_k = A_k^c$ und $A'_j = A_j$, für $j \neq k$, ebenfalls eine Familie unabhängiger Ereignisse.
- (ii) Die Familie bleibt auch dann Familie unabhängiger Ereignisse, wenn in (i) endlich viele Ereignisse A_j durch ihr Komplement ersetzt werden.
- (iii) Ist $\{A_j\}_{j \in J}$ Familie unabhängiger Ereignisse, so gilt dies auch für $\{A_j\}_{j \in J} \cup \{\emptyset, \Omega\}$.

Beweis. (ii) folgt durch vollständige Induktion aus (i). Um (i) zu zeigen, wähle $I \subseteq J$ endlich. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass $k \in I$ (sonst ist nichts zu zeigen). Dann gilt

$$\begin{aligned}
 P\left(\bigcap_{j \in I} A'_j\right) &= P\left(\bigcap_{j \in I \setminus \{k\}} A_j \cap A_k^c\right) \\
 &= P\left(\bigcap_{j \in I \setminus \{k\}} A_j\right) - P\left(\bigcap_{j \in I \setminus \{k\}} A_j \cap A_k\right) \\
 &= \prod_{j \in I \setminus \{k\}} P(A_j) - \prod_{j \in I} P(A_j) \\
 &= \prod_{j \in I \setminus \{k\}} P(A_j) [1 - P(A_k)] \\
 &= \prod_{j \in I \setminus \{k\}} P(A_j) P(A_k^c) = \prod_{j \in I} P(A'_j) .
 \end{aligned} \tag{5.4}$$

Im zweiten Schritt haben wir verwendet, dass $P(B) = P(B \cap A_k) + P(B \cap A_k^c)$ mit $B = \bigcap_{j \in I \setminus \{k\}} A_j$.

(iii) Fügt man zu einer Familie unabhängiger Ereignisse \emptyset hinzu, dann sind alle Unabhängigkeitsbedingungen (5.3), die \emptyset involvieren, trivialerweise erfüllt, da auf beiden Seiten von (5.3) der Wert 0 steht. Nach (ii) kann man nun anstelle von \emptyset auch $\Omega = \emptyset^c$ hinzufügen, ohne dass sich an der Unabhängigkeitseigenschaft der Familie etwas ändert. Fügt man nun noch einmal \emptyset zu der erweiterten Familie hinzu, ergibt sich die Aussage von (iii). \square

Zum Abschluss dieses Kapitels betrachten wir noch abzählbare Familien von Ereignissen $\{A_j\}_{j \in \mathbb{N}}$ und fragen, was man über das Eintreten unendlich vieler dieser Ereignisse sagen kann.

Definition 5.4. Sei Ω eine Menge und $A_j \subseteq \Omega$, $j \in \mathbb{N}$. Wir definieren den **Limes superior** des Mengensystems $\{A_j\}_{j \in \mathbb{N}}$ als

$$\begin{aligned}
 \limsup_{j \in \mathbb{N}} A_j &:= \{\omega \in \Omega : \omega \in A_k \text{ für unendlich viele } k \in \mathbb{N}\} \\
 &= \bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j .
 \end{aligned} \tag{5.5}$$

Korollar 5.5. Für einen messbaren Raum (Ω, \mathcal{A}) und eine abzählbare Familie von Ereignissen $\{A_j\}_{j \in \mathbb{N}} \subseteq \mathcal{A}$ gilt $\limsup_{j \in \mathbb{N}} A_j \in \mathcal{A}$.

Satz 5.6 (Null-Eins Gesetze von Borel und Cantelli). *Es sei (Ω, \mathcal{A}, P) Wahrscheinlichkeitsraum. Dann gilt*

(i) Für $\{A_j\}_{j \in \mathbb{N}} \subseteq \mathcal{A}$

$$\sum_{j=1}^{\infty} P(A_j) < \infty \quad \Rightarrow \quad P(\limsup_{j \in \mathbb{N}} A_j) = 0 . \quad (5.6)$$

(ii) Für $\{A_j\}_{j \in \mathbb{N}} \subseteq \mathcal{A}$ unabhängige Familie von Ereignissen

$$\sum_{j=1}^{\infty} P(A_j) = \infty \quad \Rightarrow \quad P(\limsup_{j \in \mathbb{N}} A_j) = 1 . \quad (5.7)$$

Beweis. (i) Es gilt Aufgrund von Isotonie und Sub-Additivität von P

$$\begin{aligned} P(\limsup_{j \in \mathbb{N}} A_j) &= P\left(\bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j\right) \\ &\leq \inf_{n \in \mathbb{N}} P\left(\bigcup_{j \geq n} A_j\right) \\ &\leq \inf_{n \in \mathbb{N}} \sum_{j=n}^{\infty} P(A_j) = 0 , \end{aligned} \quad (5.8)$$

denn das Restglied der unendlichen Reihe $\sum_{j=1}^{\infty} P(A_j)$ verschwindet nach Voraussetzung.

(ii) Wir erhalten das Resultat aus der Sub-Additivität (*), der Stetigkeit des Wahrscheinlichkeitsmaßes P (**) und Lemma 5.3 (ii) (***) wie folgt

$$\begin{aligned} 0 &\leq 1 - P\left(\limsup_{j \in \mathbb{N}} A_j\right) = P\left(\left(\limsup_{j \in \mathbb{N}} A_j\right)^c\right) \\ &= P\left(\bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j^c\right) \\ &\stackrel{(*)}{\leq} \sum_{n=1}^{\infty} P\left(\bigcap_{j \geq n} A_j^c\right) \\ &\stackrel{(**)}{=} \sum_{n=1}^{\infty} \lim_{q \rightarrow \infty} P\left(\bigcap_{j=n}^q A_j^c\right) \\ &\stackrel{(***)}{=} \sum_{n=1}^{\infty} \lim_{q \rightarrow \infty} \prod_{j=n}^q P(A_j^c) \\ &= \sum_{n=1}^{\infty} \lim_{q \rightarrow \infty} \prod_{j=n}^q [1 - P(A_j)] \\ &\leq \sum_{n=1}^{\infty} \lim_{q \rightarrow \infty} \prod_{j=n}^q \exp\{-P(A_j)\} \\ &= \sum_{n=1}^{\infty} \lim_{q \rightarrow \infty} \exp\left\{-\sum_{j=n}^q P(A_j)\right\} = \sum_{n=1}^{\infty} 0 = 0 . \end{aligned} \quad (5.9)$$

5. Unabhängigkeit

Hier haben wir verwendet, dass $\exp(-p) \geq 1 - p$ für $p \in [0, 1]$ ist. Diese Ungleichung folgt aus der Übereinstimmung beider Funktionen bei $p = 0$ und dem Umstand, dass die p -te Ableitung von $\exp(-p)$ für $p \in [0, 1]$ stets größer ist als die p -te Ableitung von $1 - p$. \square

Beispiel 5.7 ('Kann mich das Glück für immer verlassen?'). Wie groß ist die Wahrscheinlichkeit, dass ich bei einem fairen Würfelspiel ab einem bestimmten Punkt nie wieder eine 6 würfeln?

Lösung. Setze A_j = 'Würfe eine 6 beim j -ten Wurf'. Dann gilt für das Ereignis A = 'Es werden nur endlich viele Male die 6 gewürfelt' = $(\limsup_{j \in \mathbb{N}} A_j)^c$. Also

$$P(A) = 1 - P(\limsup_{j \in \mathbb{N}} A_j) . \quad (5.10)$$

Da das Würfelspiel fair ist, gilt dass $\{A_j\}_{j \in \mathbb{N}}$ eine Familie unabhängiger Ereignisse ist und $P(A_j) = \frac{1}{6}$ (genauer: so definieren wir 'fair'). Also

$$\sum_{j=1}^{\infty} P(A_j) = \sum_{j=1}^{\infty} \frac{1}{6} = \infty . \quad (5.11)$$

Deshalb gilt nach Satz 5.6 (ii) $P(\limsup_{j \in \mathbb{N}} A_j) = 1$ und somit $P(A) = 0$. Das Glück wird Sie nie für immer verlassen – und das ist jetzt sogar bewiesen. \diamond

5.1. Übungsaufgaben

1. Selbsttest

1. Konstruieren Sie ein Gegenbeispiel zu folgender Aussage: Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann ist für alle paarweise unabhängigen Ereignisse $A, B, C \in \mathcal{A}$ die Familie $\{A, B, C\}$ eine Familie unabhängiger Ereignisse. Mit paarweise unabhängig ist hier gemeint, dass A und B unabhängig, A und C unabhängig und B und C unabhängig.
2. Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B, C \in \mathcal{A}$ drei beliebige Ereignisse. Stimmen die folgenden Aussagen? Falls ja, dann geben Sie einen Beweis und falls nein ein Gegenbeispiel!
 - a) Falls A unabhängig von B , dann ist auch B unabhängig von A .
 - b) Falls A unabhängig von B und B unabhängig von C , dann ist auch A unabhängig von C .
 - c) Falls $A \cap B = \emptyset$ disjunkt, so ist A unabhängig von B .
 - d) Ω ist immer unabhängig von A .

2. Transfer

1. Sie würfeln zuerst mit einem Würfel und gewinnen, wenn Sie eine 6 würfeln. Danach würfeln Sie mit zwei Würfeln und gewinnen, wenn Sie zwei 6en würfeln. Im n -ten Schritt würfeln Sie mit n Würfeln und gewinnen, wenn Sie n 6en würfeln. Wie groß ist die Wahrscheinlichkeit, dass Sie unendlich oft gewinnen, wenn Sie dieses Spiel für alle $n \in \mathbb{N}$ genau einmal durchführen?
2. Ein Affe sitzt an einer Schreibmaschine bestehend aus 27 Tasten, dem kleingeschriebenen Alphabet und der Leertaste. Er tippt willkürlich jeweils eine Taste, d.h. alle Zeichen erscheinen mit der gleichen Wahrscheinlichkeit und unabhängig voneinander. Mit welcher Wahrscheinlichkeit tippt er unendlich oft den Satz "affen sind die coolsten", wenn er unendlich lange arbeitet?

Hinweis: Finden Sie geeignete unabhängige Ereignisse und wenden Sie das Borel-Cantelli Lemma an! Das Ergebnis das Sie erarbeiten sollen ist unter dem Namen **infinite monkey theorem** bekannt.

3. Theorie

1. Beweisen Sie die zweite Gleichheit in Formel (5.5)! Führen Sie ebenfalls den Beweis von Korollar 5.5 explizit durch!

5. Unabhängigkeit

2. Es sei Ω eine Menge und $A_j \subseteq \Omega$, $j \in \mathbb{N}$. Wir definieren den **Limes inferior** des Mengensystems $\{A_j\}_{j \in \mathbb{N}}$ als

$$\liminf_{j \in \mathbb{N}} A_j := \{\omega \in \Omega : \omega \in A_k \text{ für fast alle } k \in \mathbb{N}\} .$$

- a) Beweisen Sie, dass $\liminf_{j \in \mathbb{N}} A_j = \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j$ gilt!
 - b) Beweisen Sie die folgende Aussage (vergleichen Sie diese mit Korollar 5.5): Für einen messbaren Raum (Ω, \mathcal{A}) und eine abzählbare Familie von Ereignissen $\{A_j\}_{j \in \mathbb{N}} \subseteq \mathcal{A}$ gilt $\liminf_{j \in \mathbb{N}} A_j \in \mathcal{A}$.
 - c) Gilt $\liminf_{j \in \mathbb{N}} A_j \subseteq \limsup_{j \in \mathbb{N}} A_j$ oder $\liminf_{j \in \mathbb{N}} A_j \supseteq \limsup_{j \in \mathbb{N}} A_j$?
3. Es seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B, C \in \mathcal{A}$ drei Ereignisse, wobei $P(C) > 0$ gilt. Dann heißen die Ereignisse A und B *bedingt unabhängig unter C* , wenn gilt

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

Beweisen oder widerlegen Sie die folgende Aussage: Für einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und zwei unabhängige Ereignisse $A, B \in \mathcal{A}$ sind für alle $C \in \mathcal{A}$ mit $P(C) > 0$ die Ereignisse A und B auch bedingt unabhängig unter C .

5.2. Literatur

Das Konzept der Unabhängigkeit ist implizit schon bei J. Bernoulli (1655-1705) [Ber99], A. de Moivre (1667-1754) [dM67] oder T. Bayes [Bay70] zu finden. Die formale Definition wurde erst 1900 von G. Bohlmann (1869-1928) in einem Artikel über Lebensversicherungsmathematik für die Enzyklopädie der Mathematik im Jahre 1900 gegeben. Ebenfalls von Bohlmann wurde 1908 ein Beispiel angegeben, dass paarweise Unabhängigkeit nicht die Unabhängigkeit mehrerer Ereignisse impliziert.

Das originale Null-Eins-Gesetz von É. Borel (1871-1956) und F. Cantelli (1875-1966) enthält streng genommen nur den ersten Teil des Satzes 5.6. Die Zweite Aussage stammt von P. Erdős (1913-1996) und A. Rényi (1921-1970). Neben dem Null-Einsgesetz von Borel-Cantelli gibt es noch zahlreiche weitere Null-Eins-Gesetze, etwa von A. N. Kolmogorov, siehe z.B. [Bau91]. Moderne Varianten betreffen etwa die Eindeutigkeit eines stabilen Zustandes in der statistischen Mechanik [Geo11].

6. Diskrete Zufallsvariablen

Oft interessiert uns weniger der durch $\omega \in \Omega$ kodierte Ausgang eines Zufallsexperiments, sondern nur gewisse Größen $X(\omega)$, die von dem Ausgang abhängen, und damit ihrerseits Zufallscharakter haben. Wir möchten diese Zufallseigenschaften abgeleiteter Größen nun in unsere Theorie der Wahrscheinlichkeitsräume mit einbeziehen.

Definition 6.1 (Zufallsvariable und Verteilung). Es seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und (E, \mathcal{B}) ein messbarer Raum.

- (i) Eine Abbildung $X : \Omega \rightarrow E$ heißt **messbar**, geschrieben $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B})$, falls $\forall B \in \mathcal{B}$ gilt $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$. Eine solche messbare Abbildung wird auch E -wertige **Zufallsvariable** auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) genannt.

- (ii) Die Abbildung

$$X_*P : \mathcal{B} \rightarrow [0, 1] , \quad B \mapsto X_*P(B) = P(X^{-1}(B)) \quad (6.1)$$

heißt **Bildmaß von P unter X** oder auch **Verteilung von X** . Häufig findet man auch $\text{Vert}(X)$ als Schreibweise für X_*P .

Satz 6.2. Das Bildmaß X_*P einer Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{B})$ ist ein Wahrscheinlichkeitsmaß auf (E, \mathcal{B}) .

Beweis. Zuerst zeigen wir die Normiertheit von X_*P ,

$$X_*P(E) = P(X^{-1}(E)) = P(\{\omega \in \Omega : X(\omega) \in E\}) = P(\Omega) = 1 . \quad (6.2)$$

Es seien nun $B_j \in \mathcal{B}$, $j \in \mathbb{N}$, paarweise disjunkt. Setze $A_j := X^{-1}(B_j)$, $j \in \mathbb{N}$. Wäre für $k \neq j$ $A_j \cap A_k \neq \emptyset$, so existiert $\omega \in A_j \cap A_k$ und damit $X(\omega) \in B_j \cap B_k$, was im Widerspruch zu $B_j \cap B_k = \emptyset$ ist. Also sind auch die Ereignisse A_j , $j \in \mathbb{N}$, paarweise disjunkt. Desweiteren gilt

$$X^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right) = \bigcup_{j=1}^{\infty} X^{-1}(B_j) = \bigcup_{j=1}^{\infty} A_j \quad (6.3)$$

und deshalb

$$\begin{aligned} X_*P\left(\bigcup_{j=1}^{\infty} B_j\right) &= P\left(X^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right)\right) = P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \\ &= \sum_{j=1}^{\infty} P(X^{-1}(B_j)) = \sum_{j=1}^{\infty} X_*P(B_j) . \quad \square \end{aligned} \quad (6.4)$$

6. Diskrete Zufallsvariablen

Wir beschränken uns in diesem Kapitel auf den Fall, dass $(E, \mathcal{B}) = (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ (der Fall $(E, \mathcal{B}) = (\mathbb{N}_0, \mathbb{P}(\mathbb{N}_0))$ würde komplett analog verlaufen). Dann kann man ein Wahrscheinlichkeitsmaß auf diesem messbaren Raum besonders einfach definieren.

Satz 6.3 (Wahrscheinlichkeitsmaße auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$).

(i) Für $x \in \mathbb{Z}$ seien $p_x \geq 0$ positive reelle Zahlen mit

$$\sum_{x \in \mathbb{Z}} p_x = 1 \quad (\text{Normiertheit}) . \quad (6.5)$$

Dann definiert die folgende Vorschrift ein Wahrscheinlichkeitsmaß auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$,

$$P(A) := \sum_{x \in A} p_x , \quad \forall A \in \mathbb{P}(\mathbb{Z}) . \quad (6.6)$$

(ii) Sei umgekehrt P ein Wahrscheinlichkeitsmaß auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$. Dann erfüllt $p_x = P(\{x\})$, $x \in \mathbb{Z}$, die Bedingung $p_x \geq 0$ und (6.5). Das nach (i) definierte Wahrscheinlichkeitsmaß stimmt mit P überein.

Beweis. (i) $0 \leq P(A) \leq \sum_{x \in \mathbb{Z}} p_x$ erzwingt $P(A) \in [0, 1]$, für alle $A \in \mathbb{P}(\mathbb{Z})$. Nach Voraussetzung ist $P(\mathbb{Z}) = \sum_{x \in \mathbb{Z}} p_x = 1$. Also erfüllt P Normiertheit. Zu zeigen bleibt die σ -Additivität für A_j , $j \in \mathbb{N}$, paarweise disjunkt. Wir haben für alle $k \in \mathbb{N}$

$$\begin{aligned} P\left(\bigcup_{j=1}^{\infty} A_j\right) &= \sum_{x \in \bigcup_{j=1}^{\infty} A_j} p_x \\ &= \sum_{j=1}^{\infty} \sum_{x \in A_j} p_x = \sum_{j=1}^{\infty} P(A_j) . \end{aligned}$$

Hierbei verwenden wir den Umordnungssatz für Reihen mit positiven Summanden aus der Analysis. (ii) Dies folgt unmittelbar aus der σ -Additivität von P . \square

Definition 6.4 (Diskrete Zufallsvariablen und Dichten).

(i) Eine messbare Abbildung $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ heißt **diskrete Zufallsvariable**.

(ii) Für $x \in \mathbb{Z}$ heißt $p_x = p_x^X = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}) = X_* P(\{x\})$ die **diskrete Dichte** von X .

Aufgrund von Satz 6.3 können wir nun Wahrscheinlichkeitsmaße auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ definieren, indem wir geeignete Zahlen p_x , $x \in \mathbb{Z}$, spezifizieren. Für viele Anwendungsfälle erweisen sich hierbei spezielle Wahrscheinlichkeitsmaße als besonders nützlich, welche zu so genannten **Verteilungsfamilien** zusammengefasst sind. Die wichtigsten diskreten Verteilungsfamilien werden im Folgenden dargestellt.

$$P(\text{Regen nach } X = 0 \text{ trockenen Wochen}) = p(1-p)^0$$

$$P(\text{Regen nach } X = 1 \text{ trockenen Wochen}) = p(1-p)^1$$

$$P(\text{Regen nach } X = 2 \text{ trockenen Wochen}) = p(1-p)^2$$

$$P(\text{Regen nach } X = 3 \text{ trockenen Wochen}) = p(1-p)^3$$

$$P(\text{Regen nach } X = 4 \text{ trockenen Wochen}) = p(1-p)^4$$

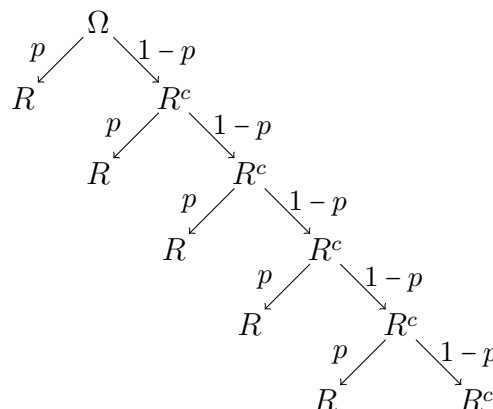


Abbildung 6.1: Wahrscheinlichkeitsbaum für Wartezeiten bei Bernoulli-Experimenten, vergleiche Beispiel 6.7.

Definition 6.5. Die **geometrische Verteilung** zum Parameter $p \in [0, 1)$ ist das Wahrscheinlichkeitsmaß $G(p)$ auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$, welches definiert ist durch die diskrete Dichte

$$p_x = \begin{cases} p(1-p)^x & , \text{ für } x \in \mathbb{N}_0 , \\ 0 & , \text{ sonst .} \end{cases} \quad (6.7)$$

Für eine diskrete Zufallsvariable X mit $X_*P = G(p)$ schreibt man auch $X \sim G(p)$, sprich ‘ X ist geometrisch verteilt zum Parameter p ’.

Bemerkung 6.6. Die Zahlen (6.7) sind offensichtlich eine diskrete Dichte, da $p_x \geq 0$ und mittels der Formel für die geometrische Reihe $\sum_{x=0}^{\infty} p_x = p \sum_{x=0}^{\infty} (1-p)^x = p \frac{1}{1-(1-p)} = 1$.

Der Anwendungsbereich der geometrischen Verteilung (genauer wäre ‘der geometrischen Verteilungen’, aber das sagt niemand) ist die Wartezeit für das Eintreffen eines Ereignisses bei beliebig häufiger Wiederholung des Zufallsexperiments. Das Ereignis tritt mit Wahrscheinlichkeit p ein und mit Wahrscheinlichkeit $1-p$ nicht ein. Dies wird durch den Wahrscheinlichkeitsbaum in Abbildung 6.1 und durch das folgende Beispiel illustriert.

Beispiel 6.7. Es sei ein trockener Sommer. Die Wahrscheinlichkeit, dass es in der kommenden Woche Regen gibt, gegeben dass diese Woche trocken ist, liege bei $p = 0.3$. Mit welcher Wahrscheinlichkeit gibt es in den nächsten 4 Wochen keinen Regen, jedoch in der 5. Woche?

Lösung. In diesem Beispiel ist R das Ereignis ‘Regen fällt in einer bestimmten Woche’ und die Wiederholungen sind die aufeinander folgenden Wochen. X sei die Anzahl der Wochen Trockenheit, bis zum ersten Mal Regen fällt. Die gerade vergangene Woche sei trocken gewesen und die gerade beginnende Woche sei Woche 0. Aus dem Wahrscheinlichkeitsbaum in Abbildung 6.1 ersehen wir, dass wir mit der geometrischen Verteilung zum Parameter

6. Diskrete Zufallsvariablen

$p = 0.3$ rechnen müssen, $X \sim G(0.3)$. R steht in dieser Abbildung für ‘Regen’ und R^c für ‘keinen Regen’. Wir erhalten für die gefragte Wahrscheinlichkeit

$$P(X = 4) = p(1 - p)^4 = 0.3 \times 0.7^4 = 0.07203 \approx 7.2\% . \quad (6.8)$$

◇

Definition 6.8. Die **Binomialverteilung** zu den Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$ ist das Wahrscheinlichkeitsmaß $B(n, p)$ auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$, welches definiert ist durch die diskrete Dichte

$$p_x = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & , \text{ für } x = 0, \dots, n , \\ 0 & , \text{ sonst .} \end{cases} \quad (6.9)$$

Eine diskrete Zufallsvariable X mit $X_*P = B(n, p)$ nennen wir binomialverteilt mit Parametern n, p und schreiben $X \sim B(n, p)$. Ist insbesondere $n = 1$, so nennen wir X auch **bernoulliverteilt** zum Parameter p , geschrieben $X \sim B(p)$.

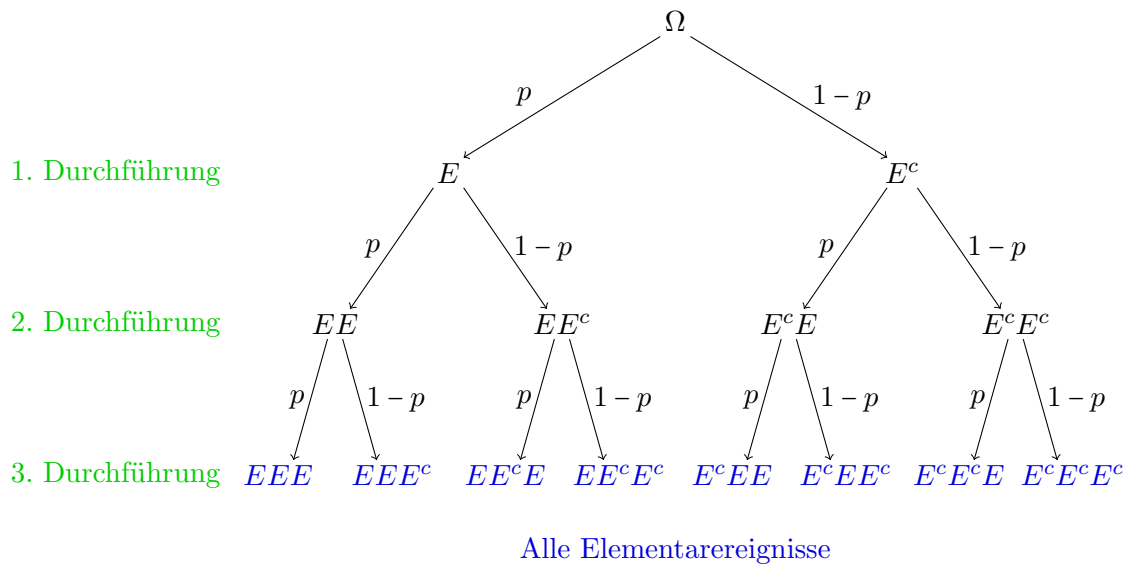
Bemerkung 6.9. Die Zahlen in (6.9) sind offensichtlich eine diskrete Dichte, da $p_x \geq 0$ und nach Proposition 1.15 gilt

$$\sum_{x=0}^{\infty} p_x = \sum_{x=0}^n \binom{n}{x} p^x (1 - p)^{n-x} = (p + (1 - p))^n = 1^n = 1 . \quad (6.10)$$

Der Anwendungsbereich der Binomialverteilung liegt bei unabhängig wiederholten Zufallsexperimenten, bei denen ein Ereignis E mit Wahrscheinlichkeit p eintritt bzw. mit Wahrscheinlichkeit $1 - p$ ausbleibt. Zur Illustration siehe Abbildung 6.2. Die Zählvariable, die bei n Wiederholungen die Anzahl des Auftretens von E zählt, ist dann binomialverteilt zu den Parametern n, p . Die Wahrscheinlichkeit in Abbildung 6.2 bei drei Versuchen zwei mal E zu sehen und einmal nicht, ist demnach $\binom{3}{2} p^2 (1 - p) = 3p^2(1 - p)$, wie man durch Multiplikation der Zweigwahrscheinlichkeiten und Abzählen in Abbildung 6.2 leicht bestätigt.

Beispiel 6.10. An einer Ski- und Snowboardfreizeit, die man im Internet buchen kann, nehmen Sie und weitere 10 Ihnen unbekannte Teilnehmer/innen teil. Sie fahren Snowboard und eins steht fest: Für Sie ist es dann und nur dann cool, wenn nicht zu viele Skilangweiler dabei sind.

Ihrer Einschätzung nach wird dieses Angebot von Einzelreisenden unabhängig voneinander gebucht und zwar ist jede Buchung zu 60% eine Snowboardbuchung und zum Glück nur zu 40% die Buchung eines/einer ultralangweiligen Skifahrer/in. Wie groß schätzen Sie die Wahrscheinlichkeit ein, dass Sie am Ende mit Ihrem Snowboard in der Minderheit sind und Ihr Coolness-Faktor schmerzlich leidet?



$$\Omega = \{EEE, EEE^c, EE^cE, EE^cE^c, E^cEE, E^cEE^c, E^cE^cE, E^cE^cE^c\}$$

Abbildung 6.2: Wahrscheinlichkeitsbaum für dreistufiges Bernoulli-Experiment.

Lösung. Aus den Annahmen folgt, dass die Anzahl der Skifahrer/innen X binomialverteilt zu $n = 10$ und $p = 0.4$ sein muss. Gefragt ist also nach $P(X > 5)$. Wir erhalten

$$\begin{aligned} P(X > 5) &= \sum_{x=6}^{10} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{10}{6} 0.4^6 0.6^4 + \binom{10}{7} 0.4^7 0.6^3 + \binom{10}{8} 0.4^8 0.6^2 \\ &\quad + \binom{10}{9} 0.4^9 0.6^1 + \binom{10}{10} 0.4^{10} 0.6^0 \\ &= 0.1662386 \approx 16.6\% . \end{aligned} \tag{6.11}$$

Die Gefahr ist durchaus ernst zu nehmen!

◇

In den Fällen, in denen das Auftreten von einer Eigenschaft E sehr selten ist, aber auch sehr viele unabhängige Wiederholungen stattfinden, können wir die Binomialverteilung oft durch die Poissonverteilung approximieren.

Definition 6.11. Die **Poissonverteilung** zum Parameter $0 < \lambda \in \mathbb{R}$ ist das Wahrscheinlichkeitsmaß $\text{Po}(\lambda)$ auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$, welches definiert ist durch die diskrete Dichte

$$p_x = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & , \text{ für } x \in \mathbb{N}_0 , \\ 0 & , \text{ sonst .} \end{cases} \tag{6.12}$$

6. Diskrete Zufallsvariablen

Eine Zufallsvariable X mit $X_*P = \text{Po}(\lambda)$ wird als poissonverteilt zum Parameter λ bezeichnet, in Kurzschreibweise $X \sim \text{Po}(\lambda)$.

Bemerkung 6.12. Die Zahlen in (6.12) sind offensichtlich eine diskrete Dichte, da $p_x \geq 0$ und $\sum_{x=0}^{\infty} p_x = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$.

Satz 6.13 (Approximation von Binomialverteilung durch Poissonverteilung). *Es sei $\lambda > 0$, $p_n := \frac{\lambda}{n}$ und n hinreichend groß, so dass $p_n \in [0, 1]$. Dann gilt*

$$\lim_{n \rightarrow \infty} B(n, p_n)(\{x\}) = \text{Po}(\lambda)(\{x\}) \quad , \quad \forall x \in \mathbb{Z} . \quad (6.13)$$

Beweis. Für $\mathbb{Z} \ni x < 0$ ist die Aussage trivial. Sei also nun $x \in \mathbb{N}_0$. Unter den gegebenen Voraussetzungen gilt wegen $\binom{n}{x} \left(\frac{\lambda}{n}\right)^x = \frac{n(n-1)\cdots(n-x+1)}{n^x} \lambda^x$ der folgende Grenzwert,

$$\begin{aligned} B(n, p_n)(\{x\}) &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \left(\prod_{k=1}^x \underbrace{\frac{n-k+1}{n}}_{\rightarrow 1}\right) \underbrace{\frac{1}{\left(1 - \frac{\lambda}{n}\right)^x}}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \frac{\lambda^x}{x!} \\ &\xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^x}{x!} = \text{Po}(\lambda)(\{x\}) . \quad \square \end{aligned} \quad (6.14)$$

Folgendes Beispiel gibt eine typische Anwendung für die Poissonverteilung:

Beispiel 6.14. Eine Lebensversicherung hat 1000 Risikolebensversicherungen mit recht jungen Menschen abgeschlossen. Das Individualrisiko für einen Todesfall in einem gegebenen Kalenderjahr betrage $p = 0.005 = \frac{5}{1000}$.

Unter der Annahme, dass die Todesfälle der Versicherten unabhängig von einander auftreten, bestimmen Sie für die Zählvariable X der Todesfälle im betreffenden Jahr die Wahrscheinlichkeit, dass das Ereignis $X \geq 10$ auftritt.

Lösung. Unter den getroffenen Voraussetzungen gilt

$$X \sim B\left(1000, \frac{5}{1000}\right) \sim_a \text{Po}(5) , \quad (6.15)$$

wobei die mathematisch ungenaue Schreibweise \sim_a für ‘asymptotisch bzw. approximativ verteilt’ im Sinne von Satz 6.13 steht. Da 1000 eine recht große Zahl ist, rechnen wir nun mit $\text{Po}(5)$ weiter und nehmen einen kleinen Fehler in Kauf. Es ist dann

$$\begin{aligned} P(X \geq 10) &\approx \text{Po}(5)(\{10, 11, 12, 13, \dots\}) \\ &= 1 - \text{Po}(5)(\{0, 1, \dots, 9\}) \\ &= 1 - e^{-5} \sum_{x=0}^9 \frac{5^x}{x!} \approx 1 - 0.9681719 \approx 3.2\% . \end{aligned} \quad (6.16)$$

◇

Die Situation ‘Stichprobenentnahme der Größe n aus einer Gruppe die aus zwei Untergruppen der Größe n_1 und n_2 besteht’, verbunden mit der Frage, wie viele Fälle aus der Stichprobe aus Gruppe 1 stammen, wird mittels der folgenden Verteilungsfamilie modelliert:

Definition 6.15. Die **hypergeometrische Verteilung** zu den Parametern $n, n_1, n_2 \in \mathbb{N}$, wobei $n \leq n_1 + n_2$, ist das Wahrscheinlichkeitsmaß $H(n, n_1, n_2)$ auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$, welches definiert ist durch die diskrete Dichte

$$p_x = \begin{cases} \frac{\binom{n_1}{x} \binom{n_2}{n-x}}{\binom{n_1+n_2}{n}} & , \text{ für } x = 0, \dots, n_1 , \\ 0 & , \text{ sonst .} \end{cases} \quad (6.17)$$

Eine diskrete Zufallsvariable mit Verteilung $X_*P = H(n, n_1, n_2)$ heißt hypergeometrisch verteilt zu den Parametern n, n_1, n_2 , geschrieben $X \sim H(n, n_1, n_2)$.

Bemerkung 6.16. Die Herleitung der Formel für die hypergeometrische Verteilung erfolgt durch die folgende kombinatorische Überlegung: Bei einer Zufallsauswahl von n Elementen aus einer Menge M mit $n_1 + n_2$ Elementen (ohne Berücksichtigung der Reihenfolge) gibt es nach Satz 1.10 $\binom{n_1+n_2}{n}$ Möglichkeiten. Nach dem Prinzip des unzureichenden Grundes sollten alle diese Möglichkeiten als gleich wahrscheinlich angesehen werden. Wir setzen also als Wahrscheinlichkeit, dass eine spezifische Stichprobe gezogen wird, $p = \frac{1}{\binom{n_1+n_2}{n}}$.

Die Frage ist nun, wie viele solcher Stichproben gibt es, so dass genau x Elemente einer Untermenge $U \subseteq M$ mit n_1 Elementen entstammen? Hierzu wählen wir x Elemente aus U und haben $\binom{n_1}{x}$ Möglichkeiten für diese Wahl. Für jede dieser Möglichkeiten können wir nun die restlichen $n - x$ Elemente der Stichprobe aus den n_2 Elementen des Komplementes U^c wählen. Hierfür gibt es wiederum $\binom{n_2}{n-x}$ Möglichkeiten. Wir erhalten somit die rechte Seite von (6.17) als die Wahrscheinlichkeit, dass genau x Elemente der Stichprobe aus U sind.

Aus dieser Herleitung ergibt sich automatisch, dass $\sum_{x=0}^n p_x = 1$. Die Bedingung $p_x \geq 0$ ist offensichtlich erfüllt.

Beispiel 6.17. Beim Zahlenlotto ‘6 aus 49’ legen Sie vor der Ziehung der Lottozahlen Ihre 6 Zahlen fest. Wie groß ist die Wahrscheinlichkeit, dass bei der darauf folgenden Ziehung von 6 Zahlen aus 49 (ohne Zurücklegen) mindestens 5 Ihrer Zahlen drankommen?

Lösung. Die von Ihnen festgelegten Zahlen bilden eine Untermenge U mit 6 Elementen der Zahlen $M = \{1, \dots, 49\}$. Nun wird eine Stichprobe mit 6 Elementen gezogen. Wir haben also eine hypergeometrische Verteilung mit $n = 6$, $n_1 = 6$ und $n_2 = 49 - 6 = 43$. Daraus ergibt sich

$$P(5 \text{ oder } 6 \text{ Richtige}) = \frac{\binom{6}{5} \binom{43}{1}}{\binom{49}{6}} + \frac{\binom{6}{6} \binom{43}{0}}{\binom{49}{6}} = 1.8521 \times 10^{-5} . \quad (6.18)$$

Also stehen die Aussichten für 5 oder 6 Richtige ziemlich schlecht. \diamond

6. Diskrete Zufallsvariablen

In vielen der Beispiele in diesem Kapitel tauchte die Frage auf, wie groß die Wahrscheinlichkeit ist, dass eine Zufallsvariable X einen Wert überhalb oder unterhalb eines bestimmten Grenzwertes $x \in \mathbb{Z}$ annimmt. Für diese Fragestellungen empfiehlt sich eine eigene Begriffsbildung:

Definition 6.18. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Für eine diskrete Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ ist die **Verteilungsfunktion** $F_X : \mathbb{R} \rightarrow [0, 1]$ definiert als

$$F_X(x) := P(X \leq x) := P(\{\omega \in \Omega : X(\omega) \leq x\}) . \quad (6.19)$$

Wir zeigen nun, dass diese Definition tatsächlich Sinn macht.

Lemma 6.19. (i) Die Verteilungsfunktion ist wohldefiniert, d.h. $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$ für alle $x \in \mathbb{R}$.

(ii) Die Funktion F_X hängt nur von der Verteilung von X ab, d.h. für X, X' diskrete Zufallsvariablen mit $X_*P = X'_*P'$ gilt $F_X = F_{X'}$.

Beweis. Beide Aussagen ergeben sich aus $\{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}(\mathbb{Z} \cap (-\infty, x])$. Denn $\mathbb{Z} \cap (-\infty, x] \in \mathbb{P}(\mathbb{Z})$ und somit $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$, also gilt (i). Die zweite Aussage folgt daraus, dass sich die Mengen $\{\omega \in \Omega : X(\omega) \leq x\}$ als Urbilder von \mathbb{Z} -Teilmengen schreiben lassen, siehe oben, und der Definition des Bildmaßes (6.1). \square

Bevor die Eigenschaften von Verteilungsfunktionen diskutiert werden, sei an den Begriff der **Rechtsstetigkeit** erinnert. Eine Funktion F auf \mathbb{R} hat diese Eigenschaft, wenn $\forall x \in \mathbb{R}$ und $\forall \{y_n\}_{n \in \mathbb{N}}, y_n \in \mathbb{R}$ mit $y_n \geq x \ \forall n \in \mathbb{N}$ sowie $\lim_{n \rightarrow \infty} y_n = x$ gilt $\lim_{n \rightarrow \infty} F(y_n) = F(x)$.

Satz 6.20 (Eigenschaften der Verteilungsfunktion). Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ eine diskrete Zufallsvariable. Dann gilt für die Verteilungsfunktion F_X von X :

(i) F_X ist rechtsstetig.

(ii) F_X ist monoton steigend auf \mathbb{R} .

$$(iii) \quad \lim_{x \rightarrow \pm\infty} F_X(x) = \begin{cases} 1 \\ 0 \end{cases} .$$

(iv) $F_X \upharpoonright_{[x, x+1)}$ ist konstant, für alle $x \in \mathbb{Z}$.

$$(v) \quad F_X(x) = \sum_{y \in \mathbb{Z}, y \leq x} p_y^X .$$

(vi) Ist umgekehrt F eine Funktion mit den Eigenschaften (i)–(iv), so existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß \tilde{P} auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$, so dass $F(x) = \tilde{P}(\mathbb{Z} \cap (-\infty, x])$, $\forall x \in \mathbb{R}$.

Beweis. Zunächst ist Aussage (iv) stärker als Aussage (i). (iv) folgt aus $X^{-1}(\mathbb{Z} \cap (-\infty, x]) = X^{-1}(\mathbb{Z} \cap (-\infty, [x]))$ mit $[x]$ der größten ganzen Zahl, die kleiner oder gleich x ist. Da die Funktion $g: x \mapsto [x]$ konstant ist auf $[x, x+1)$, für alle $x \in \mathbb{Z}$, muss dies auch für F_X gelten, denn wie wir gesehen haben gilt $F_X = F_X \circ g$.

(ii) Diese Eigenschaft folgt aus $\{\omega \in \Omega : X(\omega) \leq x\} \supseteq \{\omega \in \Omega : X(\omega) \leq y\}$ für $y \leq x$ und der Isotonie von P .

(iii) Dies ist nichts anderes als eine Anwendung der σ -Stetigkeit von P in \emptyset bzw. Ω , siehe Proposition 2.11.

(v) Folgt aus der Definition der diskreten Dichte p_x^X und der Eindeutigkeitsaussage von Satz 6.3 (ii).

(vi) Sei nun F eine Funktion mit den besagten Eigenschaften. Wegen (ii) gilt, für alle $x \in \mathbb{Z}$,

$$p_x := F(x) - F(x-1) \geq 0 \quad (6.20)$$

und außerdem mit (iii)

$$\begin{aligned} \sum_{x \in \mathbb{Z}} p_x &= \lim_{n \rightarrow \infty} \sum_{x=-n}^n p_x \\ &= \lim_{n \rightarrow \infty} \sum_{x=-n}^n (F(x) - F(x-1)) \\ &= \lim_{n \rightarrow \infty} (F(n) - F(-n-1)) = 1 - 0 = 1 . \end{aligned} \quad (6.21)$$

Also definieren die Zahlen p_x , $x \in \mathbb{Z}$, eine diskrete Dichte, die nach Satz 6.3 ein eindeutig bestimmtes Wahrscheinlichkeitsmaß \tilde{P} auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ induziert. Offensichtlich gilt wiederum für $x \geq y$

$$F(x) - F(y) = \sum_{z \in \mathbb{Z} \cap (y, x]} p_z = \tilde{P}(\mathbb{Z} \cap (y, x]) . \quad (6.22)$$

Betrachten wir nun den Grenzübergang $y \rightarrow -\infty$, so konvergiert die linke Seite wegen (iii) gegen $F(x)$ und die rechte Seite wegen der σ -Stetigkeit von \tilde{P} gegen $\tilde{P}(\mathbb{Z} \cap (-\infty, x])$. \square

Wir können nun leicht die Wahrscheinlichkeit dafür bestimmen, dass der Wert einer diskreten Zufallsvariable in einem Bereich $(y, x]$ liegt:

Lemma 6.21. *Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ eine diskrete Zufallsvariable mit Verteilungsfunktion F_X . Für $y \leq x$ gilt*

$$P(y < X \leq x) := P(\{\omega \in \Omega : X(\omega) \in (y, x]\}) = F_X(x) - F_X(y) . \quad (6.23)$$

Beweis. Der Beweis folgt aus folgender Rechnung,

$$\begin{aligned} P(y < X \leq x) &= P(\{\omega \in \Omega : X(\omega) \leq x\} \setminus \{\omega \in \Omega : X(\omega) \leq y\}) \\ &= P(\{\omega \in \Omega : X(\omega) \leq x\}) - P(\{\omega \in \Omega : X(\omega) \leq y\}) \\ &= F_X(x) - F_X(y) . \quad \square \end{aligned} \quad (6.24)$$

6. Diskrete Zufallsvariablen

Die Verteilungsfunktionen der hier vorgestellten Verteilungsfamilien (und vieler anderer mehr) wurden früher in Tabellenwerken gesammelt. Heutzutage sind sie in Form von numerischen Algorithmen hinterlegt – etwa als Befehle in der Statistik-Programmiersprache R. Für praktische Zwecke können wir also die Verteilungsfunktionen als gegeben voraussetzen. Damit vereinfachen sich viele Aufgaben. Ist etwa F_X die Verteilungsfunktion zu einer Zufallsvariable $X \sim B(0.4, 10)$, so können wir die Lösung von Beispiel 6.10 schreiben als $1 - F_X(5)$.

Die Verteilungsfunktion einer diskreten Zufallsvariable X kann man nicht invertieren. Sie hat jedoch als monotone Funktion eine so genannte Rechtsinverse. Da diese ebenfalls von zentraler Bedeutung ist, führen wir eine weitere Notation ein.

Definition 6.22. Es sei X eine diskrete Zufallsvariable mit Verteilungsfunktion F_X . Dann ist die **Quantilfunktion** zu X , geschrieben $Q_X : [0, 1] \rightarrow \mathbb{R} = [-\infty, \infty]$, definiert als

$$Q_X(p) := \inf\{x \in \mathbb{R} : F_X(x) \geq p\}, \quad \forall p \in [0, 1], \quad (6.25)$$

wobei wir die Konvention $F_X(-\infty) = 0$ und $F_X(\infty) = 1$ verwenden.

Die Quantilfunktion $Q_X(p)$ beantwortet also die Frage nach dem Schwellenwert x für X -Werte, so dass x die kleinste Zahl ist für die $P(X \leq x) \geq p$ gilt. Werte der Quantilfunktion – auch **Quantile** genannt – sind von großer Bedeutung, wenn die Auswirkungen von Risiken, die höchstens mit einer Wahrscheinlichkeit p eintreten dürfen, abgeschätzt werden sollen.

Beispiel 6.23. Eine Bank sei exponiert gegenüber einem Risiko, welches mittels der diskreten Zufallsvariablen X modelliert werde. Stark negative Werte von X entsprechen großen Verlusten der Bank. Das Risikomanagement der Bank verlangt nun, einen Deckungsbetrag V vorzuhalten, der zu mindestens $p = 95\%$ das Risiko aus X abdeckt. Andererseits soll dieser Deckungsbetrag auch nicht größer sein, als unbedingt nötig. Welchen Deckungsbetrag hält die Bank vor? Die Verteilungsfunktion F_X sei als bekannt vorausgesetzt. Aus naheliegenden Gründen wird der erfragte Betrag auch ‘Value at Risk’ zum Risiko 5% genannt, geschrieben $-V@R_{0.05}$.

Lösung. Bestimme aus F_X die Quantilfunktion Q_X , dann ist $V = -Q_X(0.05)$ die Lösung des Problems. Der Value at Risk ist also gerade ein Quantil. \diamond

Bemerkung 6.24. Die hier gezeigte Analyse von diskreten Zufallsvariablen und Wahrscheinlichkeitsmaßen läßt sich sehr einfach von \mathbb{Z} auf jede beliebige abzählbare Menge E übertragen. *Per definitionem* heißt E abzählbar, falls eine bijektive Abbildung $\phi : E \rightarrow \mathbb{Z}$ existiert. Trivialer Weise sind $\phi : (E, \mathbb{P}(E)) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ und $\phi^{-1} : (\mathbb{Z}, \mathbb{P}(\mathbb{Z})) \rightarrow (E, \mathbb{P}(E))$ messbar. Ist also ein Wahrscheinlichkeitsmaß P auf $(E, \mathbb{P}(E))$ gegeben, so entspricht diesem das Wahrscheinlichkeitsmaß $\phi_* P$ auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$. Umgekehrt, gegeben ein Wahrscheinlichkeitsmaß P auf $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$, so erhalten wir ein Wahrscheinlichkeitsmaß $\phi_*^{-1} P$ auf $(E, \mathbb{P}(E))$. Insbesondere sehen wir daraus, dass sich Wahrscheinlichkeitsmaße auf beliebigen, abzählbaren Räumen $(E, \mathbb{P}(E))$ durch ihre diskreten Dichten $\{P(\{e\})\}_{e \in E}$ beschreiben lassen.

Ebenso lassen sich E -wertige Zufallsvariablen X äquivalent durch die \mathbb{Z} -wertigen Zufallsvariablen $Y = \phi \circ X$ beschreiben und \mathbb{Z} -wertige Zufallsvariablen Y durch E -wertige $X = \phi^{-1} \circ Y$.

Zuletzt soll in diesem Kapitel auf die Visualisierung von diskreten Verteilungen eingegangen werden. Meistens verwendet man hierzu das Balkendiagramm der diskreten Dichten oder die Verteilungsfunktion, siehe Abbildung 6.3.

6. Diskrete Zufallsvariablen

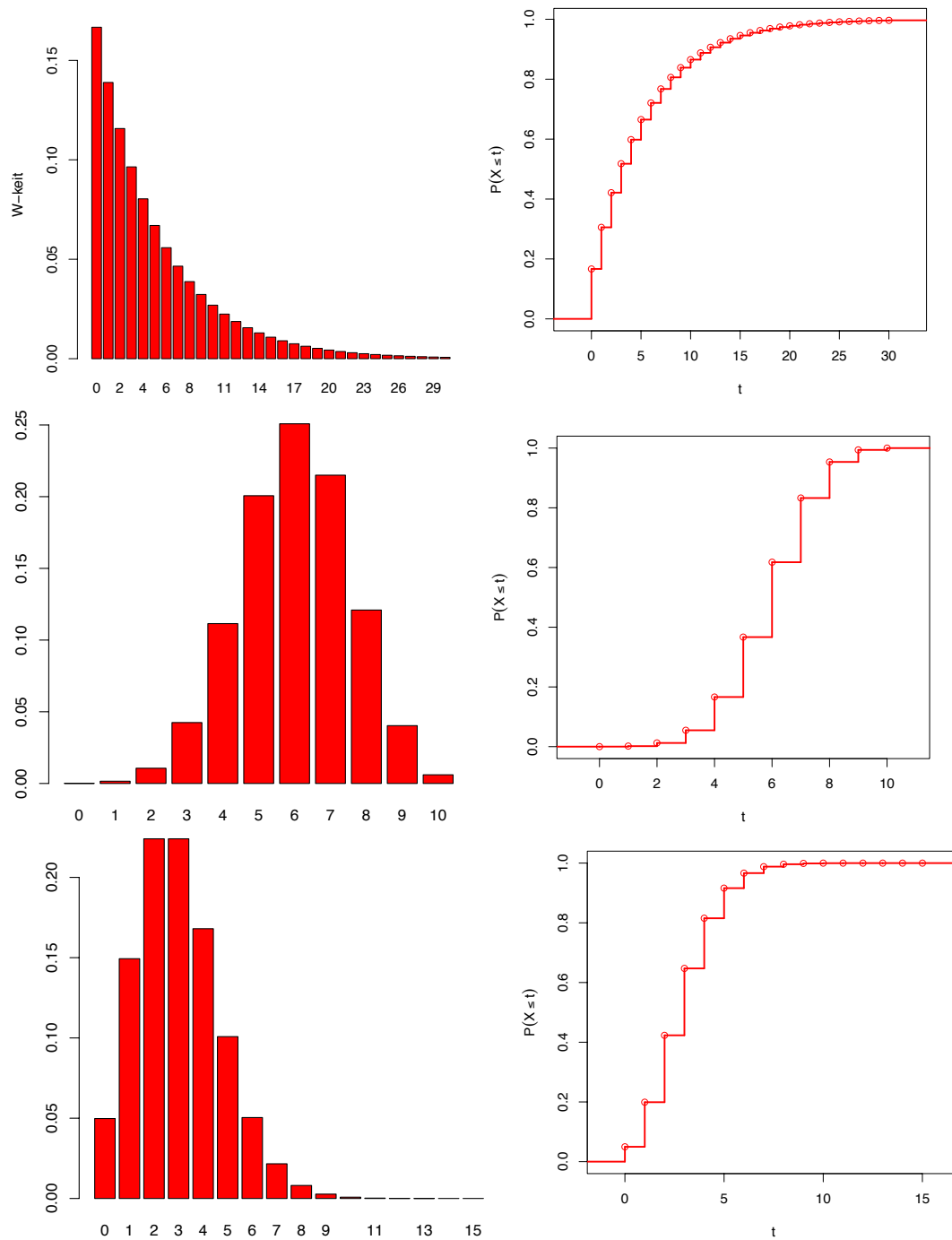


Abbildung 6.3: Dichte und Verteilungsfunktion zu verschiedenen Verteilungsfamilien. Oben: Geometrische Verteilung zum Parameter $p = 0.6$. Mitte: Binomialverteilung zu den Parametern $p = 0.6$ und $n = 10$. Unten: Poissonverteilung zum Parameter $\lambda = 3$.

6.1. Übungsaufgaben

1. Selbsttest

1. Gegeben seien zwei messbare Räume: $\Omega = \{1, 2, 3, 4, 5, 6\}$ mit der σ -Algebra $\mathcal{A} = \{\emptyset, \Omega, \{1, 2, 3\}, \{4, 5, 6\}\}$ und $E = \{0, 1\}$ mit der σ -Algebra $\mathcal{B} = \{\emptyset, E, \{0\}, \{1\}\}$. Die Abbildung $X : \Omega \rightarrow E$ ist definiert durch

$$\omega \mapsto X(\omega) = \begin{cases} 1 & , \text{ falls } \omega \text{ gerade,} \\ 0 & , \text{ sonst .} \end{cases}$$

Ist X messbar?

2. Es seien (Ω, \mathcal{A}) und (E, \mathcal{B}) zwei messbare Räume, so dass $\mathcal{A} = \mathbb{P}(\Omega)$, die Potenzmenge von Ω . Beweisen Sie, dass jede Abbildung $X : \Omega \rightarrow E$ messbar ist!
3. Gegeben seien nun zwei messbare Räume $\Omega = \{1, 2, 3, 4, 5, 6\}$ mit σ -Algebra $\mathcal{A} = \mathbb{P}(\Omega)$ und $E = \{0, 1\}$ mit σ -Algebra $\mathcal{B} = \mathbb{P}(E)$. Wir betrachten die Abbildung $X : \Omega \rightarrow E$ definiert durch

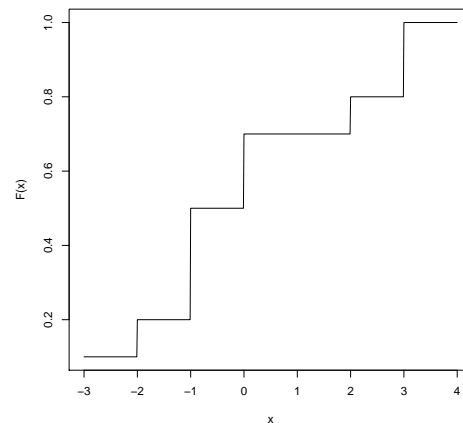
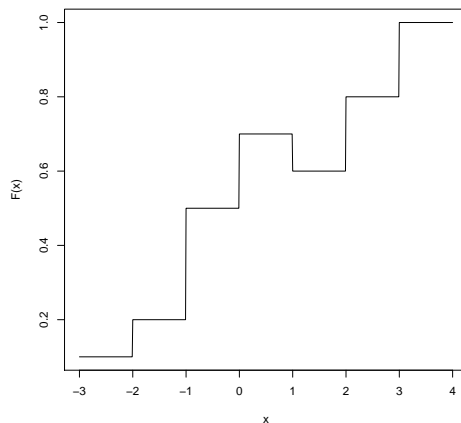
$$\omega \mapsto X(\omega) = \begin{cases} 1 & , \text{ falls } \omega \text{ gerade,} \\ 0 & , \text{ sonst .} \end{cases}$$

Wegen der vorangegangenen Aufgabe ist X automatisch messbar. Sei weiter das Wahrscheinlichkeitsmaß $P = \delta_3$ auf (Ω, \mathcal{A}) das Dirac-Maß, für alle $A \in \mathcal{A}$,

$$P(A) = \begin{cases} 1 & , \text{ falls } 3 \in A , \\ 0 & , \text{ sonst .} \end{cases}$$

Berechnen Sie das Bildmaß X_*P !

4. Welche der beiden Graphiken zeigt eine Verteilungsfunktion?



6. Diskrete Zufallsvariablen

5. Berechnen Sie die Verteilungsfunktion $F_X(x)$ einer geometrisch verteilten Zufallsvariablen $X \sim G(p)$. Bestimmen Sie ebenfalls das $\frac{1}{2}$ -Quantil $Q_X(\frac{1}{2})$.

2. Transfer

1. Eine Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{B})$ kann man sich auch vorstellen als eine Abbildung, welche aus einem (möglicherweise komplizierten) Zufallsexperiment die für uns relevanten Beobachtungsgrößen zuweist. Diese Sichtweise wollen wir nun Anhand von Beispielen intensivieren.
 - a) Es sei (Ω, \mathcal{A}, P) der Wahrscheinlichkeitsraum der das einmalige Würfeln mit einem 6-seitigen Würfel beschreibt. Geben Sie diesen an! Konstruieren Sie eine Zufallsvariable X , welche beschreibt ob die Augenzahl des Würfels gerade oder ungerade ist. Berechnen Sie das Bildmaß von X !
 - b) Es sei nun (Ω, \mathcal{A}, P) der Wahrscheinlichkeitsraum der das 2-malige Würfeln mit einem 6-seitigen Würfel beschreibt. Konstruieren Sie eine Zufallsvariable X , welche die Augenzahlsumme beider Würfelerggebnisse beschreibt und berechnen Sie das Bildmaß.
2. Ein Korrektur verzichtet auf das Korrigieren von Klausuren und ermittelt die Noten (von 1 bis 6) indem er einen sechsseitigen Würfel dreimal wirft und die kleinste auftretende Augenzahl als Endnote vergibt.
 - a) Definieren Sie eine Zufallsvariable X in der Form einer Abbildung $X : \Omega \rightarrow \mathbb{Z}$, die das Zufallsexperiment “Auswürfeln der Endnote” beschreibt. Geben Sie dazu die Menge Ω sowie die Abbildungsvorschrift $\omega \mapsto X(\omega)$ an!
 - b) Geben Sie diejenigen $\omega \in \Omega$ an, welche auf die Endnote $X = 5$ bzw. $X = 6$ abgebildet werden.
 - c) Berechnen Sie die Verteilungsfunktion der ausgewürfelten Endnote. Wie hoch ist die Wahrscheinlichkeit für das Bestehen der Klausur (d.h. für eine Endnote besser oder gleich 4)?
3. Welche der in der in diesem Kapitel eingeführten diskreten Verteilungsfamilien besitzen die folgenden Zufallsvariablen? Geben Sie jeweils eine kurze Erklärung an!
 - a) Banana Joe liefert jeden Tag 2000 Bananen von seinem Heimatdorf Amantido in die Hauptstadt Puerto Grande. Da es beim Be- und Entladen der Bananen teilweise etwas ruppig zugeht (sehen Sie sich einfach mal den Film an), erleidet jede einzelne Banane mit einer Wahrscheinlichkeit von 2% einen Schaden (unabhängig von Banane zu Banane). Es sei X_1 die Zufallsvariable die die Anzahl der beschädigten Bananen in der Lieferung beschreibt.
 - b) In einem Areal lebt eine unbekannte Anzahl N von Tieren. Um die Populationssgröße zu schätzen, verfahren Ökologen nach dem folgenden Schema: Zunächst fangen sie eine Zahl m von Tieren und markieren sie. Diese werden wieder frei

gelassen. Man wartet ab, bis sie sich mit den übrigen gut durchmischt haben und fängt dann (zufällig) n Tiere ein. Angenommen wird, dass alle Tiere mit gleicher Wahrscheinlichkeit gefangen werden. Es sei X_2 die Zufallsvariable die die Anzahl der markierten Tiere in der Stichprobe beschreibt.

- c) Sie bewerben sich nach dem folgenden Schema auf Arbeitsstellen: Sie schreiben eine Bewerbung an eine Firma, warten auf die Antwort und schreiben erst im Falle einer negativen Antwort die nächste Bewerbung. Da Sie in der Stochastik Vorlesung immer super aufgepasst haben und Sie somit für die Firmen wertvoll sind, schätzen Sie ihre Erfolgsquote auf 90% für jede Bewerbung. Die Zufallsvariable X_3 beschreibt die Anzahl der abgelehnten Bewerbungen bis der erste Erfolg eintritt.

3. Theorie

Gegeben sei die hypergeometrische Verteilung $H(n, n_1, n_2)$ und die Binomialverteilung $B(n, p)$, wobei $n, n_1, n_2 \in \mathbb{N}$, $n \leq n_1 + n_2$ und $p \in (0, 1)$. Beweisen Sie, dass für alle $j \in \mathbb{Z}$ im Limes $n_1, n_2 \rightarrow \infty$, so dass $\frac{n_1}{n_1 + n_2} = p$, $H(n, n_1, n_2)(\{j\})$ in $B(n, p)(\{j\})$ übergeht.

4. Aktivität

Für den messbaren Raum $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ sind Wahrscheinlichkeitsmaße, Verteilungsfunktionen und diskrete Dichten äquivalente Konzepte. Zeichnen Sie eine Mindmap und geben Sie die Verbindungen dieser Konzepte durch Pfeile an. Schreiben Sie ebenfalls an diese Pfeile die relevante Formel die den Übergang erlaubt. Geben Sie auch Referenzen zu den relevanten Sätzen oder Formeln in diesem Kapitel.

6.2. Literatur

Die Binomialverteilung wurde zuerst von Jacob Bernoulli (1655-1705) beschrieben, [Ber99].

Die Poissonverteilung wurde von S. Poisson (1781-1840) im Jahr 1837 eingeführt.

Neben den hier genannten Verteilungen, existieren noch weitere wichtige diskrete Verteilungsfamilien, etwa die negative Binomialverteilung oder allgemeiner die Panjer-Verteilungsfamilie, welche Geometrische, Binomiale und Negativ-Binomiale Verteilungen einheitlich beschreibt [Sch09].

7. Kontinuierliche Zufallsvariablen – Grundlagen

Wir möchten nun kontinuierliche Zufallsvariablen mit Wertebereich \mathbb{R} untersuchen. Hier stellt sich zunächst die Frage, welches die geeignete σ -Algebra auf \mathbb{R} ist. Wir haben schon gesehen, dass die Potenzmenge $\mathbb{P}(\mathbb{R})$ ‘zu groß’ ist, vergleiche die Übungen in Kapitel 2. Andererseits ist die σ -Algebra die durch $\{\mathbb{R}, \mathbb{R} \setminus \{x_j\}, \{x_j\}, \emptyset : \{x_j\} \subseteq \mathbb{R} \text{ Folge}\}$ definiert ist, zu klein, denn ein jedes Wahrscheinlichkeitsmaß mit $P(\{x\}) = 0, \forall x \in \mathbb{R}$, nimmt auf dieser σ -Algebra nur die Werte 1 oder 0 an.

Für reellwertige Zufallsvariablen X ist es auf jeden Fall sinnvoll, von der Wahrscheinlichkeit zu sprechen, dass der Wert von X in einem bestimmten Intervall liegt. Also sollten Intervalle in einer solchen σ -Algebra liegen. Andererseits wird schon aus den Intervallen eine sehr große σ -Algebra ‘erzeugt’, die für alle praktischen Zwecke ausreicht.

Zunächst präzisieren wir den Begriff des ‘Erzeugens’:

Lemma 7.1. *Zu jedem Mengensystem $\mathcal{A}^\circ \subseteq \mathbb{P}(\Omega)$ existiert eine eindeutig bestimmte σ -Algebra $\sigma(\mathcal{A}^\circ)$ über Ω , so dass*

$$(i) \quad \mathcal{A}^\circ \subseteq \sigma(\mathcal{A}^\circ),$$

(ii) *für alle σ -Algebren \mathcal{A} über Ω mit $\mathcal{A}^\circ \subseteq \mathcal{A}$ gilt $\sigma(\mathcal{A}^\circ) \subseteq \mathcal{A}$ (d.h. $\sigma(\mathcal{A}^\circ)$ ist minimal).*

Beweis. Es sei $\Sigma(\mathcal{A}^\circ)$ die Menge aller σ -Algebren über Ω , die \mathcal{A}° enthalten. Setze

$$\sigma(\mathcal{A}^\circ) = \bigcap_{\mathcal{A} \in \Sigma(\mathcal{A}^\circ)} \mathcal{A}. \quad (7.1)$$

Wir wollen zeigen, dass $\sigma(\mathcal{A}^\circ)$ eine σ -Algebra ist. Da Ω und \emptyset in allen $\mathcal{A} \in \Sigma(\mathcal{A}^\circ)$ enthalten sind, gilt dies auch für den Durchschnitt $\sigma(\mathcal{A}^\circ)$. Ist $A \in \sigma(\mathcal{A}^\circ) \Rightarrow A \in \mathcal{A} \forall \mathcal{A} \in \Sigma(\mathcal{A}^\circ) \Rightarrow A^c \in \mathcal{A} \forall \mathcal{A} \in \Sigma(\mathcal{A}^\circ) \Rightarrow A^c \in \sigma(\mathcal{A}^\circ)$. Ist $A_j \in \sigma(\mathcal{A}^\circ), j \in \mathbb{N}, \Rightarrow A_j \in \mathcal{A} \forall \mathcal{A} \in \Sigma(\mathcal{A}^\circ), j \in \mathbb{N}, \Rightarrow \bigcup_{j \in \mathbb{N}} A_j \in \mathcal{A} \forall \mathcal{A} \in \Sigma(\mathcal{A}^\circ) \Rightarrow \bigcup_{j \in \mathbb{N}} A_j \in \sigma(\mathcal{A}^\circ)$. \square

Definition 7.2. (i) Die in Lemma 7.1 definierte σ -Algebra $\sigma(\mathcal{A}^\circ)$ wird die **von \mathcal{A}° erzeugte σ -Algebra** genannt.

(ii) Ist \mathcal{A} eine σ -Algebra über Ω und gilt $\sigma(\mathcal{A}^\circ) = \mathcal{A}$ für ein $\mathcal{A}^\circ \subseteq \mathcal{A}$, dann heißt \mathcal{A}° **Erzeuger** oder **erzeugendes System** für \mathcal{A} .

(iii) Ein Erzeuger \mathcal{A}° von \mathcal{A} heißt **schnittstabil**, falls

$$A, B \in \mathcal{A}^\circ \Rightarrow A \cap B \in \mathcal{A}^\circ. \quad (7.2)$$

7. Kontinuierliche Zufallsvariablen – Grundlagen

Wir geben nun bestimmte Intervalle als Erzeuger unserer σ -Algebra über \mathbb{R} vor.

Definition 7.3. Wir betrachten das System der links offenen und rechts geschlossenen Intervalle in \mathbb{R} ,

$$\mathcal{I} := \{(y, x] \cap \mathbb{R} : y \in [-\infty, \infty), x \in (-\infty, \infty], y \leq x\}. \quad (7.3)$$

(i) $\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{I})$ wird die **Borelsche σ -Algebra** über \mathbb{R} genannt.

(ii) Eine messbare Abbildung $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ wird **reellwertige Zufallsvariable** genannt.

Bemerkung 7.4. Da die Einbettung von \mathbb{Z} in \mathbb{R} , $i : (\mathbb{Z}, \mathbb{P}(\mathbb{Z})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, eine messbare Abbildung ist, können wir jede \mathbb{Z} -wertige Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ auch mit der reellwertigen Zufallsvariablen $i \circ X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ identifizieren.

Die nächste Aufgabe wird es sein, Verteilungen (d.h. Bildmaße reellwertiger Zufallsvariablen) auf der Borelschen σ -Algebra zu charakterisieren. Dazu zitieren wir ohne Beweis den folgenden Eindeutigkeitssatz aus der Maßtheorie:

Satz 7.5. *Es sei \mathcal{A}° ein schnittstabiler Erzeuger der σ -Algebra \mathcal{A} über Ω und $\Omega \in \mathcal{A}^\circ$. Es seien P_1 und P_2 zwei Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) , so dass $P_1(A) = P_2(A)$ für alle $A \in \mathcal{A}^\circ$. Dann sind beide Wahrscheinlichkeitsmaße gleich, d.h. es gilt $P_1(A) = P_2(A)$ für alle $A \in \mathcal{A}$.*

Beweis. Siehe H. Bauer, Maß- und Integrationstheorie, 2. Auflage, de Gruyter 1992, Satz 5.4. □

Aus diesem allgemeinen Sachverhalt ergibt sich sofort folgendes

Korollar 7.6. *Zwei Wahrscheinlichkeitsmaße P_1 und P_2 auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ stimmen überein, wenn sie auf \mathcal{I} übereinstimmen.*

Beweis. Der Schnitt zweier links offener und rechts geschlossener Intervalle ist entweder leer oder ein ebensolches Intervall. Nach unserer Definition ist für $y = x$ auch $\emptyset = (y, x] \in \mathcal{I}$. Zuletzt ist auch $\mathbb{R} = (-\infty, \infty] \cap \mathbb{R} \in \mathcal{I}$. Damit erfüllt \mathcal{I} die Voraussetzungen des Satzes 7.5. □

Diese Eindeutigkeitsaussage führt dazu, dass man Wahrscheinlichkeitsmaße auch auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ über die zugehörige Verteilungsfunktion charakterisieren kann.

Definition 7.7. (i) Es sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Funktion, die die Bedingungen (i)–(iii) aus Satz 6.20 erfüllt. Dann heißt F **Verteilungsfunktion** auf \mathbb{R} .

(ii) Für ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist $F_P(x) := P((-\infty, x])$, $x \in \mathbb{R}$, die **zu P assoziierte Verteilungsfunktion**.

(iii) Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine reellwertige Zufallsvariable. Dann ist $F_X(x) := P(X \leq x) := F_{X*P}(x)$ die **zu X assoziierte Verteilungsfunktion**.

Korollar 7.8. *Ist P ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, dann ist P durch seine Verteilungsfunktion eindeutig bestimmt. Insbesondere ist auch die Verteilung einer reellwertigen Zufallsvariablen durch ihre Verteilungsfunktion festgelegt.*

Beweis. Für $(y, x] \in \mathcal{I}$ ist

$$P((y, x]) = P((-\infty, x] \setminus (-\infty, y]) = P((-\infty, x]) - P((-\infty, y]) = F_P(x) - F_P(y) . \quad (7.4)$$

Also bestimmt F_P das Wahrscheinlichkeitsmaß P eindeutig auf \mathcal{I} . Die Behauptung folgt nun aus dem Korollar 7.6. \square

Lemma 7.9. *Es sei F_P die Verteilungsfunktion assoziiert zu einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann ist F_P eine Verteilungsfunktion auf \mathbb{R} im Sinne von Definition 7.7.*

Beweis. Der Beweis für die Eigenschaften (ii) und (iii) kann analog zu dem Nachweis der entsprechenden Eigenschaften in Satz 6.20 geführt werden. Die Eigenschaft (i) ist einmal mehr eine Folge von Proposition 2.11. \square

Was uns jetzt noch fehlt ist ein einfaches Rezept, wie man Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definieren kann. Es stellt sich heraus, dass der Begriff der Verteilungsfunktion das Problem löst. In der Tat existiert zu jeder Verteilungsfunktion F genau ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so dass $F_P = F$. Der Beweis ist leider um Einiges schwieriger als im diskreten Fall und sprengt den Rahmen unserer Behandlung. Deshalb zeigen wir hier nur kurz, wie sich diese Aussage in den allgemeinen maßtheoretischen Rahmen einfügt und geben Zitate für die benötigten Sätze der Maßtheorie.

Definition 7.10. Das Mengensystem

$$\mathcal{F} := \left\{ A \subseteq \mathbb{R} : A = \bigcup_{j=1}^n I_j, \ I_j \in \mathcal{I}, \ j = 1, \dots, n, \ n \in \mathbb{N} \right\} \quad (7.5)$$

heißt **Algebra der Figuren** über \mathbb{R} .

Lemma 7.11. *\mathcal{F} ist eine Algebra in dem Sinne, dass \mathcal{F} die σ -Algebra Axiome aus Definition 2.2 (i) und (ii) erfüllt, und statt (iii) die Geschlossenheit unter endlichen Vereinigungen*

$$A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F} . \quad (7.6)$$

Beweis. Da $\emptyset, \Omega \in \mathcal{I}$ gilt auch $\emptyset, \Omega \in \mathcal{F}$. Also gilt Definition 2.2 (i).

Für $A = \bigcup_{j=1}^n I_j$ und $B = \bigcup_{j=1}^m J_j$, mit $I_j, J_j \in \mathcal{I}$, gilt

$$A \cup B = \bigcup_{j=1}^{n+m} K_j, \quad \text{mit } K_j = \begin{cases} I_j & , \text{ für } j = 1, \dots, n, \\ J_{j-n} & , \text{ für } j = n+1, \dots, n+m. \end{cases} \in \mathcal{I} \quad (7.7)$$

7. Kontinuierliche Zufallsvariablen – Grundlagen

und somit die Geschlossenheit unter endlichen Vereinigungen.

Zu zeigen bleibt $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$. Es sei zunächst $I \in \mathcal{I}$, d.h. $I = (y, x]$, mit $y \leq x$. Dann ist $I^c = J \cup K$, mit $J = (-\infty, y] \in \mathcal{I}$ und $K = (x, \infty] \cap \mathbb{R} \in \mathcal{I}$. Nun gilt für beliebiges $A \in \mathcal{F}$, $A = \bigcup_{j=1}^n I_j$, mit $I_j \in \mathcal{I}$, und somit

$$A^c = \left(\bigcup_{j=1}^n I_j \right)^c = \bigcap_{j=1}^n (K_j \cup J_j) = \bigcup_{L \subseteq \{1, \dots, n\}} \left(\bigcap_{j \in L} J_j \cap \bigcap_{j \in L^c} K_j \right). \quad (7.8)$$

Im letzten Schritt haben wir Lemma 7.12 vorgegriffen, welches wir im Anschluss beweisen wollen. Wegen der Schnittstabilität von \mathcal{I} , $K_j, J_j \in \mathcal{I}$ und der Endlichkeit der Potenzmenge von $\{1, \dots, n\}$ besteht A^c aus einer endlichen Vereinigung von Intervallen in \mathcal{I} . Also gilt $A^c \in \mathcal{F}$. \square

Lemma 7.12. Für Teilmengen $A_j, B_j \subseteq \Omega$, $j = 1, \dots, n$ gilt

$$\bigcap_{j=1}^n (A_j \cup B_j) = \bigcup_{L \subseteq \{1, \dots, n\}} \left(\bigcap_{j \in L} A_j \cap \bigcap_{j \in L^c} B_j \right). \quad (7.9)$$

Hierbei verwenden wir die Konvention $\bigcap_{j \in \emptyset} A_j = \bigcap_{j \in \emptyset} B_j = \Omega$.

Beweis. Der Beweis erfolgt durch vollständige Induktion über n , wobei der Induktionsanfang

$$A_1 \cup B_1 = (A_1 \cap \Omega) \cup (\Omega \cap B_1) \quad (7.10)$$

lautet. Der Induktionsschritt $n \Rightarrow n+1$ wird folgendermaßen ausgeführt

$$\begin{aligned} \bigcap_{j=1}^{n+1} (A_j \cup B_j) &= (A_{n+1} \cup B_{n+1}) \cap \left(\bigcap_{j=1}^n (A_j \cup B_j) \right) \\ &= (A_{n+1} \cup B_{n+1}) \cap \bigcup_{L \subseteq \{1, \dots, n\}} \left(\bigcap_{j \in L} A_j \cap \bigcap_{j \in L^c} B_j \right) \\ &= \bigcup_{L \subseteq \{1, \dots, n\}} \left(A_{n+1} \cap \bigcap_{j \in L} A_j \cap \bigcap_{j \in L^c} B_j \right) \cup \bigcup_{L \subseteq \{1, \dots, n\}} \left(\bigcap_{j \in L} A_j \cap \bigcap_{j \in L^c} B_j \cap B_{n+1} \right) \\ &= \bigcup_{L \subseteq \{1, \dots, n+1\}} \left(\bigcap_{j \in L} A_j \cap \bigcap_{j \in L^c} B_j \right). \end{aligned} \quad (7.11)$$

\square

Definition 7.13. Gegeben sei eine Algebra \mathcal{A}° über einer Menge Ω . Eine Abbildung $P^\circ : \mathcal{A}^\circ \rightarrow [0, 1]$ heißt **Inhalt**, falls

- (i) $P^\circ(\Omega) = 1$, (Normiertheit)
- (ii) für $A_1, \dots, A_n \in \mathcal{A}^\circ$ paarweise disjunkt gilt

$$P^\circ\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n P^\circ(A_j) \quad . \quad (\text{endliche Additivität}) \quad (7.12)$$

Gilt darüber hinaus σ -Additivität für alle paarweise disjunkten Folgen $A_j \in \mathcal{A}^\circ$, $j \in \mathbb{N}$, mit $\bigcup_{j=1}^\infty A_j \in \mathcal{A}^\circ$, dann heißt P° ein **Prämaß**.

Uns interessiert insbesondere, wann ein Inhalt auf der Algebra der Figuren zu einem Wahrscheinlichkeitsmaß auf der Borelschen σ -Algebra erweitert werden kann. Aus Konsistenzgründen müßte zumindest die σ -Additivität für solche paarweise disjunkten Ereignisse gelten, deren abzählbare Vereinigung in der Algebra selbst liegt. Die zusätzliche Prämaß-Eigenschaft ist also notwendig für die Fortsetzung auf die von der Algebra erzeugten σ -Algebra. Der Fortsetzungssatz von Constantin Carathéodory besagt, dass dies auch schon hinreichend ist.

Satz 7.14 (Fortsetzungssatz von Carathéodory). *Es sei \mathcal{A}° eine Algebra über einer Menge Ω und $P^\circ : \mathcal{A}^\circ \rightarrow [0, 1]$ ein Prämaß. Dann existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P auf $(\Omega, \sigma(\mathcal{A}^\circ))$, so dass die Einschränkung von P auf \mathcal{A}° mit P° übereinstimmt.*

Beweis. Siehe H. Bauer, Maß- und Integrationstheorie, 2. Auflage, de Gruyter 1992, Satz 5.3. \square

Mit Hilfe von diesen technischen Werkzeugen wollen wir nun zeigen, dass es für die Definition eines Wahrscheinlichkeitsmaßes auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ genügt, eine Verteilungsfunktion anzugeben. Wir zeigen zunächst, dass eine Verteilungsfunktion F die Definition eines Inhaltes P° erlaubt.

Lemma 7.15. *Es sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion auf \mathbb{R} . Für ein Element $A \in \mathcal{F}$ aus der Algebra der Figuren \mathcal{F} über \mathbb{R} existiert eine Darstellung $A = \bigcup_{j=1}^n I_j$ mit $I_j = (y_j, x_j]$, $j = 1, \dots, n$, paarweise disjunkt. Setze nun*

$$P^\circ(A) := \sum_{j=1}^n (F(x_j) - F(y_j)) . \quad (7.13)$$

Dann ist P° ein Inhalt auf \mathcal{F} .

Beweis. Zunächst einmal ist für eine Figur $A = \bigcup_{j=1}^n I_j$ mit $I_j \in \mathcal{I}$ beliebig eine Darstellung möglich, bei der die Intervalle I_j paarweise disjunkt und nicht leer sind (es sei denn $A = \emptyset$). Denn sind etwa I_j und I_k nicht disjunkt, dann ist $I_j \cup I_k \in \mathcal{I}$. Also können wir die zwei Intervalle I_j und I_k in der Darstellung von A durch ein neues Intervall ersetzen. Dann hat A eine Darstellung mit $n - 1$ Intervallen. Indem wir so fortfahren, erhalten wir schließlich nach endlich vielen Schritten eine Darstellung durch paarweise disjunkte Intervalle. Im Folgenden nennen wir diese Darstellung wieder $\bigcup_{j=1}^n I_j$ mit möglicher Weise redefinierten n und I_j .

Diese Darstellung ist nicht notwendiger Weise eindeutig: Indem wir die Anfangspunkte der Intervalle y_j ordnen, d.h. $y_1 < y_2 < \dots < y_n$, erhält man aus der Disjunktheit, dass $x_1 \leq y_2, \dots, x_{n-1} \leq y_n$. In dem Fall, dass für ein $j = 1, \dots, n - 1$ gilt $x_j = y_{j+1}$, kann man $I_j = (y_j, x_j]$ und $I_{j+1} = (y_{j+1}, x_{j+1}]$ durch $(y_j, x_{j+1}]$ ersetzen. Nach endlich vielen Schritten

7. Kontinuierliche Zufallsvariablen – Grundlagen

erhält man nun eine eindeutige Darstellung durch Intervalle $\tilde{I}_j = (\tilde{y}_j, \tilde{x}_j]$, $j = 1, \dots, \tilde{n}$, so dass $\tilde{x}_j < \tilde{y}_{j+1}$.

Für jede Darstellung $A = \bigcup_{j=1}^n I_j$ durch paarweise disjunkte I_j gilt nun

$$P^\circ(A) = \sum_{j=1}^n P^\circ(I_j) = \sum_{j=1}^n (F(x_j) - F(y_j)) = \sum_{j=1}^{\tilde{n}} P^\circ(\tilde{I}_j) = \sum_{j=1}^{\tilde{n}} (F(\tilde{x}_j) - F(\tilde{y}_j)) . \quad (7.14)$$

Hier verwenden wir wieder die Konvention $F(-\infty) = 0$ und $F(\infty) = 1$. Möglicherweise zusätzlich auftretende Terme $F(x_j)$ im mittleren Ausdruck, die zu einem j mit $x_j = y_{j+1}$ gehören, werden durch den $-F(y_{j+1})$ Term im darauffolgenden Summanden kompensiert. Insbesondere ist P° wohldefiniert auf \mathcal{F} .

Wegen der Monotonie von F gilt $F(x_j) - F(y_j) \geq 0$ und deshalb $P^\circ(A) \geq 0$, $\forall A \in \mathcal{F}$. Ebenfalls ist $P^\circ(\mathbb{R}) = F(\infty) - F(-\infty) = 1 - 0 = 1$, also gilt Normiertheit. Wenn wir nun zeigen können, dass P° endlich additiv ist, dann erhalten wir automatisch $P^\circ(A) \leq P^\circ(\mathbb{R}) = 1$, denn Isotonie ist eine Folge der Additivität, vergleiche Proposition 2.8 (iii). Dies würde also implizieren, dass wie gefordert $P^\circ : \mathcal{F} \rightarrow [0, 1]$.

Um die endliche Additivität zu zeigen, betrachte $A = \bigcup_{j=1}^n I_j$, $B = \bigcup_{j=1}^m J_j \in \mathcal{F}$ mit $A \cap B = \emptyset$. Es folgt, dass $K_1 = I_1, \dots, K_n = I_n, K_{n+1} = J_1, \dots, K_{n+m} = J_m$ eine Zerlegung von $A \cup B$ in disjunkte Intervalle ist. Deshalb gilt

$$P^\circ(A \cup B) = \sum_{j=1}^{n+m} P^\circ(K_j) = \sum_{j=1}^n P^\circ(I_j) + \sum_{j=1}^m P^\circ(J_j) = P^\circ(A) + P^\circ(B) . \quad (7.15)$$

Hieraus ergibt sich nun durch vollständige Induktion die endliche Additivität. \square

Die Rechtsstetigkeit einer Verteilungsfunktion F ist entscheidend dafür, dass der folgende Sachverhalt gilt.

Satz 7.16. *Ist P° der zu einer Verteilungsfunktion F gehörende Inhalt auf $(\mathbb{R}, \mathcal{F})$, dann ist P° ein Prämaß. Insbesondere existiert zu jeder Verteilungsfunktion F ein eindeutiges Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so dass $F(x) = P((-\infty, x])$ für alle $x \in \mathbb{R}$.*

Beweis. Die zweite Aussage folgt aus der ersten und Satz 7.14. Für die erste Aussage verweisen wir z.B. auf Satz 6.5 in H. Bauer, Maß- und Integrationstheorie, 2. Auflage, de Gruyter 1992. Beachten Sie, dass sich die Konventionen in diesem Buch von den hier gewählten Konventionen dahingehend unterscheiden, dass mit den links stetigen Funktionen $F(x) = P(X < x)$ gerechnet wird. An den Beweisideen ändert das aber nichts. \square

Im Weiteren ist es oft bequem, Verteilungsfunktionen über kontinuierliche Dichten zu beschreiben.

Definition 7.17. Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine reellwertige Zufallsvariable mit Verteilungsfunktion F_X .

- (i) Falls $F_X(x)$ stetig ist, dann heißt X **kontinuierlich**.

- (ii) Ist $F_X(x)$ sogar stetig und differenzierbar für $x \in \mathbb{R} \setminus \mathcal{N}$, mit \mathcal{N} einer abzählbaren Menge, so ist die **kontinuierliche Wahrscheinlichkeitsdichte** (kurz: kontinuierliche Dichte) von X gegeben durch

$$f_X(x) := \begin{cases} F'_X(x) & , \text{ für } x \in \mathbb{R} \setminus \mathcal{N} , \\ 0 & , \text{ sonst .} \end{cases} \quad (7.16)$$

Bemerkung 7.18. Natürlich kann man sich eine kontinuierliche Dichte auch ebenso als assoziiert zu einem Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ vorstellen, zu welchem eine differenzierbare Verteilungsfunktion gehört.

Satz 7.19. *Es sei $F_X : \mathbb{R} \rightarrow [0, 1]$ eine differenzierbare Verteilungsfunktion und $f_X : \mathbb{R} \rightarrow \mathbb{R}$ die zugehörige kontinuierliche Dichte. Dann gilt*

- (i) $f_X(x) \geq 0, \forall x \in \mathbb{R}$,
- (ii) $\int_{-\infty}^{\infty} f_X(x) dx = 1$, (Normiertheit)
- (iii) $F_X(x) = \int_{-\infty}^x f_X(y) dy, \forall x \in \mathbb{R}$.

Sei umgekehrt $f : \mathbb{R} \rightarrow \mathbb{R}$ eine (Lebesgue-) integrierbare Funktion mit den Eigenschaften (i) und (ii). Dann definiert (iii) eine Verteilungsfunktion.

Beweis. (i) Wegen der Monotonie von F_X gilt

$$\frac{F_X(x+h) - F_X(x)}{h} \geq 0, \quad \forall h \in \mathbb{R} \setminus \{0\} . \quad (7.17)$$

Die Ungleichung bleibt stabil unter der Limes-Bildung $0 \neq h \rightarrow 0$.

(ii) folgt aus (iii) und $\lim_{x \rightarrow \infty} F_X(x) = 1$.

(iii) Für $y < x$ ist nach dem Hauptsatz der Differential- und Integralrechnung $F_X(x) - F_X(y) = \int_y^x f_X(z) dz$. Die Behauptung ergibt sich nun aus $\lim_{y \rightarrow -\infty} F_X(y) = 0$.

Es sei nun umgekehrt eine nichtnegative und normierte Funktion f gegeben. Dann ist $F(x) = \int_{-\infty}^x f(y) dy$ differenzierbar, also stetig und insbesondere rechtsstetig. F ist monoton, denn für $x > y$ gilt

$$F(x) - F(y) = \int_y^x f(z) dz \geq 0 . \quad (7.18)$$

Die Bedingung $\lim_{x \rightarrow \pm\infty} F(x) = \begin{cases} 1 \\ 0 \end{cases}$ folgt aus der Definition des uneigentlichen Integrals für Riemann-integrierbare Funktionen f und aus dem Satz über die dominierte Konvergenz für den allgemeineren Fall der Lebesgue-integrierbaren Funktionen. Zum Beweis des Satzes über die dominierte Konvergenz siehe wiederum H. Bauer, Maß- und Integrationstheorie, 2. Auflage, de Gruyter 1992, Satz 15.6. \square

7. Kontinuierliche Zufallsvariablen – Grundlagen

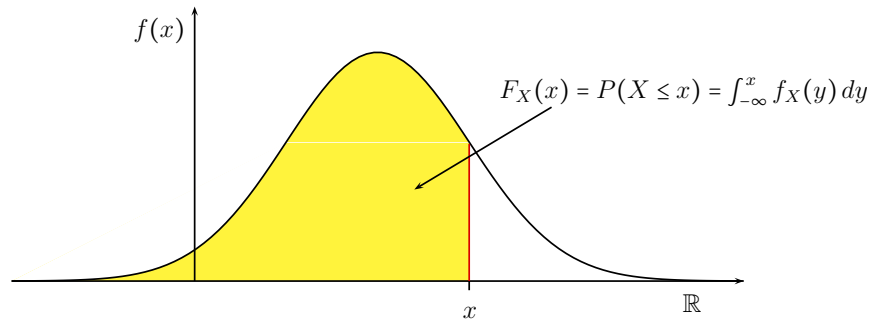


Abbildung 7.1: Verteilungsfunktion und kontinuierliche Dichte.

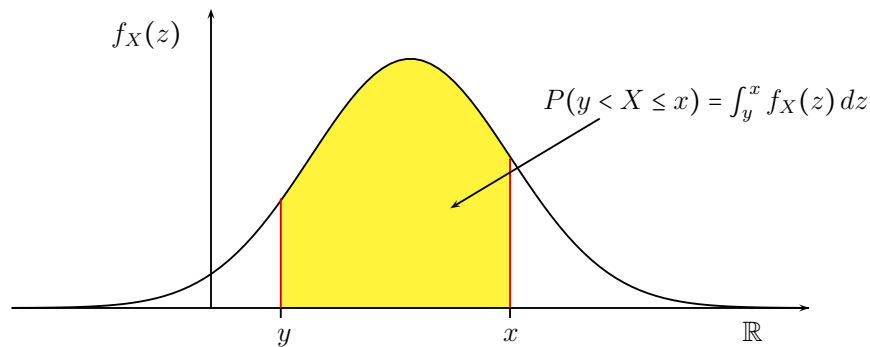


Abbildung 7.2: Bestimmung von Bereichswahrscheinlichkeiten mittels der kontinuierlichen Dichte.

Der Zusammenhang zwischen der kontinuierlichen Dichte und der Verteilungsfunktion wird in [Abbildung 7.1](#) illustriert. Die [Abbildung 7.2](#) zeigt, wie man aus der kontinuierlichen Dichte die Wahrscheinlichkeit bestimmt, dass X Werte in einem gegebenen Bereich annimmt, vergleiche [Lemma 6.21](#), das auch für reellwertige Zufallsvariablen gilt.

Zuletzt beschäftigen wir uns mit der Transformation von Zufallsvariablen, Wahrscheinlichkeitsmaßen, Verteilungsfunktionen und Dichten. Da wir bereits mit dem Begriff des Bildmaßes vertraut sind, ist die Transformation eines Maßes unter einer messbaren Abbildung bereits abgehandelt. Uns interessiert jedoch der Fall, in dem mehr als nur einmal transformiert wird, d.h.

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (E, \mathcal{B}) \xrightarrow{\phi} (E', \mathcal{B}') . \quad (7.19)$$

Hierbei sind (Ω, \mathcal{A}) , (E, \mathcal{B}) und (E', \mathcal{B}') messbare Räume und X, ϕ messbare Abbildungen zwischen den jeweiligen Räumen. Leicht zeigt man: Die Hintereinanderausführung von zwei messbaren Abbildungen bleibt messbar.

Uns interessiert nun die Verteilung von $Y = \phi \circ X : (\Omega, \mathcal{A}, P) \rightarrow (E', \mathcal{B}')$, $\omega \mapsto Y(\omega) =$

$\phi(X(\omega))$. Offensichtlich gilt

$$Y_*P = (\phi \circ X)_*P = \phi_*(X_*P) . \quad (7.20)$$

Doch was bedeutet dies nun für die Verteilungsfunktionen und Dichten, wenn $(E, \mathcal{B}) = (E', \mathcal{B}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$? Zumindest für den folgenden, einfachen Fall kann man dies sehr schön beantworten:

Satz 7.20 (Transformationssatz). *Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine reellwertige Zufallsvariable und $\phi : \mathbb{R} \rightarrow \mathbb{R}$ eine streng monoton wachsende und differenzierbare Funktion. Da eine solche Funktion insbesondere stetig ist und beliebige stetige Abbildungen Borel-messbar sind, können wir $\phi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ auch als messbare Funktion auffassen (siehe Bauer, MI, S. 57). Dann gilt:*

(i) *Die Verteilungsfunktion der transformierten Zufallsvariablen Y ist*

$$F_Y(y) = \begin{cases} 0 & , \text{ für } y < \text{Bild}(\phi) , \\ F_X(\phi^{-1}(y)) & , \text{ für } y \in \text{Bild}(\phi) , \\ 1 & , \text{ für } y > \text{Bild}(\phi) . \end{cases} \quad (7.21)$$

Hierbei wurde verwendet, dass $\text{Bild}(\phi) \subseteq \mathbb{R}$ ein Intervall sein muss und $\phi : \mathbb{R} \rightarrow \text{Bild}(\phi)$ bijektiv (also umkehrbar) ist, da $\phi : \mathbb{R} \rightarrow \mathbb{R}$ als streng monotone Funktion injektiv ist.

(ii) *Hat X darüber hinaus die kontinuierliche Dichte $f_X(x)$, dann gilt für $x \in \mathbb{R}$*

$$f_Y(y) = \begin{cases} 0 & , \text{ für } y < \text{Bild}(\phi) , \\ \frac{f_X(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))} & , \text{ für } y \in \text{Bild}(\phi) , \\ 0 & , \text{ für } y > \text{Bild}(\phi) . \end{cases} \quad (7.22)$$

Beweis. (i) Beachte, dass mit $\phi : \mathbb{R} \rightarrow \text{Bild}(\phi)$ streng monoton steigend und differenzierbar auch $\phi^{-1} : \text{Bild}(\phi) \rightarrow \mathbb{R}$ streng monoton steigend und differenzierbar.

Für $y < \text{Bild}(\phi)$ gilt $\phi^{-1}((-\infty, y]) = \emptyset$ und somit

$$Y^{-1}((-\infty, y]) = (\phi \circ X)^{-1}((-\infty, y]) = X^{-1}(\phi^{-1}((-\infty, y])) = X^{-1}(\emptyset) = \emptyset . \quad (7.23)$$

Deshalb gilt für diesen Fall $F_Y(y) = Y_*P((-\infty, y]) = P(\emptyset) = 0$. Andererseits, für $y > \text{Bild}(\phi)$ gilt $\phi^{-1}((0, y]) = \mathbb{R}$ und somit

$$Y^{-1}((-\infty, y]) = (\phi \circ X)^{-1}((-\infty, y]) = X^{-1}(\phi^{-1}((-\infty, y])) = X^{-1}(\mathbb{R}) = \Omega . \quad (7.24)$$

Deshalb gilt für diesen Fall $F_Y(x) = Y_*P((-\infty, y]) = P(\Omega) = 1$. Sei nun zuletzt $x \in \text{Bild}(\phi)$. Dann ist

$$\begin{aligned} F_Y(y) &= P(\{\omega \in \Omega : Y(\omega) \leq y\}) \\ &= P(\{\omega \in \Omega : \phi^{-1}(Y(\omega)) \leq \phi^{-1}(y)\}) \\ &= P(\{\omega \in \Omega : X(\omega) \leq \phi^{-1}(y)\}) = F_X(\phi^{-1}(y)) . \end{aligned} \quad (7.25)$$

7. Kontinuierliche Zufallsvariablen – Grundlagen

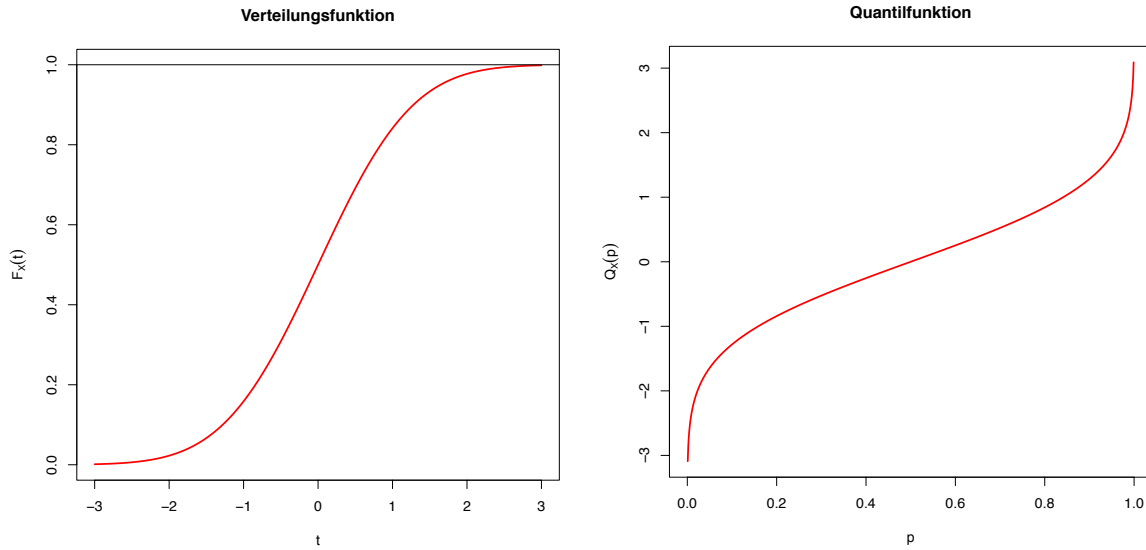


Abbildung 7.3: Verteilungs- und Quantilfunktion im kontinuierlichen Fall.

(ii) Für $y \in \text{Bild}(\phi)$ haben wir $y = \phi(\phi^{-1}(y))$ und wegen der Kettenregel gilt

$$1 = (\phi(\phi^{-1}(y)))' = (\phi^{-1})'(y) \phi'(\phi^{-1}(y)) . \quad (7.26)$$

Wenden wir nun die Kettenregel für $F_Y(y) = F_X(\phi^{-1}(y))$ an der Stelle $y \in \text{Bild}(\phi)$ an, so dass $\phi^{-1}(y) \notin \mathcal{N}$, so erhalten wir

$$f_Y(y) = F_Y'(y) = (F_X(\phi^{-1}(y)))' = F_X'(\phi^{-1}(y))(\phi^{-1})'(y) = \frac{f_X(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))} . \quad (7.27)$$

Die Ausnahmemenge $\tilde{\mathcal{N}} = \{y \in \mathbb{R} : \phi^{-1}(y) \in \mathcal{N}\}$ ist jedoch als Bild einer abzählbaren Menge abzählbar und es gilt $\tilde{\mathcal{N}} = \phi(\mathcal{N})$.

Schließlich sind noch die Fälle $y < \inf \text{Bild}(\phi)$ und $y > \sup \text{Bild}(\phi)$ zu betrachten. Hier ist $F_Y(y)$ konstant und somit überall (bis auf eventuell an den Endpunkten $\inf \text{Bild}(\phi)$ bzw. $\sup \text{Bild}(\phi)$) differenzierbar. \square

Auch im allgemeinen Fall kann man die Quantilfunktion definieren, vergleiche auch Abbildung 7.3.

Definition 7.21. Es sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion.

- (i) $Q(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}$ ist die **Quantilfunktion** zu F .
- (ii) Ist F streng monoton steigend und stetig, dann ist $Q(p) = F^{-1}(p)$.

Beispiel 7.22. In Simulationen möchte man oft Zufallszahlen aus einer vorgegebenen Verteilungsfunktion F generieren. F sei als streng monoton und stetig vorausgesetzt. Die

Algorithmen für die Erzeugung von (pseudo-) Zufallszahlen auf dem Computer geben zunächst jedoch $[0, 1]$ -gleichverteilte Zufallszahlen heraus, d.h. Zufallszahlen aus einer Dichte

$$f(x) = \begin{cases} 1 & , \text{ für } 0 \leq x \leq 1 , \\ 0 & , \text{ sonst .} \end{cases} \quad (7.28)$$

Diese Zufallszahlen sehen wir als Realisierungen einer $[0, 1]$ -gleichverteilten Zufallsvariablen X . Wie müssen wir diese Zahlen transformieren, um Zufallszahlen zu erhalten, die P_F -verteilt sind, wobei P_F das Wahrscheinlichkeitsmaß zur Verteilung F ist?

Lösung. Wir betrachten eine $[0, 1]$ -gleichverteilte Zufallsvariable X und setzen $Y = Q \circ X$ mit $Q = F^{-1} : (0, 1) \rightarrow (-\infty, \infty)$. Für $F_X(x)$ haben wir

$$F_X(x) = \int_{-\infty}^x f(y) dy = \begin{cases} 0 & , \text{ für } x < 0 , \\ x & , \text{ für } 0 \leq x \leq 1 , \\ 1 & , \text{ für } x > 1 . \end{cases} \quad (7.29)$$

Nach dem Transformationssatz gilt

$$F_Y(x) = F_X(Q^{-1}(x)) = F_X(F(x)) = F(x) , \quad (7.30)$$

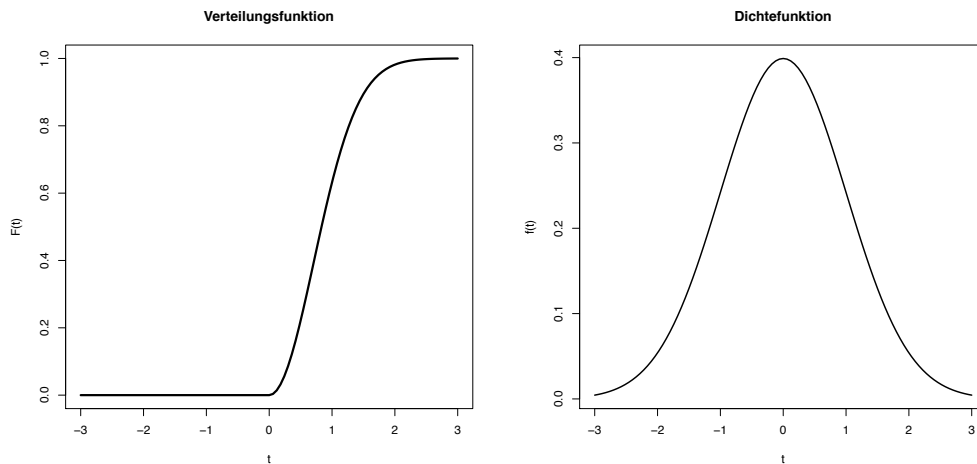
wobei wir im letzten Schritt (7.29) verwendet haben und dass $F(x) \in [0, 1]$ für alle $x \in \mathbb{R}$. Also ist $F_Y = F$ wie gewünscht.

Zusammenfassung: Wir transformieren die $[0, 1]$ -gleichverteilten (pseudo-) Zufallszahlen mit der Quantilfunktion Q um P_F (pseudo-) Zufallszahlen zu erhalten. \diamond

7.1. Übungsaufgaben

1. Selbsttest

1. Wir möchten nun etwas besser verstehen, welche Teilmengen $A \subseteq \mathbb{R}$ in der Borelschen σ -Algebra $\mathcal{B}(\mathbb{R})$ enthalten sind. Beantworten Sie die folgenden Fragen mit Ja oder Nein. Geben Sie entsprechend Ihrer Antwort einen Beweis oder ein Gegenbeispiel an, um ihre Antwort zu bekräftigen!
 - a) Sind die einelementigen Mengen $\{x\} \subset \mathbb{R}$, mit $x \in \mathbb{R}$, in $\mathcal{B}(\mathbb{R})$ enthalten?
 - b) Sind alle endlichen Mengen $\{x_1, \dots, x_n\} \subset \mathbb{R}$, mit $x_1, \dots, x_n \in \mathbb{R}$, in $\mathcal{B}(\mathbb{R})$ enthalten?
 - c) Sind alle Intervall-Typen (x, y) , $(x, y]$, $[x, y)$ und $[x, y]$, mit $x \leq y$, in $\mathcal{B}(\mathbb{R})$ enthalten?
 - d) Ist die Cantor Menge in $\mathcal{B}(\mathbb{R})$ enthalten?
 - e) Ist jede Teilmenge $A \subseteq \mathbb{R}$ in $\mathcal{B}(\mathbb{R})$ enthalten?
2. Die unten stehende Graphik zeigt eine Verteilungsfunktion und eine kontinuierliche Dichte.



Können beide von derselben Verteilung her stammen?

3. Die logistische Verteilung ist definiert durch die Verteilungsfunktion

$$F_X(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

Mit welcher Wahrscheinlichkeit nimmt eine logistisch verteilte Zufallsvariable X Werte

- a) kleiner oder gleich 0,
- b) im Intervall $(2, 5]$,

c) größer als 1 an?

4. Prüfen Sie ob die folgende Funktion eine kontinuierliche Dichte ist

$$f(x) = \begin{cases} 1 - |x| & , \text{ für } -1 \leq x \leq 1 , \\ 0 & , \text{ sonst ,} \end{cases}$$

wobei $|\cdot|$ die Betragsfunktion ist. Berechnen Sie ebenfalls die zu f gehörige Verteilungsfunktion F und skizzieren Sie deren Verlauf. Wie groß ist $P(|X| \leq 0.3)$?

2. Theorie

1. Wir betrachten einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und eine reellwertige Zufallsvariable $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
 - a) Sei $X_*(P)(\{x\}) = P(X = x) \neq 0$ für ein $x \in \mathbb{R}$. Zeigen Sie, dass dann die Verteilungsfunktion F_X *nicht* stetig in diesem x ist.
 - b) Diese Aussage kann man noch verschärfen. Beweisen Sie, dass F_X dann und nur dann stetig ist, falls $P(X = x) = 0$ für alle $x \in \mathbb{R}$.
2. Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine kontinuierliche Zufallsvariable mit Verteilungsfunktion F_X . Im Transformationssatz haben wir gelernt, wie man die Verteilungsfunktion F_Y der transformierten Zufallsvariablen $Y := \phi \circ X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ bestimmt, falls $\phi : \mathbb{R} \rightarrow \mathbb{R}$ eine streng monoton steigende und differenzierbare Funktion ist. Dies beinhaltet *nicht* den oftmals relevanten Fall, dass $\phi : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \phi(x) = x^{2n}$ mit $n \in \mathbb{N}$. Bestimmen Sie eigenständig eine Formel, die die Verteilungsfunktion F_Y der Zufallsvariablen $Y = \phi \circ X = X^{2n}$ durch F_X ausdrückt!

Hinweis: Setzen Sie $F_Y(y) = P(Y^{-1}((-\infty, y]))$ an und versuchen Sie dies durch F_X auszudrücken. Eine Skizze von ϕ kann hilfreich sein!

3. Aktivität

Für den messbaren Raum $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sind unter gewissen Voraussetzungen Wahrscheinlichkeitsmaße, Verteilungsfunktionen und kontinuierliche Dichten äquivalente Konzepte. Zeichnen Sie eine Mindmap und geben Sie die Verbindungen dieser Konzepte durch Pfeile an. Spezifizieren Sie ebenfalls unter welchen Voraussetzungen diese Pfeile gelten und schreiben Sie diese die relevante Formel die den Übergang erlaubt. Geben Sie auch Referenzen zu den relevanten Sätzen oder Formeln in diesem Kapitel.

7.2. Literatur

8. Kontinuierliche Verteilungen – Beispiele

In diesem Abschnitt wird eine Anzahl von kontinuierlichen Zufallsverteilungen vorgestellt. Wir beginnen mit dem einfachsten Fall:

Definition 8.1. Die **Gleichverteilung** (oder auch **uniforme Verteilung**) $U(a, b)$ zu den Parametern $a, b \in \mathbb{R}$, $a < b$, ist das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, welches definiert ist durch die kontinuierliche Dichte

$$f_X(x) = \begin{cases} 0 & , \text{ für } x \leq a , \\ \frac{1}{b-a} & , \text{ für } a < x \leq b , \\ 0 & , \text{ für } x > b . \end{cases} \quad (8.1)$$

Eine Zufallsvariable X heißt $U(a, b)$ -verteilt, geschrieben $X \sim U(a, b)$, falls $\text{Vert}(X) = U(a, b)$.

Die Gleichverteilung führt in der Tat zu einem Wahrscheinlichkeitsmaß, denn $f_X(x) \geq 0$ und $\int_{\mathbb{R}} f_X(x) dx = \frac{1}{b-a} \int_a^b dx = 1$. Auch die Verteilungsfunktion zu $U(a, b)$ läßt sich leicht berechnen als

$$F_X(x) = \int_{-\infty}^x f_X(y) dy = \begin{cases} 0 & , \text{ für } x \leq a , \\ \frac{x-a}{b-a} & , \text{ für } a < x \leq b , \\ 1 & , \text{ für } x > b . \end{cases} \quad (8.2)$$

Siehe auch Abbildung 8.1 für den graphischen Verlauf von der kontinuierlichen Dichte und Verteilungsfunktion. Die uniforme Verteilung wird oft dann zur Modellierung herangezogen, wenn das Prinzip vom unzureichenden Grund auf ein Zufallsexperiment mit kontinuierlichem Ausgang im Intervall $[a, b]$ angewendet werden soll. Es ist jedoch zu beachten, dass eine naive Anwendung dieses Prinzips schnell zu Widersprüchen führen kann. Das folgende Beispiel ist auch als Buffonsches Nadelparadoxon bekannt:

Beispiel 8.2. Durch einen Kreis mit Radius r wird rein zufällig eine Sehne gezogen. Mit welcher Wahrscheinlichkeit ist diese Sehne länger als die Seite des einbeschriebenen Dreiecks?

Lösung 1. Ohne Beschränkung der Allgemeinheit können wir annehmen, dass ein Punkt der Sehne auf $(-r, 0)$ liegt. Nach dem Prinzip vom unzureichenden Grund wählen wir den Winkel φ der Sehne zur (verschobenen) y -Achse gleichverteilt aus $[0, \pi]$, vergleiche

8. Kontinuierliche Verteilungen – Beispiele

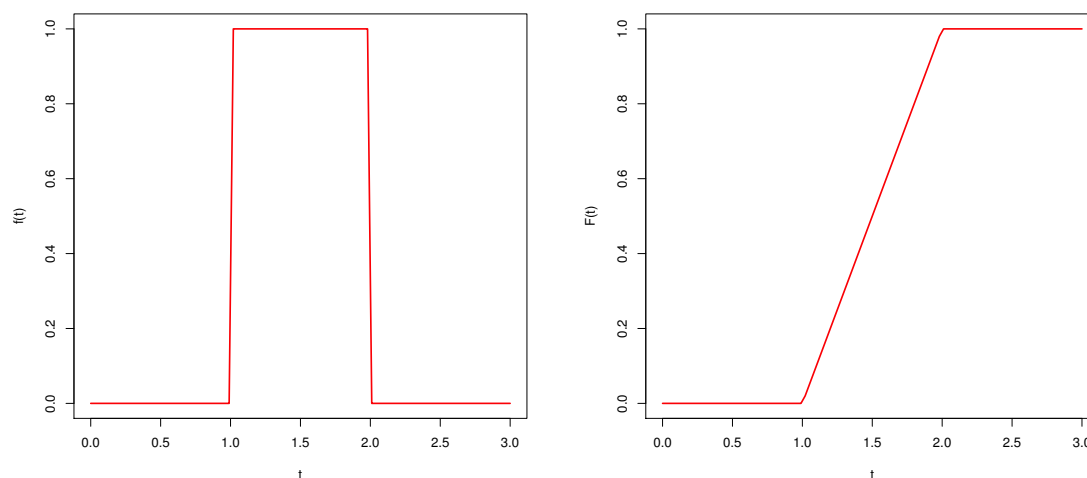


Abbildung 8.1: Kontinuierliche Dichte und Verteilungsfunktion von $U(1, 2)$.

Abbildung 8.2 (links). Das Ereignis A = ‘Sehne ist länger als einbeschriebene Seite’ ist dann gleichzusetzen mit $\varphi \in (\frac{\pi}{3}, \frac{2\pi}{3})$. Die Wahrscheinlichkeit für dieses Ereignis beträgt somit $P(A) = \frac{1}{\pi} \int_{\frac{\pi}{3}}^{\frac{2\pi}{3}} dx = \frac{1}{3}$. \diamond

Lösung 2. Eben so gut können wir den Abstand des Sehnenmittelpunktes zum Kreismittelpunkt als rein zufällig gegeben ansehen. Wieder nach dem Prinzip vom unzureichenden Grund wählen wir den Abstand $d \sim U(0, r)$, siehe Abbildung 8.2 (rechts). Das Ereignis A = ‘Sehne ist länger als einbeschriebene Seite’ entspricht $d < r/2$, denn die Sehne ist dann länger, wenn ihr Mittelpunkt näher am Kreismittelpunkt liegt, als der einer Seite eines einbeschriebenen Dreiecks. Für diesen Abstand gilt jedoch $d_{\text{Dreieck}} = r \sin(\pi/6) = \frac{r}{2}$. Also $P(A) = \frac{1}{r} \int_0^{\frac{r}{2}} dx = \frac{1}{2}$. \diamond

Bemerkung 8.3. Dieser scheinbare Widerspruch erklärt sich dadurch, dass die Formulierung ‘rein zufällig’ in der Fragestellung zu ungenau ist, da man sie auf sehr verschiedene Vorschriften, wie die Sehne zufällig zu platzieren ist, beziehen kann. Aus mathematischer Sicht führt dies z.B. dazu, dass die Transformation $\phi(\varphi) = r|\sin(\frac{\pi}{2} - \varphi)| = d$ nichtlinear ist und deshalb nach dem Transformationssatz die uniforme φ -Verteilung nicht auf eine uniforme d -Verteilung abgebildet wird. Das Prinzip vom unzureichenden Grund für φ ist deswegen nicht äquivalent zum Prinzip des unzureichenden Grundes für d .

Unsere nächste Verteilungsfamilie ist der Klassiker schlechthin und sie findet Anwendung in vielen praktischen relevanten Aufgaben.

Definition 8.4. Die **Normalverteilung** (oder auch **Gaussverteilung**) $N(\mu, \sigma^2)$ zu den Parametern $\mu \in \mathbb{R}$, $\sigma^2 \in (0, \infty)$ ist das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, welches

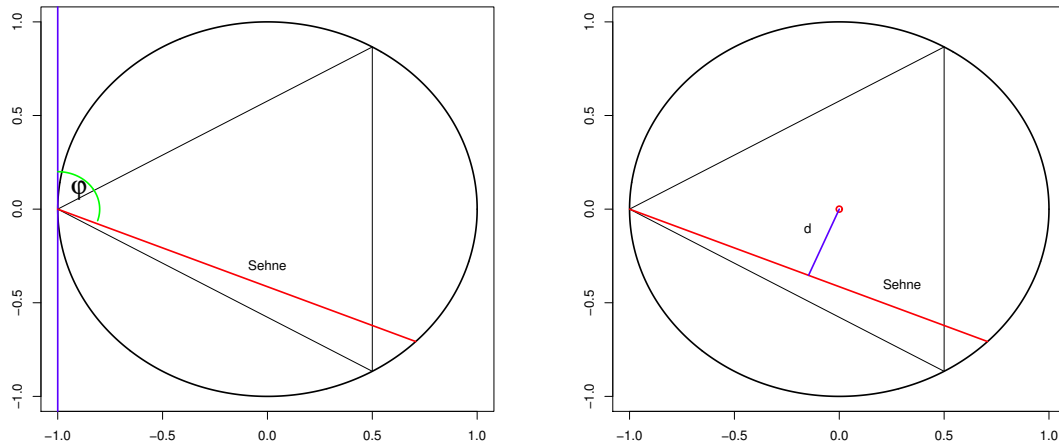


Abbildung 8.2: Zwei verschiedene Arten eine Sehne zufällig durch einen Kreis zu legen.

definiert ist durch die kontinuierliche Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} . \quad (8.3)$$

Ist eine Zufallsvariable X normalverteilt zu $N(\mu, \sigma^2)$, so schreibe $X \sim N(\mu, \sigma^2)$.

Lemma 8.5. Die Gaussdichte (8.3) ist eine kontinuierliche Dichte.

Beweis. Offensichtlich gilt $f_X(x) \geq 0$. Zu zeigen ist die Normiertheit $\int_{-\infty}^{\infty} f_X(x) dx = 1$. Hierzu setzen wir zunächst $\mu = 0$ und $\sigma^2 = 1$. Der klassische, auf Gauss zurückgehende, Trick quadriert diese Gleichung und verwendet dann Polarkoordinaten $dx dy = r dr d\varphi$:

$$\begin{aligned} \left(\int_{-\infty}^{\infty} f_X(x) dx \right)^2 &= \int_{-\infty}^{\infty} f_X(x) dx \int_{-\infty}^{\infty} f_X(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x) f_X(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dx dy \\ &= \int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r dr d\varphi \\ &= \int_0^{\infty} -\frac{d}{dr} e^{-\frac{1}{2}r^2} dr \\ &= -\left[e^{-\frac{1}{2}r^2} \right]_{r=0}^{\infty} = 1 . \end{aligned} \quad (8.4)$$

Für den allgemeinen Fall verwenden wir den Transformationssatz für kontinuierliche Dichten angewandt auf die $N(0,1)$ -Dichte und die Transformation $\phi(x) = \sigma x + \mu \Leftrightarrow \phi^{-1}(x) = \left(\frac{x-\mu}{\sigma}\right)$.

8. Kontinuierliche Verteilungen – Beispiele

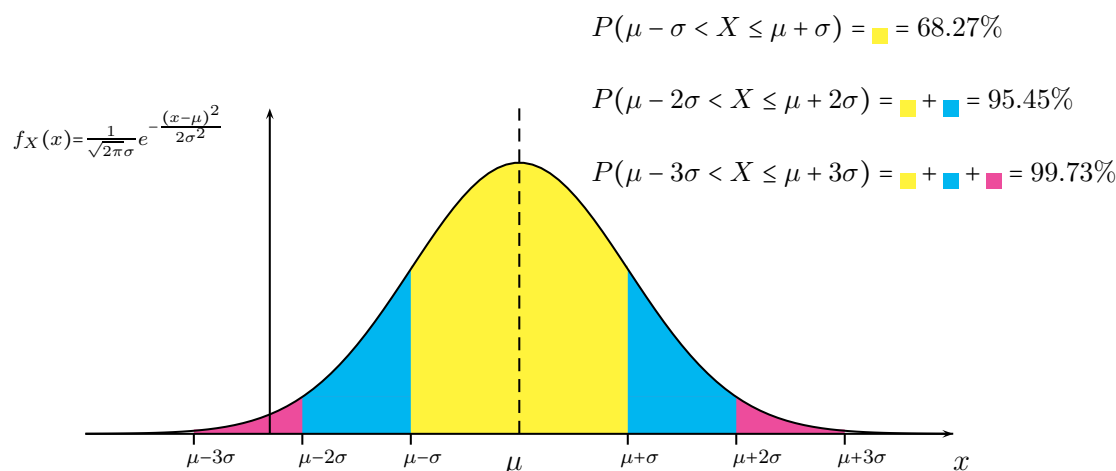


Abbildung 8.3: Einfache, zweifache und dreifache σ -Umgebung um μ bei der Normalverteilung $N(\mu, \sigma^2)$.

Somit ist $\phi'(x) = \sigma$ konstant und

$$f_{\phi(X)}(x) = \frac{f_X(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} . \quad \square \quad (8.5)$$

Der Parameter μ in der Normalverteilung entspricht dem Zentrum der Verteilung. Wie wir später sehen, kann der umgangssprachliche Begriff ‘Zentrum’ einerseits als 0.5-Quantil und andererseits als Erwartungswert verstanden werden. Der Parameter σ^2 steuert die Breite der Verteilung. σ^2 wird im nächsten Kapitel mit der Varianz einer normalverteilten Zufallsvariable und $\sigma = \sqrt{\sigma^2}$ mit ihrer Standardabweichung identifiziert. In einem Bereich von $1 \times \sigma$, $2 \times \sigma$ und $3 \times \sigma$ um den Wert μ liegen jeweils 68.27%, 95.45% bzw. 99.73% (gerundet) der gesamten Wahrscheinlichkeitsmasse, siehe Abbildung 8.3. Nach dem Transformationssatz gilt diese Aussage unabhängig von den Werten von μ und σ^2 .

Normalverteilungen werden in der Regel dann angewandt, wenn viele gleichartige kleine Zufallseffekte sich aufaddieren, siehe das Kapitel über den zentralen Grenzwertsatz für eine theoretische Begründung der zentralen Stellung der Normalverteilung. In vielen technischen Prozessen treten Normalverteilungen als Fehlerverteilungen auf.

Beispiel 8.6. Eine Maschine soll Stanzbleche mit der vorgeschriebenen Länge 10cm produzieren. Da es unmöglich ist, Metallteile herzustellen, die ganz genau 10cm lang sind, akzeptiert die abnehmende Firma auch Teile, die um ± 0.1 cm von 10cm abweichen (Toleranzband). Die Länge der Teile, die die Maschine tatsächlich herstellt, sei als Normalverteilung mit Parametern $\mu = 10.05$ cm und $\sigma^2 = (0.05\text{cm})^2$ gegeben. Wie groß ist die Ausschußrate in diesem Herstellungsprozess?

Lösung. Es sei $X \sim N(10.05, 0.05^2)$ die zufällige Länge eines von der Maschine produzierten Teils. Für ein akzeptables Teil gilt $9.9 < X \leq 10.1$ und darum ist die Ausbringungsrate des

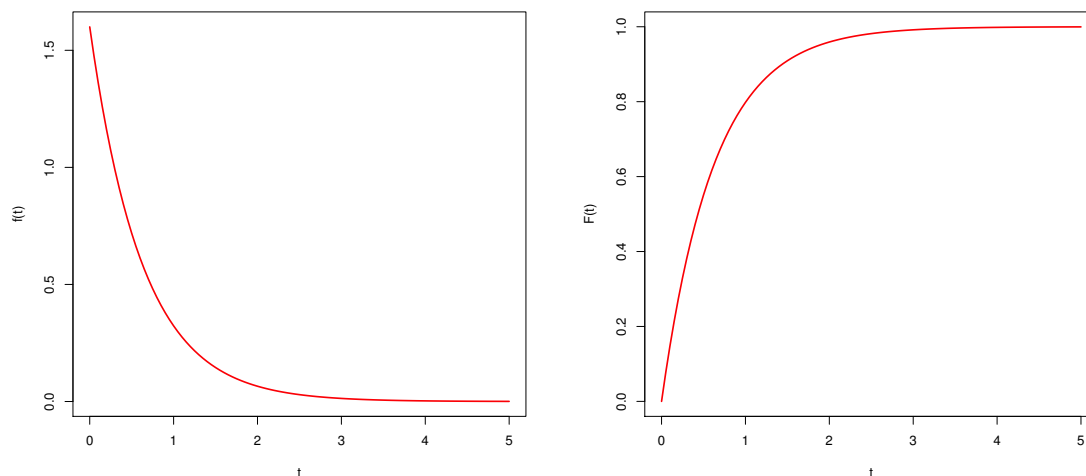


Abbildung 8.4: Kontinuierliche Dichte und Verteilungsfunktion der Exponentialverteilung zum Parameter $\alpha = 1.6$.

Prozesses

$$\begin{aligned}
 P(9.9 < X \leq 10.1) &= P(-0.15 < X - 10.05 \leq 0.05) \\
 &= P\left(\frac{-0.15}{0.05} < \frac{X - \mu}{\sigma} \leq \frac{0.05}{0.05}\right) \\
 &= P(-3 < Z \leq 1) \\
 &= F_{N(0,1)}(1) - F_{N(0,1)}(-3) \\
 &\approx 0.84133 - 0.00135 = 0.84,
 \end{aligned} \tag{8.6}$$

wobei gemäß dem Transformationssatz $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$. Also beträgt die Ausschußrate 16%. Die Werte von $F_{N(0,1)}$ sind in Anhang A tabellarisiert oder auch in Programmpaketen wie R implementiert.

Unternimmt die Firma nicht sehr schnell etwas, dürfte sie bald weg vom Fenster sein. In der Zulieferer-Metallindustrie werden Fehlerraten von 1 zu 1Mio angestrebt, um konkurrenzfähig zu sein (sogenannte Null Fehler bzw. six-sigma Programme). \diamond

Nun wenden wir uns Verteilungen zu, die typischer Weise als Verteilung für Wartezeiten bis zum Eintritt eines Ereignisses genutzt werden.

Definition 8.7. Die **Exponentialverteilung** $\text{Exp}(\alpha)$ zum Parameter $\alpha > 0$ ist das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, welches definiert ist durch die Verteilungsfunktion

$$F_X(x) = \begin{cases} 1 - e^{-\alpha x} & , \text{ für } x \geq 0, \\ 0 & , \text{ sonst,} \end{cases} \tag{8.7a}$$

8. Kontinuierliche Verteilungen – Beispiele

bzw. die kontinuierliche Dichte

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & , \text{ für } x \geq 0 , \\ 0 & , \text{ sonst .} \end{cases} \quad (8.7b)$$

Siehe auch Abbildung 8.4. Ist X exponentialverteilt zum Parameter α , so schreibe $X \sim \text{Exp}(\alpha)$.

Die Exponentialverteilung ist eine Wartezeitverteilung, die verwendet wird, wenn ein Ereignis zu jedem Zeitpunkt im Kontinuum eintreten kann und die Länge der bereits vergangenen Wartezeit keinen Rückschluss auf die noch abzuwartende Zeit zulässt. Diesen Sachverhalt können wir mit Hilfe der Risikofunktion mathematisch erfassen.

Definition 8.8. Es sei X eine Zufallsvariable mit kontinuierlicher Dichte f_X .

(i) Die **Überlebensfunktion** (englisch: *survival function*) zu X ist definiert als

$$S_X(x) := P(X > x) = 1 - F_X(x) . \quad (8.8)$$

(ii) Die **Risikofunktion** (englisch: *hazard function*) zu X ist definiert als

$$h_X(x) := \begin{cases} \lim_{\Delta x \searrow 0} \frac{1}{\Delta x} P(X \in (x, x + \Delta x] | X > x) & , \text{ für } x \text{ mit } S_X(x) > 0 , \\ 0 & , \text{ sonst .} \end{cases} \quad (8.9)$$

Die Risikofunktion versteht man am Besten, wenn man zunächst vom Limes $\Delta x \searrow 0$ und dem Vorfaktor $\frac{1}{\Delta x}$ absieht. Dann ist nämlich $P(X \in (x, x + \Delta x] | X > x)$ die Wahrscheinlichkeit für das Eintreten des Ereignisses im nächsten Zeitabschnitt der Länge Δx , vorausgesetzt es ist bis zur Zeit x noch nicht eingetreten. Für $\Delta x \searrow 0$ erhalten wir also die Eintrittswahrscheinlichkeit im allernächsten Moment. Da diese natürlich 0 ist, müssen wir durch $\frac{1}{\Delta x}$ teilen, um die ‘Eintrittsrates’ $h_X(x)$ zu erhalten (Einheit: $[\text{Zeit}]^{-1}$ falls X die Einheit $[\text{Zeit}]$ hat).

Beachten Sie auch, dass der deutsche Begriff ‘Risikofunktion’ inkonsistent mit der oft gebrauchten Definition von “Risiko = Wahrscheinlichkeit \times Auswirkung” ist. Nehmen wir z.B. an, für einen Schadensfall sei die Auswirkung aus Sicht einer Versicherung in Geld zu bemessen, so ist das Risiko durch den möglichen Eintritt eines Schadens ein Geldbetrag und keine Zeirate. Deswegen ist der englische Begriff *hazard function* weniger widersprüchlich und wird auch oft eingedeutscht.

Den Zusammenhang zwischen Dichte, Überlebensfunktion und Risikofunktion gibt die folgende

Proposition 8.9. *Gegeben seien dieselben Voraussetzungen wie in Definition 8.8. Dann gilt:*

(i) $h_X(x) = \frac{f_X(x)}{S_X(x)}$, für $x \in \mathbb{R}$ mit $S_X(x) > 0$.

(ii) $S_X(x) = \exp\{-\int_{-\infty}^x h_X(y) dy\}$, für $x \in \mathbb{R}$ mit $S_X(x) > 0$.

Beweis. (i) Es sei $x \in \mathbb{R}$ mit $S_X(x) > 0$. Dann gilt

$$\begin{aligned} \lim_{\Delta x \searrow 0} \frac{1}{\Delta x} P(X \in (x, x + \Delta x] | X > x) &= \lim_{\Delta x \searrow 0} \frac{1}{\Delta x} \frac{P(\{X \in (x, x + \Delta x]\} \cap \{X > x\})}{P(X > x)} \\ &= \lim_{\Delta x \searrow 0} \frac{1}{\Delta x} \frac{P(X \in (x, x + \Delta x])}{S_X(x)} \\ &= \lim_{\Delta x \searrow 0} \frac{1}{\Delta x} \frac{F_X(x + \Delta x) - F_X(x)}{S_X(x)} \\ &= \frac{f_X(x)}{S_X(x)}. \end{aligned} \quad (8.10)$$

(ii) Wir setzen (i) ein in die rechte Seite von (ii) und erhalten wegen $S'_X(x) = -f_X(x)$ und somit $-\frac{f_X(x)}{S_X(x)} = [\log S_X(x)]'$,

$$\begin{aligned} \exp\left\{-\int_{-\infty}^x h_X(y) dy\right\} &= \exp\left\{-\int_{-\infty}^x \frac{f_X(y)}{S_X(y)} dy\right\} \\ &= \exp\left\{\int_{-\infty}^x [\log S_X(y)]' dy\right\} \\ &= \exp\{[\log S_X(y)]_{-\infty}^x\} \\ &= \exp\{\log S_X(x) - \log(1)\} = S_X(x). \quad \square \end{aligned} \quad (8.11)$$

Lemma 8.10. Die Risikofunktion einer Zufallsvariable $X \sim \text{Exp}(\alpha)$ ist

$$h_X(x) = \begin{cases} \alpha & , \text{ für } x \geq 0, \\ 0 & , \text{ sonst.} \end{cases} \quad (8.12)$$

Beweis. Für $x < 0$ ist per Definition $f_X(x) = 0$ und $S_X(x) = 1$, siehe (8.7). Also gilt mit Proposition 8.9 $h_X(x) = 0$. Für $x \geq 0$ erhält man wieder mittels (8.7) und Proposition 8.9

$$h_X(x) = \frac{\alpha e^{-\alpha x}}{e^{-\alpha x}} = \alpha. \quad \square \quad (8.13)$$

Man ist häufig auch an zufälligen Eintrittszeiten interessiert, deren Risikofunktion nicht gleichmäßig verläuft, wie es bei der Exponentialverteilung der Fall ist. Hier bietet die Familie der Weibullverteilungen eine flexible Möglichkeit der Modellierung.

Definition 8.11. Die **Weibullverteilung** $\text{Wei}(\eta, \beta)$ zu den Parametern $\eta, \beta > 0$ ist das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, welches definiert ist durch die Verteilungsfunktion

$$F_X(x) = \begin{cases} 1 - \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\} & , \text{ für } x \geq 0, \\ 0 & , \text{ sonst.} \end{cases} \quad (8.14)$$

Für eine weibullverteilte Zufallsvariable X schreibe $X \sim \text{Wei}(\eta, \beta)$.

8. Kontinuierliche Verteilungen – Beispiele

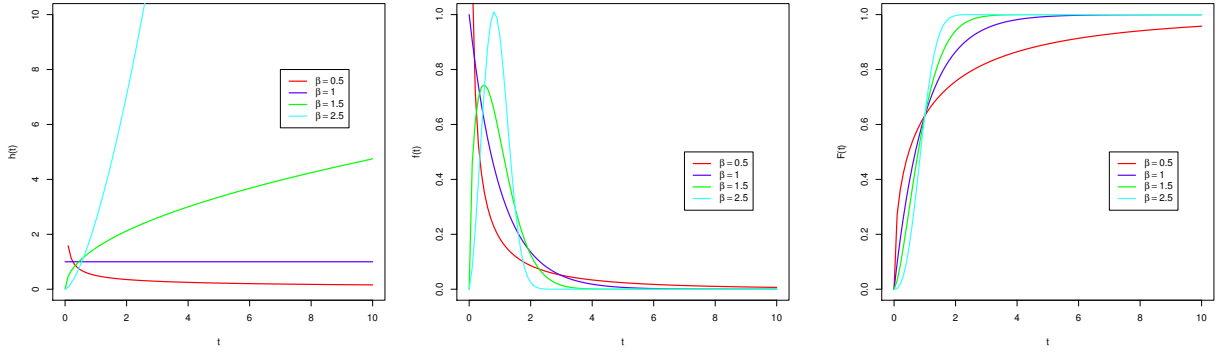


Abbildung 8.5: Risikofunktion, kontinuierliche Dichte und Verteilungsfunktion der Weibull-Verteilung zu $\eta = 1$ und verschiedenen β -Werten.

Um die Rolle der Parameter η, β besser verstehen zu können, berechnen wir zunächst einige abgeleitete Größen der Weibullverteilung.

Lemma 8.12. Für $X \sim \text{Wei}(\eta, \beta)$ gilt:

- (i) $S_X(x) = \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\}$, für $x \geq 0$, und sonst $S_X(x) = 1$.
- (ii) $f_X(x) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\}$, für $x \geq 0$, und sonst $f_X(x) = 0$.
- (iii) $h_X(x) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1}$, für $x \geq 0$, und sonst $h_X(x) = 0$.

Beweis. (i) Dieser Punkt folgt aus $S_X(x) = 1 - F_X(x)$.

(ii) Ableiten von $F_X(x)$ führt für $x < 0$ zu $f_X(x) = 0$ und für $x \geq 0$ zu

$$\begin{aligned}
 f_X(x) &= \left[1 - \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\}\right]' \\
 &= -\left[-\left(\frac{x}{\eta}\right)^\beta\right]' \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\} \\
 &= \frac{\beta x^{\beta-1}}{\eta^\beta} \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\} \\
 &= \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\}.
 \end{aligned} \tag{8.15}$$

(iii) Dieser Punkt folgt aus (i), (ii) und $h_X(x) = \frac{f_X(x)}{S_X(x)}$. □

Wir können nun die Parameter β und η interpretieren: Für $\beta < 1$ sinkt das Risiko mit zunehmendem x und für $\beta > 1$ steigt es. Für $\beta = 1$ ist die Weibullverteilung identisch zur

Exponentialverteilung mit $\alpha = \frac{1}{\eta}$. Abbildung 8.5 zeigt einige Verläufe der Risikofunktion, kontinuierlichen Dichte und Verteilungsfunktion für verschiedene β -Werte. Da die Verläufe der Funktionen qualitativ sehr unterschiedlich sind, heißt β auch **Formparameter**. Verschiedene η -Werte führen nur zu einer Skalenveränderung der x -Achse (Verteilungsfunktion) oder x und f -bzw. h -Achse. Deswegen heißt η auch **Skalenparameter**.

Beispiel 8.13. Der Betrieb eines Kraftwerks bringt pro Stunde 1 000 Euro Gewinn. Ein Versagen des Kraftwerks im laufenden Betrieb verursacht Mehrkosten von 10 Mio Euro gegenüber einem planmäßigen Halt zur Wartung. Aus statistischen Untersuchungen sei bekannt, dass die Zeit X bis zum nächsten Versagen die Verteilung $X \sim \text{Wei}(8760[h], 1.5)$ aufweist. Wie lange soll man das Kraftwerk laufen lassen? Hierbei ist eine risikoneutrale Herangehensweise anzuwenden, d.h. das Kraftwerk soll gestoppt werden, wenn der Gewinn in einer Zeiteinheit die Risikofunktion \times den Schaden übersteigt. Von sekundären Effekten wie höhere Stillstandszeiten bei vermehrter Wartung, Imageschaden beim Kraftwerksausfall etc. sei in erster Näherung abzusehen.

Lösung. Für einen Zeitpunkt x nach Anfahren des Kraftwerks ist die Risikofunktion $h_X(x) = \frac{1.5}{8760[h]} \left(\frac{x}{8760[h]} \right)^{1.5-1}$. Dies muss mit der Schadenssumme (1 Mio € = 10^7 €) multipliziert und mit der Gewinnrate $\frac{1000\text{€}}{[h]}$ gleichgesetzt werden, um den Punkt anzupassen, an dem die Risikorate die Gewinnrate übersteigt:

$$\frac{10^3\text{€}}{[h]} = \frac{1.5}{8760[h]} \left(\frac{x}{8760[h]} \right)^{0.5} \times 10^7\text{€} . \quad (8.16)$$

Auflösen nach x liefert

$$x = 8760[h] \times \left(\frac{8760}{1.5 \times 10000} \right)^2 \approx 2988[h] . \quad (8.17)$$

Ein Servicestopp sollte also nach ca. 4 Monaten Non-Stop Betrieb durchgeführt werden. \diamond

Zuletzt soll noch eine weitere Verteilungsfamilie eingeführt werden.

Definition 8.14. Die **Gammaverteilung** $\Gamma(\alpha, \beta)$ zu den Parametern $\alpha, \beta > 0$ ist das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, welches definiert ist durch die kontinuierliche Dichte

$$f_X(x) = \begin{cases} \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x} & , \text{ für } x \geq 0 , \\ 0 & , \text{ sonst } . \end{cases} \quad (8.18)$$

Für eine gammaverteilte Zufallsvariable schreibe $X \sim \Gamma(\alpha, \beta)$.

Dass die in (8.18) angegebene Funktion tatsächlich eine kontinuierliche Dichte darstellt ergibt sich für $\alpha = 1$ aus der Definition der Funktion $\Gamma(\beta) := \int_0^\infty x^{\beta-1} e^{-x} dx$. Für $\alpha > 0$ beliebig wende nun den Transformationssatz mit $\phi(x) = x/\alpha$ auf $\Gamma(1, \beta)$ an. Wiederum kann man $1/\alpha$ als Skalen- und β als Formparameter ansehen, siehe Abbildung 8.6.

8. Kontinuierliche Verteilungen – Beispiele

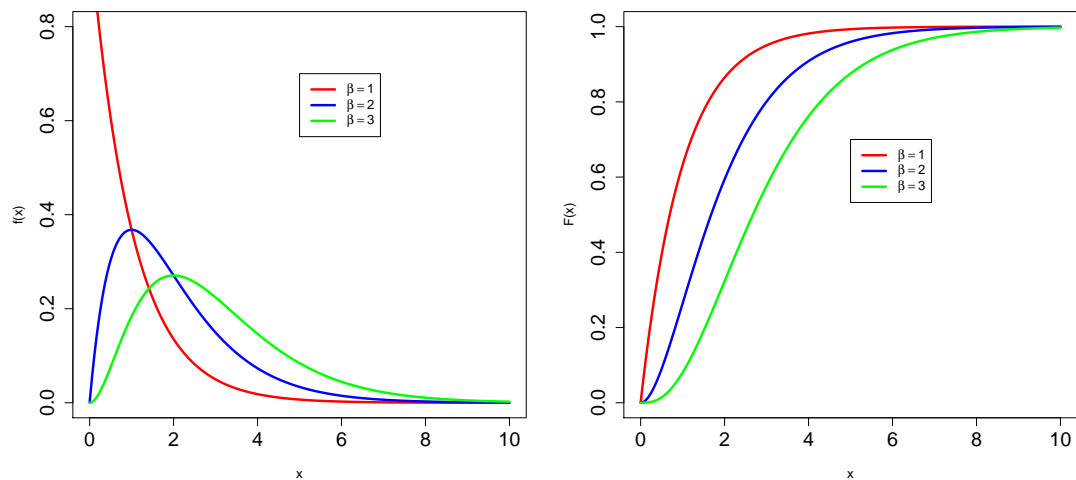


Abbildung 8.6: Kontinuierliche Dichte und Verteilungsfunktion der Gammaverteilung zu $\alpha = 1$ und verschiedenen β -Werten.

Die Gammaverteilung wird in der z.B. in der Versicherungsmathematik zur Modellierung von Schadenshöhen eingesetzt. Spezialfälle der Gammaverteilung spielen in der Statistik eine große Rolle, namentlich die χ^2 -Verteilungen. Für $\beta = 1$ fällt die Gammaverteilung mit der Exponentialverteilung zusammen.

8.1. Übungsaufgaben

1. Selbsttest

1. Betrachten Sie die erste der Tabellen in Anhang A. Für viele der folgenden Aufgaben ist es relevant, dass Sie diese Tabelle gut verstehen und anwenden können. Deshalb möchten wir dies hier ausgiebig üben.
 - a) Erklären Sie kurz welche Informationen in der Tabelle enthalten sind! Kann man aus dieser Tabelle auch die Quantile ablesen? Wenn ja, wie?
 - b) Bestimmen Sie mittels der Tabelle die Werte der Verteilungsfunktion $F_Z(z)$ einer Zufallsvariable $Z \sim N(0, 1)$ an den Stellen $z = 1.37$, $z = -2.91$ und $z = 3.12$. Erklären Sie insbesondere wie Sie $F_Z(-2.91)$ bestimmen!
 - c) Sie haben nun eine Zufallsvariable $X \sim N(\mu, \sigma^2)$ und möchten die Verteilungsfunktion $F_X(x)$ an einer Stelle $x \in \mathbb{R}$ bestimmen. Können Sie hierzu auch die Tabelle verwenden? Geben Sie die Relation zwischen $F_X(x)$ und $F_Z(z)$ für geeignetes z an! (*Tipp*: Machen Sie eine geschickte Transformation!)
 - d) Bestimmen Sie nun die Quantile $Q_Z(p)$ von $Z \sim N(0, 1)$ an den Stellen $p = 0.28$, $p = 0.41$ und $p = 0.72$.
 - e) Geben Sie einen Zusammenhang zwischen den Quantilen $Q_X(p)$ der Zufallsvariablen $X \sim N(\mu, \sigma^2)$ und den Quantilen $Q_Z(p)$ von $Z \sim N(0, 1)$ an! (*Tipp*: Machen Sie wieder eine geschickte Transformation!)
2. Es sei $X \sim N(0, 1)$ eine standardnormalverteilte Zufallsvariable. Berechnen Sie die kontinuierliche Dichte der Zufallsvariablen $Y = X^2$. Die Verteilung von Y nennt man χ^2 -Verteilung mit einem Freiheitsgrad. .
3. Sie sind auf einer Hochzeit und die Braut wirft rein zufällig den Brautstrauß hinter sich. Die Zufallsvariable $X \sim U(0, 180)$ beschreibt die Wurfriechung (in Grad) in welcher der Strauß abgefeuert wird. Sie selbst können einen Winkelbereich von 30 Grad abdecken in welchem Sie den Strauss fangen können. Mit welcher Wahrscheinlichkeit werden Sie heiraten (d.h. den Strauss fangen)?
4. Auf der Landstraße fahren Sie im Mittel 100 km/h mit einer Standardabweichung von 20 km/h. Die Zufallsvariable $X \sim N(100, 20^2)$ beschreibt Ihre Geschwindigkeit. Nun wurden Sie geblitzt, d.h. Sie waren schneller als 100 km/h. Mit welcher Wahrscheinlichkeit wird Ihr Führerschein eingezogen, in anderen Worten mit welcher Wahrscheinlichkeit waren Sie schneller als 140 km/h? Beachten Sie, dass es sich hier um eine bedingte Wahrscheinlichkeit handelt!
5. X sei weibullverteilt zu den Parametern $\beta_X = 2$ und $\eta_X = 2$, d.h. $X \sim \text{Wei}(2, 2)$. Es sei $Y = \sqrt{X}$. Zeigen Sie, dass Y ebenfalls weibullverteilt ist und bestimmen Sie die Parameter β_Y und η_Y von Y !

2. Transfer

1. Eine Bank hält Aktien eines DAX-Unternehmens und möchte sie in einem Jahr verkaufen. Die Risikoabteilung der Bank modelliert die Rendite X dieser Papiere für diesen Zeitraum als normalverteilt $X \sim N(\mu, \sigma^2)$ mit μ , der erwarteten Rendite, gleich 5% und σ , der Standardabweichung, bzw. Volatilität auf ein Jahr, gleich 3%. (Alle Zinssätze seien *effektiv*, d.h. schon korrekt auf das Jahr bezogen.)
 - a) Skizzieren Sie die kontinuierliche Dichte der Renditeverteilung X und zeichnen Sie die einfache, zweifache und dreifache σ -Umgebung um die erwartete Rendite ein.
 - b) Wie groß ist die Wahrscheinlichkeit, dass die Rendite für das Papier zwischen 2% und 8% liegt?
 - c) Wie groß ist die Wahrscheinlichkeit, dass die Rendite aus dem Aktienpapier den Referenzzinssatz EURIBOR übertrifft? Der EURIBOR betrage 2% pro Jahr.
 - d) Wie groß ist die Wahrscheinlichkeit, dass die Bank mit dem Papier Verlust macht, also die Rendite X negativ wird?
 - e) Im Bonusprogramm der Bank wird ein/e Händler/in für ihr/sein Investment dann belohnt, wenn die tatsächliche Jahresrendite x ihrer/seiner Investition im oberen 10%-Band der Vorhersage der Risikoabteilung liegt, d.h. $0.1 = P(X > x)$. Welchen Renditewert x muss die Investition nach einem Jahr übertreffen, damit die/der Händler/in einen Bonus bekommt?

Bei der Beantwortung der Fragen setzen Sie bitte voraus, dass die Risikoabteilung ihre Aufgabe korrekt gelöst und die Renditenverteilung zutreffen spezifiziert hat. Verwenden Sie außerdem die Werte der Verteilungsfunktion der Standardnormalverteilung sowie deren Quantile den Tabellen aus Anhang A.

2. Eine Aufgabe der Chemie ist es die Umwandlung von Molekülen in andere zu untersuchen. Die einfachste dieser Reaktionen ist die spontane Umwandlung eines Moleküls A in die Moleküle B, C, \dots beschrieben durch die Reaktionsgleichung $A \rightarrow B + C + \dots$. Diese Reaktionen erfolgen mit einer Rate X und die Konzentration n_A von A folgt der Differentialgleichung $\frac{d}{dt}n_A(t) = -X n_A(t)$, wobei t die Zeit ist.
 - a) Lösen Sie mittels einem geschickt gewählten Ansatz die Differentialgleichung mit Anfangsbedingung $n_A(0) = 1$. (*Tipp*: Welche Funktion die Sie kennen reproduziert sich selbst unter Ableitung?)
 - b) Nun betrachten wir die Konzentration von A zur Zeit $t = 1$, d.h. $Y := n_A(1)$. Wie sie sehen hängt diese von der Rate X ab. Da die Rate X von Umwelt- und Temperaturschwankungen beeinflusst wird, beschreiben wir diese durch eine Zufallsvariable mit Verteilung $X \sim N(10, 1)$. Bestimmen Sie die kontinuierliche Dichte der Zufallsvariablen $Y = n_A(1)$. (*Hinweis*: Diese Verteilung nennt man log-Normalverteilung.)

3. Finden Sie heraus, welche der in diesem Kapitel vorgestellten Verteilungsfamilien auf die folgenden Probleme passen.
- a) Die Größe einer Frau/eines Mannes hängt von vielen kleinen Zufallseffekten ab, die sich wie wir hier annehmen aufaddieren. Die Durchschnittsgröße einer deutschen Frau ist ungefähr 165 cm und die Varianz ist ca. $(5 \text{ cm})^2$. Welcher (approximativen) Verteilung folgt die Zufallsvariable X_1 = “Größe einer deutschen Frau”?
 - b) Sie haben am Wochenende einen netten Mann/eine nette Frau getroffen und haben ihm/ihr Ihre Handynummer gegeben. Er/sie hat gesagt, dass er/sie Sie irgendwann mal anruft. Welcher Verteilung folgt die Zufallsvariable X_2 = “Wartezeit bis Anruf”?
 - c) Im wahren Leben ist es aber so, dass man mit höherer Wahrscheinlichkeit kurz nach dem ersten Treffen angerufen wird, als viele Jahre später. Welche Verteilungsfamilie eignet sich somit besser für Aufgabe b)?

8.2. Literatur

9. Erwartungswert, Varianz und Momente

Wenn eine Größe eine statistische Streuung aufweist, dann ist es natürlich danach zu fragen, welchen Wert diese Größe ‘im Mittel’ annimmt. Weiterhin möchte man oft die Streuung um dieses ‘Mittel’ quantifizieren. Eine Lösung für diese Aufgabe wäre die Angabe des **Medians**, also des 50%-Quantils $Q(0.5)$, als ‘Mittel’ und des **Interquantilabstandes** $Q(0.75) - Q(0.25)$ als Maß für die Streuung. Auch wenn daran nichts auszusetzen ist, ist es üblicher, den Erwartungswert als ‘Mittel’ und die Standardabweichung als Streumaß anzugeben. Diese Begriffe werden in diesem Kapitel eingeführt.

Die Idee hinter dem Erwartungswert ist, alle auftretenden Werte gemäß ihres Wahrscheinlichkeitsgewichtes zu berücksichtigen. Besonders einfach lässt sich dieses Konzept für Funktionen realisieren, die nur endlich viele verschiedene Werte annehmen. In dem entgegengesetzten Fall, dass ein ganzes Kontinuum an Werten auftritt und die Wahrscheinlichkeit für jeden einzelnen Wert Null beträgt, muss das Konzept des Erwartungswertes jedoch von der gewichteten Summe hin zu einem Integral bezüglich eines Wahrscheinlichkeitsmaßes erweitert werden. Um eine allgemeine Theorie zu erhalten, benötigen wir also ein paar Resultate aus der **Integrationstheorie**.

Definition 9.1. Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Zufallsvariable.

- (i) X heißt **elementar**, wenn $\text{Bild}(X) = \{X(\omega) : \omega \in \Omega\}$ eine endliche Menge ist.
- (ii) Ist X elementar mit $\text{Bild}(X) = \{\alpha_1, \dots, \alpha_n\}$ und $A_j = X^{-1}(\{\alpha_j\}) = \{\omega \in \Omega : X(\omega) = \alpha_j\}$, $j = 1, \dots, n$, dann definiere den **Erwartungswert** von $X = \sum_{j=1}^n \alpha_j 1_{A_j}$ bzw. gleichbedeutend das **Lebesgue-Integral** von X über das Wahrscheinlichkeitsmaß P als

$$\mathbb{E}[X] := \int_{\Omega} X dP := \int_{\Omega} X(\omega) dP(\omega) := \sum_{j=1}^n \alpha_j P(A_j) . \quad (9.1)$$

Hierbei sind die ersten drei Ausdrücke häufig verwendete unterschiedliche Symbole für dieselbe Sache.

In Definition 9.1 haben wir 1_A für die **charakteristische Funktion** der Menge $A \in \mathcal{A}$ geschrieben. Diese ist definiert durch

$$1_A(\omega) = \begin{cases} 1 & , \text{ für } \omega \in A , \\ 0 & , \text{ für } \omega \in A^c . \end{cases} \quad (9.2)$$

9. Erwartungswert, Varianz und Momente

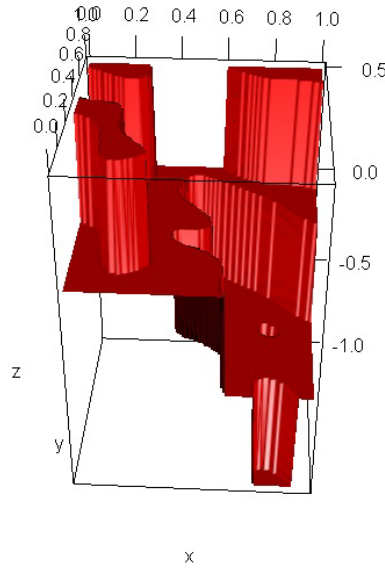


Abbildung 9.1: Ein Beispiel für eine elementare Funktion über \mathbb{R}^2 .

Offensichtlich sind charakteristische Funktionen elementar und Sie sind für $A \in \mathcal{A}$ auch messbar, denn $\forall B \in \mathcal{B}(\mathbb{R})$

$$X^{-1}(B) = \begin{cases} \Omega & , \text{ für } 1, 0 \in B , \\ A & , \text{ für } 1 \in B , 0 \notin B , \\ A^c & , \text{ für } 1 \notin B , 0 \in B , \\ \emptyset & , \text{ für } 0, 1 \notin B . \end{cases} \quad (9.3)$$

Die Abbildung 9.1 zeigt ein typisches Beispiel für eine elementare Funktion.

Definition 9.2. Eine nicht-negative Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $X(\omega) \geq 0 \forall \omega \in \Omega$, heißt **obere Einhüllende** einer monotonen Folge von nicht-negativen elementaren Zufallsvariablen $X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, falls $\sup_{n \in \mathbb{N}} X_n(\omega) \stackrel{(*)}{=} X(\omega)$ für P -fast alle ω gilt, d.h. $\exists \mathcal{N} \in \mathcal{A}$ mit $P(\mathcal{N}) = 0$ so dass $(*)$ gilt $\forall \omega \in \Omega \setminus \mathcal{N}$.

Offensichtlich ist jede nicht-negative Zufallsvariable eine obere Einhüllende im Sinne von Definition 9.2, denn

$$X_n(\omega) = \sum_{j=1}^{2^{2n}} \frac{j}{2^n} 1_{\{\frac{j}{2^n} < X \leq \frac{j+1}{2^n}\}}(\omega) \quad (9.4)$$

definiert eine monotone Folge von nicht-negativen elementaren Zufallsvariablen, so dass $\sup_{n \in \mathbb{N}} X_n(\omega) = X(\omega)$.

Definition 9.3. Gegeben eine nicht-negative Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, definiere das **Lebesgue-Integral von X über P** als **Erwartungswert**

$$\mathbb{E}[X] := \int_{\Omega} X dP := \int_{\Omega} X(\omega) dP(\omega) := \sup_{n \in \mathbb{N}} \int_{\Omega} X_n dP, \quad (9.5)$$

wobei $X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine monotone Folge von nicht-negativen elementaren Zufallsvariablen ist, so dass $\sup_{n \in \mathbb{N}} X_n(\omega) = X(\omega)$ P -fast überall.

Diese Definition ist sinnvoll aufgrund des folgenden

Lemma 9.4. *Der Wert des Lebesgue-Integrals in (9.5) hängt nicht von der gewählten monotonen Folge von elementaren Zufallsvariablen X_n ab.*

Bemerkung 9.5. Bei dem folgenden Beweis spielt die σ -Additivität von P die entscheidende Rolle. Dies zeigt, dass ein Axiomensystem für die Wahrscheinlichkeitstheorie, welches, auch Erwartungswerte mit einschließt, nicht ohne diese technische Eigenschaft auskommt. Dass Erwartungswerte auf Basis des Axiomensystems der Wahrscheinlichkeitstheorie konstruiert werden können, war von D. Hilbert in seiner Formulierung des 23. Jahrhundertproblems gefordert worden.

Beweis von Lemma 9.4. Zuerst weisen wir darauf hin, dass für X und Y elementare Zufallsvariablen mit $X \geq Y$ gilt $\int_{\Omega} X dP \geq \int_{\Omega} Y dP$ (siehe Übungsaufgabe).

Im nächsten Schritt wird gezeigt, dass für eine elementare Zufallsvariable $Y = \sum_{j=1}^m \beta_j 1_{B_j}$ und X_n eine monotone Folge von elementaren Zufallsvariablen mit $X(\omega) = \sup_{n \in \mathbb{N}} X_n(\omega) \geq Y(\omega)$ P -fast sicher mit \mathcal{N} der Ausnahmemenge vom P -Maß Null, gilt $\sup_{n \in \mathbb{N}} \int_{\Omega} X_n dP \geq \int_{\Omega} Y dP$. Hierfür definieren wir für $1 > a > 0$ eine Menge $A_n := \{X_n \geq aY\}$. Es gilt wegen der Monotonie der X_n , dass $A_n \cup \mathcal{N} \uparrow \Omega$ und mittels der Unterhalbstetigkeit von P (welche äquivalent zur σ -Additivität ist) folgt

$$\begin{aligned} \sup_{n \in \mathbb{N}} \int_{\Omega} X_n dP &\geq \sup_{n \in \mathbb{N}} \int_{\Omega} aY 1_{A_n} dP \\ &= a \sup_{n \in \mathbb{N}} \sum_{j=1}^m \beta_j P(B_j \cap A_n) \\ &= a \sup_{n \in \mathbb{N}} \sum_{j=1}^m \beta_j P(B_j \cap (A_n \cup \mathcal{N})) \\ &= a \sum_{j=1}^m \beta_j P(B_j) = a \int_{\Omega} Y dP. \end{aligned} \quad (9.6)$$

Da $1 > a > 0$ beliebig gewählt war, gilt $\sup_{n \in \mathbb{N}} \int_{\Omega} X_n dP \geq \int_{\Omega} Y dP$. Es seien nun X_n und X'_n zwei monotone Folgen von elementaren Zufallsvariablen mit gemeinsamer oberer Einhüllenden X . Da wie eben gesehen $\sup_{n \in \mathbb{N}} \int_{\Omega} X_n dP \geq \int_{\Omega} X'_{n'} dP$ und auch $\sup_{n \in \mathbb{N}} \int_{\Omega} X'_n dP \geq \int_{\Omega} X_{n'} dP \quad \forall n' \in \mathbb{N}$, gilt schließlich

$$\sup_{n \in \mathbb{N}} \int_{\Omega} X_n dP = \sup_{n \in \mathbb{N}} \int_{\Omega} X'_n dP. \quad \square \quad (9.7)$$

9. Erwartungswert, Varianz und Momente

Definition 9.6. Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Zufallsvariable.

- (i) Ist X nicht-negativ, dann heißt X **Lebesgue integrierbar** bezüglich P , falls

$$\mathbb{E}[X] = \int_{\Omega} X dP < \infty . \quad (9.8)$$

- (ii) Für X nicht notwendiger Weise nicht-negativ, definiere den **Positiv-** bzw. **Negativ-**anteil von X als

$$X^+ := \max(X, 0) \geq 0 , \quad X^- := -\min(X, 0) \geq 0 . \quad (9.9)$$

Ist nun der Positivanteil und der Negativanteil von X Lebesgue integrierbar, so heißt X Lebesgue integrierbar. In diesem Fall definieren wir den **Erwartungswert** von X als

$$\mathbb{E}[X] := \int_{\Omega} X dP := \int_{\Omega} X(\omega) dP(\omega) := \int_{\Omega} X^+ dP - \int_{\Omega} X^- dP . \quad (9.10)$$

- (iii) Es sei $\phi_p(x) = x^p$, $p \geq 0$. $\phi_p : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist als stetige Funktion messbar. Dann heißt X im p -ten Mittel Lebesgue integrierbar, falls $X^p := \phi_p \circ X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ Lebesgue integrierbar ist. Definiere für ein solches X das **p -te Moment** von X als

$$\mathbb{M}_p[X] := \mathbb{E}[X^p] . \quad (9.11)$$

- (iv) Ist $p \geq 1$ und X im p -ten Mittel Lebesgue integrierbar, so schreiben wir auch $X \in L^p((\Omega, \mathcal{A}, P); \mathbb{R})$. Die **L^p -Norm** von X ist dann definiert als

$$\|X\|_p := (\mathbb{E}[|X|^p])^{\frac{1}{p}} = \left(\int_{\Omega} |X|^p dP \right)^{\frac{1}{p}} . \quad (9.12)$$

Bemerkung 9.7 (L^p -Räume). Die Schreibweise in Definition 9.6 (iv) ist zwar allgemein üblich, jedoch nicht präzise. Zunächst einmal ist $\|\cdot\|_p$ offensichtlich im Allgemeinen keine Norm auf der Menge aller im p -ten Mittel integrierbarer Zufallsvariablen. Nehmen wir z.B. den Wahrscheinlichkeitsraum $([0, 1], \mathcal{B}(\mathbb{R}) \cap [0, 1], U(0, 1))$ und die Zufallsvariable $X(\omega) = 1$ für $\omega = \frac{1}{2}$ und $X(\omega) = 0$ sonst, so ist offensichtlich $X \neq 0$, jedoch $\|X\|_p = 0 \quad \forall p \in [1, \infty)$.

Die präzise Definition des Raumes $L^p((\Omega, \mathcal{A}, P); \mathbb{R})$ ist der Raum aller im p -ten Mittel integrierbaren Zufallsvariablen modulo der Äquivalenzrelation \sim_P : $X \sim_P Y$ falls $X = Y$ P -fast sicher, d.h.

$$\exists \mathcal{N} \in \mathcal{A}, P(\mathcal{N}) = 0, \text{ so dass } X(\omega) = Y(\omega) \quad \forall \omega \in \Omega \setminus \mathcal{N} . \quad (9.13)$$

Man sieht mit der Definition des Lebesgue-Integrals unmittelbar ein, dass $X \sim_P Y \Rightarrow \|X\|_p = \|Y\|_p$ und $\|X\|_p = 0 \Rightarrow X \sim_P 0$. Also definiert $\|\cdot\|_p$ eine Norm auf $L^p((\Omega, \mathcal{A}, P); \mathbb{R})$. Die

Dreiecksungleichung heißt hierbei Minkowski-Ungleichung und ist eine Folge der Hölderschen Ungleichung, vergleiche Bauer, MI, Satz 14.1 und 14.2. Man kann sogar zeigen, dass $L^p((\Omega, \mathcal{A}, P); \mathbb{R})$ mit der Norm $\| \cdot \|_p$ ein Banachraum ist, also ein vollständiger (d.h. jede Cauchyfolge konvergiert) normierter Raum.

Trotzdem ist die Schreibweise $X \in L^p((\Omega, \mathcal{A}, P); \mathbb{R})$ kein Versehen, sondern volle Absicht. Denn wer kann sich schon viel unter einer Äquivalenzklasse von Zufallsvariablen vorstellen? Spätestens wenn man in der Statistik über Realisierungen (‘Würfelergebnisse’) von kontinuierlichen Zufallsvariablen redet, kommt man hier ins Schleudern. Die Wahrscheinlichkeitstheorie hat hier völlig zu Recht die intuitiven Schreibweisen entgegen einem Übermaß an formaler Spitzfindigkeit beibehalten. Wichtig ist aber für die/den Mathematiker/in, die *genaue* Definition von L^p -Räumen und Normen zu kennen und im Zweifel korrekt anzuwenden, obwohl die Notation zum Teil *absichtlich ungenau* ist.

Was die Spitzfindigkeit angeht, so wäre es vom Gesichtspunkt der reinen Mathematik sinnvoll, die σ -Algebren \mathcal{A} als grundlegendes Objekt der Wahrscheinlichkeitstheorie anzusehen und diese σ -Algebra (genauer: σ -Ring) durch den σ -Ring der P -Nullmengen zu quotientieren. Zufallsvariablen X sind dann σ -Ring-Homomorphismen auf diesen Quotientenringen (vergleiche den Klassiker der Maßtheorie P. R. Halmos, Measure Theory, Springer 1974). Die Ereignismenge Ω wird in dieser Sprache obsolet. Wir wagen zu bezweifeln, dass jemand, die/der in dieser abstrakten Weise über Wahrscheinlichkeitstheorie nachdenkt, in der Lage wäre, signifikante Resultate zu dieser Disziplin beizutragen, geschweige denn, sie erfolgreich in der Praxis anzuwenden. Das soll natürlich nicht heißen, dass es schlecht wäre, sich diese strukturellen Dinge einmal klar zu machen.

Definition 9.8 (Lebesgue-Integral über eine messbare Menge). Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine Lebesgue integrierbare Zufallsvariable und $A \in \mathcal{A}$ eine messbare Menge.

(i) Dann ist das **Lebesgue-Integral von X über A** definiert als

$$\int_A X dP := \int_{\Omega} 1_A X dP . \quad (9.14)$$

(ii) Ist $P(A) > 0$, dann definiere die **bedingte Erwartung** von X gegeben A als

$$\mathbb{E}[X|A] := \frac{\mathbb{E}[X 1_A]}{\mathbb{E}[1_A]} = \frac{\int_A X dP}{P(A)} . \quad (9.15)$$

Satz 9.9 (Eigenschaften des Erwartungswertes). Es seien $X, Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $L^1((\Omega, \mathcal{A}, P); \mathbb{R})$ -Zufallsvariablen.

(i) Der Erwartungswert ist **linear**

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \quad \forall a, b \in \mathbb{R} . \quad (9.16)$$

(ii) Der Erwartungswert ist **isoton**

$$X \geq Y \quad \Rightarrow \quad \mathbb{E}[X] \geq \mathbb{E}[Y] . \quad (9.17)$$

9. Erwartungswert, Varianz und Momente

(iii) (Satz über dominierte Konvergenz bzw. Satz von Lebesgue) Falls $X_n \rightarrow X$ P -fast sicher und es existiert $Y \in L^1((\Omega, \mathcal{A}, P); \mathbb{R})$, so dass $Y \geq |X_n|$ P -fast sicher $\forall n \in \mathbb{N}$, dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X] . \quad (9.18)$$

(iv) (Transformationssatz für das Lebesgue-Integral) Für

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (\tilde{\Omega}, \tilde{\mathcal{A}}) \xrightarrow{\phi} (\mathbb{R}, \mathcal{B}(\mathbb{R})) \quad (9.19)$$

mit $Y := \phi \circ X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ Lebesgue integrierbar gilt

$$\mathbb{E}[Y] = \int_{\Omega} Y dP = \int_{\Omega} \phi(X(\omega)) dP(\omega) = \int_{\tilde{\Omega}} \phi(x) dX_*P(x) . \quad (9.20)$$

(v) Insbesondere gilt mit

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R})) \xrightarrow{\text{id}} (\mathbb{R}, \mathcal{B}(\mathbb{R})) , \quad (9.21)$$

dass $\mathbb{E}[X]$ nur von der Verteilung $\text{Vert}(X) = X_*P$ abhängt, d.h.

$$\mathbb{E}[X] = \int_{\Omega} X dP = \int_{\Omega} \text{id} \circ X dP = \int_{\mathbb{R}} x dX_*P(x) = \int_{\mathbb{R}} x d\text{Vert}(X)(x) . \quad (9.22)$$

(vi) Gegeben sei dieselbe Situation wie in (iv) mit $(\tilde{\Omega}, \tilde{\mathcal{A}}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Des Weiteren sei ϕ stetig, also insbesondere Riemann integrierbar bezüglich der Verteilungsfunktion F_X . Dann gilt

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} \phi(x) dF_X(x) = \lim_{\Delta \searrow 0} \sum_{j \in \{\Delta q : q \in \mathbb{Z}\}} \phi\left(j + \frac{\Delta}{2}\right) [F_X(j + \Delta) - F_X(j)] . \quad (9.23)$$

Hier ist die rechte Seite gerade die Definition des Riemann-Integrals.

Beweis. Ein voller Beweis ist Sache der Maß- und Integrationstheorie, siehe etwa (i) H. Bauer, MI, Gleichung (12.10), (ii) ibda Gleichung (12.9) (iii) ibda Satz 15.6 (iv) ibda Korollar 19.2. (v) ist direkte Folgerung aus (iv). Alle diese Sachverhalte beweist man zunächst für elementare Funktionen durch direkte Rechnung, geht dann über zu nicht-negativen bzw. nicht-positiven Funktionen mittels der Definition des Lebesgue-Integrals und schließt letztendlich auf integrierbare Funktionen.

(vi) Da wir $\phi(x) = \phi^+(x) - \phi^-(x)$ in Positiv- und Negativanteil zerlegen können und ϕ^{\pm} wiederum stetig sind, dürfen wir ohne Beschränkung der Allgemeinheit $\phi(x) \geq 0$ voraussetzen. Es sei $Y_n := 1_{\{-n < X \leq n\}} \phi(X)$, dann ist $\phi(X)$ eine L^1 -Dominante von Y_n und $Y_n \rightarrow Y$ P -fast überall. Also existiert für $\epsilon > 0$ ein $n_0(\epsilon)$, so dass

$$\begin{aligned} \int_{\Omega} |\phi(X) - Y_n| dP &= \int_{\Omega} (1 - 1_{\{-n < X \leq n\}}) |\phi(X)| dP \\ &= \int_{(-\infty, -n]} |\phi(x)| dX_*P(x) + \int_{(n, \infty)} |\phi(x)| dX_*P(x) \\ &< \frac{\epsilon}{5} \quad \forall n > n_0(\epsilon) . \end{aligned} \quad (9.24)$$

Aus der Definition des uneigentlichen Riemann-Integrals entnehmen wir weiter, dass für hinreichend große n

$$\left| \int_{-\infty}^{\infty} \phi(x) dF_X(x) - \int_{-n}^n \phi(x) dF_X(x) \right| \leq \frac{\epsilon}{5}. \quad (9.25)$$

Aufgrund der Stetigkeit von ϕ existiert des Weiteren ein $\Delta_0 > 0$ hinreichend klein, so dass $\phi(x) - \phi_j^\Delta < \frac{\epsilon}{5}$ mit $\phi_j^\Delta = \inf\{\phi(y) : y \in (j, j + \Delta]\}$ für $x \in (j, j + \Delta]$, $j \in \{\Delta q : q \in \mathbb{Z}\} \cap [-n, n]$, $0 < \Delta < \Delta_0$. Daher folgt

$$\begin{aligned} \left| \int_{\Omega} \phi(X) dP - \int_{-\infty}^{\infty} \phi(x) dF_X(x) \right| &\leq \left| \int_{\Omega} (\phi(X) - Y_n) dP \right| \\ &+ \left| \int_{-\infty}^{\infty} \phi(x) dF_X(x) - \int_{-n}^n \phi(x) dF_X(x) \right| \\ &+ \left| \int_{-n}^n \phi(x) dF_X(x) \right. \\ &\quad \left. - \sum_{j \in \{\Delta q : q \in \mathbb{Z}\} \cap (-n, n]} \phi\left(j + \frac{\Delta}{2}\right) [F_X(j + \Delta) - F_X(j)] \right| \\ &+ \left| \sum_{j \in \{\Delta q : q \in \mathbb{Z}\} \cap (-n, n]} \phi\left(j + \frac{\Delta}{2}\right) [F_X(j + \Delta) - F_X(j)] \right. \\ &\quad \left. - \sum_{j \in \{\Delta q : q \in \mathbb{Z}\} \cap (-n, n]} \phi_j^\Delta [F_X(j + \Delta) - F_X(j)] \right| \\ &+ \left| \sum_{j \in \{\Delta q : q \in \mathbb{Z}\} \cap (-n, n]} \phi_j^\Delta [F_X(j + \Delta) - F_X(j)] \right. \\ &\quad \left. - \int_{\mathbb{R}} 1_{(-n, n]}(x) \phi(x) dX_* P(x) \right|. \end{aligned} \quad (9.26)$$

Wir haben bereits gesehen, dass für n hinreichend groß und Δ hinreichend klein jeder der ersten 4 Terme kleiner als $\frac{\epsilon}{5}$ ist. Der Summenterm im 5. Ausdruck ist jedoch gerade das Integral einer monotonen Folge von Elementarfunktionen, die $1_{(-n, n]}(x)\phi(x)$ von unten approximieren. Folglich wird auch die Differenz im letzten Ausdruck für Δ hinreichend klein beschränkt durch $\frac{\epsilon}{5}$. Also ist die linke Seite kleiner als ϵ . Da dies für beliebige $\epsilon > 0$ gilt, folgt die Aussage. \square

Der folgende Satz ist von immenser praktischer Relevanz, da er es ermöglicht die Bestimmung von Erwartungswerten und Momenten auf einfachere Rechnungen mit der diskreten oder kontinuierlichen Dichte von Zufallsvariablen zurückzuführen.

9. Erwartungswert, Varianz und Momente

Satz 9.10. (i) Für $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ mit diskreter Dichte p_x^X gilt

$$\mathbb{E}[X] = \sum_{x \in \mathbb{Z}} x p_x^X, \quad (9.27a)$$

$$\mathbb{M}_p[X] = \sum_{x \in \mathbb{Z}} x^p p_x^X, \quad p \in \mathbb{N}. \quad (9.27b)$$

$$\mathbb{E}[\phi(X)] = \sum_{x \in \mathbb{Z}} \phi(x) p_x^X, \quad (9.27c)$$

$$f_{\mathbb{Z}}^{\mathbb{R}} \phi : (\mathbb{Z}, \mathbb{P}(\mathbb{Z})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

(ii) Für $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit kontinuierlicher Dichte $f_X(x)$ gilt

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx, \quad (9.28a)$$

$$\mathbb{M}_p[X] = \int_{\mathbb{R}} x^p f_X(x) dx. \quad (9.28b)$$

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x) f_X(x) dx \quad (9.28c)$$

$$f_{\mathbb{R}}^{\mathbb{R}} \phi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

Beweis. (i) Nach dem Transformationssatz gilt

$$\mathbb{E}[X] = \int_{\mathbb{Z}} x dX_* P(x), \quad \mathbb{M}_p[X] = \int_{\mathbb{Z}} x^p dX_* P(x). \quad (9.29)$$

Den Integranden $\text{id}(x) = x$ bzw. $\text{id}(x)^p = x^p$ können wir durch elementare Zufallsvariablen $\text{id}(x)1_{(-n,n]}(x)$ bzw. $\text{id}(x)^p 1_{(-n,n]}(x)$, wobei $x \in \mathbb{Z}$, approximieren. Das Lebesgue-Integral dieser elementaren Zufallsvariablen entspricht gerade den rechten Seiten von (9.27) mit der Summation nur von $-n+1$ bis n . Die Aussage ergibt sich nun durch den Grenzübergang $n \rightarrow \infty$.

(ii) Dies folgt direkt aus Punkt (vi) in Satz 9.9 für $\phi(x) = x$ bzw. $\phi(x) = x^p$ und aus dem Riemann-Kalkül, für das gilt $dF_X(x) = f_X(x)dx$. \square

Nachdem wir mit dem Erwartungswert etwas über den ‘Schwerpunkt’ bzw. das ‘Mittel’ der Verteilung aussagen konnten, wollen wir nun ein Maß für die ‘Streubreite’ um diesen Schwerpunkt herum entwickeln.

Definition 9.11. Es sei $X \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$.

(i) Die **Varianz** von X ist definiert als

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]. \quad (9.30)$$

(ii) Die **Standardabweichung** von X ist definiert als

$$\sigma[X] = \sqrt{\text{Var}[X]}. \quad (9.31)$$

Wie wir im Folgenden sehen werden, hat die Varianz schönere mathematische Eigenschaften als die Standardabweichung. Sie hat jedoch auch den folgenden praktischen Nachteil: Trägt die Zufallsvariable X eine Einheit wie z.B. ‘Geld’, ‘Zeit’ oder ‘Länge’, so hat die Varianz diese Einheit im Quadrat, was oft unanschaulich ist. Hingegen hat die Standardabweichung dieselbe Einheit wie X und ist von daher einfacher zu interpretieren und an nicht-Mathematiker/innen zu kommunizieren.

Proposition 9.12. *Es sei $X \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$. Dann gilt*

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{M}_2[X] - \mathbb{E}[X]^2 . \quad (9.32)$$

Beweis. Die Aussage folgt durch die folgende Rechnung,

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 . \quad \square \end{aligned} \quad (9.33)$$

Beispiel 9.13. Wir möchten den Erwartungswert und Varianz einer $B(n, p)$ -verteilten Zufallsvariablen X berechnen.

Lösung. Nach Satz 9.10 ist

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} \\ &= q \frac{d}{dq} \sum_{x=0}^n \binom{n}{x} q^x (1-p)^{n-x} \Big|_{q=p} \\ &= q \frac{d}{dq} (q + (1-p))^n \Big|_{q=p} \\ &= nq (q + (1-p))^{n-1} \Big|_{q=p} = np \times 1^{n-1} = np . \end{aligned} \quad (9.34)$$

Analog kann man zeigen, dass

$$\begin{aligned} \mathbb{E}[X^2] &= \left(q \frac{d}{dq} \right)^2 (q + (1-p))^n \Big|_{q=p} \\ &= np + np(n-1)p \\ &= np(1-p) + (np)^2 . \end{aligned} \quad (9.35)$$

Also gilt

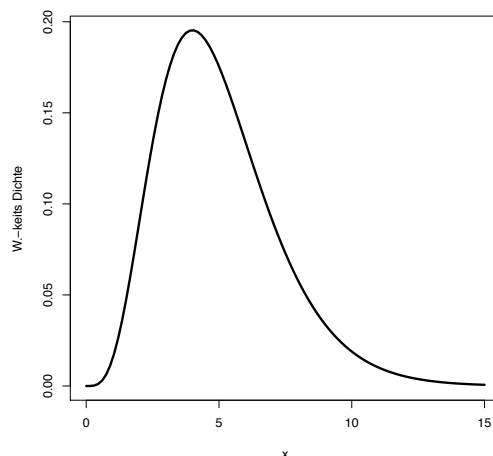
$$\text{Var}[X] = np(1-p) + (np)^2 - (np)^2 = np(1-p) . \quad (9.36)$$

◇

9.1. Übungsaufgaben

1. Selbsttest

- Die unten stehende Graphik zeigt die kontinuierliche Dichte einer Zufallsvariablen X . Welchen ungefähren Wert hat der Erwartungswert μ und die Standardabweichung σ der Zufallsvariablen X ? Zur Auswahl stehen $\mu = 1, 5, 10, 20$ und $\sigma = 0.1, 2.3, 8.5, 20$.



- Betrachten Sie nochmals die kontinuierliche Dichte aus der obigen Aufgabe. Ist für diese Verteilung das 50%-Quantil größer als der Mittelwert, oder kleiner?
- Berechnen Sie den Erwartungswert und die Varianz von folgenden Zufallsvariablen insofern diese existieren:
 - $X_1 \sim U(a, b)$ ist $U(a, b)$ -gleichverteilt
 - $X_2 \sim G(p)$ ist geometrisch verteilt zum Parameter p
 - $X_3 \sim \text{Exp}(\alpha)$ ist exponentialverteilt zum Parameter α
 - $X_4 = e^Y$ für eine Zufallsvariable $Y \sim N(0, 1)$
 - X_5 hat die Cauchy-Dichte $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ (*Hinweis:* Passen Sie hier besonders gut auf!)
- Wir betrachten eine geometrisch verteilte Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$, d.h. $X \sim G(p)$. Welche der folgenden Zufallsvariablen Y_1, Y_2, Y_3 liegen in $L^1((\Omega, \mathcal{A}, P); \mathbb{R})$?

$$Y_1 := X^2, \quad Y_2 := \sin(X), \quad Y_3 := e^{X^2}$$

Untersuchen Sie ebenfalls weitere Beispiele für Zufallsvariablen Y (ausgedrückt durch X) und entscheiden Sie ob diese Lebesgue integrierbar sind oder nicht!

5. Es sei X eine kontinuierliche Zufallsvariable deren kontinuierliche Dichte $f_X(x)$ um einen Punkt $\mu \in \mathbb{R}$ achsensymmetrisch ist, d.h.

$$f_X(\mu + s) = f_X(\mu - s) \quad , \quad \forall s \in \mathbb{R} \quad .$$

Zeigen Sie, dass $\mathbb{E}[X] = \mu$ gilt!

6. Zeigen Sie, dass für $X \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$ und $a \in \mathbb{R}$ gilt $\text{Var}[aX] = a^2 \text{Var}[X]$!

2. Transfer

Aus einer Urne mit 5 Kugeln - nummeriert von 0 bis 4 - wird zweimal mit Zurücklegen gezogen.

- Bestimmen Sie die Verteilungsfunktion der Summe der Zahlen auf den gezogenen Kugeln. Wie groß ist die Wahrscheinlichkeit, dass die Summe (i) kleiner als 2, (ii) mindestens 4 ist?
- Berechnen Sie den Erwartungswert der Summe der gezogenen Zahlen.
- Berechnen Sie die Varianz der Summe der gezogenen Zahlen.
- Ein spielsüchtiger Bekannter auf der Suche nach weiteren Einnahmequellen schlägt Ihnen folgendes Glücksspiel vor: Sie zahlen 2.5 Euro Einsatz und ziehen zweimal wie oben beschrieben aus der Urne. Anschließend wird die Summe der gezogenen Zahlen bestimmt und er zahlt Ihnen die Hälfte dieser Summe in Euro als Gewinn aus. Berechnen Sie Erwartungswert und Varianz Ihres Nettogewinns (Gewinn abzüglich Einsatz) für ein Einzelspiel. Für wen lohnt sich das Spiel langfristig mehr?

3. Theorie

- Beweisen Sie die folgenden Aussagen über elementare Zufallsvariablen:
 - Der Raum der elementaren Zufallsvariablen ist ein Vektorraum, d.h. mit X, Y elementaren Zufallsvariablen und $a, b \in \mathbb{R}$ ist auch $aX + bY$ elementare Zufallsvariable.
 - Der Raum der elementaren Zufallsvariablen ist sogar eine Algebra, d.h. mit X, Y elementaren Zufallsvariablen ist auch XY eine elementare Zufallsvariable.
 - Das Lebesgue-Integral für elementare Zufallsvariablen ist linear, d.h. $\int_{\Omega} (aX + bY) dP = a \int_{\Omega} X dP + b \int_{\Omega} Y dP$.
 - Das Lebesgue-Integral einer positiven elementaren Zufallsvariablen ist positiv.
 - Für X, Y elementare Zufallsvariablen mit $X \geq Y$ gilt $\int_{\Omega} X dP \geq \int_{\Omega} Y dP$.

9. Erwartungswert, Varianz und Momente

2. Eine Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißt **deterministisch**, falls ein $\mu \in \mathbb{R}$ existiert, so dass $X = \mu$ P -fast sicher, d.h. $P(X = \mu) = 1$. Zeigen Sie für quadratintegrierbare X die folgende Aussage: X ist deterministisch genau dann wenn $\text{Var}[X] = 0$.

9.2. **Literatur**

10. Erzeugende Funktionen

Schon in Beispiel 9.13 haben wir gesehen, dass es oft nützlich ist, die Berechnung von den Momenten einer Zufallsvariablen auf die Berechnung von Ableitungen zurückzuführen. Mittels Laplace- und Fouriertransformationen wird dieser Ansatz nun systematisiert.

Definition 10.1. Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine reellwertige Zufallsvariable.

(i) Für X nicht-negativ, definiere die **Laplacetransformation** von X als

$$\mathcal{L}_X(\alpha) := \mathbb{E}[e^{-\alpha X}] \quad , \quad \alpha \in [0, \infty) . \quad (10.1)$$

(ii) Für X beliebig, definiere die **Fouriertransformation** von X als

$$\mathcal{F}_X(\alpha) := \mathbb{E}[e^{i\alpha X}] \quad , \quad \alpha \in \mathbb{R} . \quad (10.2)$$

Hierbei ist für $H : \mathbb{R} \rightarrow \mathbb{C}$ mit messbaren Real- und Imaginärteil $\Re H, \Im H : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $\Re H(X), \Im H(X) \in L^1((\Omega, \mathcal{A}, P); \mathbb{R})$ das Lebesgue-Integral definiert als

$$\int_{\Omega} H(X) dP := \int_{\Omega} \Re H(X) dP + i \int_{\Omega} \Im H(X) dP . \quad (10.3)$$

Aus Satz 9.9 folgt, dass für eine diskrete Zufallsvariable X mit diskreter Dichte p_x^X die Fourier- bzw. die Laplacetransformation wie folgt ausgedrückt werden können

$$\mathcal{L}_X(\alpha) = \sum_{x=0}^{\infty} e^{-\alpha x} p_x^X \quad , \quad \text{für } X \geq 0 \quad , \quad \alpha \geq 0 \quad , \quad (10.4a)$$

$$\mathcal{F}_X(\alpha) = \sum_{x \in \mathbb{Z}} e^{i\alpha x} p_x^X \quad , \quad \text{für } X \text{ beliebig} \quad , \quad \alpha \in \mathbb{R} . \quad (10.4b)$$

Die analogen Formeln für X mit kontinuierlicher Dichte $f_X(x)$ lauten

$$\mathcal{L}_X(\alpha) = \int_{[0, \infty)} e^{-\alpha x} f_X(x) dx \quad , \quad \text{für } X \geq 0 \quad , \quad \alpha > 0 \quad , \quad (10.5a)$$

$$\mathcal{F}_X(\alpha) = \int_{\mathbb{R}} e^{i\alpha x} f_X(x) dx \quad , \quad \text{für } X \text{ beliebig} \quad , \quad \alpha \in \mathbb{R} . \quad (10.5b)$$

Eine schöne Anwendung der Laplace- bzw. Fouriertransformation ist, dass wir die Momente von X relativ einfach aus diesen Transformationen gewinnen können.

Satz 10.2. Es sei $X \in L^p((\Omega, \mathcal{A}, P); \mathbb{R})$, mit $p \in \mathbb{N}$.

10. Erzeugende Funktionen

(i) Falls $X \geq 0$, dann ist $\mathcal{L}_X(\alpha)$ p -fach differenzierbar auf $[0, \infty)$ und es gilt

$$\mathbb{M}_p[X] = (-1)^p \frac{d^p}{d\alpha^p} \mathcal{L}_X(\alpha) \upharpoonright_{\alpha=0} . \quad (10.6)$$

(ii) Für X beliebig ist $\mathcal{F}_X(\alpha)$ p -fach differenzierbar auf \mathbb{R} und es gilt

$$\mathbb{M}_p[X] = (-i)^p \frac{d^p}{d\alpha^p} \mathcal{F}_X(\alpha) \upharpoonright_{\alpha=0} . \quad (10.7)$$

Beweis. Hier wird nur der allgemeine Fall (ii) bewiesen. Das Argument für (i) folgt analog. Die Beweisidee ist ganz einfach, dass $(-i)^p \frac{d^p}{d\alpha^p} e^{i\alpha x} = e^{i\alpha x} x^p$. Um Ableitung und Erwartungswert zu vertauschen, soll nun der Satz über die dominierte Konvergenz herangezogen werden. Dazu verfahren wir induktiv wie folgt: Für $j < p$ und $\alpha_n \rightarrow \alpha$ betrachte die Zufallsvariable

$$Y_n := (-i) \frac{e^{i\alpha_n X} - e^{i\alpha X}}{\alpha_n - \alpha} X^j . \quad (10.8)$$

Da $\left| \frac{e^{i\alpha_n X} - e^{i\alpha X}}{\alpha_n - \alpha} \right| \leq |X|$, gilt $|Y_n| \leq |X|^{j+1} \in L^1((\Omega, \mathcal{A}, P); \mathbb{R})$. Also haben der Real- und der Imaginärteil von Y_n eine L^1 -Dominante. Nun können wir mit

$$Y_n(\omega) \rightarrow X^{j+1}(\omega) \quad \forall \omega \in \Omega \quad (10.9)$$

und dem Satz von Lebesgue für Real- und Imaginärteil schließen, dass

$$(-i) \frac{d}{d\alpha} \mathbb{E}[e^{i\alpha X} X^j] = (-i) \lim_{n \rightarrow \infty} \frac{\mathbb{E}[e^{i\alpha_n X} X^j] - \mathbb{E}[e^{i\alpha X} X^j]}{\alpha_n - \alpha} = \lim_{n \rightarrow \infty} \mathbb{E}[Y_n] = \mathbb{E}[e^{i\alpha X} X^{j+1}] . \quad (10.10)$$

Insbesondere existiert dieser Grenzwert für beliebige $\alpha_n \rightarrow \alpha$ und damit existiert die Ableitung. Durch vollständige Induktion über p (Induktionsanfang $p = 0$ ist trivial) erhalten wir, dass p Ableitungen existieren und dass

$$(-i)^p \frac{d^p}{d\alpha^p} \mathbb{E}[e^{i\alpha X}] = \mathbb{E}[e^{i\alpha X} X^p] \quad (10.11)$$

gilt. Setze nun $\alpha = 0$, um das Argument zu beenden. \square

Beispiel 10.3. Wir berechnen die Laplacetransformation, den Erwartungswert und die Varianz für $X \sim \text{Po}(\lambda)$.

Lösung. Mit $X \sim \text{Po}(\lambda)$ gilt

$$\mathcal{L}_X(\alpha) = \sum_{x=0}^{\infty} e^{-\alpha x} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{-\alpha})^x}{x!} = e^{-\lambda} e^{\lambda e^{-\alpha}} = e^{\lambda(e^{-\alpha}-1)} . \quad (10.12)$$

Die erste und zweite Ableitung von $\mathcal{L}_X(\alpha)$ berechnen sich nun wie folgt:

$$\mathbb{E}[X] = -\frac{d}{d\alpha} \mathcal{L}_X(\alpha) \upharpoonright_{\alpha=0} = \lambda e^{-\alpha} e^{\lambda(e^{-\alpha}-1)} \upharpoonright_{\alpha=0} = \lambda \quad (10.13)$$

und

$$\mathbb{M}_2[X] = \frac{d^2}{d\alpha^2} \mathcal{L}_X(\alpha) \upharpoonright_{\alpha=0} = \left(\lambda e^{-\alpha} e^{\lambda(e^{-\alpha}-1)} + \lambda^2 e^{-\alpha} e^{\lambda(e^{-\alpha}-1)} \right) \upharpoonright_{\alpha=0} = \lambda + \lambda^2 . \quad (10.14)$$

Deswegen

$$\text{Var}[X] = \mathbb{M}_2[X] - \mathbb{E}[X]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda . \quad (10.15)$$

◇

Definition 10.4. Für α hinreichend klein ist $\psi_X(\alpha) := \log \mathcal{F}_X(\alpha)$ definiert. Wir definieren für $p \in \mathbb{N}$ und $X \in L^p((\Omega, \mathcal{A}, P); \mathbb{R})$ die **p -te Kumulante** von X als

$$\mathcal{C}_p[X] := (-i)^p \frac{d^p}{d\alpha^p} \psi_X(\alpha) \upharpoonright_{\alpha=0} . \quad (10.16a)$$

Ist ferner $X \geq 0$, so gilt

$$\mathcal{C}_p[X] = (-1)^p \frac{d^p}{d\alpha^p} \log \mathcal{L}_X(\alpha) \upharpoonright_{\alpha=0} . \quad (10.16b)$$

Satz 10.5. Für $m \in \mathbb{N}$ und $X \in L^m((\Omega, \mathcal{A}, P); \mathbb{R})$ gilt

$$\mathbb{M}_m[X] = \sum_{k=1}^m \sum_{\substack{n_1+\dots+n_k=m \\ n_j > 0}} \frac{1}{k!} \binom{m}{n_1 \dots n_k} \prod_{l=1}^k \mathcal{C}_{n_l}[X] . \quad (10.17)$$

Insbesondere gilt

$$\mathcal{C}_1[X] = \mathbb{E}[X] \quad \text{und} \quad \mathcal{C}_2[X] = \text{Var}[X] . \quad (10.18)$$

Beweis. In einer Umgebung von 0 gilt $\mathcal{F}_X(\alpha) = \exp(\psi_X(\alpha))$. Nach Korollar 1.24 und (10.7) folgt daher (10.17).

Da es für $m = 1$ auf der rechten Seite von (10.17) nur einen einzigen Summanden gibt, nämlich $k = 1$ und $n_1 = 1$, so gilt

$$\mathbb{E}[X] = \mathbb{M}_1[X] = \frac{1}{1!} \binom{1}{1} \mathcal{C}_1[X] = \mathcal{C}_1[X] . \quad (10.19)$$

Für $m = 2$ stehen auf der rechten Seite von (10.17) die Summanden für $k = 1$ und $n_1 = 2$ sowie $k = 2$, $n_1 = n_2 = 1$. Also

$$\text{Var}[X] + \mathbb{E}[X]^2 = \mathbb{M}_2[X] = \frac{1}{1!} \binom{2}{2} \mathcal{C}_2[X] + \frac{1}{2!} \binom{2}{1 \ 1} \mathcal{C}_1[X]^2 = \mathcal{C}_2[X] + \mathcal{C}_1[X]^2 . \quad (10.20)$$

Verwende nun $\mathcal{C}_1[X] = \mathbb{E}[X]$ und ziehe oben $\mathbb{E}[X]^2$ von beiden Seiten ab. □

10. Erzeugende Funktionen

Korollar 10.6. *Es sei $X \sim \text{Po}(1)$, dann gilt*

$$\mathbb{M}_m[X] = B_m , \quad (10.21)$$

wobei B_m die m -te Bell-Zahl ist.

Beweis. Dies folgt aus $\mathcal{C}_n[X] = (-1)^n \frac{d^n}{d\alpha^n} (e^{-\alpha} - 1) \upharpoonright_{\alpha=0} = 1 \quad \forall n \in \mathbb{N}$. Hier haben wir die Rechnung aus Beispiel 10.3 und (10.16b) verwendet. \square

Wir können nun die Parameter μ und σ^2 der Normalverteilung als Erwartungswert und Varianz identifizieren.

Satz 10.7. *Sei $X \sim N(\mu, \sigma^2)$ eine normalverteilte Zufallsvariable. Dann gilt*

(i) *Die Fouriertransformation von X ist*

$$\mathcal{F}_X(\alpha) = e^{i\mu\alpha - \frac{1}{2}\sigma^2\alpha^2} . \quad (10.22)$$

(ii) *Erwartungswert, Varianz und höhere Kumulanten der Normalverteilung sind gegeben durch*

$$\mathbb{E}[X] = \mathcal{C}_1[X] = \mu , \quad \text{Var}[X] = \mathcal{C}_2[X] = \sigma^2 , \quad \mathcal{C}_n[X] = 0 , \quad n > 2 . \quad (10.23)$$

Beweis. Wir berechnen zunächst die sogenannte zweiseitige Laplacetransformation von X ,

$$\begin{aligned} \mathcal{L}_X(\alpha) = \mathbb{E}[e^{\alpha X}] &= \int_{\mathbb{R}} e^{\alpha x} dN(\mu, \sigma^2)(x) \\ &= \int_{-\infty}^{\infty} e^{\alpha x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &\stackrel{\phi(x)=\frac{x-\mu}{\sigma}}{=} \int_{-\infty}^{\infty} e^{\alpha(\sigma x + \mu)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= e^{\alpha\mu + \frac{1}{2}\sigma^2\alpha^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 + 2\sigma\alpha x + \sigma^2\alpha^2)} dx \\ &= e^{\alpha\mu + \frac{1}{2}\sigma^2\alpha^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+\sigma\alpha)^2} dx \\ &\stackrel{\phi(x)=x+\sigma\alpha}{=} e^{\alpha\mu + \frac{1}{2}\sigma^2\alpha^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx}_{=1} . \end{aligned} \quad (10.24)$$

Betrachte nun die Fortsetzung von $\mathcal{L}_X(\alpha)$ auf \mathbb{C} vermöge

$$\mathcal{L}_X(z) = \int_{\mathbb{R}} e^{zx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx , \quad z \in \mathbb{C} . \quad (10.25)$$

Mit dem Satz über dominierte Konvergenz angewandt auf einen komplexen Differenzenquotienten sieht man leicht, dass $\mathcal{L}_X(z)$ analytisch ist in ganz \mathbb{C} . Folglich gilt

$$\mathcal{L}_X(z) = e^{\mu z + \frac{1}{2}\sigma^2 z^2}, \quad (10.26)$$

denn diese Relation ist korrekt für z mit Imaginärteil Null und setzt sich somit auf ganz \mathbb{C} fort. Wähle nun $\alpha \in \mathbb{R}$ und $z = i\alpha$, dann gilt

$$\mathcal{F}_X(\alpha) = \mathcal{L}_X(i\alpha) = e^{i\mu\alpha - \frac{1}{2}\sigma^2\alpha^2}. \quad (10.27)$$

Dies beweist (i). Der Punkt (ii) folgt unmittelbar aus Definition 10.4 sowie (i). \square

Definition 10.8. (i) Es sei $X \in L^3((\Omega, \mathcal{A}, P); \mathbb{R})$ mit $\text{Var}[X] > 0$. Dann ist die **Schiefe** (englisch: *skewness*) von X definiert als

$$\text{Skew}[X] := \frac{\mathcal{C}_3[X]}{\text{Var}[X]^{\frac{3}{2}}}. \quad (10.28)$$

(ii) Es sei $X \in L^4((\Omega, \mathcal{A}, P); \mathbb{R})$ mit $\text{Var}[X] > 0$. Dann ist die **Wölbung** (englisch: *kurtosis*) von X definiert als

$$\text{Kurt}[X] = \frac{\mathcal{C}_4[X]}{\text{Var}[X]^2}. \quad (10.29)$$

Offensichtlich ist eine normalverteilte Zufallsvariable X weder schief ($\text{Skew}[X] = 0$) noch gewölbt ($\text{Kurt}[X] = 3$). Die Abbildung 10.1 zeigt **linkssteile** kontinuierliche Dichten mit $\text{Skew}[X] > 0$ und **rechtssteile** mit $\text{Skew}[X] < 0$. In Abbildung 10.2 sieht man positiv und negativ gewölbte kontinuierliche Dichten.

In linkssteilen Verteilungen ist der Median in der Regel kleiner als der Erwartungswert, in rechtssteilen Verteilungen ist es umgekehrt. Dies sind jedoch keine mathematischen Aussagen, da man (ziemlich merkwürdig aussehende) Gegenbeispiele konstruieren kann.

Für die Schiefe und Wölbung kann man mittels Satz 10.5 zeigen, dass

$$\text{Skew}[X] = \frac{\mathbb{E}[(X - \mathbb{E}[X])^3]}{\mathbb{E}[(X - \mathbb{E}[X])^2]^{\frac{3}{2}}}, \quad (10.30a)$$

$$\text{Kurt}[X] = \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{\mathbb{E}[(X - \mathbb{E}[X])^2]^2} - 3. \quad (10.30b)$$

Zuletzt möchten wir noch für diskrete nicht-negative Zufallsvariablen $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{N}_0, \mathbb{P}(\mathbb{N}_0))$ eine weitere erzeugende Funktion einführen.

Definition 10.9. Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{N}_0, \mathbb{P}(\mathbb{N}_0))$ eine diskrete nicht-negative Zufallsvariable mit diskreter Dichte p_x^X , $x \in \mathbb{N}_0$. Dann heißt

$$\mathcal{M}_X(\beta) := \sum_{x=0}^{\infty} \beta^x p_x^X, \quad \beta \in [0, 1] \quad (10.31)$$

die **wahrscheinlichkeitserzeugende Funktion** von X .

10. Erzeugende Funktionen

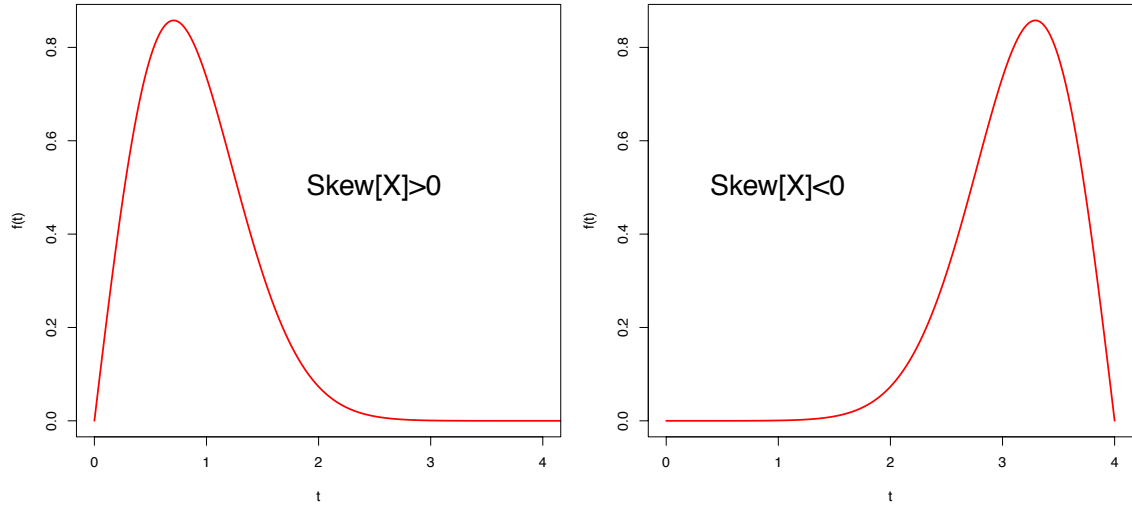


Abbildung 10.1: Beispiele für kontinuierliche Dichten mit positiver und negativer Schiefe.

Proposition 10.10. Sei X wie in Definition 10.9 und $\mathcal{M}_X(\beta)$ die zugehörige Wahrscheinlichkeitserzeugende Funktion. Dann gilt

(i) $\mathcal{M}_X(\beta)$ erzeugt die diskrete Dichte von X durch Ableiten bei 0, d.h. für alle $x \in \mathbb{N}_0$

$$p_x^X = \frac{1}{x!} \frac{d^x}{d\beta^x} \mathcal{M}_X(\beta) \upharpoonright_{\beta=0} . \quad (10.32)$$

(ii) Für $\beta \in (0, 1]$ und $\alpha = -\log(\beta)$ gilt

$$\mathcal{M}_X(\beta) = \mathcal{L}_X(\alpha) . \quad (10.33)$$

Beweis. (i) Offensichtlich ist $\mathcal{M}_X(\beta)$ analytisch für $|\beta| < 1$ und daher Term für Term ableitbar. Bei der Auswertung der x -ten Ableitung an der Stelle 0 bleibt folglich nur $x! p_x^X$ übrig.

Um (ii) zu zeigen, wähle $0 < \beta \leq 1$ und schließe wie folgt:

$$\mathcal{M}_X(\beta) = \sum_{x=0}^{\infty} \beta^x p_x^X = \sum_{x=0}^{\infty} \left(e^{\log(\beta)} \right)^x p_x^X = \sum_{x=0}^{\infty} e^{-x\alpha} p_x^X = \mathcal{L}_X(\alpha) . \quad \square \quad (10.34)$$

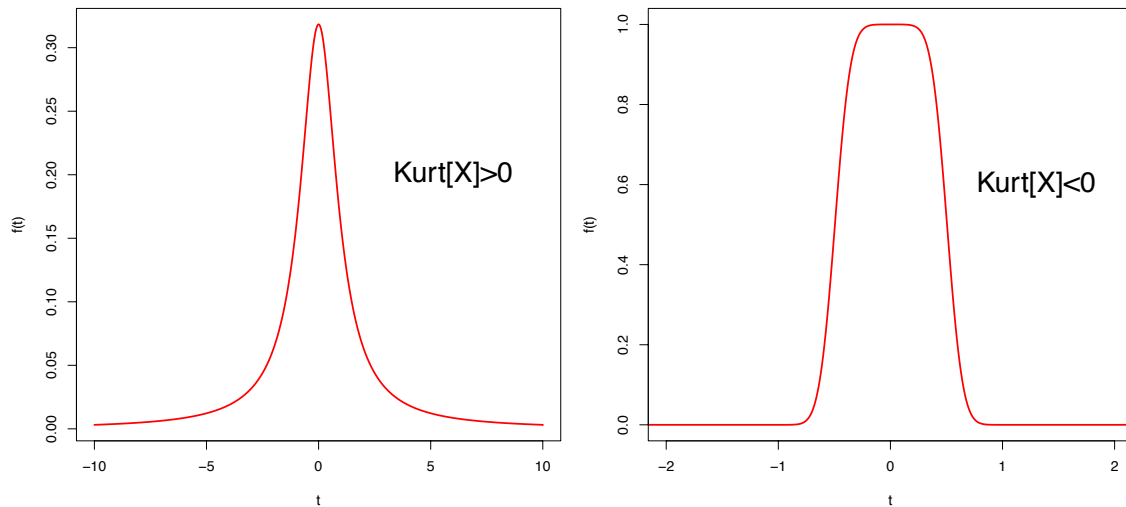


Abbildung 10.2: Beispiele für kontinuierliche Dichten mit positiver und negativer Wölbung.

10.1. Übungsaufgaben

1. Selbsttest

- Bestimmen Sie die Laplacetransformation von $X \sim \text{Exp}(\alpha)$ und $Y \sim G(p)$!

Hinweis: α tritt hier in zwei unterschiedlichen Rollen auf, nämlich als Argument der Laplacetransformation und als Parameter der Exponentialverteilung. Bringen Sie das nicht durcheinander.

- Eine Zufallsvariable X habe eine bestimmte zusammengesetzte (compound) Poisson-Verteilung mit Fouriertransformation

$$\mathcal{F}_X(\alpha) = \mathbb{E}[e^{i\alpha X}] = \exp(\cos(\alpha) - 1).$$

Berechnen Sie Erwartungswert, Varianz, Schiefe und Wölbung von X !

- Welche der folgenden Funktionen ist eine wahrscheinlichkeitserzeugende Funktion? Geben Sie eine Begründung an!

- $m_X(\beta) \equiv 1$
- $m_X(\beta) = e^{-\lambda} e^{\beta\lambda}$, mit $\lambda > 0$
- $m_X(\beta) = \frac{1}{3} - \frac{1}{3}\beta + \frac{1}{3}\beta^5$
- $m_X(\beta) = \frac{1}{5} + \frac{2}{5}\beta + \frac{1}{5}\beta^9 + \frac{1}{5}\beta^{16054}$
- $m_X(\beta) = \frac{1-p}{1-\beta^p}$, mit $p \in (0, 1)$

Hinweis: Erinnern Sie sich, dass es zwei Bedingungen an die Zahlen p_x^X gibt, so dass sie eine diskrete Dichte beschreiben!

2. Transfer

Sie arbeiten in einem Versicherungsunternehmen und bekamen die Aufgabe die Schadenssumme X der an Silvester entstehenden Unfälle zu kalkulieren. Da Sie in der Stochastik-Vorlesung gut aufgepasst haben, wählen Sie als Verteilungsmodell für X die Gammaverteilung, d.h. $X \sim \Gamma(\alpha, \beta)$.

- a) Bestimmen Sie die Laplacetransformation von X !
- b) Bestimmen Sie aus der Laplacetransformation den Erwartungswert und die Varianz von X !
- c) Aus langjähriger Erfahrung wissen Sie, dass im Mittel ein Schaden von 150000 Euro entsteht. Die Schwankung um den mittleren Schaden sei mit 50000 Euro bemessen. Bestimmen Sie die freien Parameter α, β der Gammaverteilung indem Sie den Erwartungswert gleich dem mittleren Schaden und die Varianz gleich der quadrierten Schwankung setzen!
- d) Wie groß ist die Wahrscheinlichkeit, dass mehr als 200000 Euro Schaden entstehen?

Hinweis: Um dieses Problem zu lösen, müssen Sie ein Integral berechnen, welches der unvollständigen Gammafunktion entspricht. Informieren Sie sich hierzu im Internet! Alternativ können Sie auch die gesuchte Wahrscheinlichkeit mittels R ausrechnen. Der relevante Befehl ist `dgamma`, über welchen Sie sich mittels der Hilfsseiten `R>dgamma` mehr erfahren können.

10.2. Literatur

11. Multivariate Verteilungen

In vielen statistischen Untersuchungen wird pro Untersuchungseinheit mehr als nur eine Größe erhoben. Dementsprechend muss die Wahrscheinlichkeitstheorie auch Modelle mit mehr als nur einer Zufallsvariable X zur Verfügung stellen. Solche Modelle nennt man **multivariat**. Des Weiteren möchten wir in der Lage sein, die ‘unabhängige’ Wiederholung von Zufallsexperimenten auf der Theorie-Seite abzubilden. Auch dies geschieht durch die gleichzeitige Betrachtung von mehreren Zufallsvariablen.

Zwei Betrachtungsweisen sind hier gleichermaßen legitim: Wir können d reellwertige Zufallsvariablen $X_j : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $j = 1, \dots, d$, betrachten, oder eine \mathbb{R}^d -wertige Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, also einen Zufallsvektor

$$X = (X_1, \dots, X_d) . \quad (11.1)$$

Die Definition einer geeigneten σ -Algebra $\mathcal{B}(\mathbb{R}^d)$ auf \mathbb{R}^d lautet wie folgt:

Definition 11.1. Es seien $x, y \in \bar{\mathbb{R}}^d := [-\infty, \infty]^d$. Wir schreiben $y \leq x$ falls $y_1 \leq x_1, \dots, y_d \leq x_d$ gilt, wobei $y = (y_1, \dots, y_d)$ und $x = (x_1, \dots, x_d)$. Die Notation $y < x$ ist analog zu verstehen. Wir verwenden die Konvention $\pm\infty \in \bar{\mathbb{R}}^d$ für $(\pm\infty, \dots, \pm\infty)$ (d Mal). Wir definieren die folgenden Objekte:

- (i) Die **halboffenen Intervalle** in \mathbb{R}^d sind gegeben für $y, x \in \bar{\mathbb{R}}^d$, $y \leq x$, durch

$$(y, x] := \{q \in \mathbb{R}^d : y < q \leq x\} . \quad (11.2)$$

Mit $\mathcal{I}(\mathbb{R}^d)$ bezeichnen wir das \cap -stabile System von allen halboffenen Intervallen in \mathbb{R}^d .

- (ii) Die Algebra der **Figuren in \mathbb{R}^d** (vergleiche Abbildung 11.1) ist gegeben als

$$\mathcal{F}(\mathbb{R}^d) := \left\{ A = \bigcup_{j=1}^n I_j : I_j \in \mathcal{I}(\mathbb{R}^d) \right\} . \quad (11.3)$$

Der Beweis zur Algebra-Eigenschaft dieses Mengensystems ist analog zu Lemma 7.11.

- (iii) Die **Borelsche σ -Algebra $\mathcal{B}(\mathbb{R}^d)$** ist die von $\mathcal{F}(\mathbb{R}^d)$ erzeugte σ -Algebra. Offensichtlich ist auch $\mathcal{I}(\mathbb{R}^d)$ ein Erzeuger für $\mathcal{B}(\mathbb{R}^d)$.

11. Multivariate Verteilungen

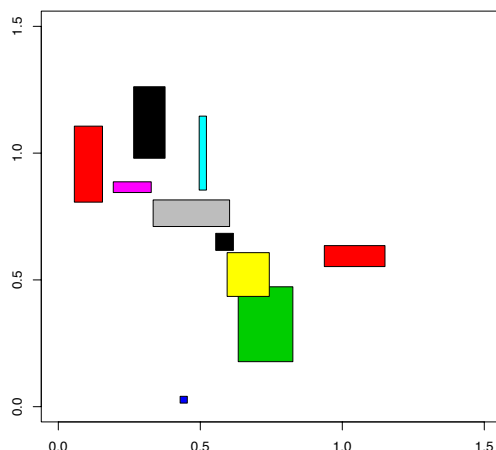


Abbildung 11.1: Eine Figur in \mathbb{R}^2 .

Bemerkung 11.2. Die Borelsche σ -Algebra ist ein Spezialfall der **Tensor σ -Algebra**, die wie folgt definiert ist: Es sei I eine beliebige Indexmenge und $(\Omega_i, \mathcal{A}_i)$, $i \in I$, eine Familie messbarer Räume. Definiere die Menge $\Omega := \times_{i \in I} \Omega_i$ und bezeichne deren Elemente mit $\omega = \times_{i \in I} \omega_i$, wobei $\omega_i \in \Omega_i$ für alle $i \in I$. Definiere weiter die Projektionen, für alle $j \in I$,

$$\pi_j : \Omega \rightarrow \Omega_j, \quad \omega \mapsto \pi_j(\omega) = \omega_j. \quad (11.4)$$

Dann ist $\otimes_{i \in I} \mathcal{A}_i$ per Definition die kleinste σ -Algebra über Ω , die sämtliche Mengensysteme $\pi_j^{-1}(\mathcal{A}_j)$, $j \in I$ enthält. Insbesondere gilt $\mathbb{R}^d = \mathbb{R} \times \dots \times \mathbb{R}$ (d Mal) und

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d} = \underbrace{\mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})}_{d \text{ Mal}}. \quad (11.5)$$

Um dies einzusehen, genügt es zu zeigen, dass für $(y, x] \in \mathcal{I}(\mathbb{R})$ gilt $\pi_j^{-1}((y, x]) = \mathbb{R} \times \dots \times (y, x] \times \dots \times \mathbb{R} \in \mathcal{I}(\mathbb{R}^d)$, was offensichtlich der Fall ist. Also liegt ein erzeugendes System der d -fachen Tensor σ -Algebra in $\mathcal{B}(\mathbb{R}^d)$ und damit die Tensor σ -Algebra selbst. Die umgekehrte Inklusion folgt nun daraus, dass für $y, x \in \mathbb{R}^d$ das halboffene Intervall $(y, x] = \bigcap_{j=1}^d \pi_j^{-1}((y_j, x_j])$ als d -facher Durchschnitt von Mengen aus der Tensor σ -Algebra dargestellt werden kann. Also liegt ein erzeugendes System von $\mathcal{B}(\mathbb{R}^d)$ in der d -fachen Tensor σ -Algebra $\mathcal{B}(\mathbb{R})^{\otimes d}$ und damit $\mathcal{B}(\mathbb{R}^d)$ selbst.

Aus diesen Beobachtungen entnehmen wir, dass eine Abbildung $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ genau dann messbar ist, wenn jede Komponente $X_j = \pi_j \circ X$, $j = 1, \dots, d$ messbar ist.

Wie in Bauer, MI, Satz 6.4 gezeigt, ist $\mathcal{B}(\mathbb{R}^d)$ identisch mit der von der \mathbb{R}^d -Topologie erzeugten σ -Algebra. Insbesondere ist damit jede stetige Abbildung $\phi : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q))$ messbar. Dies, und der Umstand, dass jede Funktion die punktweise durch messbare Funktionen approximiert wird messbar ist, enthebt uns in der Regel der lästigen Pflicht, Messbarkeitseigenschaften aufwändig ‘Menge für Menge’ nachzuprüfen.

Definition 11.3. (i) Eine messbare Abbildung $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ heißt \mathbb{R}^d -wertige Zufallsvariable.

(ii) Das Bildmaß $\text{Vert}(X) = X_*P$ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ heißt **(multivariate) Verteilung** von X .

(iii) Die **(multivariate) Verteilungsfunktion** von X ist für $x \in \mathbb{R}^d$, $x = (x_1, \dots, x_d)$, gegeben durch

$$\begin{aligned} F_X(x) &:= P(X \leq x) \\ &:= P(\{\omega \in \Omega : X_1(\omega) \leq x_1, \dots, X_d(\omega) \leq x_d\}) \\ &= P(\{X_1 \leq x_1\} \cap \dots \cap \{X_d \leq x_d\}) \\ &= X_*P((-\infty, x]) = \text{Vert}(X)((-\infty, x]) . \end{aligned} \quad (11.6)$$

Satz 11.4 (Eigenschaften der multivariaten Verteilungsfunktion). Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ eine \mathbb{R}^d -wertige Zufallsvariable. Dann gilt für die Verteilungsfunktion $F_X(x)$:

(i) $F_X(x)$ ist monoton, d.h.

$$F_X(y) \leq F_X(x) , \quad \text{für } y, x \in \mathbb{R}^d , \quad y \leq x . \quad (11.7)$$

Weiterhin gilt die Positivitätsbedingung

$$\sum_{S \subseteq \{1, \dots, d\}} (-1)^{\#S} F_X(\tau(y, x, S)) \geq 0 \quad \forall x, y \in \mathbb{R}^d , \quad \text{mit } y \leq x , \quad (11.8)$$

wobei

$$\tau_j(y, x, S) := \begin{cases} y_j & , \quad \text{für } j \in S , \\ x_j & , \quad \text{für } j \notin S . \end{cases} \quad (11.9)$$

(ii) $F_X(x)$ ist rechtsstetig, d.h.

$$\lim_{y \searrow x} F_X(y) := \lim_{y_1 \searrow x_1, \dots, y_d \searrow x_d} F_X(y_1, \dots, y_d) = F_X(x) , \quad \forall x = (x_1, \dots, x_d) \in \mathbb{R}^d . \quad (11.10)$$

(iii) $F_X(x)$ ist normiert im folgenden Sinne

$$\lim_{x \rightarrow \infty} F_X(x) := \lim_{x_1 \rightarrow \infty, \dots, x_d \rightarrow \infty} F_X(x_1, \dots, x_d) = 1 , \quad (11.11a)$$

und

$$\lim_{x_j \rightarrow -\infty} F(x_1, \dots, x_j, \dots, x_d) = 0 , \quad \forall j = 1, \dots, d . \quad (11.11b)$$

11. Multivariate Verteilungen

Beweis. (i) Monotonie folgt aus $(-\infty, y] \subseteq (-\infty, x]$ und der Isotonie von X_*P . Die Positivitätsbedingung entspricht gerade der Positivität des Wahrscheinlichkeitsmaßes auf halboffenen Intervallen $(y, x] \subseteq \mathbb{R}^d$. Anstelle eines allgemeinen Beweises betrachten wir die Situation in $d = 2$ Dimensionen, vergleiche Abbildung 11.2. Hier gilt offensichtlich für $y \leq x$

$$P((y, x]) = F(x_1, x_2) - F(y_1, x_2) - F(x_1, y_2) + F(y_1, y_2) . \quad (11.12)$$

Dies ist genau die Positivitätsbedingung in $d = 2$. Für den allgemeinen Fall verweisen wir auf Bauer, MI, Seite 38.

(ii) und (iii) folgen aus der Stetigkeit dieses Wahrscheinlichkeitsmaßes, vergleiche Lemma 2.11. \square

Auch im mehrdimensionalen Fall können **multivariate Verteilungsfunktionen**, d.h. Funktionen $F : \mathbb{R}^d \rightarrow [0, 1]$, welche die Eigenschaften (i)–(iii) aus Satz 11.4 erfüllen, zur Definition von Wahrscheinlichkeitsmaßen auf \mathbb{R}^d herangezogen werden.

Satz 11.5. *Es sei $F : \mathbb{R}^d \rightarrow [0, 1]$ eine multivariate Verteilungsfunktion. Dann existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, so dass $F(x) = P((-\infty, x])$ für alle $x \in \mathbb{R}^d$.*

Beweis. Wir geben nur eine Skizze des Beweises und verweisen auf Bauer, MI, Satz 6.5 für einen vollen Beweis.

Im Allgemeinen definieren wir für $y, x \in \mathbb{R}^d$ mit $y \leq x$

$$P^\circ((y, x]) = \sum_{S \subseteq \{1, \dots, d\}} (-1)^{\#S} F(\tau(y, x, S)) , \quad (11.13)$$

wobei $\tau(y, x, S)$ in (11.9) definiert wurde.

Insbesondere haben wir mit Eigenschaft (iii) aus Satz 11.4

$$P^\circ(\mathbb{R}^d) = \sum_{S \subseteq \{1, \dots, d\}} (-1)^{\#S} F(\tau(-\infty, \infty, S)) = F(\infty) + 0 = 1 \quad (11.14)$$

und

$$P^\circ(\emptyset) = \sum_{S \subseteq \{1, \dots, n\}} (-1)^{\#S} F(\tau(-\infty, -\infty, S)) = 0 . \quad (11.15)$$

Mittels der Siebformel können wir nun P° konsistent und additiv auf die Algebra der Figuren fortsetzen,

$$P^\circ(A) := \sum_{I \subseteq \{1, \dots, n\}} (-1)^{\#I+1} P^\circ\left(\bigcap_{j \in I} I_j\right) , \quad A = \bigcup_{j=1}^n I_j , \quad I_j \in \mathcal{I}(\mathbb{R}^d) , \quad (11.16)$$

denn $\bigcap_{j \in I} I_j \in \mathcal{I}(\mathbb{R}^d)$ aufgrund der \cap -Stabilität dieses Mengensystems.

Zuletzt muß noch die Prämaß-Eigenschaft für den Inhalt P° nachgewiesen werden. Hierfür ist wieder die Rechtsstetigkeit von entscheidender Bedeutung. Wir verweisen auf Bauer, MI, für diesen Beweis. \square

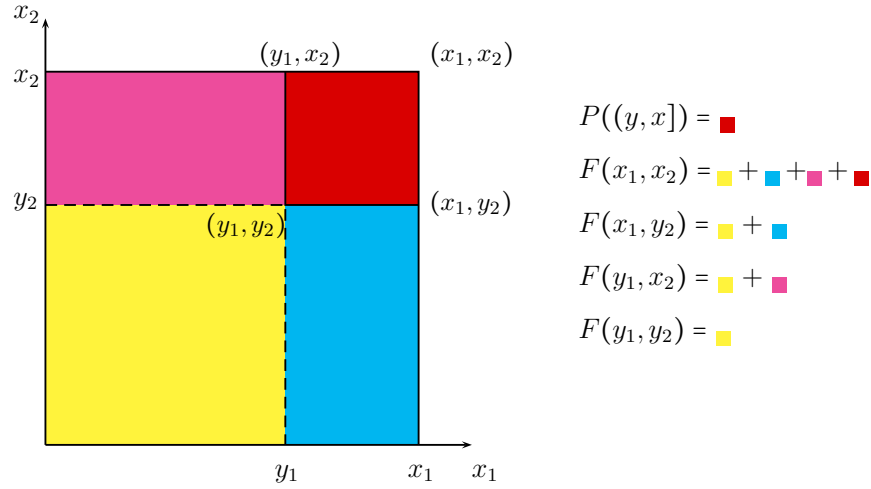


Abbildung 11.2: Definition von Wahrscheinlichkeiten für halboffene Intervalle durch die Verteilungsfunktion auf \mathbb{R}^2 .

Definition 11.6. (i) Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}^d, \mathbb{P}(\mathbb{Z}^d))$ eine **multivariate diskrete Zufallsvariable**. Dann ist die **diskrete Dichte** von X definiert durch

$$p_x^X := P(X = x) := P(\{\omega \in \Omega : X_1(\omega) = x_1, \dots, X_d(\omega) = x_d\}) , \quad (11.17)$$

für $x = (x_1, \dots, x_d) \in \mathbb{Z}^d$.

(ii) Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ eine **multivariate kontinuierliche Zufallsvariable** (d.h. F_X ist stetig auf \mathbb{R}^d) und die folgenden partiellen Ableitungen von F_X existieren (bis auf einer Lebesgue-Nullmenge $\mathcal{N} \subset \Omega$):

$$f_X(x) := \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F_X(x_1, \dots, x_d) , \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d \setminus \mathcal{N}. \quad (11.18)$$

Dann heißt $f_X(x)$ die **kontinuierliche Dichte** von X .

In Analogie zum univariaten Fall (also Dichten auf \mathbb{R} oder \mathbb{Z}) gelten folgende Eigenschaften für die multivariaten diskreten bzw. kontinuierlichen Dichten:

Satz 11.7. (i) Es sei $p_x^X \geq 0$ die diskrete Dichte einer multivariaten diskreten Zufallsvariablen $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{Z}^d, \mathbb{P}(\mathbb{Z}^d))$. Dann gilt für $A \in \mathbb{P}(\mathbb{Z}^d)$

$$X_* P(A) = \sum_{x \in A} p_x^X \quad (11.19)$$

und insbesondere

$$\sum_{x \in \mathbb{Z}^d} p_x^X = X_* P(\mathbb{Z}^d) = P(\Omega) = 1 . \quad (11.20)$$

11. Multivariate Verteilungen

Also ist die Verteilung von X durch ihre diskrete Dichte eindeutig bestimmt. Des Weiteren existiert zu jeder Familie nichtnegativer Zahlen p_x , $x \in \mathbb{Z}^d$, mit $\sum_{x \in \mathbb{Z}^d} p_x = 1$ genau ein Wahrscheinlichkeitsmaß P auf $(\mathbb{Z}^d, \mathbb{P}(\mathbb{Z}^d))$, so dass $P(\{x\}) = p_x$ für alle $x \in \mathbb{Z}^d$.

- (ii) Es sei $f_X(x)$ die kontinuierliche Dichte einer multivariaten kontinuierlichen Zufallsvariablen $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Dann gilt $f_X(x) \geq 0$ P -fast überall und für $A \in \mathcal{B}(\mathbb{R}^d)$

$$X_*P(A) = \int_A f_X(x) dx \quad (11.21)$$

und insbesondere

$$\int_{\mathbb{R}^d} f_X(x) dx = X_*P(\mathbb{R}^d) = P(\Omega) = 1. \quad (11.22)$$

Also bestimmt die kontinuierliche Dichte die Verteilung von X eindeutig.

- (iii) Ist des Weiteren eine messbare Funktion $f(x) \geq 0$ Lebesgue fast überall gegeben mit $\int_{\mathbb{R}^d} f(x) dx = 1$, dann existiert genau ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, so dass $f(x) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} P((0, x])$ Lebesgue fast überall.

Beweis. (i) \mathbb{Z}^d ist abzählbar, d.h. es existiert eine bijektive Abbildung $\phi : \mathbb{Z}^d \rightarrow \mathbb{Z}$. Diese Abbildung und ihre Inverse sind trivialerweise messbar bezüglich den σ -Algebren $\mathbb{P}(\mathbb{Z}^d)$ und $\mathbb{P}(\mathbb{Z})$. Also können wir jede Aussage über den messbaren Raum $(\mathbb{Z}^d, \mathbb{P}(\mathbb{Z}^d))$ äquivalent als Aussage über $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ formulieren. Die Charakterisierung von Wahrscheinlichkeitsmaßen über ihre diskreten Dichten ist deshalb schon in Satz 6.3 erfolgt.

(ii) Dass die kontinuierliche Dichte $f_X(x)$ bereits die Verteilung X_*P bestimmt ergibt sich aus dem Hauptsatz der Differential- und Integralrechnung

$$F_X(x) = \int_{(-\infty, x]} f_X(y) dy = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_X(y_1, \dots, y_d) dy_d \dots dy_1 \quad (11.23)$$

und Satz 11.5. Dass $f_X(x)$ keine negativen Werte annehmen kann folgt aus Gleichung (11.21).

(iii) Indem wir die obige Gleichung (11.23) als Definition einer Funktion $F(x)$ verwenden, zeigt man mit Hilfe der Isotonie des Lebesgue-Integrals auf \mathbb{R}^d , dass $F(x)$ monoton ist. Positivität (11.8) von $F(x)$ folgt aus $\int_{(y, x]} f(x) dx \geq 0$. Stetigkeit von $F(x)$ und $\lim_{x \rightarrow \infty} F(x) = 1$ sowie $\lim_{x_j \rightarrow -\infty} F(x_1, \dots, x_j, \dots, x_d) = 0$ zeigt man mit dem Satz von Lebesgue angewandt auf das Maß dx mit dominierender Funktion $f(x)$. Die Behauptung folgt dann aus Satz 11.5. \square

Im Folgenden verwenden wir folgende Konventionen: Für $x, y \in \mathbb{R}^d$ sei $|x| = \sqrt{\langle x, x \rangle}$ die Euklidische Norm von x und $\langle x, y \rangle = \sum_{j=1}^d x_j y_j$ das Euklidische Skalarprodukt. Mit $|C|$ bezeichnen wir die Determinante einer Matrix C und mit C^{-1} die inverse Matrix. Ebenfalls

benötigen wir für das nun folgende Beispiel den Begriff der strikt positiv definiten Matrix, den wir aus der linearen Algebra wiederholen: Eine symmetrische Matrix $C \in \text{Mat}_{d \times d}(\mathbb{R})$ heißt positiv definit, falls $\forall x \in \mathbb{R}^d$ gilt, dass $\langle x, Cx \rangle \geq 0$. Folgt zudem aus $\langle x, Cx \rangle = 0$, dass $x = 0$ ist, dann heißt C strikt positiv definit. Positiv definite Matrizen kann man diagonalisieren, d.h. $\exists O, D \in \text{Mat}_{d \times d}(\mathbb{R})$ mit $C = ODO'$, wobei O Orthogonalmatrix ist ($O' = O^{-1}$) und D Diagonalmatrix, d.h.

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \lambda_d \end{pmatrix}. \quad (11.24)$$

Ist C strikt positiv definit, so gilt $\lambda_j > 0$, $j = 1, \dots, d$, und $C^{-\frac{1}{2}} := OD^{-\frac{1}{2}}O'$ existiert, wobei $D^{-\frac{1}{2}}$ die Diagonalmatrix mit $\frac{1}{\sqrt{\lambda_j}}$ im j -ten Diagonaleintrag ist.

Definition 11.8 (Multivariate Normalverteilung). Es sei $C \in \text{Mat}_{d \times d}(\mathbb{R})$ eine strikt positiv definite Matrix und $\mu \in \mathbb{R}^d$. Dann heißt

$$f(x) := f(x|\mu, C) := \frac{1}{(2\pi)^{\frac{d}{2}} |C|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \langle x - \mu, C^{-1}(x - \mu) \rangle \right\} \quad (11.25)$$

die kontinuierliche Dichte der **multivariaten Normalverteilung** zu den ‘Parametern’ μ, C . Für $C = \mathbb{I}_d$, der Einheitsmatrix in $\text{Mat}_{d \times d}$, und $\mu = 0$ sprechen wir von der kontinuierlichen Dichte der **multivariaten Standardnormalverteilung**.

Das Aussehen des Graphen von $f(x)$ für $d = 2$ gleicht dem einer ‘Glocke’, siehe Abbildung 11.3. Abweichungen von C^{-1} von der Einheitsmatrix strecken oder stauchen die Achsen um einen Faktor $\sqrt{\lambda_j}$ und verdrehen ‘diese’ anschließend mittels der orthogonalen Matrix O . Wir müssen nun noch zeigen, dass $f(x|\mu, C)$ tatsächlich eine multivariate kontinuierliche Dichte ist.

Satz 11.9. Für C, μ wie in Definition 11.8 ist $f(x|\mu, C)$ eine kontinuierliche Dichte im Sinne von Satz 11.7. Insbesondere existiert ein eindeutiges Wahrscheinlichkeitsmaß $N(\mu, C)$ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, so dass $f(x|\mu, C)$ die zugehörige kontinuierliche Dichte ist. $N(\mu, C)$ heißt multivariate Normalverteilung zu den Parametern C und μ . Insbesondere heißt $N(0, \mathbb{I}_d)$ die Standardnormalverteilung auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Beweis. Die Bedingung $f(x|\mu, C) \geq 0$ ist offensichtlich. Zu zeigen ist demnach nur die

11. Multivariate Verteilungen

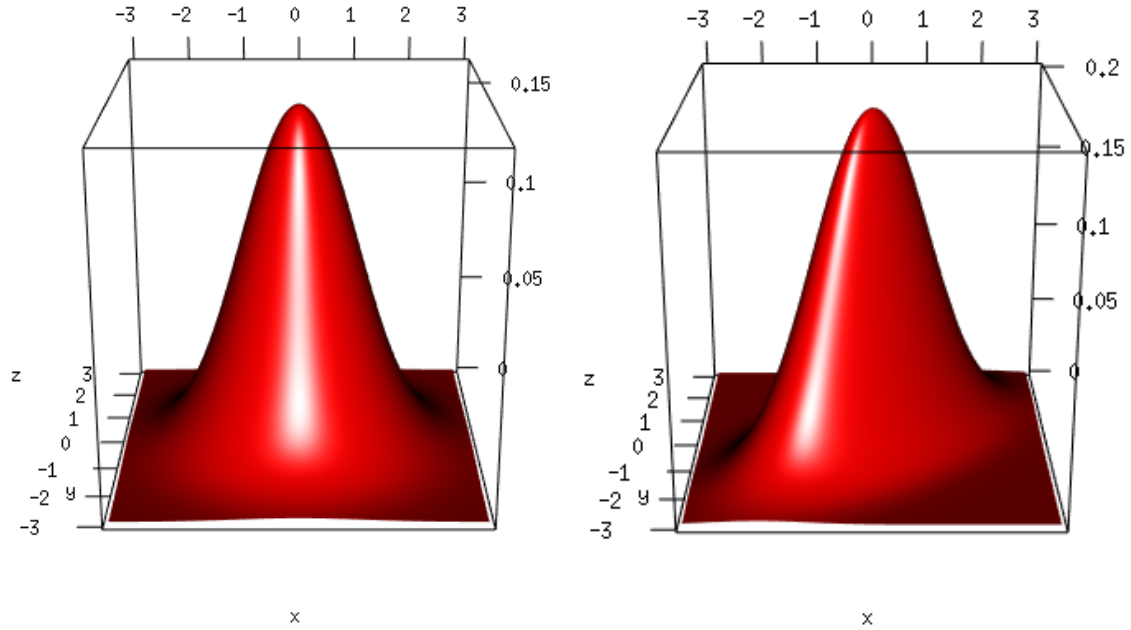


Abbildung 11.3: Links: Kontinuierliche Dichte der Standardnormalverteilung in zwei Dimensionen. Rechts: Kontinuierliche Dichte einer nicht-Standard-Normalverteilung.

Normiertheit

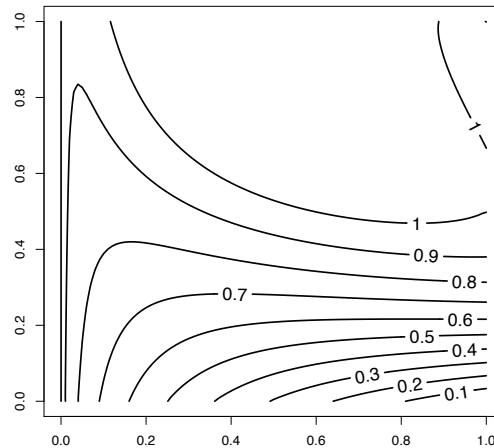
$$\begin{aligned}
 \int_{\mathbb{R}^d} f(x|\mu, C) dx &= \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{\frac{d}{2}} |C|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \langle x - \mu, C^{-1}(x - \mu) \rangle \right\} dx \\
 &= \frac{1}{(2\pi)^{\frac{d}{2}} |C|^{\frac{1}{2}}} \int_{\mathbb{R}^d} \exp \left\{ -\frac{1}{2} \langle C^{-\frac{1}{2}}(x - \mu), C^{-\frac{1}{2}}(x - \mu) \rangle \right\} dx \\
 &\stackrel{y=C^{-\frac{1}{2}}(x-\mu)}{=} \frac{1}{(2\pi)^{\frac{d}{2}} |C|^{\frac{1}{2}}} \int_{\mathbb{R}^d} \exp \left\{ -\frac{1}{2} \langle y, y \rangle \right\} |C^{\frac{1}{2}}| dy \\
 &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{j=1}^d e^{-\frac{1}{2} y_j^2} dy_1 \cdots dy_d \\
 &= \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2} y^2} dy \right)^d = (1)^d = 1 .
 \end{aligned} \tag{11.26}$$

Hierbei verwenden wir in der dritten Identität den Transformationssatz für d -dimensionale Lebesgue-Integrale und in der darauf folgenden Gleichung die Identität $|C^{\frac{1}{2}}| = |C|^{\frac{1}{2}}$ aus der linearen Algebra. \square

11.1. Übungsaufgaben

1. Selbsttest

1. Betrachten Sie die Höhenlinien der Funktion $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ in der unten stehenden Grafik. Kann die hier dargestellte Funktion F eine Verteilungsfunktion auf \mathbb{R}^2 sein? Geben Sie eine kurze Begründung an!



2. Ist die Funktion

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto F(x, y) = \begin{cases} 1 - e^{-xy} & , \text{ für } x > 0, y > 0, \\ 0 & , \text{ sonst,} \end{cases}$$

eine multivariate Verteilungsfunktion?

Hinweis: Falls Sie nicht weiterkommen lohnt es sich die multivariate kontinuierliche Dichte zu berechnen!

3. Wir betrachten die Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto f(x, y) = \begin{cases} N e^{-\alpha x - \beta y} & , \text{ für } x > 0, y > 0, \\ 0 & , \text{ sonst,} \end{cases}$$

wobei $\alpha, \beta > 0$ Parameter sind und $N \in \mathbb{R}$ eine Normierungskonstante.

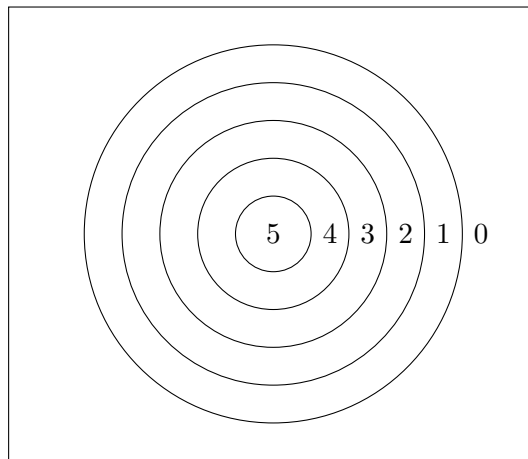
- a) Für welchen Wert von N ist f eine multivariate kontinuierliche Dichte?
 - b) Bestimmen Sie die multivariate Verteilungsfunktion die zu f korrespondiert!
 - c) Wenn Sie das Ergebnis aus b) mit der Verteilungsfunktion der (univariaten) Exponentialverteilung vergleichen, was erkennen Sie?
4. Wir betrachten eine $N(\mu, C)$ -verteilte \mathbb{R}^d -wertige Zufallsvariable $X = (X_1, \dots, X_d)$, wobei $\mu \in \mathbb{R}^d$ und C eine strikt positiv definite $d \times d$ -Matrix. Die **Randverteilung** von $X' = (X_1, \dots, X_{d-1})$ ist definiert durch $F_{X'}(x_1, \dots, x_{d-1}) = \lim_{x_d \rightarrow \infty} F_X(x_1, \dots, x_d)$.

11. Multivariate Verteilungen

- a) Interpretieren Sie was die Randverteilung von X' beschreibt!
- b) Bestimmen Sie die multivariate kontinuierliche Dichte der Randverteilung von X' in unserem Fall $X \sim N(\mu, C)$!
- c) Was passiert wenn Sie den Schritt zur Randverteilung iterieren, d.h. die Randverteilung der Randverteilung der Randverteilung ... bestimmen?

2. Transfer

Ein Bogenschütze schießt Pfeile auf eine Zielscheibe mit Radius 50cm. Die Zielscheibe ist in 5 Ringe mit je 10cm Durchmesser eingeteilt: Für einen Treffer im innersten Ring gibt es 5 Punkte, im nächsten Ring 4, dann 3, 2, 1 und schließlich gibt es keine Punkte wenn er die Zielscheibe verfehlt. Die folgende Graphik visualisiert die Anordnung der Zielscheibe



Natürlich versucht der Bogenschütze immer genau in die Mitte zu treffen, was ihm aber nicht immer gelingt. Vielmehr ist der Auftreffpunkt des Pfeiles eine \mathbb{R}^2 -wertige Zufallsvariable (X, Y) , wobei X die horizontale Koordinate und Y die vertikale Koordinate in einem Koordinatensystem durch den Mittelpunkt der Scheibe darstellt. Wir wollen im Folgenden annehmen, dass die multivariate Verteilung von (X, Y) einer multivariaten Normalverteilung $N(\mu, C)$ entspricht, wobei $\mu = (0, 0)$ und

$$C = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

mit $\sigma = 20\text{cm}$.

- a) Geben Sie die multivariate kontinuierliche Dichte von (X, Y) in Polarkoordinaten (r, ϕ) an!
- b) Berechnen Sie damit die Wahrscheinlichkeit, dass der Pfeil in einem Bereich mit n Punkten auftrifft, für alle $n = 0, \dots, 5$.

11.1. Übungsaufgaben

- c) Wie groß ist die Wahrscheinlichkeit, dass der Bogenschütze mit einem Schuss 2 oder weniger Punkte erzielt?
- d) Bestimmen Sie die mittlere Punktzahl die der Schütze durch einen Schuss erzielt!

11. Multivariate Verteilungen

11.2. Literatur

12. Unabhängige und unkorrelierte Zufallsvariablen

Wir hatten bereits in Kapitel 5 den Begriff der Unabhängigkeit von Ereignissen bzw. von Ereignisfamilien kennengelernt. Diese Begriffsbildung möchten wir nun auf Zufallsvariablen übertragen. Dafür weisen wir jeder Zufallsvariable ein (minimales) Mengensystem zu.

Definition 12.1. Es sei $X : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{B})$ eine E -wertige Zufallsvariable (für uns ist E meistens eine der Mengen \mathbb{Z} , \mathbb{Z}^d , \mathbb{R} oder \mathbb{R}^d und \mathcal{B} die kanonische σ -Algebra). Die **von X erzeugte σ -Algebra** $\sigma(X)$ ist definiert als die kleinste unter- σ -Algebra von \mathcal{A} , bezüglich derer X als Abbildung $X : (\Omega, \sigma(X)) \rightarrow (E, \mathcal{B})$ messbar ist.

Die Bedeutung der σ -Algebra $\sigma(X)$ ist die, dass sie alle Ereignisse enthält, deren Eintreten bzw. nicht-Eintreten allein durch eine X -Beobachtung festgestellt werden kann. Wir können diese σ -Algebra in gewisser Weise mit der Information identifizieren, die wir erhalten können, wenn wir nur X beobachten. Diese Einsicht wird durch das folgende Lemma präzisiert.

Lemma 12.2. Die σ -Algebra $\sigma(X)$ besteht aus den Urbildern von allen Mengen aus \mathcal{B} unter X . In Formeln,

$$\sigma(X) = X^{-1}(\mathcal{B}) = \{X^{-1}(B) : B \in \mathcal{B}\} , \quad (12.1)$$

wobei $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ das Urbild von B unter X ist.

Beweis. Aufgrund der Messbarkeit von X liegt das Mengensystem auf der rechten Seite von (12.1) in \mathcal{A} . Offensichtlich ist es minimal für die Messbarkeit von X , denn es enthält ja gerade alle X -Urbilder von Mengen aus \mathcal{B} und sonst nichts.

Zu zeigen ist demnach, dass $X^{-1}(\mathcal{B})$ eine σ -Algebra ist. Wählen wir $E \in \mathcal{B}$, so ist $X^{-1}(E) = \Omega \in X^{-1}(\mathcal{B})$. Für jedes $A \in X^{-1}(\mathcal{B})$ existiert ein $B \in \mathcal{B}$, so dass $A = X^{-1}(B)$. Damit ist $B^c \in \mathcal{B}$ und deshalb $A^c = X^{-1}(B^c) \in X^{-1}(\mathcal{B})$. Seien schließlich $A_n \in X^{-1}(\mathcal{B})$, d.h. es existieren $B_n \in \mathcal{B}$ mit $A_n = X^{-1}(B_n)$. Dann gilt $\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{B}$ und deshalb $X^{-1}(\bigcup_{n \in \mathbb{N}} B_n) = \bigcup_{n \in \mathbb{N}} X^{-1}(B_n) = \bigcup_{n \in \mathbb{N}} A_n \in X^{-1}(\mathcal{B})$. \square

Wir führen die Unabhängigkeit von (Familien von) Zufallsvariablen auf die zugehörigen (Familien von) erzeugten σ -Algebren zurück. Ausgehend von Definition 5.1 motiviert dies die folgende

Definition 12.3 (Unabhängigkeit von Zufallsvariablen).

12. Unabhängige und unkorrelierte Zufallsvariablen

- (i) Gegeben seien zwei Zufallsvariablen $X_j : (\Omega, \mathcal{A}, P) \rightarrow (E_j, \mathcal{B}_j)$, $j = 1, 2$. Dann heißen X_1 und X_2 **unabhängig** voneinander, falls für alle $A \in \mathcal{B}_1$ und $B \in \mathcal{B}_2$ gilt

$$P(\{X_1 \in A\} \cap \{X_2 \in B\}) = P(\{X_1 \in A\}) P(\{X_2 \in B\}) . \quad (12.2)$$

- (ii) Ist allgemeiner J eine beliebige Indexmenge und $\{X_j : (\Omega, \mathcal{A}, P) \rightarrow (E_j, \mathcal{B}_j)\}_{j \in J}$, eine Familie von Zufallsvariablen, dann heißt diese Familie **unabhängig**, falls für alle $A_j \in \mathcal{B}_j$ und $I \subseteq J$ endlich gilt

$$P\left(\bigcap_{j \in I} \{X_j \in A_j\}\right) = \prod_{j \in I} P(\{X_j \in A_j\}) . \quad (12.3)$$

- (iii) Es sei $\{X_j : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{B})\}_{j \in J}$ eine Familie von Zufallsvariablen mit identischem Ergebnisraum (E, \mathcal{B}) und es gelte $X_{i*}P = X_{j*}P$ für alle $i, j \in J$. Dann heißt die Familie $\{X_j\}_{j \in J}$ **identisch verteilt**.
- (iv) Die Familie von Zufallsvariablen $\{X_j\}_{j \in J}$ heißt **unabhängig identisch verteilt** (englische Abkürzung i.i.d. für *independently identically distributed*), falls sie sowohl unabhängig als auch identisch verteilt ist.

In Bemerkung 11.2 hatten wir bereits die Tensor σ -Algebra diskutiert. Eine analoge Begriffsbildung existiert auch für Wahrscheinlichkeitsmaße. Motivierend ist hierbei die Faktorisierungseigenschaft (12.2) bzw. (12.3). Um den Notationsaufwand in Grenzen zu halten, beschränken wir uns auf endliche Tensorprodukte von Wahrscheinlichkeitsmaßen.

Definition 12.4. Es seien $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, \dots, d$, Wahrscheinlichkeitsräume

- (i) Wir bezeichnen mit $\mathcal{Z}(\Omega) := \{A_1 \times \dots \times A_d : A_i \in \mathcal{A}_i, i = 1, \dots, d\}$ das Mengensystem der **Zylindermengen** über $\Omega := \times_{i=1}^d \Omega_i$. Dann ist die **Tensor σ -Algebra** $\otimes_{i=1}^d \mathcal{A}_i$ definiert als die von den Zylindermengen erzeugte σ -Algebra, d.h.

$$\bigotimes_{i=1}^d \mathcal{A}_i := \sigma(\mathcal{Z}(\Omega)) . \quad (12.4)$$

- (ii) Das **Produktmaß** $\otimes_{i=1}^d P_i$ der Wahrscheinlichkeitsmaße P_i ist definiert als das (eindeutig bestimmte) Wahrscheinlichkeitsmaß auf $(\Omega, \otimes_{i=1}^d \mathcal{A}_i)$, so dass für $A = A_1 \times \dots \times A_d \in \mathcal{Z}(\Omega)$ gilt

$$\left[\bigotimes_{i=1}^d P_i \right] (A) = \prod_{i=1}^d P_i(A_i) . \quad (12.5)$$

Bemerkung 12.5. Bei dieser Definition wurde die Existenz des Produktmaßes vorausgesetzt. Einen Beweis für die Existenz von Produktmaßen findet man z.B. in Bauer, MI, Satz 23.9. Die Eindeutigkeit ist wieder eine Folge der \cap -Stabilität von $\mathcal{Z}(\Omega)$. Existenz und Eindeutigkeit von Produktmaßen gilt jedoch nicht nur für endliche Tensorprodukte, sondern auch für Tensorprodukte mit beliebiger Indexmenge I , siehe etwa Bauer, Wahrscheinlichkeitstheorie (4. Auflage), de Gruiter 1990, Satz 9.2.

Satz 12.6. Es sei $X_i : (\Omega, \mathcal{A}, P) \rightarrow (E_i, \mathcal{B}_i)$, $i = 1, \dots, d$, eine Familie von Zufallsvariablen. Wir setzen $X := (X_1, \dots, X_d) : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{B})$, wobei $E := \times_{i=1}^n E_i$ und $\mathcal{B} := \otimes_{i=1}^d \mathcal{B}_i$. Dann sind folgende Aussagen äquivalent:

(i) Die Familie $\{X_i\}_{i=1, \dots, d}$ ist unabhängig.

(ii) Die Verteilung von X ist gleich dem Produktmaß der Verteilungen der X_i , d.h.

$$X_* P = \bigotimes_{i=1}^d X_{i*} P . \quad (12.6)$$

(iii) Für beliebige $\phi_i : (E_i, \mathcal{B}_i) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ messbar und beschränkt für alle $i = 1, \dots, d$ gilt

$$\mathbb{E} \left[\prod_{i=1}^d \phi_i(X_i) \right] = \prod_{i=1}^d \mathbb{E}[\phi_i(X_i)] . \quad (12.7)$$

Beweis. (i) \Leftrightarrow (ii): Es sei $A = A_1 \times \dots \times A_d \in \mathcal{Z}(E)$. Dann gilt

$$\{X \in A\} = \bigcap_{i=1}^d \{X_i \in A_i\} . \quad (12.8)$$

Gilt (i), so folgt

$$\begin{aligned} X_* P(A) &= P(\{X \in A\}) \\ &= P\left(\bigcap_{i=1}^d \{X_i \in A_i\}\right) \\ &= \prod_{i=1}^d P(X_i \in A_i) = \left[\bigotimes_{i=1}^d X_{i*} P \right](A) . \end{aligned} \quad (12.9)$$

Wegen der \cap -Stabilität von $\mathcal{Z}(E)$ folgt nun (ii). Gilt umgekehrt (ii), so stellen wir die obige Gleichungskette folgendermaßen um

$$\begin{aligned} P\left(\bigcap_{i=1}^d \{X_i \in A_i\}\right) &= P(\{X \in A\}) \\ &= X_* P(A) \\ &= \left[\bigotimes_{i=1}^d X_{i*} P \right](A) = \prod_{i=1}^d P(\{X_i \in A_i\}) . \end{aligned} \quad (12.10)$$

(ii) \Leftrightarrow (iii): Wegen der Linearität des Erwartungswertes und

$$\prod_{i=1}^d \phi_i(X_i) = \sum_{J \subseteq \{1, \dots, d\}} (-1)^{d-\#J} \prod_{i=1}^d \phi_i^J(X_i) \quad (12.11)$$

12. Unabhängige und unkorrelierte Zufallsvariablen

mit

$$\phi_i^J(X_i) = \begin{cases} \phi^+(X_i) & , \text{ für } i \in J , \\ \phi^-(X_i) & , \text{ für } i \notin J , \end{cases} \quad (12.12)$$

können wir ohne Beschränkung der Allgemeinheit annehmen, dass ϕ_i nicht-negativ ist. Denn gilt (iii) für solche Funktionen, dann gilt wegen

$$\prod_{i=1}^d \mathbb{E}[\phi_i(X_i)] = \sum_{J \subset \{1, \dots, d\}} (-1)^{d-\#J} \prod_{i=1}^d \mathbb{E}[\phi_i^J(X_i)] \quad (12.13)$$

(iii) auch für beliebige ϕ_i .

Es sei nun $\chi_{i,n}$ eine von unten approximierende Sequenz von nicht-negativen Elementarfunktionen für ϕ_i auf (E_i, \mathcal{B}_i) . Dann ist $\chi_n = \bigotimes_{i=1}^d \chi_{i,n}$ eine von unten approximierende Sequenz von Elementarfunktionen, wobei das Tensorprodukt von Funktionen definiert ist als

$$\chi_n(e) = \left[\bigotimes_{i=1}^d \chi_{i,n} \right] (e) = \prod_{i=1}^d \chi_{i,n}(e_i) , \quad e = \bigtimes_{i=1}^n e_i = (e_1, \dots, e_d) \in E . \quad (12.14)$$

Die Messbarkeit dieser Funktion ergibt sich aus der Stetigkeit der d -fachen Produktbildung und daher der Messbarkeit von $\prod : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Daher ist $\chi_n(X)$ eine approximierende Sequenz von elementaren Zufallsvariablen für $\prod_{i=1}^d \phi_i(X_i)$. Gilt (iii) für elementare Zufallsvariablen, so gilt (iii) auch sonst, denn

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^d \phi_i(X_i) \right] &= \lim_{n \rightarrow \infty} \mathbb{E}[\chi_n(X)] \\ &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\prod_{i=1}^d \chi_{i,n}(X_i) \right] \\ &= \lim_{n \rightarrow \infty} \prod_{i=1}^d \mathbb{E}[\chi_{i,n}(X_i)] = \prod_{i=1}^d \mathbb{E}[\phi_i(X_i)] . \end{aligned} \quad (12.15)$$

Es gelte nun (ii) und wir betrachten ohne Beschränkung der Allgemeinheit (positive)

Elementarfunktionen $\phi_i = \sum_{j=1}^{n_i} \alpha_{i,j} 1_{A_{i,j}}$. Es folgt (iii), da

$$\begin{aligned}
\mathbb{E} \left[\prod_{i=1}^d \phi_i(X_i) \right] &= \sum_{j_1, \dots, j_d=1}^{n_1, \dots, n_d} \prod_{i=1}^d \alpha_{i,j_i} \mathbb{E} \left[\prod_{i=1}^d 1_{A_{i,j_i}}(X_i) \right] \\
&= \sum_{j_1, \dots, j_d=1}^{n_1, \dots, n_d} \prod_{i=1}^d \alpha_{i,j_i} \mathbb{E} [1_{A_{1,j_1} \times \dots \times A_{d,j_d}}(X)] \\
&= \sum_{j_1, \dots, j_d=1}^{n_1, \dots, n_d} \prod_{i=1}^d \alpha_{i,j_i} X_* P(A_{1,j_1} \times \dots \times A_{d,j_d}) \\
&= \sum_{j_1, \dots, j_d=1}^{n_1, \dots, n_d} \prod_{i=1}^d \alpha_{i,j_i} \left[\bigotimes_{i=1}^d X_{i*} P \right] (A_{1,j_1} \times \dots \times A_{d,j_d}) \\
&= \sum_{j_1, \dots, j_d=1}^{n_1, \dots, n_d} \prod_{i=1}^d \alpha_{i,j_i} \prod_{i=1}^d X_{i*} P(A_{i,j_i}) \\
&= \prod_{i=1}^d \left[\sum_{j_i=1}^{n_i} \alpha_{i,j_i} X_{i*} P(A_{i,j_i}) \right] = \prod_{i=1}^d \mathbb{E} [\phi_i(X_i)] . \tag{12.16}
\end{aligned}$$

Gilt umgekehrt (iii), so erhalten wir für $A = A_1 \times \dots \times A_d \in \mathcal{Z}(E)$ und $\phi_i = 1_{A_i}$

$$\begin{aligned}
X_* P(A) &= \mathbb{E} [1_A(X)] \\
&= \mathbb{E} \left[\prod_{i=1}^d 1_{A_i}(X_i) \right] \\
&= \prod_{i=1}^d \mathbb{E} [1_{A_i}(X_i)] \\
&= \prod_{i=1}^d X_{i*} P(A_i) \\
&= \left[\bigotimes_{i=1}^d X_{i*} P \right] (A) . \tag{12.17}
\end{aligned}$$

Wieder genügt dies und die \cap -Stabilität von $\mathcal{Z}(E)$, um $X_* P = \bigotimes_{i=1}^d X_{i*} P$, also (ii), zu zeigen. \square

Satz 12.6 charakterisiert die Unabhängigkeit von Zufallsvariablen in einem sehr abstrakten Kontext. Wir wollen nun weitere Schlussfolgerungen ziehen, falls alle X_i \mathbb{R}^q bzw. \mathbb{Z}^q -wertige Zufallsvariablen sind, d.h. $(E_i, \mathcal{B}_i) = (\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q))$ bzw. $(E_i, \mathcal{B}_i) = (\mathbb{Z}^q, \mathcal{P}(\mathbb{Z}^q))$, $i = 1, \dots, d$.

Satz 12.7. *Es seien X_i , $i = 1, \dots, d$, \mathbb{R}^q oder \mathbb{Z}^q -wertige Zufallsvariablen und dementsprechend $X := (X_1, \dots, X_d)$ eine \mathbb{R}^{qd} bzw. \mathbb{Z}^{qd} -wertige Zufallsvariable. Dann gelten die folgenden Charakterisierungen von Unabhängigkeit:*

(i) *Die Familie $\{X_i\}_{i=1, \dots, d}$ ist genau dann unabhängig, wenn*

$$F_X(x) = \prod_{i=1}^d F_{X_i}(x_i) , \quad x = (x_1, \dots, x_d) \in \mathbb{R}^{qd} . \tag{12.18}$$

12. Unabhängige und unkorrelierte Zufallsvariablen

(ii) Sind alle Zufallsvariablen X_i \mathbb{Z}^q -wertig mit diskreten Dichten $p_{x_i}^{X_i}$, dann ist die Familie $\{X_i\}_{i=1,\dots,d}$ unabhängig genau dann wenn

$$p_x^X = \prod_{i=1}^d p_{x_i}^{X_i} \quad , \quad x = (x_1, \dots, x_q) \in \mathbb{Z}^{qd} \quad . \quad (12.19)$$

(iii) Sind alle Zufallsvariablen X_i \mathbb{R}^q -wertig und kontinuierlich verteilt mit (multivariater) kontinuierlicher Dichte $f_{X_i}(x_i)$, dann ist die Familie $\{X_i\}_{i=1,\dots,d}$ unabhängig genau dann wenn

$$f_X(x) = \prod_{i=1}^d f_{X_i}(x_i) \quad , \quad x = (x_1, \dots, x_q) \in \mathbb{R}^{qd} \quad . \quad (12.20)$$

Beweis. (i) Sei zunächst $\{X_i\}_{i=1,\dots,d}$ eine unabhängige Familie. Dann gilt

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\bigcap_{i=1}^d \{X_i \leq x_i\}\right) \\ &= \prod_{i=1}^d P(X_i \leq x_i) = \prod_{i=1}^d F_{X_i}(x_i) \quad . \end{aligned} \quad (12.21)$$

Für die Gegenrichtung bemerken wir zunächst, dass F_X die Verteilungsfunktion von X_*P ist und $\prod_{i=1}^d F_{X_i}(x_i)$ die Verteilungsfunktion von $\bigotimes_{i=1}^d X_{i*}P$. Stimmen beide überein, so gilt auch $X_*P = \bigotimes_{i=1}^d X_{i*}P$, vergleiche Satz 11.5. Dies ist jedoch nach Satz 12.6 (ii) äquivalent zur Unabhängigkeit der X_i .

(ii) Setze wieder zunächst die Unabhängigkeit der Familie $\{X_i\}_{i=1,\dots,d}$ voraus. Dann gilt

$$\begin{aligned} p_x^X &= P(X = x) \\ &= P\left(\bigcap_{i=1}^d \{X_i = x_i\}\right) \\ &= \prod_{i=1}^d P(\{X_i = x_i\}) = \prod_{i=1}^d p_{x_i}^{X_i} \quad . \end{aligned} \quad (12.22)$$

Beachte zur Beweisführung im umgekehrten Fall, dass p_x^X die diskrete Dichte zu X_*P und $\prod_{i=1}^d p_{x_i}^{X_i}$ die diskrete Dichte zu $\bigotimes_{i=1}^d X_{i*}P$ ist. Stimmen diese diskreten Dichten überein, so gilt nach Satz 11.7 (i), dass die Wahrscheinlichkeitsmaße übereinstimmen. Nach Satz 12.6 (ii) ist dies äquivalent zur Unabhängigkeit der X_i .

(iii) Zunächst gilt für eine unabhängige Familie $\{X_i\}_{i=1,\dots,d}$ nach (i)

$$\begin{aligned} f_X(x) &= \frac{\partial^{qd}}{\partial x_{1,1} \cdots \partial x_{1,q} \cdots \partial x_{d,1} \cdots \partial x_{d,q}} F_X((x_{1,1}, \dots, x_{d,q})) \\ &= \prod_{i=1}^d \frac{\partial^q}{\partial x_{i,1} \cdots \partial x_{i,q}} F_{X_i}((x_{i,1}, \dots, x_{i,q})) \\ &= \prod_{i=1}^d f_{X_i}(x_i) . \end{aligned} \quad (12.23)$$

Gilt umgekehrt (12.20) so stimmen nach dem Hauptsatz der Differential- und Integralrechnung $F_X(x)$ und $\prod_{i=1}^d F_{X_i}(x_i)$ überein und deswegen auch $X_*P = \bigotimes_{i=1}^d X_{i*}P$, vergleiche Satz 11.7 (ii). Nach Satz 12.6 (ii) garantiert dies wieder die Unabhängigkeit der X_i . \square

Wir möchten nun eine einfache Kennzahl für den Zusammenhang von zwei Zufallsvariablen entwickeln. Diese Kennzahl setzt die Existenz gemischter 2-ter Momente voraus, daher ist das folgende Lemma hilfreich.

Lemma 12.8. *Es seien X, Y Zufallsvariablen aus $L^2((\Omega, \mathcal{A}, P); \mathbb{R})$. Dann gilt $XY \in L^1((\Omega, \mathcal{A}, P); \mathbb{R})$.*

Beweis. Das Produkt aus zwei Zufallsvariablen ist messbar. Wegen $|XY| \leq \frac{1}{2}(X^2 + Y^2)$ gilt

$$\mathbb{E}[|XY|] \leq \frac{1}{2}(\mathbb{E}[X^2] + \mathbb{E}[Y^2]) < \infty . \quad (12.24)$$

\square

Definition 12.9. Es seien $X, Y \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$.

(i) Die **Kovarianz** von X und Y ist definiert als

$$\text{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] . \quad (12.25)$$

(ii) Haben zudem X und Y eine nicht verschwindende Standardabweichung, d.h. $\sigma[X] > 0$ und $\sigma[Y] > 0$, so definieren wir den **Korrelationskoeffizient** von X und Y als

$$\text{Cor}[X, Y] := \frac{\text{Cov}[X, Y]}{\sigma[X]\sigma[Y]} . \quad (12.26)$$

(iii) X und Y heißen **unkorreliert**, falls $\text{Cov}[X, Y] = 0$ gilt.

Die Kovarianz misst die (im Mittel) gleich- oder gegenläufige Schwankung der Zufallsvariablen X und Y . So ergibt sich ein positiver Beitrag zur Kovarianz, falls X und Y ihren Erwartungswert gleichzeitig über- bzw. unterschreiten (gleichläufige Schwankung). Überschreitet jedoch X seinen Erwartungswert, während Y den seinigen unterschreitet,

12. Unabhängige und unkorrelierte Zufallsvariablen

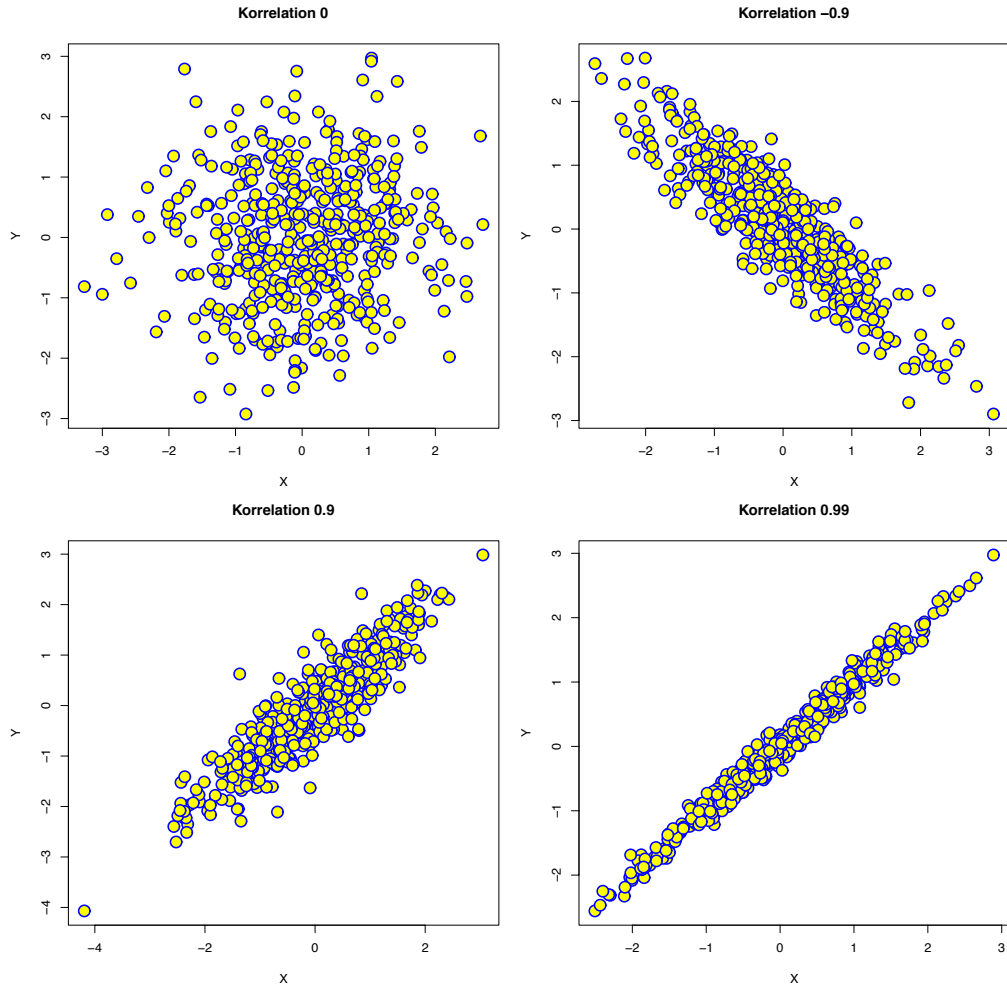


Abbildung 12.1: (Pseudo-) Realisierung von korrelierten Zufallsvariablen X und Y mit unterschiedlichen Korrelationskoeffizienten.

oder umgekehrt, (gegenläufige Schwankung), dann ergibt sich ein negativer Beitrag zur Kovarianz.

Der Vorteil des Korrelationskoeffizienten gegenüber der Kovarianz liegt darin, dass dies eine dimensionslose Größe ist und man daher die Stärke der Korrelation auch zwischen grundsätzlich verschiedenen Situationen miteinander vergleichen kann. Abbildung 12.1 vermittelt ein Gefühl für die Stärke der Korrelation aufgrund von Realisierungen einer bivariaten Zufallsvariablen (X, Y) . Des Weiteren liegt der Wert von $\text{Cor}[X, Y]$ stets zwischen -1 und 1 .

Lemma 12.10. *Es seien $X, Y \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$ mit $\sigma[X] > 0$ und $\sigma[Y] > 0$. Dann gilt*

$$-1 \leq \text{Cor}[X, Y] \leq 1. \quad (12.27)$$

Ist $\text{Cor}[X, Y] = \pm 1$, so gilt $X = \pm bY + a$ P -fast sicher mit $b > 0$ und $a \in \mathbb{R}$.

Beweis. Für X, Y wie oben ist nach Lemma 12.8

$$\langle X, Y \rangle = \mathbb{E}[XY] \quad (12.28)$$

ein positiv definites inneres Produkt (hier müssen wir wieder Zufallsvariablen, die P -fast sicher übereinstimmen sind, identifizieren, sonst ist das innere Produkt nur semidefinit, vergleiche auch Bemerkung 9.7). Deshalb gilt nach der Cauchy-Schwarz Ungleichung

$$\begin{aligned} |\text{Cor}[X, Y]| \sigma[X] \sigma[Y] &= |\text{Cov}[X, Y]| \\ &= |\langle X - \mathbb{E}[X], Y - \mathbb{E}[Y] \rangle| \\ &\leq \|X - \mathbb{E}[X]\|_2 \|Y - \mathbb{E}[Y]\|_2 = \sigma[X] \sigma[Y] \end{aligned} \quad (12.29)$$

mit Gleichheit genau dann, wenn $X - \mathbb{E}[X] = c(Y - \mathbb{E}[Y])$ P -fast sicher für ein $c \in \mathbb{R} \setminus \{0\}$. Ist $c > 0$, so ist

$$\text{Cov}[X, Y] = c \|Y - \mathbb{E}[Y]\|_2^2 \geq 0 \quad (12.30)$$

und daher $\text{Cor}[X, Y] = 1$, ist $c < 0$ so sieht man analog, dass $\text{Cor}[X, Y] = -1$. Setze nun $b = |c|$ und $a = -c \mathbb{E}[Y] + \mathbb{E}[X]$. \square

Die Unabhängigkeit von zwei Zufallsvariablen impliziert das Verschwinden der Kovarianz bzw. der Korrelation.

Lemma 12.11. Sind $X, Y \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$ unabhängig, dann sind X und Y unkorreliert, d.h.

$$\text{Cov}[X, Y] = 0 \quad \text{und} \quad \text{Cor}[X, Y] = 0 \quad \text{falls} \quad \sigma[X] > 0, \quad \sigma[Y] > 0. \quad (12.31)$$

Beweis. Nach Satz 12.6 und dem Satz von Fubini (Bauer, MI, Korollar 23.7) gilt

$$\begin{aligned} \text{Cov}[X, Y] &= \int_{\mathbb{R}^2} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) d(X_* P \otimes Y_* P)(x, y) \\ &= \int_{\mathbb{R}} (x - \mathbb{E}[X]) dX_* P(x) \int_{\mathbb{R}} (y - \mathbb{E}[Y]) dY_* P(y) \\ &= (\mathbb{E}[X] - \mathbb{E}[X]) (\mathbb{E}[Y] - \mathbb{E}[Y]) = 0. \end{aligned} \quad (12.32)$$

\square

Bemerkung 12.12. Die Umkehrung von Lemma 12.11 gilt nicht, d.h. eine verschwindende Korrelation von zwei Zufallsvariablen X, Y impliziert **nicht** deren Unabhängigkeit. Dies ist am einfachsten zu verstehen, wenn wir als Spezialfall nur Zufallsvariablen X, Y mit endlicher Bildmenge $X(\Omega) := \{X(\omega) : \omega \in \Omega\} = \{x_1, \dots, x_n\} \subset \mathbb{R}$ bzw. $Y(\Omega) = \{y_1, \dots, y_m\} \subset \mathbb{R}$ betrachten. Die multivariate Verteilung von (X, Y) ist somit durch eine diskrete Dichte $p_{(x,y)}^{(X,Y)}$ mit endlich vielen freien Parametern bestimmt (explizit: $n \times m$ Parameter 'minus' eine Nebenbedingung für die Normiertheit der diskreten, bivariaten Dichte). Verlangt man $\text{Cov}[X, Y] = 0$, so erhält man eine Nebenbedingung, welche die Anzahl der Parameter um

12. Unabhängige und unkorrelierte Zufallsvariablen

1 reduziert. Die Unabhängigkeit von X und Y , siehe (12.2) und (12.19), liefert jedoch $(n-1) \times (m-1)$ viele Nebenbedingungen (betrachten Sie $A = \{x_i\}$ und $B = \{y_j\}$, für $i = 1, \dots, n-1$ und $j = 1, \dots, m-1$). Für $n > 2$ oder $m > 2$ sind dies mehr Bedingungen als die eine Nebenbedingung für Unkorreliertheit, also kann man Beispiele konstruieren, so dass X, Y unkorreliert ist, aber nicht unabhängig.

Wir sind nun in der Lage die Varianzadditionsgesetze aufzustellen, die bei den Grenzwertsätzen im folgenden Kapitel eine entscheidende Rolle spielen.

Satz 12.13. *Es seien $X_1, \dots, X_n \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$.*

(i) *Es gilt*

$$\text{Var}[X_1 + \dots + X_n] = \sum_{i,j=1}^n \text{Cov}[X_i, X_j] = \sum_{i=1}^n \text{Var}[X_i] + 2 \sum_{\substack{i,j=1 \\ i < j}}^n \text{Cov}[X_i, X_j] . \quad (12.33)$$

(ii) *Sind die Zufallsvariablen X_i paarweise unkorreliert, dann gilt*

$$\text{Var}[X_1 + \dots + X_n] = \sum_{i=1}^n \text{Var}[X_i] . \quad (12.34)$$

(iii) *Sind die Zufallsvariablen X_i paarweise unkorreliert mit derselben Varianz, so gilt*

$$\text{Var}[X_1 + \dots + X_n] = n \text{Var}[X_1] . \quad (12.35)$$

Dies trifft insbesondere zu, falls X_i i.i.d. sind.

Beweis. (i) Wir erhalten durch eine direkte Rechnung

$$\begin{aligned} \text{Var}[X_1 + \dots + X_n] &= \mathbb{E}[(X_1 + \dots + X_n - \mathbb{E}[X_1 + \dots + X_n])^2] \\ &= \mathbb{E}\left[\sum_{i,j=1}^n (X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] \\ &= \sum_{i,j=1}^n \text{Cov}[X_i, X_j] . \end{aligned} \quad (12.36)$$

Wir bemerken außerdem, dass $\text{Cov}[X_i, X_i] = \text{Var}[X_i]$ und erhalten dadurch den Ausdruck auf der rechten Seite von (12.33).

(ii) folgt aus (i) und (iii) folgt unmittelbar aus (ii). □

12.1. Übungsaufgaben

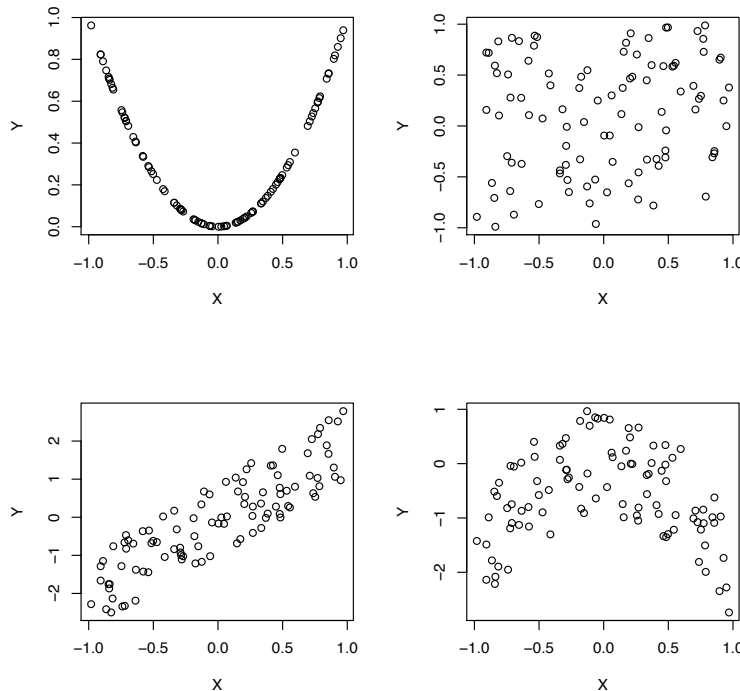
1. Selbsttest

1. Gegeben seien zwei unabhängige Zufallsvariablen $X, Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und zwei messbare Abbildungen $\phi, \psi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sind damit die Zufallsvariablen $X' := \phi \circ X$ und $Y' := \psi \circ Y$ ebenfalls unabhängig?
2. In Lemma 12.11 haben wir gezeigt, dass aus der Unabhängigkeit von zwei Zufallsvariablen $X, Y \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$ deren Unkorreliertheit folgt. Die Umkehrung gilt jedoch im Allgemeinen *nicht*, d.h. es gibt Zufallsvariablen $X, Y \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$ welche unkorreliert aber nicht unabhängig sind. Finden Sie ein Beispiel für diesen Fall!
3. Es sei $X = (X_1, \dots, X_d)$ eine multivariat normalverteilte Zufallsvariable, d.h. $X \sim N(\mu, C)$. Zeigen Sie, dass $\mathbb{E}[X] = \mu$ und

$$\left(\text{Cov}(X_i, X_j) \right)_{i,j=1,\dots,d} = C.$$

Dies rechtfertigt die häufig verwendete Bezeichnung von C als *Kovarianzmatrix*.

4. Gegeben seien zwei Zufallsvariablen $X, Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Erwartungswert $\mathbb{E}[X] = \mathbb{E}[Y] = 2$, Varianz $\text{Var}[X] = \text{Var}[Y] = 1$ und Kovarianz $\text{Cov}[X, Y] = -\frac{1}{2}$. Berechnen Sie den Erwartungswert und die Varianz der Zufallsvariablen $Z = X + Y$!
5. Die untenstehenden Graphiken zeigen jeweils 100 Realisierungen einer bivariaten Zufallsvariablen (X, Y) . Geben Sie eine Abschätzung für die Korrelationskoeffizienten $\text{Cor}(X, Y)$!



12. Unabhängige und unkorrelierte Zufallsvariablen

2. Transfer

1. In einem Experiment, welches zufälligen Störungen unterworfen ist, möchten Sie eine gewisse reellwertige Messgröße X bestimmen. Ihre Strategie dazu ist die folgende: Sie messen die gewünschte Größe mehrmals (n -mal) in unabhängigen Versuchsdurchführungen und bezeichnen mit X_i das Resultat der i -ten Messung. Als Endresultat nehmen Sie den Mittelwert $X = \frac{1}{n} \sum_{i=1}^n X_i$. Nun wollen wir dieses Problem theoretisch modellieren und eine Abschätzung für den Messfehler von X machen. Da das Resultat der i -ten Messung zufälligen Fluktuationen unterworfen ist, modellieren wir die X_i durch Zufallsvariablen mit Erwartungswert μ_i und Varianz σ_i^2 (die höheren Momente spielen im folgenden keine Rolle). Die Familie $\{X_i\}_{i=1,\dots,n}$ nehmen wir als unabhängig an, da wir ja unabhängige Versuchsdurchführungen annehmen.
 - a) Bestimmen Sie den Erwartungswert und die Varianz von $X = \frac{1}{n} \sum_{i=1}^n X_i$!
 - b) Nehmen Sie im folgenden an, dass die X_i i.i.d. Zufallsvariablen sind. Der Erwartungswert ist $\mu_i = 1000$ und die Varianz ist $\sigma_i^2 = 10000$. Bestimmen Sie den Erwartungswert $\mu = \mathbb{E}[X]$ und Varianz $\sigma^2 = \text{Var}[X]$ für diesen Spezialfall!
 - c) Mit den selben Zahlen wie in b), wie viele Versuche n müssen Sie mindestens durchführen, so dass der relative Fehler $\frac{\sigma}{\mu}$ kleiner als 0.01 (also 1%) ist?
2. Nach den Newtonschen Gesetzen ist ein zur Zeit 0 am Ort 0 startendes Teilchen, welches sich mit einer konstanten Geschwindigkeit V bewegt, nach einer Zeit T am Ort $X = VT$. Wenn wir in der realen Welt Zeiten messen oder die Geschwindigkeit des Teilchens präparieren, so machen wir unvermeidbar Mess- bzw. Präparationsfehler und führen eine zufällige Komponente in sowohl die Zeit T als auch die Geschwindigkeit V ein. Wenn wir V und T als Zufallsvariablen beschreiben, dann ist auch X eine Zufallsvariable. Wir nehmen im Folgenden vereinfachend an, dass $V \sim U(0, 1)$ und $T \sim U(0, 1)$ unabhängig und uniform verteilt sind.
 - a) Bestimmen Sie die Verteilungsfunktion der Zufallsvariablen $X = VT$!
Hinweis: Starten Sie mit $F_X(x) = P(X \leq x)$ und verwenden Sie die multivariate kontinuierliche Dichte $f_{V,T}(v, t)$.
 - b) Berechnen Sie ebenfalls die kontinuierliche Dichte $f_X(x)$, den Erwartungswert und die Varianz von X !

3. Theorie

Zeigen Sie, dass zwei normalverteilte Zufallsvariablen $X, Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ genau dann unabhängig sind wenn sie unkorreliert sind!

12.2. Literatur

13. Faltung von Verteilungen

Im vorangegangenen Abschnitt haben wir unabhängige Zufallsvariablen kennen gelernt und für die Summen von unabhängigen bzw. unkorrelierten Zufallsvariablen die Varianzen bestimmt. Die Erwartungswerte solcher Summen sind ja ohnehin als Summen der Erwartungswerte bekannt. In diesem Abschnitt möchten wir nun genauere Informationen über die Verteilung einer Summe von unabhängigen Zufallsvariablen herausfinden.

Definition 13.1. Es seien P_1 und P_2 Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Betrachte die Operation “Addition” als messbare Abbildung

$$\phi_+ : (\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) , \quad (x, y) \mapsto \phi_+(x, y) = x + y . \quad (13.1)$$

Dann ist die **Faltung** $P_1 * P_2$ von P_1 und P_2 definiert als das Bildmaß des Produktmaßes unter ϕ_+

$$P_1 * P_2 := \phi_{+*}(P_1 \otimes P_2) . \quad (13.2)$$

Die Relevanz der Faltung ergibt sich aus dem folgenden

Satz 13.2. *Es seien $X, Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ unabhängige \mathbb{R}^d -wertige Zufallsvariablen. Dann ist die Verteilung von $X + Y$ gerade die Faltung der Verteilung von X mit der Verteilung von Y , d.h.*

$$(X + Y)_*P = X_*P * Y_*P . \quad (13.3)$$

Beweis. Es gilt $X + Y = \phi_+(X, Y)$. Mit Satz 12.6 (ii) folgt aus der Unabhängigkeit von X und Y

$$\begin{aligned} (X + Y)_*P &= (\phi_+(X, Y))_*P \\ &= \phi_{+*}(X, Y)_*P \\ &= \phi_{+*}(X_*P \otimes Y_*P) = X_*P * Y_*P . \quad \square \end{aligned} \quad (13.4)$$

Nachdem wir uns von der Nützlichkeit der Faltung überzeugt haben, möchten wir nun Formeln für die diskreten bzw. kontinuierlichen Dichten von gefalteten Wahrscheinlichkeitsmaßen ableiten. Dabei ändert sich die Definition von Faltung nicht, wenn wir statt \mathbb{R}^d -wertigen Zufallsvariablen \mathbb{Z}^d -wertige Zufallsvariablen betrachten – wir können uns ja z.B. \mathbb{Z}^d als in den \mathbb{R}^d eingebettet vorstellen.

13. Faltung von Verteilungen

Satz 13.3. *Es seien X, Y unabhängige Zufallsvariablen auf (Ω, \mathcal{A}, P) mit Werten in \mathbb{Z}^d bzw. \mathbb{R}^d .*

- (i) *Sind X und Y jeweils \mathbb{Z}^d -wertige Zufallsvariablen mit diskreten Dichten p_x^X und p_x^Y , dann erhalten wir für die diskrete Dichte von $X + Y$*

$$p_x^{X+Y} = (p^X * p^Y)_x := \sum_{y \in \mathbb{Z}^d} p_{x-y}^X p_y^Y . \quad (13.5)$$

- (ii) *Sind X und Y jeweils \mathbb{R}^d -wertige Zufallsvariablen mit kontinuierlichen Dichten $f_X(x)$ und $f_Y(x)$, dann erhalten wir für die kontinuierliche Dichte von $X + Y$*

$$f_{X+Y}(x) = (f_X * f_Y)(x) := \int_{\mathbb{R}^d} f_X(x-y) f_Y(y) dy . \quad (13.6)$$

Hierbei sind die Ausdrücke im Mittelteil der Gleichungen (13.5) und (13.6) jeweils definiert durch die rechte Seite.

Beweis. (i) Für $x \in \mathbb{Z}^d$ erhalten wir

$$\begin{aligned} p_x^{X+Y} &= P(\{X + Y = x\}) \\ &= P(\{X = x - Y\}) \\ &= P\left(\bigcup_{y \in \mathbb{Z}^d} (\{X = x - y\} \cap \{Y = y\})\right) \\ &= \sum_{y \in \mathbb{Z}^d} P(\{X = x - y\} \cap \{Y = y\}) \\ &= \sum_{y \in \mathbb{Z}^d} P(\{X = x - y\}) P(\{Y = y\}) = \sum_{y \in \mathbb{Z}^d} p_{x-y}^X p_y^Y . \end{aligned} \quad (13.7)$$

- (ii) Mit Satz 13.2, dem Satz von Fubini und der Translationsinvarianz des Lebesgue-Maßes erhalten wir für $x \in \mathbb{R}^d$

$$\begin{aligned} F_{X+Y}(x) &= [(X + Y)_* P]((-\infty, x]) \\ &= X_* P * Y_* P((-\infty, x]) \\ &= X_* P \otimes Y_* P(\phi_+^{-1}((-\infty, x])) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_{\{y+z \leq x\}}(y, z) f_X(y) f_Y(z) dy dz \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_{\{y+z \leq x\}}(y, z) f_X(y) f_Y(z) dy dz \\ &\stackrel{y \rightarrow y+z}{=} \int_{(-\infty, x]} \left(\int_{\mathbb{R}^d} f_X(y-z) f_Y(z) dz \right) dy . \end{aligned} \quad (13.8)$$

Durch den Hauptsatz der Differential- und Integralrechnung erhalten wir nun durch d -faches Ableiten

$$f_{X+Y}(x) = \int_{\mathbb{R}^d} f_X(x-y) f_Y(y) dy = (f_X * f_Y)(x) . \quad (13.9)$$

Insbesondere existiert die Dichte von $X + Y$. □

Beispiel 13.4. Es sei P das Laplacesche Wahrscheinlichkeitsmaß auf $\{1, 2, 3, 4, 5, 6\}$, also das Wahrscheinlichkeitsmaß des einfachen Würfels. Indem wir $p_x = \frac{1}{6}$, für $x = 1, \dots, 6$, und $p_x = 0$ sonst setzen, so können wir P als ein Wahrscheinlichkeitsmaß auf \mathbb{Z} auffassen. Der unabhängige Wurf mit zwei fairen Würfeln hat demnach die Verteilung $P * P$. Die Dichte von $P * P$ wurden hierbei schon in Kapitel 1 berechnet, vergleiche auch Abbildung 1.1.

Besonders einfach transformieren sich die erzeugenden Funktionen unter der Addition von unabhängigen Zufallsvariablen bzw. unter der Faltung der zugehörigen Verteilungen.

Proposition 13.5. *Es sei $\{X_i : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}_{i=1, \dots, n}$ eine unabhängige Familie von Zufallsvariablen.*

(i) *Es gilt*

$$\mathcal{F}_{X_1 + \dots + X_n}(\alpha) = \prod_{j=1}^n \mathcal{F}_{X_j}(\alpha) . \quad (13.10)$$

(ii) *Insbesondere gilt für X_1, \dots, X_n i.i.d.*

$$\mathcal{F}_{X_1 + \dots + X_n}(\alpha) = [\mathcal{F}_{X_1}(\alpha)]^n . \quad (13.11)$$

Analoge Aussagen gelten auch für die Laplacetransformation und die wahrscheinlichkeits-erzeugende Funktion, falls diese definiert sind.

Beweis. (ii) ist triviale Konsequenz aus (i). (i) wiederum ist eine unmittelbare Folgerung aus Satz 12.6 (iii) mit $\phi_i(x) = e^{i\alpha x}$ (dieser Satz läßt sich unmittelbar auf komplexe Funktionen erweitern). \square

Korollar 13.6. *Es seien X_1, \dots, X_n wie in Proposition 13.5.*

(i) *Die p -te Kumulante einer Summe unabhängiger Zufallsvariablen ist die Summe der p -ten Kumulanten der einzelnen Zufallsvariablen, d.h.*

$$\mathcal{C}_p[X_1 + \dots + X_n] = \sum_{j=1}^n \mathcal{C}_p[X_j] . \quad (13.12)$$

(ii) *Insbesondere gilt für X_1, \dots, X_n i.i.d.*

$$\mathcal{C}_p[X_1 + \dots + X_n] = n \mathcal{C}_p[X_1] . \quad (13.13)$$

Beweis. Wieder folgt (ii) aus (i). Für α aus einer hinreichend kleinen Umgebung von 0 gilt $\mathcal{F}_{X_i}(\alpha) = \exp \psi_{X_i}(\alpha)$, $i = 1, \dots, n$. Dann gilt für solche α nach Proposition 13.5 (i)

$$\begin{aligned} \psi_{X_1 + \dots + X_n}(\alpha) &= \log \mathcal{F}_{X_1 + \dots + X_n}(\alpha) \\ &= \log \left[\prod_{j=1}^n \mathcal{F}_{X_j}(\alpha) \right] \\ &= \sum_{j=1}^n \log \mathcal{F}_{X_j}(\alpha) = \psi_{X_1}(\alpha) + \dots + \psi_{X_n}(\alpha) . \end{aligned} \quad (13.14)$$

Durch p -faches ableiten nach α an der Stelle $\alpha = 0$ erhalten wir nun die Aussage. \square

13. *Faltung von Verteilungen*

Beachten Sie, dass Korollar 13.6 die direkte Verallgemeinerung des Varianzadditionssatzes ist, siehe Satz 12.13.

13.1. Übungsaufgaben

1. Selbsttest

1. Bestimmen Sie die Faltung für zwei unabhängige Zufallsvariablen $X, Y \sim G(p)$!
2. Bestimmen Sie die Faltung für zwei unabhängige Zufallsvariablen $X \sim N(\mu, \sigma^2)$ und $Y \sim N(\mu', \sigma'^2)$!
3. Können Sie aus den Resultaten der Aufgabe zuvor die Verteilung von $X = \sum_{j=1}^n X_j$ berechnen, wobei $X_j \sim N(\mu, \sigma^2)$ i.i.d. Zufallsvariablen.

2. Theorie

Die Faltung der Verteilungen von zwei unabhängigen Zufallsvariablen $X, Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ wurde in der Vorlesung gerade so definiert, dass sie die Verteilung von der Summe der Zufallsvariablen $X + Y$ beschreibt. Es seien nun X, Y zwei unabhängige und **positive** Zufallsvariablen, d.h. $X > 0$ und $Y > 0$. Entwickeln Sie eine Strategie um die Verteilung vom Produkt $X \cdot Y$ der Zufallsvariablen zu bestimmen!

Hinweis: Verwenden Sie eine geeignete Transformation, welche das Produkt auf eine Summe zurückführt!

13. Faltung von Verteilungen

13.2. Literatur

14. Gesetz der großen Zahlen

Das Gesetz der großen Zahlen besagt, dass das arithmetische Mittel von n identisch verteilten und paarweise unkorrelierten Zufallsvariablen gegen den Erwartungswert der Zufallsvariablen konvergiert, sofern $n \rightarrow \infty$. Für eine mathematische Formulierung müssen wir zunächst klären, was hier ‘konvergiert’ bedeutet. Leider gibt es in der Wahrscheinlichkeitstheorie nicht den einen universellen Konvergenzbegriff, der für alle Anwendungsfälle passt. Die folgende Definition liefert statt dessen eine Auswahl an Konvergenzbegriffen:

Definition 14.1. Es seien $X, X_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ Zufallsvariablen, für alle $n \in \mathbb{N}$.

- (i) Ist weiter $X, X_n \in L^p((\Omega, \mathcal{A}, P); \mathbb{R})$, für alle $n \in \mathbb{N}$, dann konvergiert X_n im **p -ten Mittel** gegen X bzw. im L^p -Sinne, falls

$$\|X_n - X\|_p \longrightarrow 0 \quad \text{für } n \rightarrow \infty . \quad (14.1)$$

- (ii) X_n konvergiert **P -fast sicher** gegen X falls $\exists \mathcal{N} \in \mathcal{A}$ mit $P(\mathcal{N}) = 0$, so dass

$$X_n(\omega) \longrightarrow X(\omega) \quad \text{für } n \rightarrow \infty \quad \forall \omega \in \Omega \setminus \mathcal{N} . \quad (14.2)$$

- (iii) X_n konvergiert **nach Wahrscheinlichkeit** bzw. stochastisch gegen X , falls für jede beliebig kleine Schranke $\epsilon > 0$ gilt

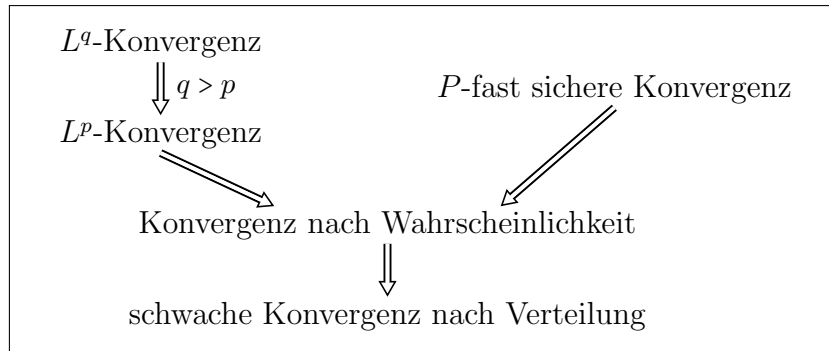
$$P(|X_n - X| > \epsilon) \longrightarrow 0 \quad \text{für } n \rightarrow \infty . \quad (14.3)$$

- (iv) X_n konvergiert **schwach nach Verteilung** gegen X , falls für alle $\phi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ stetig und beschränkt gilt

$$\mathbb{E}[\phi(X_n)] \longrightarrow \mathbb{E}[\phi(X)] \quad \text{für } n \rightarrow \infty . \quad (14.4)$$

Es gibt noch mehr Konvergenzbegriffe, etwa die starke Konvergenz nach Verteilung, aber wir wollen es mal bei diesen Begriffen belassen. Einige der hier vorgestellten Konvergenzbegriffe sind stärker als andere. Das folgende Diagramm visualisiert dies und die nötigen Beweise finden Sie in Bauer, Wahrscheinlichkeitstheorie, S. 34 ff..

14. Gesetz der großen Zahlen



Je schwächer der Konvergenzbegriff ist, desto einfacher ist die Konvergenz zu beweisen. Mit dem Gesetz der großen Zahlen beginnen wir deshalb mit der schwachen Form, die auf der Konvergenz nach Wahrscheinlichkeit beruht. Der Beweis ist eine Folge aus der Chebychev-Ungleichung, die wir zunächst betrachten wollen.

Satz 14.2 (Markov- und Chebychev-Ungleichung).

(i) Es sei $X \in L^1((\Omega, \mathcal{A}, P); \mathbb{R})$ nicht negativ, dann gilt für $c > 0$ beliebig

$$P(X > c) \leq \frac{1}{c} \mathbb{E}[X] . \quad (\text{Markov-Ungleichung}) \quad (14.5)$$

(ii) Es sei $X \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$, dann gilt für $\epsilon > 0$ beliebig

$$P(|X - \mathbb{E}[X]| > \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}[X] . \quad (\text{Chebychev-Ungleichung}) \quad (14.6)$$

Beweis. (i) Dies kann man mit einer geschickten Anwendung der charakteristischen Funktionen 1_A von Mengen $A \in \mathcal{A}$ wie folgt beweisen:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X 1_{\{X < c\}}] + \mathbb{E}[X 1_{\{X \geq c\}}] \\ &\geq \mathbb{E}[X 1_{\{X \geq c\}}] \\ &\geq c \mathbb{E}[1_{\{X \geq c\}}] = c P(X > c) . \end{aligned} \quad (14.7)$$

Division durch $c > 0$ führt nun zum Ergebnis.

(ii) Wir wenden die Markov-Ungleichung auf $|X - \mathbb{E}[X]|^2$ und $c = \epsilon^2$ an und erhalten

$$\begin{aligned} P(|X - \mathbb{E}[X]| > \epsilon) &= P(|X - \mathbb{E}[X]|^2 > \epsilon^2) \\ &\leq \frac{1}{\epsilon^2} \mathbb{E}[|X - \mathbb{E}[X]|^2] = \frac{1}{\epsilon^2} \text{Var}[X] . \end{aligned} \quad (14.8)$$

□

Satz 14.3 (Schwaches Gesetz der großen Zahlen). Es seien $X_j \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$, $j \in \mathbb{N}$, paarweise unkorreliert und identisch verteilt. Dann konvergiert das arithmetische Mittel von n Zufallsvariablen

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j \quad (14.9)$$

nach Wahrscheinlichkeit gegen den Erwartungswert $\mathbb{E}[X_1]$ der Zufallsvariable.

Beweis. Aufgrund der Linearität des Erwartungswertes gilt

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{j=1}^n \mathbb{E}[X_j] = \frac{1}{n} n \mathbb{E}[X_1] = \mathbb{E}[X_1] . \quad (14.10)$$

Nach der Chebychev-Ungleichung ist deswegen

$$\begin{aligned} P(|\bar{X}_n - \mathbb{E}[X_1]| > \epsilon) &\leq \frac{1}{\epsilon^2} \text{Var} \left[\frac{1}{n} \sum_{j=1}^n X_j \right] \\ &= \frac{1}{\epsilon^2 n^2} \text{Var} \left[\sum_{j=1}^n X_j \right] \\ &= \frac{1}{\epsilon^2 n^2} n \text{Var}[X_1] \longrightarrow 0 \quad \text{für } n \rightarrow \infty . \end{aligned} \quad (14.11)$$

Hierbei wurden ebenfalls die Varianzadditionsgesetze aus Satz 12.13 und $\text{Var}[aX] = a^2 \text{Var}[X]$ verwendet. \square

Beispiel 14.4 (Monte-Carlo Integration). Es sei $D = [0, 1]^d$ der d -dimensionale Einheitswürfel und $\phi : D \rightarrow \mathbb{R}$ eine stetige Funktion. Wir möchten das Integral $\int_D \phi(x) dx$ numerisch berechnen. Ist d groß (sagen wir $d > 5$), dann sind klassische Quadraturmethoden ineffizient, da diese ein Skalierungsverhalten k^d aufweisen (‘Fluch der Dimension’).

Lösung. Wir betrachten $d \times n$ i.i.d. Zufallsvariablen mit der Verteilung $U(0, 1)$ und fassen jeweils d von ihnen zu einem Vektor $X^{(j)} = (X_1^{(j)}, \dots, X_d^{(j)})$ zusammen. $X^{(j)} \sim U(0, 1)^{\otimes n}$ ist dann i.i.d. multivariat gleichverteilt auf D . Die Zufallsvariablen $Y_j := \phi(X^{(j)})$ sind beschränkt und deshalb in $L^2((\Omega, \mathcal{A}, P); \mathbb{R})$. Nach dem schwachen Gesetz der großen Zahlen gilt für jede beliebige Fehlerschranke $\epsilon > 0$

$$P \left(\left| \frac{1}{n} \sum_{j=1}^n \phi(X^{(j)}) - \int_D \phi(x) dx \right| > \epsilon \right) = P(|\bar{Y}_n - \mathbb{E}[Y_1]| > \epsilon) \longrightarrow 0 \quad \text{für } n \rightarrow \infty . \quad (14.12)$$

Wir können also den Ausdruck $\frac{1}{n} \sum_{j=1}^n \phi(X^{(j)}(\omega))$ als eine Approximation von $\int_D \phi(x) dx$ verstehen, und machen für große n mit großer Wahrscheinlichkeit nur einen kleinen Fehler. Beachten Sie, dass die Konvergenzrate in n im Gesetz der großen Zahlen nur von der Varianz von $\phi(X^{(1)})$ abhängt, und deshalb vom Maximum von ϕ zum Quadrat beschränkt ist. Insbesondere hängt die Konvergenzgeschwindigkeit nicht von d ab.

In der praktischen Umsetzung generieren wir $X^{(j)}(\omega)$ indem wir d $U(0, 1)$ -verteilte Zufallsvariablen mit dem (pseudo-) Zufallszahlengenerator erzeugen.

Die Monte-Carlo Integration ist weiterhin ein aktives Forschungsgebiet und wurde in den letzten Jahren noch erheblich verfeinert, wie zum Beispiel durch ‘Importance Sampling’. \diamond

Zum schwachen Gesetz der großen Zahlen gibt es noch die starke Version. Diese beruht auf P -fast sicherer Konvergenz. Die schwache Version wird demnach von der starken impliziert. Hier geben wir eine Version des starken Gesetzes der großen Zahlen an, die von der Existenz der 2. Momente ausgeht. Dies ist nicht unbedingt notwendig (vergleiche Satz von Etemadi), erleichtert aber den Beweis.

14. Gesetz der großen Zahlen

Satz 14.5 (Starkes Gesetz der großen Zahlen). *Unter den Voraussetzungen von Satz 14.3 konvergiert $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ sogar P -fast sicher gegen $\mathbb{E}[X_1]$.*

Beweis. Hier folgen wir dem Beweis in Bauer, Maß- und Integrationstheorie, (4. Auflage) de Gruyter 1990. Wir müssen zeigen, dass

$$Y_n = \frac{1}{n} \sum_{j=1}^n (X_j - \mathbb{E}[X_j]) \longrightarrow 0 \quad \text{für } n \rightarrow \infty \quad P\text{-fast sicher} . \quad (14.13)$$

Wir betrachten zuerst die Unterfolge Y_{n^2} und zeigen, dass

$$P(|Y_{n^2}| \geq \epsilon \text{ für unendlich viele } n \in \mathbb{N}) = P\left(\limsup_n \{|Y_{n^2}| > \epsilon\}\right) = 0 . \quad (14.14)$$

Nach der Chebychev-Ungleichung aus Satz 14.2 (ii) gilt

$$\sum_{n=1}^{\infty} P(|Y_{n^2}| > \epsilon) \leq \sum_{n=1}^{\infty} \frac{\text{Var}[X_1]}{\epsilon^2 n^2} < \infty \quad (14.15)$$

und mit dem Satz 5.6 (i) von Borel-Cantelli folgt (14.14). Als Konsequenz gilt auch

$$P(Y_{n^2} \text{ konvergiert nicht gegen } 0) = P\left(\bigcup_{k \in \mathbb{N}} \limsup_n \{|Y_{n^2}| > \frac{1}{k}\}\right) = 0 . \quad (14.16)$$

Also gilt $Y_{n^2} \longrightarrow 0$ P -fast sicher.

Das nächste Ziel ist es, ein beliebiges Y_m mit einem Element aus der Unterfolge Y_{n^2} zu vergleichen und zu zeigen, dass die Differenz klein ist. Dazu wählen wir zu $m \in \mathbb{N}$ ein $n = n(m) \in \mathbb{N}$, so dass $n^2 \leq m < (n+1)^2$. Für die Partialsumme der ersten k zentrierten Zufallsvariablen $X_j - \mathbb{E}[X_j]$, d.h. für $S_k = \sum_{j=1}^k (X_j - \mathbb{E}[X_j]) = k Y_k$, gilt für die so bestimmten m und n

$$P(|S_m - S_{n^2}| \geq \epsilon n^2) \leq \frac{1}{\epsilon^2 n^4} \text{Var} \left[\sum_{j=n^2+1}^m X_j \right] = \frac{m - n^2}{n^4 \epsilon^2} \text{Var}[X_1] . \quad (14.17)$$

Deshalb gilt auch

$$\begin{aligned} \sum_{m=1}^{\infty} P(|S_m - S_{n^2}| > \epsilon n^2) &\leq \sum_{n=1}^{\infty} \sum_{m=n^2}^{(n+1)^2-1} \frac{m - n^2}{n^4 \epsilon^2} \text{Var}[X_1] \\ &\stackrel{l=m-n^2}{=} \frac{\text{Var}[X_1]}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^4} \sum_{l=0}^{(n+1)^2-1-n^2} l \\ &= \frac{\text{Var}[X_1]}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^4} \sum_{l=0}^{2n} l \\ &= \frac{\text{Var}[X_1]}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^4} \frac{(2n+1)(2n)}{2} \\ &< \infty . \end{aligned} \quad (14.18)$$

Mit dem Satz von Borel-Cantelli können wir nun wieder folgern, dass

$$P\left(\left|\frac{S_m}{n(m)} - Y_{n(m)}\right| > \epsilon^2 \text{ unendlich oft}\right) = 0. \quad (14.19)$$

Wie zuvor folgt daraus, dass

$$P\left(\frac{S_m}{n(m)^2} \text{ konvergiert nicht gegen } 0\right) = P\left(\bigcup_{k \in \mathbb{N}} \limsup_m \left\{\left|\frac{S_m}{n(m)^2} - Y_{n(m)^2}\right| > \frac{1}{k}\right\}\right) = 0. \quad (14.20)$$

Also $\frac{S_m}{n(m)^2} \rightarrow 0$ für $m \rightarrow \infty$ P -fast sicher. Wegen $|Y_m| = \left|\frac{S_m}{m}\right| \leq \left|\frac{S_m}{n(m)^2}\right|$ konvergiert dann auch Y_m P -fast sicher gegen Null. \square

Zuletzt möchten wir in diesem Kapitel das Gesetz der großen Zahlen noch mittels einer Simulation visualisieren. In der Stochastik-Programmiersprache R stehen Zufallszahlengeneratoren für eine große Anzahl von Verteilungsfamilien zur Verfügung. Da das Gesetz der großen Zahlen unabhängig von den Details der Verteilung funktioniert, wählen wir der Einfachheit halber uniform verteilte Zufallsvariablen $U(0,1)$ für unsere Simulation. Der R Befehl für den $U(0,1)$ Zufallszahlengenerator lautet `runif(n)`, wobei `n` die Anzahl der zu erzeugenden Zufallszahlen ist.

```
# simulation to visualize the law of large numbers
# note that we do not set a seed, so every simulation looks different

n=1000                # set number of repetitions
x=runif(n)             # generates n pseudo random no.s U(0,1)
plot(x,type="l")       # just noise
X=cumsum(x)            # cumulative sum - random walk
plot(X,type="l")       # random walk with upward trend
XC=X-1:n*(1/2)         # center - subtract expected values
plot(XC,type="l")      # random walk without trend
barX=XC/1:n            # now take average values
plot(barX,type="l")    # 'converges' to 0
abline(h=0)            # draw line at 0
```

Die Abbildung 14.1 reproduziert die vier Plots aus dem obigen Skript.

14. Gesetz der großen Zahlen

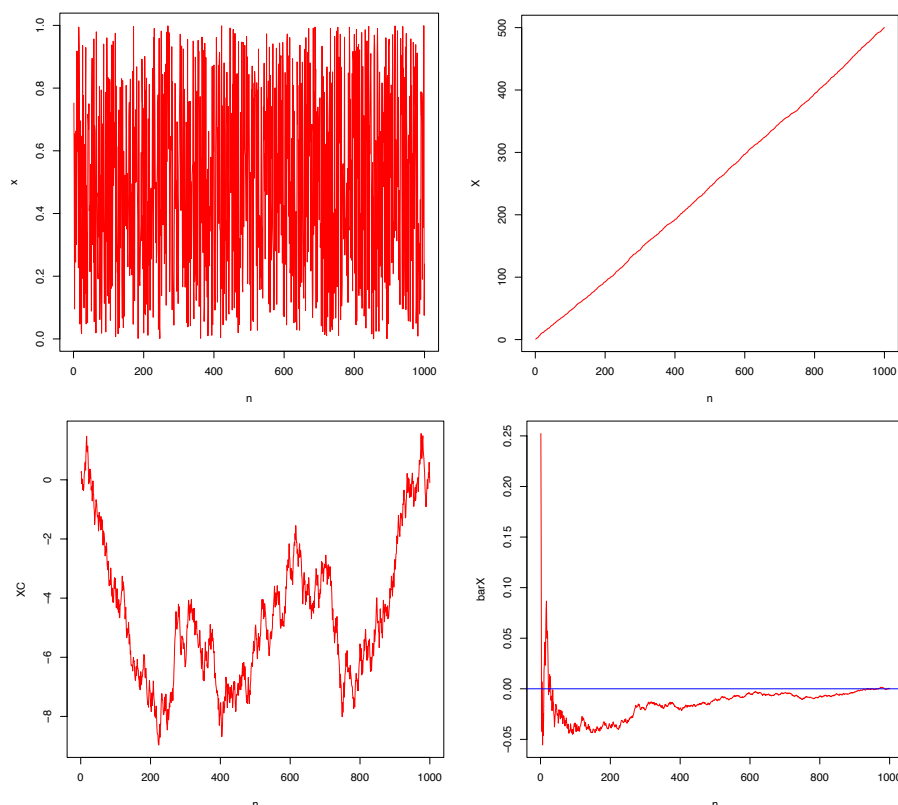


Abbildung 14.1: Visualisierung zum Gesetz der großen Zahlen. Oben links: Rauschen. Oben Rechts: Irrfahrt mit Drift. Unten links: Zentrierte Irrfahrt. Unten rechts: Gesetz der großen Zahlen.

14.1. Übungsaufgaben

1. Selbsttest

- Es seien $X_j \sim N(0,1)$, $j \in \mathbb{N}$, eine unabhängige Familie von normalverteilten Zufallsvariablen. Wir betrachten das arithmetische Mittel $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$.
 - Begründen Sie Mittels des Gesetzes der großen Zahlen ob und wenn ja gegen was die Zufallsvariablen \bar{X}_n für $n \rightarrow \infty$ konvergieren.
 - Was ist hier der relevante Konvergenzbegriff?
 - Geben Sie einen direkten Beweis für das Resultat in a) indem Sie die n -fache Faltung von standardnormalverteilten Zufallsvariablen betrachten! Verwenden Sie hierzu die Resultate aus den Übungen zu Kapitel 13.
- Es sei X eine Zufallsvariable über die Sie nichts mehr wissen als dass $\mathbb{E}[X] = 1$ und $\text{Var}[X] = 1$. Geben Sie eine obere Schranke für die Wahrscheinlichkeit $P(X > 11)$

indem Sie einen die Markov- und zum anderen die Chebychev-Ungleichung verwenden! Welche Abschätzung ist besser?

2. Transfer

1. Ein Versicherungskonzern hält einen Bestand an Versicherungspolice, der im Mittel zu 1 000 000 Euro/Jahr Auszahlungen führt. Die Standardabweichung der zufälligen Auszahlungsbeträge X sei 100 000 Euro/Jahr. Geben Sie ohne jede weitere Verteilungsannahme eine obere Abschätzung für die Wahrscheinlichkeit, dass der Auszahlungsbetrag für diesen Bestand in einem bestimmten Jahr 1 500 000 Euro übersteigt!
2. Betrachten Sie nochmals die erste Transferaufgabe aus Kapitel 12, welche eine Behandlung von zufälligen Messfehlern in Experimenten gibt. Wenn Sie die Anzahl der Experimente n mehr und mehr erhöhen, was passiert dann mit den stochastischen Fluktuationen des Endresultats $X = \frac{1}{n} \sum_{i=1}^n X_i$?

3. Aktivität

Das schöne am Gesetz der großen Zahlen ist, das man das ‘ausprobieren’ kann. Nehmen Sie einen 6-seitigen Würfel und würfeln Sie mindestens 30 Mal hintereinander. Zeichnen Sie die Würfelergebnisse auf und berechnen Sie \bar{X}_n für $n = 1, \dots, 30$. Wenn Sie das mit dem Taschenrechner oder einem spreadsheet Kalkulator machen, verwenden Sie hierfür die rekursive Formel

$$\bar{X}_n = \frac{1}{n}((n-1)\bar{X}_{n-1} + X_n) . \quad (14.21)$$

In R können Sie das mit den untenstehenden Befehlen realisieren:

```
X=c('W1','W2',...,'W30')      # replace Wxx by your dicing result
barX=cumsum(X)/(1:length(X))    # cumsum does partial sums
plot(barX)                      # visualize
```

Diskutieren Sie Ihre Beobachtungen und vergleichen Sie diese mit der theoretischen Vorhersage für den fairen Würfel.

14. Gesetz der großen Zahlen

14.2. Literatur

15. Der zentrale Grenzwertsatz

Wenn wir die Konvergenz im Gesetz der großen Zahlen für quadratintegrierbare i.i.d. Zufallsvariablen X_i betrachten, so können wir uns zusätzlich ein Bild machen von der Fluktuation der zufälligen Mittelwerte $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ um den wahren Erwartungswert für ein gegebenes n . Dabei hängt die Verteilung von \bar{X}_n natürlich von der Verteilung der X_i ab. Die Überraschung ist, dass diese Abhängigkeit für große n immer schwächer wird und die Verteilungen immer mehr der Normalverteilung ähneln. Diesen Effekt nennt man den zentralen Grenzwertsatz. Er begründet die zentrale Stellung der Normalverteilung in der Wahrscheinlichkeitstheorie und Statistik.

Bevor wir den Satz formulieren, schauen wir uns zuerst das Resultat in der Simulation an. Dazu simulieren wir 10000 zentrierte Irrfahrten wie in Abbildung 14.1 c). Wieder verwenden wir gleichverteilte Zufallsvariablen $X_i \sim U(-\frac{1}{2}, \frac{1}{2})$. Für $n = 1, 3, 10$ und $n = 100$ schneiden wir nun vertikal durch die Irrfahrten, vergleiche Abbildung 15.1, und bilden Histogramme für die Schnittwerte, welche Realisierungen von Summen von n unabhängigen $U(-\frac{1}{2}, \frac{1}{2})$ -Zufallsvariablen darstellen. Die Abbildung 15.2 zeigt die Resultate.

Würden wir diese Simulation mit beliebigen Verteilungen (mit existenten 2. Momenten) wiederholen, so würde sich das Bild jeweils nicht stark von der hier gezeigten Situation unterscheiden. Wiederum erhalten wir für große n glockenförmige Verteilungen, deren Breite proportional \sqrt{n} zunimmt. Teilen wir also die zentrierten Irrfahrten durch die Wurzel aus der Anzahl der Schritte, d.h. durch \sqrt{n} , dann sollte sich sowohl die Form als auch die Breite der Verteilung stabilisieren.

Eine andere Sichtweise auf diesen Sachverhalt ist, dass sich die Fluktuationen des arithmetischen Mittels von i.i.d. Zufallsvariablen um den Erwartungswert, d.h. die Fluktuationen von $\bar{X}_n - \mathbb{E}[X_1]$, einer Normalverteilung mit abnehmender Varianz $\sim \frac{1}{\sqrt{n}}$ annähern. Die mathematisch präzise Aussage zu diesem Sachverhalt liefert der folgende

Satz 15.1 (Zentraler Grenzwertsatz). *Es seien $X_j \in L^2((\Omega, \mathcal{A}, P); \mathbb{R})$, $j \in \mathbb{N}$, i.i.d. Zufallsvariablen. Dann gilt mit $\mu = \mathbb{E}[X_1]$ und $\sigma^2 = \text{Var}[X_1]$*

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mu) \xrightarrow{n \rightarrow \infty} X \sim N(0, \sigma^2) \quad \text{schwach nach Verteilung} . \quad (15.1)$$

Bemerkung 15.2. Man beachte, dass X nicht notwendiger Weise auf (Ω, \mathcal{A}, P) definiert sein muss.

Beweis von Satz 15.1. Der Beweis des zentralen Grenzwertsatzes ist kompliziert. Hier geben wir nur den Beweis an unter der stärkeren Annahme, dass $X_j \in L^3((\Omega, \mathcal{A}, P); \mathbb{R})$.

15. Der zentrale Grenzwertsatz

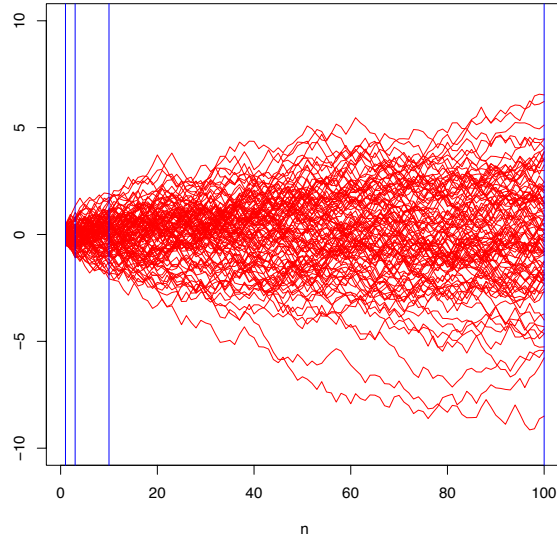


Abbildung 15.1: Simulation von zentrierten Irrfahrten - nur 100 Irrfahrten wurden dargestellt.

Zudem setzen wir einen Satz von Paul Lévy voraus, nach dem für Zufallsvariablen Y_n und Y gilt

$$\mathcal{F}_{Y_n}(\alpha) \xrightarrow{n \rightarrow \infty} \mathcal{F}_Y(\alpha) \quad \forall \alpha \in \mathbb{R} \quad \Leftrightarrow \quad Y_n \xrightarrow{n \rightarrow \infty} Y \text{ schwach nach Verteilung.} \quad (15.2)$$

Hierbei ist die Implikation \Leftarrow unmittelbar klar, da $\phi_\alpha(x) = e^{i\alpha x} \in C_b(\mathbb{R}, \mathbb{R})$ messbar und beschränkt. Die Rückrichtung kann man intuitiv verstehen, wenn man sich klar macht, dass Linearkombinationen der komplexen Exponentialfunktionen in geeigneter Weise dicht in $C_b(\mathbb{R}, \mathbb{R})$ liegen. Der Beweis des Satzes von Lévy wird im Anschluss an diesen Beweis nachgetragen, siehe Satz 15.3.

Von nun an setzen wir ohne Beschränkung der Allgemeinheit $\mathbb{E}[X_1] = \mu = 0$ voraus und betrachten die Fouriertransformation von $\sqrt{n}\bar{X}_n$

$$\begin{aligned} \mathcal{F}_{\sqrt{n}\bar{X}_n}(\alpha) &= \mathcal{F}_{\frac{1}{\sqrt{n}} \sum_{j=1}^n X_j}(\alpha) \\ &= \prod_{j=1}^n \mathcal{F}_{\frac{1}{\sqrt{n}} X_j}(\alpha) \\ &= \prod_{j=1}^n \mathcal{F}_{X_j} \left(\frac{\alpha}{\sqrt{n}} \right) \\ &= \left(\mathcal{F}_{X_1} \left(\frac{\alpha}{\sqrt{n}} \right) \right)^n \\ &= e^{n\psi_{X_1} \left(\frac{\alpha}{\sqrt{n}} \right)} \end{aligned} \quad (15.3)$$

für n hinreichend groß, so dass $\frac{\alpha}{\sqrt{n}}$ in einer Umgebung von 0 liegt, auf der ψ_{X_1} eindeutig definiert ist.

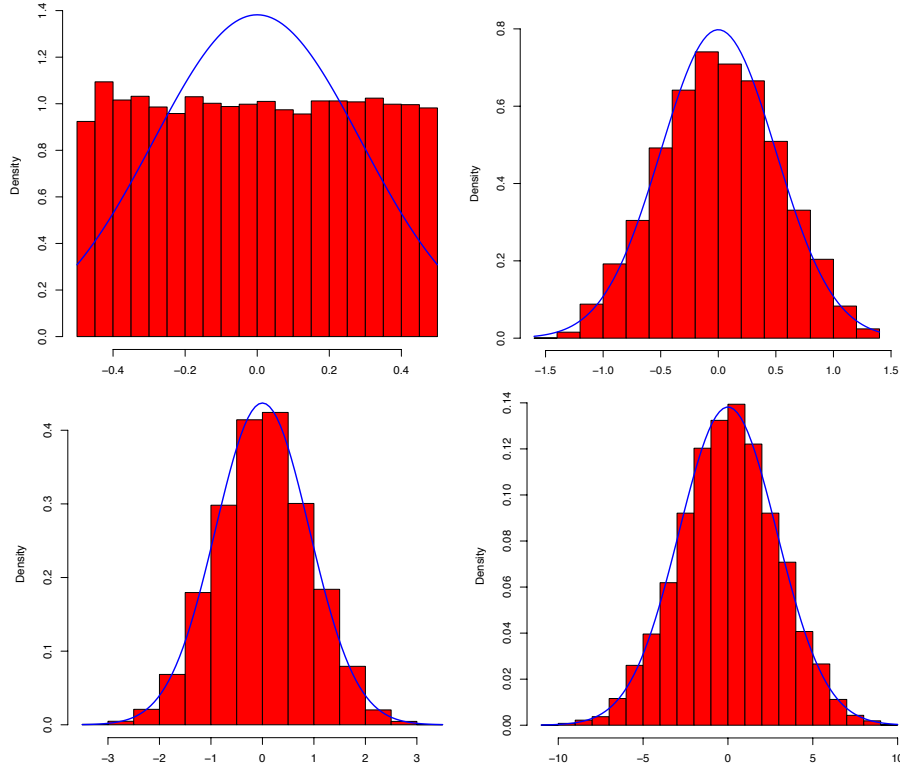


Abbildung 15.2: Histogramme der Summe von n $U(-\frac{1}{2}, \frac{1}{2})$ -verteilten Zufallsvariablen für $n = 1, 3, 10$ und $n = 100$ (oben links, oben rechts, unten links und unten rechts). Die blaue Linie stellt jeweils die Dichte der zentrierten Normalverteilung mit Varianz $\sigma^2 = \frac{n}{12}$ dar.

$\mathcal{F}_{X_1}(\alpha)$ ist wegen $X_1 \in L^3((\Omega, \mathcal{A}, P); \mathbb{R})$ drei mal differenzierbar, vergleiche Satz 10.2 (ii). Aus dem Beweis dieses Satzes sieht man, dass die dritte Ableitung uniform beschränkt ist durch

$$\left| \frac{d^3}{d\alpha^3} \mathcal{F}_{X_1}(\alpha) \right| \leq \mathbb{E}[|X_1|^3] . \quad (15.4)$$

Analog zeigt man die uniforme Beschränktheit von $\frac{d^2}{d\alpha^2} \mathcal{F}_{X_1}(\alpha)$ und $\frac{d}{d\alpha} \mathcal{F}_{X_1}(\alpha)$. In einer kleinen Umgebung der 1 ist auch der Logarithmus beliebig oft differenzierbar mit beschränkten Ableitungen. Nach Satz 1.22 ist dann auch $\left| \frac{d^3}{d\alpha^3} \psi_{X_1}(\alpha) \right|$ in einer Umgebung der 0 uniform durch eine Konstante C beschränkt.

Nun wenden wir die Taylorsche Formel (Taylor Entwicklung um $\alpha = 0$) auf ψ_{X_1} an und erhalten für hinreichend kleine α

$$\begin{aligned} \psi_{X_1}(\alpha) &= \psi_{X_1}(0) + \frac{d}{d\alpha} \psi_{X_1}(\alpha) \upharpoonright_{\alpha=0} \alpha + \frac{1}{2} \frac{d^2}{d\alpha^2} \psi_{X_1}(\alpha) \upharpoonright_{\alpha=0} \alpha^2 + R_3(\alpha) \\ &= -\frac{1}{2} \sigma^2 \alpha^2 + R_3(\alpha) . \end{aligned} \quad (15.5)$$

15. Der zentrale Grenzwertsatz

Hier haben wir $\psi_{X_1}(0) = \log(\mathcal{F}_{X_1}(0)) = \log(1) = 0$ und Satz 10.5 verwendet. Für das Restglied $R_3(\alpha)$ verwenden wir die Lagrange Darstellung, d.h. es existiert ein $\theta \in [0, 1]$ so dass für α hinreichend klein

$$|R_3(\alpha)| = \left| \frac{1}{3!} \frac{d^3}{d\alpha^3} \psi_{X_1}(\theta\alpha) \alpha^3 \right| \leq \frac{1}{6} C |\alpha|^3 \quad (15.6)$$

und deswegen

$$\left| n\psi_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right) - \left(-\frac{1}{2}\sigma^2\alpha^2\right) \right| = \left| nR_3\left(\frac{\alpha}{\sqrt{n}}\right) \right| \leq \frac{n}{6} C \left|\frac{\alpha}{\sqrt{n}}\right|^3 \rightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (15.7)$$

Also gilt

$$\mathcal{F}_{\sqrt{n}X_n}(\alpha) = e^{n\psi_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right)} \rightarrow e^{-\frac{1}{2}\sigma^2\alpha^2} \quad \text{für } n \rightarrow \infty \quad \forall \alpha \in \mathbb{R}. \quad (15.8)$$

Gemäß Satz 10.7 (i) haben wir für $X \sim N(0, \sigma^2)$, dass $\mathcal{F}_X(\alpha) = e^{-\frac{1}{2}\sigma^2\alpha^2}$ und deshalb

$$\mathcal{F}_{\sqrt{n}X_n}(\alpha) \rightarrow \mathcal{F}_X(\alpha) \quad \text{für } n \rightarrow \infty \quad \forall \alpha \in \mathbb{R}. \quad (15.9)$$

Nun appellieren wir an den Satz von Lévy 15.3, um den Beweis zu beenden. \square

Wir müssen noch den folgenden Satz nachreichen, um den Beweis des zentralen Grenzwertsatzes zu vervollständigen.

Satz 15.3 (P. Lévy). *Eine hinreichende und notwendige Bedingung für $X_n \xrightarrow{n \rightarrow \infty} X$ schwach nach Verteilung ist die punktweise Konvergenz der charakteristischen Funktionen*

$$\mathcal{F}_{X_n}(\alpha) \rightarrow \mathcal{F}_X(\alpha) \quad \text{für } n \rightarrow \infty \quad \forall \alpha \in \mathbb{R}. \quad (15.10)$$

Als Vorbereitung für den Beweis dieses Satzes benötigen wir zwei Lemmas.

Lemma 15.4. *Eine hinreichende und notwendige Bedingung für $X_n \xrightarrow{n \rightarrow \infty} X$ schwach nach Verteilung ist*

$$\mathbb{E}[\phi(X_n)] \rightarrow \mathbb{E}[\phi(X)] \quad \text{für } n \rightarrow \infty \quad (15.11)$$

für alle stetigen und beschränkten Funktionen $\phi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, die außerhalb eines kompakten Intervalls $[-K, K]$ verschwinden.

Beweis. Offensichtlich ist die hier gestellte Bedingung schwächer als die Definition der schwachen Konvergenz nach Verteilung in Definition 14.1 (iv). Also ist sie notwendig.

Wir beweisen nun, dass sie auch hinreichend ist. Dazu sei ϕ stetig und beschränkt. Es sei $\epsilon > 0$ beliebig und a hinreichend groß, so dass $P(X \notin [-a, a]) < \epsilon/(2B)$ gilt mit B einer

oberen Schranke von $|\phi(x)|$. Des Weiteren sei $\chi(x)$ eine stetige Funktion, so dass $\chi(x) = 1$ für $|x| \leq a$ und $\chi(x) = 0$ für $|x| \geq 2a$. Dann gilt

$$\begin{aligned} |\mathbb{E}[\phi(X_n)] - \mathbb{E}[(\chi \cdot \phi)(X_n)]| &= |\mathbb{E}[\phi(X_n)(1 - \chi(X_n))]| \\ &\leq B \mathbb{E}[1 - \chi(X_n)] = B(1 - \mathbb{E}[\chi(X_n)]) . \end{aligned} \quad (15.12)$$

$\chi(x)$ ist nach Konstruktion stetig, beschränkt und verschwindet außerhalb eines kompakten Intervalls. Deswegen gilt nach Voraussetzung für n hinreichend groß

$$|B(1 - \mathbb{E}[\chi(X_n)]) - B(1 - \mathbb{E}[\chi(X)])| < \frac{\epsilon}{2} . \quad (15.13)$$

Da aber $0 \leq B(1 - \mathbb{E}[\chi(X)]) \leq \epsilon/2$ nach Konstruktion der Abschneidefunktion χ , folgt aus (15.13), dass die rechte Seite von (15.12) für hinreichend große n durch ϵ beschränkt ist. Daraus können wir nun für hinreichend große n folgendes schließen:

$$|\mathbb{E}[\phi(X_n)] - \mathbb{E}[\phi(X)]| \leq |\mathbb{E}[(\chi \cdot \phi)(X_n)] - \mathbb{E}[(\chi \cdot \phi)(X)]| + 2\epsilon . \quad (15.14)$$

Der erste Term auf der rechten Seite verschwindet für $n \rightarrow \infty$, da $(\chi \cdot \phi)(x) = \chi(x)\phi(x)$ stetig, beschränkt ist und außerhalb von $[-2a, 2a]$ verschwindet. Da ϵ beliebig gewählt war, muss auch die linke Seite für hinreichend große n beliebig klein werden. \square

Lemma 15.5. *Es sei $\phi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine stetige und von B betragsmäßig beschränkte Funktion, die außerhalb eines kompakten Intervalls $[-K, K]$ verschwindet. Weiter sei $\chi_\epsilon(x)$ die kontinuierliche Dichte der Normalverteilung mit Erwartungswert 0 und Varianz ϵ^2 , vergleiche Definition 8.4. Dann gilt mit $\phi_\epsilon(x)$ definiert durch*

$$\phi_\epsilon(x) := (\chi_\epsilon * \phi)(x) = \int_{\mathbb{R}} \chi_\epsilon(x - y)\phi(y) dy , \quad (15.15)$$

dass $\phi_\epsilon(x)$ stetig ist, $|\phi_\epsilon(x)| \leq B \forall x \in \mathbb{R}$ und $\phi_\epsilon(x) \rightarrow \phi(x)$ für $\epsilon \searrow 0$ uniform in $x \in \mathbb{R}$.

Beweis. Die Stetigkeit folgt aus dem Satz über dominierte Konvergenz, vergleiche Satz 9.9 (iii), denn $\phi_\epsilon(x) = \mathbb{E}[\phi(x + X)]$ mit $X \sim N(0, \epsilon^2)$ einer Zufallsvariable und $\phi(x_n + X) \rightarrow \phi(x + X)$ falls $x_n \rightarrow x$, aufgrund der Stetigkeit von $\phi(x)$.

Die zweite Aussage ergibt sich aus

$$|\phi_\epsilon(x)| \leq \int_{\mathbb{R}} \chi_\epsilon(x - y)|\phi(y)| dy \leq B \int_{\mathbb{R}} \chi_\epsilon(x - y) dy = B . \quad (15.16)$$

Um die dritte Aussage zu begründen, wähle $\delta > 0$ beliebig und γ hinreichend klein, so dass $|\phi(x) - \phi(y)| < \delta/2$ für $|x - y| < \gamma$. Dies ist stets möglich, denn stetige Funktionen, die außerhalb eines Intervalls verschwinden, sind gleichmäßig stetig. Es sei ϵ hinreichend klein, so dass für eine Zufallsvariable X wie oben $P(|X| > \gamma) \leq \delta/(4B)$ gilt. Dann haben wir für solche $\epsilon > 0$

$$\begin{aligned} |\phi_\epsilon(x) - \phi(x)| &= |\mathbb{E}[\phi(x + X) - \phi(x)]| \\ &\leq \frac{\delta}{2} P(|X| \leq \gamma) + 2B P(|X| > \gamma) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta . \end{aligned} \quad (15.17)$$

Diese Abschätzung ist offensichtlich uniform in $x \in \mathbb{R}$. \square

15. Der zentrale Grenzwertsatz

Beweis von Satz 15.3. Da der Real- und Imaginärteil von $\phi_\alpha(x) = e^{i\alpha x}$ jeweils eine stetige und beschränkte Funktion in x darstellen, folgt aus der schwachen Konvergenz nach Verteilung die punktweise Konvergenz der charakteristischen Funktion.

Nun nehmen wir an, dass die charakteristischen Funktionen punktweise konvergieren. Zuerst argumentieren wir, dass wir das Problem mit Hilfe der beiden vorangegangenen Lemmas in eine für den Beweis günstigere Form überführen können, in der die charakteristischen Funktionen explizit auftreten.

Nach Lemma 15.4 genügt es, (15.11) für $\phi(x)$ stetig, beschränkt und verschwindend außerhalb von einem kompakten Intervall nachzuweisen. Es sei $\delta > 0$ und $\epsilon > 0$ hinreichend klein, so dass $|\phi(x) - \phi_\epsilon(x)| < \delta \quad \forall x \in \mathbb{R}$, vergleiche Lemma 15.5. $\phi_\epsilon(x)$ ist hierbei definiert wie in (15.15). Dann gilt

$$|\mathbb{E}[\phi(X_n)] - \mathbb{E}[\phi(X)]| \leq |\mathbb{E}[\phi_\epsilon(X_n)] - \mathbb{E}[\phi_\epsilon(X)]| + 2\delta. \quad (15.18)$$

Da δ beliebig ist, genügt es zu zeigen, dass der erste Term auf der rechten Seite für $n \rightarrow \infty$ verschwindet. Dafür rechnet man mittels Gleichung (10.22) nach

$$\begin{aligned} \phi_\epsilon(x) &= \frac{1}{\sqrt{2\pi\epsilon}} \int_{\mathbb{R}} \exp\left\{-\frac{1}{2}\left(\frac{x-y}{\epsilon}\right)^2\right\} \phi(y) dy \\ &= \frac{1}{\sqrt{2\pi\epsilon}} \int_{\mathbb{R}} \left[\frac{\epsilon}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\alpha(x-y)} \exp\left\{-\frac{1}{2}(\epsilon\alpha)^2\right\} d\alpha \right] \phi(y) dy \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\alpha x} \exp\left\{-\frac{1}{2}(\epsilon\alpha)^2\right\} \hat{\phi}(\alpha) d\alpha \end{aligned} \quad (15.19)$$

mit $\hat{\phi}(\alpha) = \int_{\mathbb{R}} e^{-i\alpha y} \phi(y) dy$. Daher kann man den Satz von Fubini auf das $d\alpha$ und das dP -Integral des Erwartungswertes anwenden und erhält

$$\begin{aligned} \mathbb{E}[\phi_\epsilon(X_n)] &= \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}[e^{i\alpha X_n}] \exp\left\{-\frac{1}{2}(\epsilon\alpha)^2\right\} \hat{\phi}(\alpha) d\alpha \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}_{X_n}(\alpha) \exp\left\{-\frac{1}{2}(\epsilon\alpha)^2\right\} \hat{\phi}(\alpha) d\alpha. \end{aligned} \quad (15.20)$$

Man beachte nun, dass $\exp\left\{-\frac{1}{2}(\epsilon\alpha)^2\right\} d\alpha$ bis auf die Normierung eine kontinuierliche Dichte in α ist. Zudem ist $|\hat{\phi}(\alpha)| \leq \int_{\mathbb{R}} |\phi(y)| dy$ und $|\mathcal{F}_{X_n}(\alpha)| \leq 1 \quad \forall \alpha \in \mathbb{R}$. Also ist der Integrand des Maßes $\exp\left\{-\frac{1}{2}(\epsilon\alpha)^2\right\} d\alpha$ in (15.20) uniform beschränkt und nach Voraussetzung punktweise konvergent gegen $\mathcal{F}_X(\alpha)\hat{\phi}(\alpha)$. Deshalb gilt nach dem Satz über dominierte Konvergenz, siehe Satz 9.9 (iii),

$$\begin{aligned} \mathbb{E}[\phi_\epsilon(X_n)] &\longrightarrow \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}_X(\alpha) \exp\left\{-\frac{1}{2}(\epsilon\alpha)^2\right\} \hat{\phi}(\alpha) d\alpha \\ &= \mathbb{E}[\phi_\epsilon(X)] \quad \text{für } n \rightarrow \infty. \end{aligned} \quad (15.21)$$

Hier wurden zuletzt die Rechenschritte aus (15.20) in umgekehrter Reihenfolge angewendet. Dies setzt den Schlussstein auf den etwas längeren Beweis dieses Satzes. \square

15.1. Übungsaufgaben

1. Selbsttest

1. Berechnen Sie eine approximative Verteilung, welche eine Summe aus 100 i.i.d. Zufallsvariablen $X_j \sim \text{Exp}(\alpha)$ für $\alpha = 2$ approximiert.
2. Wir summieren $n = 100$ i.i.d. L^2 -Zufallsvariablen X_j mit unbekannter Verteilung und erhalten eine Zufallsvariable $X = \sum_{j=1}^n X_j$ mit Erwartungswert 100 und Varianz 200. Ist es denkbar, dass $X_j \sim \text{Po}(\lambda)$ gilt? Falls ja, bestimmen Sie λ ! Falls nein, argumentieren Sie, warum nicht!

2. Transfer

1. Die Firma Müller stellt Autoteile her. In der Produktion, Verpackung und Lieferung eines Teiles geht dieses zufällig und unabhängig von anderen Teilen mit der Wahrscheinlichkeit $p \in (0, 1)$ kaputt. Betrachten Sie für eine Lieferung von $n \in \mathbb{N}$ Teilen die folgenden unabhängigen Zufallsvariablen, für alle $i \in \mathbb{N}$ mit $1 \leq i \leq n$,

$$X_i = \begin{cases} 1 & \text{falls Teil } i \text{ kaputt} \\ 0 & \text{sonst} \end{cases}$$

- a) Welche Zufallsvariable X beschreibt die Anzahl der kaputten Teile in der Lieferung von n Teilen?
 - b) Bestimmen Sie den Erwartungswert und die Varianz von X . Wie ist X für große n approximativ verteilt?
 - c) Ein Kunde bestellt $n = 1000$ Teile (nehmen Sie $n = 1000$ als groß an und verwenden Sie b)). Außerdem sei die Defektwahrscheinlichkeit als $p = 0.1$ gegeben. Mit welcher Wahrscheinlichkeit sind in der Lieferung mehr als 950 Teile intakt? Verwenden Sie für Ihre Berechnungen die Tabellen im Anhang!
 - d) Ein anderer, sehr wichtiger und auch sehr schnell verärrgerter Kunde, benötigt unbedingt 1000 **funktionsstüchtige** Teile. Wie viele Teile müssen Sie diesem Kunden mindestens schicken, so dass dieser mit einer Wahrscheinlichkeit von 95% seine benötigten Teile erhält?
2. Die Firma **Tasaco** stellt scharfe Gewürzsaucen her. Die Chilischoten, welche der Hauptbestandteil dieser Saucen sind, bekommt **Tasaco** von n unabhängigen Zulieferern, welche im Mittel eine Schärfe von 10^5 Scoville mit einer Standardabweichung von 2×10^4 Scoville garantieren können. Um den Schärfegrad und somit die Qualität der Saucen konstant zu halten, wendet die Firma folgenden Trick an: Es werden die Chilischoten aller Zulieferer in gleichen Anteilen gemischt und daraus die Sauce produziert. Die Zufallsvariable X_j beschreibe den Schärfegrad der Chilischoten von Zulieferer j und wir nehmen an, dass die X_j i.i.d. sind.

15. Der zentrale Grenzwertsatz

- a) Welche Zufallsvariable X beschreibt den Schärfegrad der gemischten Chilischoten?
- b) Bestimmen Sie Erwartungswert und Varianz von X . Wie ist X für große n approximativ verteilt?
- c) Angenommen es gibt $n = 25$ Zulieferer (betrachten Sie n als groß). Wie groß ist die Wahrscheinlichkeit, dass das Gemisch von Chilischoten einen Schärfegrad von 1.2×10^5 Scoville überschreitet und somit die Sauce zu scharf wird?
- d) Welchen Mindestschärfegrad kann **Tasaco** zu 95% den Kunden garantieren?

3. Theorie

Wir möchten zeigen, dass sich für großes n eine binomialverteilte Zufallsvariable $X_n \sim B(n, p)$ durch eine normalverteilte Zufallsvariable $X \sim N(np, np(1-p))$ “approximieren” lässt. Die mathematisch präzise Aussage ist als der Satz von de Moivre-Laplace bekannt und lautet wie folgt: Für $X_n \sim B(n, p)$, definiere $Y_n := \frac{X_n - np}{\sqrt{npq}}$. Dann konvergiert Y_n für $n \rightarrow \infty$ schwach nach Verteilung gegen $Z \sim N(0, 1)$. Aufgrund des Satzes von Lévy (siehe Satz 15.3) ist dies äquivalent zur Aussage, dass die Fouriertransformation $\mathcal{F}_{Y_n}(\alpha)$ punktweise gegen die Fouriertransformation $\mathcal{F}_Z(\alpha)$ konvergiert. Dies wollen wir nun zeigen!

- a) Berechnen Sie die Fouriertransformation von Y_n ! *Hinweis:* Als Ergebnis sollten Sie

$$\mathcal{F}_{Y_n}(\alpha) = \left(e^{i\alpha\sqrt{\frac{1-p}{np}}} p + e^{-i\alpha\sqrt{\frac{p}{n(1-p)}}} (1-p) \right)^n \quad (15.22)$$

erhalten. Wenn Sie diese Rechnung nicht schaffen, arbeiten Sie einfach mit diesem Zwischenresultat weiter!

- b) Entwickeln Sie den Term in der Klammer in Gleichung (15.22) bis zur zweiten Ordnung in α ! Vergessen Sie hier einfach den Restterm.
- c) Führen Sie den Limes $n \rightarrow \infty$ durch und vergleichen Sie das Resultat mit der Fouriertransformation von $Z \sim N(0, 1)$.
- d) *Bonusaufgabe:* Geben Sie ein Argument warum das Vergessen des Restterms in der Taylor Entwicklung in b) legitim war.

15.2. Literatur

16. Statistische Modelle

In diesem Abschnitt definieren wir zunächst statistische Modelle. Hierbei wird feinsinnig unterschieden zwischen einer tatsächlichen, aber unbekannten Zufallsverteilung eines Merkmals X unter mehrfacher Messung und dem Modell, dass wir uns aufgrund der Beobachtungen von X von dieser Verteilung machen.

Definition 16.1. Gegeben sei der Ergebnisraum des einfachen Zufallsexperimentes, also ein messbarer Raum, (E, \mathcal{B}) und eine Folge von messbaren Abbildungen $X_j : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B})$. X_j repräsentiert hierbei das Ergebnis des Zufallsexperimentes bei der j -ten Wiederholung.

- (i) Ein (parametrisches) **statistisches Modell** $(\Omega, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \{X_j\}_{j \in \mathbb{N}})$ besteht neben den schon genannten Daten aus einer Parametermenge $\Theta \subseteq \mathbb{R}^q$ und einer Familie von W.-Maßen $\{P_\theta\}_{\theta \in \Theta}$ auf (Ω, \mathcal{A}) .
- (ii) Für $\omega \in \Omega$ betrachten wir $\chi_n(\omega) = (X_1(\omega), \dots, X_n(\omega))$ als Ergebnis einer Zufallsstichprobe vom Umfang n . $\chi_n : (\Omega, \mathcal{A}) \rightarrow (E^{\times n}, \mathcal{B}^{\otimes n})$ ist somit messbar.
- (iii) Ein statistisches Modell ist ein **Produktmodell**, falls es auf (E, \mathcal{B}) eine Familie von W.-Maßen $\{\bar{P}_\theta\}_{\theta \in \Theta}$ gibt so, dass

$$\chi_{n*} P_\theta = \bar{P}_\theta^{\otimes n} \quad \forall \theta \in \Theta$$

gilt. In einem Produktmodell sind somit die Ergebnisse der Wiederholungen X_j des Zufallsexperimentes stets i.i.d..

Bemerkung 16.2. (i) In der Statistik soll oft ein Merkmal $X : (G, \mathcal{A}_G) \rightarrow (E, \mathcal{B})$ untersucht werden, das jedem Mitglied g der Grundgesamtheit G die in der Befragung erhobenen Daten $X(g) \in E$ zuweist. Der Einfachheit halber nehmen wir an, dass G endlich ist. Mit P_G bezeichnen wir das Laplacesche W.-Maß auf $(G, \mathbb{P}(G))$. Wenn wir nun in jedem Schritt unabhängig von den anderen Schritten und ohne Präferenz, also gemäß P_G , eine zu befragende Person g auswählen, so kann diese theoretisch unbegrenzt wiederholbare **Zufallsauswahl** durch den messbaren Raum

$$\Omega = G^{\mathbb{N}} = \{\omega = (g_1, g_2, \dots) : g_j \in G\}$$

mit der σ -Algebra \mathcal{A} erzeugt von den Mengen $G \times G \times \dots \times A \times G \times \dots \subseteq \Omega$, $A \subseteq G$, modelliert werden. Das W.-Maß auf Ω ist gegeben als das Produktmaß $P^{\otimes \mathbb{N}} = \bar{P}_G \otimes \bar{P}_G \otimes \dots$. Unter diesem W.-Maß sind die Zufallsvariablen $X_j(\omega) := X_j((g_1, g_2, \dots)) = X(g_j)$ unabhängig und identisch verteilt. Essentiell dafür, dass das Produktmodell ein gutes Modell für die

16. Statistische Modelle

vorliegende Untersuchung ist, wird demnach sein, dass das Ausfallverfahren tatsächlich unabhängig und gleichverteilt erfolgt. Falls nicht, werden die Ergebnisse der statistischen Untersuchung verfälscht.

(ii) Der aufmerksamen Leserin und dem aufmerksamen Leser ist sicher nicht entgangen, dass die eben genannte Vorgehensweise nur zu einem W.-Maß P auf (Ω, \mathcal{A}) führt, und nicht zu einer Familie von W-Maßen $\{P_\theta\}_{\theta \in \Theta}$. Wir nehmen in der Folge jedoch an, dass wir $\{P_\theta\}_{\theta \in \Theta}$ so gut gewählt haben, dass $\exists \theta_0 \in \Theta$ mit $P = P_{\theta_0}$. Diese Annahme ist natürlich in der Regel zu optimistisch und ihre Verletzung führt auf das Thema des so genannten **Modellfehlers**. Dieser kann innerhalb der gewählten, parametrischen Modelle selbst nicht quantifiziert werden. An dieser Stelle sind der Abgleich mit **nicht parametrischen Schätzverfahren** und die so genannte **Modelldiagnostik** notwendig, um die Güte und Zuverlässigkeit des Modellansatzes zu beurteilen.

(iii) In der Regel wird die Familie der Wahrscheinlichkeitsmaße $\{P_\theta\}_{\theta \in \Theta}$ dann doch nicht direkt auf $(\Omega, \mathcal{A}) = G^{\mathbb{N}}$ konstruiert, sondern auf $(\Omega, \mathcal{A}) = (E^{\mathbb{N}}, \mathcal{B}^{\otimes \mathbb{N}})$, dem Raum der Ergebnisfolgen. Dies ist dadurch begründet, dass wir ja an den Individuen $g \in G$ nur in sofern interessiert sind, als dass die Merkmalsträger von $X(g)$ sind. Es liegt also nahe, das ursprüngliche W-Maß P_θ auf $G^{\mathbb{N}}$ zu identifizieren mit dem Bildmaß $X_*^{\times \mathbb{N}} P_\theta$ auf $E^{\mathbb{N}}$, wobei $X^{\times \mathbb{N}}((g_1, g_2, \dots)) = (X(g_1), X(g_2), \dots)$ jeder Auswahlfolge (g_1, g_2, \dots) die Folge der Ergebnisse der Befragung oder des Experimentes zuweist. Natürlich ist $X_*^{\times \mathbb{N}} P_\theta$ nun in der Regel nicht mehr gleichverteilt in jeder Komponente E der j -ten Wiederholung, sondern gibt korrekt die Präferenz für bestimmte Ausprägungen $e \in E$ wieder.

(iv) Beispiele für parametrische Produktmodelle mit $(E, \mathcal{B}_E) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sind z.B. Gleichverteilung $\bar{P}_\theta = U(a, b)$ ($\theta = (a, b) \in \{(s, t) \in \mathbb{R}^2, s < t\} = \Theta$), Normalverteilung $\bar{P}_\theta = N(\mu, \sigma^2)$ ($\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty) = \Theta$), Exponentialverteilung $\bar{P}_\theta = \text{Exp}(\lambda)$ ($\theta = \lambda \in (0, \infty) = \Theta$) oder die Weibullverteilung $\bar{P}_\theta = \text{Wei}(\eta, \beta)$ ($\theta = (\eta, \beta) \in (0, \infty) \times (0, \infty) = \Theta$).

Für $(E, \mathcal{B}_E) = (\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ haben wir z.B. die Binomialverteilung $\bar{P}_\theta = B(n, p)$ ($\theta = (n, p) \in \mathbb{N}_0 \times [0, 1] = \Theta$), die hypergeometrische Verteilung $\bar{P}_\theta = H(n, m, N)$ ($\theta = (n, m, N) \in \{(a, b, c) \in \mathbb{N}_0^3 : a \leq b + c\} = \Theta$), die geometrische Verteilung $\bar{P}_\theta = \text{Geo}(p)$ ($\theta = p \in (0, 1) = \Theta$) oder die Poissonverteilung $\bar{P}_\theta = \text{Poi}(\lambda)$ ($\theta = \lambda \in (0, \infty) = \Theta$).

Neben dem Modellfehler besteht bei den parametrischen statistischen Modellen das Problem, wie die Parameter geeignet bestimmt. Ist die parametrische Modellklasse erst einmal gewählt, so kann man mit Hilfe von Schätzverfahren die freien Parameter bestimmen. Dies geschieht mit Hilfe von **Schätzern**, die nun als Nächstes behandelt werden.

Definition 16.3. Es sei $(\Omega, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \{X_j\}_{j \in \mathbb{N}})$ ein parametrisches statistisches Modell mit Ergebnisraum (E, \mathcal{B}) , $\Theta \subseteq \mathbb{R}^q$.

- (i) Ein **Parameterschätzer** $S = \{S^{(n)}\}_{n \in \mathbb{N}}$ für den Parameter θ_j , $\theta = (\theta_1, \dots, \theta_q)$, besteht aus einer Familie von messbaren Abbildungen

$$S^{(n)} : (E^{\times n}, \mathcal{B}^{\otimes n}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) . \quad (16.1)$$

- (ii) Der Schätzer S heißt **erwartungstreu** für den Parameter θ_j , falls $\forall \theta \in \Theta$ und $n \in \mathbb{N}$,

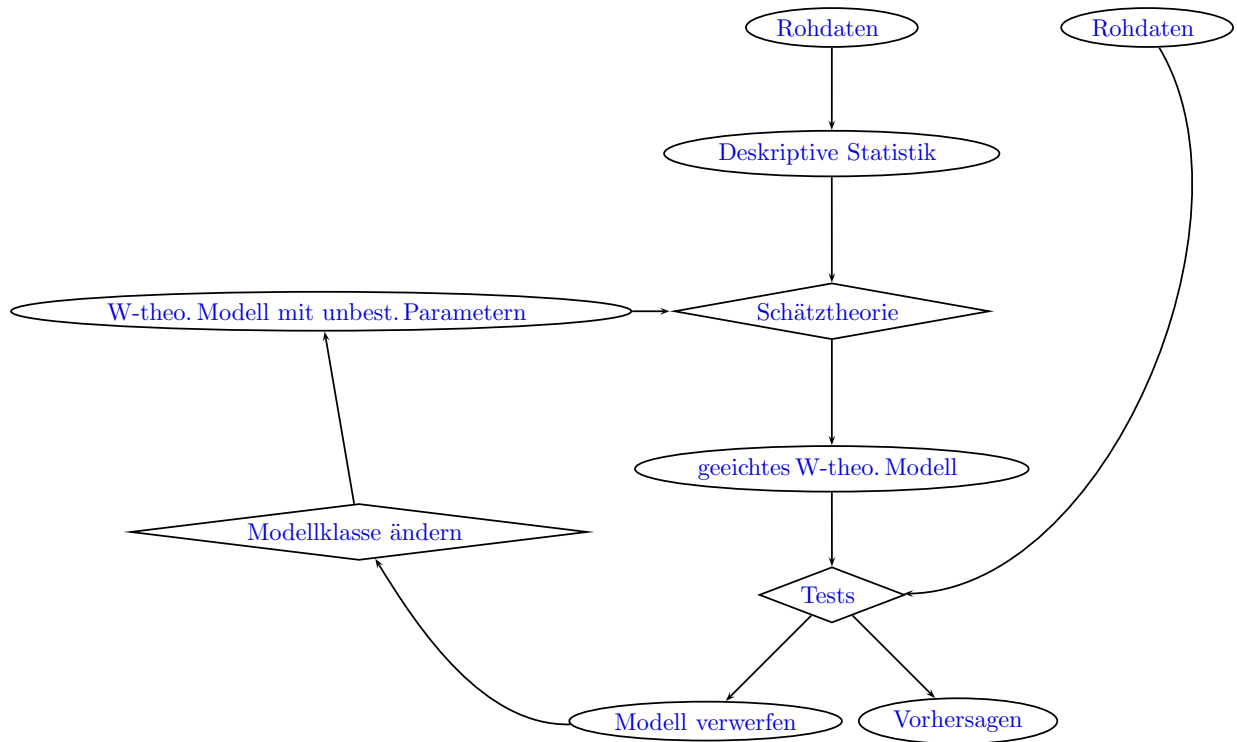


Abbildung 16.1: Parametrische statistische Modelle.

$S^{(n)}(X_1, \dots, X_n) \in L^1(\Omega, \mathcal{A}, P_\theta; \mathbb{R})$ und

$$\mathbb{E}_\theta [S^{(n)}(X_1, \dots, X_n)] = \int_{E^{\times n}} S^{(n)}(x_1, \dots, x_n) dP_\theta^{(n)}(x_1, \dots, x_n) = \theta_j, \quad (16.2)$$

wobei $\mathbb{E}_\theta[\cdot]$ den Erwartungswert bezüglich P_θ bezeichnet und $P_\theta^{(n)} = \chi_{n*} P_\theta = (X_1, \dots, X_n)_* P_\theta$.

- (iii) Ein Schätzer heißt **asymptotisch erwartungstreu** für den Parameter θ_j , falls $\forall \theta \in \Theta$, $S^{(n)}(X_1, \dots, X_n) \in L^1((\Omega, \mathcal{A}, P_\theta); \mathbb{R})$ und

$$\mathbb{E}_\theta [S^{(n)}(X_1, \dots, X_n)] \longrightarrow \theta_j \quad \text{für } n \rightarrow \infty. \quad (16.3)$$

- (iv) Ein Schätzer heißt **konsistent** für den Parameter θ_j , falls $\forall \theta \in \Theta$ und $n \in \mathbb{N}$ sowie für alle $X_j : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{B}_E)$, $j = 1, \dots, n$ so, dass $(X_1, \dots, X_n) \sim P_\theta^{(n)}$

$$S^{(n)}(X_1, \dots, X_n) \longrightarrow \theta_j \quad \text{nach W.-keit für } n \rightarrow \infty. \quad (16.4)$$

- (v) Ein **Mehrparameterschätzer** $S = \{S^{(n)}\}_{n \in \mathbb{N}}$ für die Parameter $(\theta_1, \dots, \theta_q)$, $\theta = (\theta_1, \dots, \theta_q)$, besteht aus einer Familie von messbaren Abbildungen

$$S^{(n)} : (E^{\times n}, \mathcal{B}^{\otimes n}) \rightarrow (\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q)). \quad (16.5)$$

S heißt (asymptotisch) erwartungstreu bzw. konsistent für den Parametervektor $\theta \in \Theta$, falls die j -te Komponente von $S^{(n)}$ diese Eigenschaft für θ_j aufweist für alle $j = 1, \dots, q$.

16. Statistische Modelle

Ein mit dem Parameterschätzer eng verwandter Begriff ist die **Statistik**. Mathematisch ist eine Statistik genau so definiert wie ein Schätzer. Der einzige Unterschied besteht in einem anderen Zweck. Eine Statistik ist nicht unbedingt daran gebunden, einen bestimmten Parameter zu schätzen, sondern kann irgendeine theoretische Größe (Erwartungswert, Varianz etc.) ohne Bezug auf Modellparameter schätzen. Im Folgenden unterscheiden wir nicht zwischen Parameterschätzer und Statistik. Wir verwenden insbesondere Ausdrücke wie

$$\mathbb{E}_\mu[S^{(n)}(X_1, \dots, X_n)] \quad \text{bzw.} \quad \mathbb{E}_{\sigma^2}[S^{(n)}(X_1, \dots, X_n)] \quad (16.6)$$

für Erwartungswerte mittels beliebiger W.-Maße P_θ so dass $\mathbb{E}_\theta[X_j] = \mu$ und $\text{Var}_{P_\theta}[X_j] = \sigma^2$, $j \in \mathbb{N}$.

Das einfachste Beispiel für einen 'guten' Schätzer ist das arithmetische Mittel.

Satz 16.4. *Es sei $(\Omega, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \{X_j\}_{j \in \mathbb{N}})$ ein parametrisches statistisches Modell wie oben mit Ergebnisraum $(E, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ so dass $X_j \in L^1((\Omega, \mathcal{A}, P_\theta); \mathbb{R}^d) \quad \forall \theta \in \Theta$.*

Das arithmetische Mittel $S^{(n)}(x_1, \dots, x_n) = \bar{x}_n = \frac{1}{n}(x_1 + \dots + x_n)$ hat als Schätzer über dem Ergebnisraum $(E, \mathcal{B}_E) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ folgende Eigenschaften:

(i) *Das arithmetische Mittel erwartungstreu Schätzer für den Erwartungswert $\mu(\theta) = \mathbb{E}_\theta[X_1] = \mathbb{E}_\mu[X_1]$, d.h.*

$$\mathbb{E}_\mu[\bar{X}_n] = \mu \quad \forall \mu \in \mathbb{R}^d.$$

(ii) *In jedem beliebigen Produktmodell mit $X_j \in L^2((\Omega, \mathcal{A}, P_\theta); \mathbb{R}^d) \quad \forall \theta \in \Theta$ ist das arithmetische Mittel ein konsistenter Schätzer für den Erwartungswert μ .*

(iii) *Im Bernoulli-Produktmodell mit $\bar{P}_p = B(1, p)$ über $(E, \mathcal{B}) = (\{0, 1\}, \mathbb{P}(\{0, 1\}))$ mit $X_j \sim B(1, p)$ ist das arithmetische Mittel ein konsistenter und erwartungstreu Schätzer für den Anteilswert p .*

Beweis. (i) Wir betrachten $(X_1, \dots, X_n) \sim P_\mu^{(n)}$ in einem Modell wie in den Voraussetzungen beschrieben. Linearität des Erwartungswertes impliziert dann

$$\mathbb{E}_\mu[S^{(n)}(X_1, \dots, X_n)] = \frac{1}{n} \sum_{j=1}^n \mathbb{E}_\mu[X_j] = \frac{1}{n} n \mu = \mu. \quad (16.7)$$

(ii) Die Konsistenz im Produktmodell ist nichts Anderes als das schwache Gesetz der großen Zahlen Satz 14.3.

(iii) Wir haben $\mathbb{E}_p[X_1] = p \cdot 1 + (1-p) \cdot 0 = p$. Da $|X_1| \leq 1$ ist $X_1 \in L^q(\Omega, \mathcal{A}, P_p; \mathbb{R})$ für alle $p \in [0, 1]$ und $q \geq 1$. Also gelten die Voraussetzungen von (i) bzw. (ii) und die Aussage (iii) folgt aus diesen beiden Punkten. \square

Ein 'natürlicher' Ansatz für die Schätzung der Varianz wäre $\bar{x}^2 - \bar{x}^2$. Doch es stellt sich heraus, dass dieser Schätzer nicht erwartungstreu ist. Deswegen definiert man die empirische Varianz und Standardabweichung wie folgt:

Definition 16.5. (i) Der Schätzer $\hat{\sigma}^2 = \{\hat{\sigma}^{2(n)}\}_{n \in \mathbb{N}}$, $\hat{\sigma}^{2(n)} : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, für die Varianz σ^2 ist definiert als

$$\hat{\sigma}^2 = \hat{\sigma}^2(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{n}{n-1} (\bar{x}^2 - \bar{x}^2) \quad (16.8)$$

und heißt **empirische Varianz**.

(ii) Der Schätzer $\hat{\sigma} = \{\hat{\sigma}^{(n)}\}_{n \in \mathbb{N}}$ definiert durch $\hat{\sigma}^{(n)} = \sqrt{\hat{\sigma}^{2(n)}}$ heißt **empirische Standardabweichung**. Explizit haben wir

$$\hat{\sigma} = \hat{\sigma}(x_1, \dots, x_n) = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2} = \sqrt{\frac{n}{n-1} (\bar{x}^2 - \bar{x}^2)} . \quad (16.9)$$

Satz 16.6. (i) In jedem beliebigen Produktmodell mit $X_j \in L^2(\Omega, \mathcal{A}, P_\theta; \mathbb{R}^d) \forall \theta \in \Theta$ ist die empirische Varianz ein erwartungstreu Schätzer für die Varianz σ^2 .

(ii) In jedem beliebigen Produktmodell¹ mit $X_j \in L^4(\Omega, \mathcal{A}, P_\theta; \mathbb{R}^d) \forall \theta \in \Theta$ ist die empirische Varianz ein konsistenter Schätzer für die Varianz σ^2 .

Beweis. (i) Durch direkte Rechnung sehen wir unter Zuhilfenahme von $\text{Cov}[X_i, X_j] = 0$ für $i \neq j$

$$\begin{aligned} \mathbb{E}_{(\mu, \sigma^2)}[\hat{\sigma}^2(X_1, \dots, X_n)] &= \mathbb{E}_{(\mu, \sigma^2)} \left[\frac{1}{n-1} \sum_{j=1}^n [(X_j - \bar{X})]^2 \right] \\ &= \frac{1}{n-1} \mathbb{E}_{(\mu, \sigma^2)} \left[\sum_{j=1}^n \left[(X_j - \mu) - \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \right]^2 \right] \\ &= \frac{1}{n-1} \left\{ \sum_{j=1}^n \mathbb{E}_{(\mu, \sigma^2)} [(X_j - \mu)^2] \right. \\ &\quad - \frac{2}{n} \sum_{i,j=1}^n \mathbb{E}_{(\mu, \sigma^2)} [(X_i - \mu)(X_j - \mu)] \\ &\quad \left. + \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}_{(\mu, \sigma^2)} [(X_i - \mu)(X_j - \mu)] \right\} \\ &= \frac{1}{n-1} \left\{ \sum_{j=1}^n \text{Var}_{\sigma^2}[X_j] - \frac{1}{n} \sum_{i,j=1}^n \text{Cov}_{\sigma^2}[X_i, X_j] \right\} \\ &= \frac{1}{n-1} \left\{ n\sigma^2 - \frac{1}{n} n\sigma^2 \right\} = \sigma^2 . \end{aligned} \quad (16.10)$$

¹Tatsächlich reicht hier L^2 schon aus, vergl. die Bemerkungen zum Starken Gesetz der gr. Zahlen!

16. Statistische Modelle

(ii) Wir schreiben $\hat{\sigma}^2(X_1, \dots, X_n)$ als

$$\begin{aligned}\hat{\sigma}^2(X_1, \dots, X_n) &= \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2 \\ &= \frac{n}{n-1} \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 \\ &= \frac{n}{n-1} (\overline{X^2} - \bar{X}^2) .\end{aligned}\tag{16.11}$$

Sind die wahren Parameter μ und σ^2 gegeben, dann konvergiert \bar{X}_n nach dem starken Gesetz der großen Zahlen $P_{(\mu, \sigma^2)}$ -fast sicher gegen $\mu = \mathbb{E}_\mu[X]$. Also konvergiert \bar{X}^2 $P_{(\mu, \sigma^2)}$ -fast sicher gegen μ^2 . Weiter können wir das starke Gesetz der großen Zahlen auch auf $\overline{X^2}_n$ anwenden und erhalten $P_{(\mu, \sigma^2)}$ -fast sichere Konvergenz dieses Ausdrucks gegen $\mathbb{E}_{(\mu, \sigma^2)}[X^2]$. Also

$$\frac{n}{n-1} (\overline{X^2} - \bar{X}^2) \longrightarrow \mathbb{E}_{(\mu, \sigma^2)}[X^2] - \mathbb{E}_{(\mu, \sigma^2)}[X]^2 = \sigma^2 \quad \text{für } n \rightarrow \infty \text{ } P_{(\mu, \sigma^2)}\text{-fast sicher} .\tag{16.12}$$

Dies impliziert die Konvergenz nach W.-keit. \square

Da der Erwartungswert linear ist, wird Erwartungstreue unter linearen Transformationen erhalten. Die nicht lineare Transformation eines erwartungstreuen Schätzers ist jedoch im Allgemeinen nicht erwartungstreu. Als Beispiel betrachten wir die empirische Standardabweichung, die als Wurzel aus der (erwartungstreuen) empirischen Varianz definiert ist. Nach der Ungleichung von Jensen gilt

$$\begin{aligned}\mathbb{E}_{(\mu, \sigma^2)}[\hat{\sigma}(X_1, \dots, X_n)] &= \mathbb{E}_{(\mu, \sigma^2)}[\sqrt{\hat{\sigma}^2(X_1, \dots, X_n)}] \\ &\leq \sqrt{\mathbb{E}_{(\mu, \sigma^2)}[\hat{\sigma}^2(X_1, \dots, X_n)]} \\ &= \sqrt{\sigma^2} = \sigma .\end{aligned}\tag{16.13}$$

Also wird die Standardabweichung systematisch unterschätzt.

Ist ein Schätzer nicht erwartungstreu, d.h. über- oder unterschätzt er den wahren Parameterwert systematisch, so ist er **verzerrt** (engl. *biased*).

Definition 16.7. Es seien $S = \{S^{(n)}\}_{n \in \mathbb{N}}$ und $T = \{T^{(n)}\}_{n \in \mathbb{N}}$ Schätzer für den Parameter θ_j in einem statistischen Modell $(\Omega, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \{X_j\}_{j \in \mathbb{N}})$.

(i) Der **Bias** des Schätzers S ist

$$\text{Bias}_\theta[S^{(n)}] = \mathbb{E}_\theta[S^{(n)}(X_1, \dots, X_n)] - \theta_j .\tag{16.14}$$

(ii) Der **mittlere quadratische Fehler** (engl. *mean square error – MSE*) ist

$$\text{MSE}_\theta[S^{(n)}] = \mathbb{E}_\theta \left[(S^{(n)}(X_1, \dots, X_n) - \theta_j)^2 \right] .\tag{16.15}$$

(iii) Ein Schätzer S ist **effizient** als ein Schätzer T falls

$$\text{MSE}_\theta[S^{(n)}] \leq \text{MSE}_\theta[T^{(n)}] \quad \forall n \in \mathbb{N}, \theta \in \Theta. \quad (16.16)$$

Lemma 16.8. *Es sei $S = \{S^{(n)}\}_{n \in \mathbb{N}}$ ein Schätzer für den Parameter θ_j in einem statistischen Modell $(\Omega, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \{X_j\}_{j \in \mathbb{N}})$.*

(i) Dann gilt

$$\text{MSE}_\theta[S^{(n)}] = \text{Var}_\theta[S^{(n)}] + \text{Bias}_\theta[S^{(n)}]^2. \quad (16.17)$$

(ii) Konvergiert der MSE für $n \rightarrow \infty$ gegen 0 $\forall \theta \in \Theta$ fest, dann ist S konsistent.

(iii) Ein Schätzer S ist insbesondere dann konsistent, wenn er asymptotisch erwartungstreu ist und $\text{Var}_\theta[S^{(n)}] \rightarrow 0$ für $n \rightarrow \infty \quad \forall \theta \in \Theta$.

Beweis. (i) Wir rechnen nach, indem wir für $S^{(n)}(X_1, \dots, X_n)$ die Abkürzung $S^{(n)}$ verwenden

$$\begin{aligned} \text{MSE}_\theta[S^{(n)}] &= \mathbb{E}_\theta \left[(S^{(n)} - \theta_j)^2 \right] \\ &= \mathbb{E}_\theta \left[(S^{(n)} - \mathbb{E}[S^{(n)}] + \mathbb{E}[S^{(n)}] - \theta_j)^2 \right] \\ &= \text{Var}_\theta[S^{(n)}] - 2\text{Bias}_\theta[S^{(n)}]\mathbb{E}_\theta[(S^{(n)} - \mathbb{E}[S^{(n)}])] + \text{Bias}_\theta[S^{(n)}]^2 \\ &= \text{Var}_\theta[S^{(n)}] + \text{Bias}_\theta[S^{(n)}]^2. \end{aligned} \quad (16.18)$$

(ii) Nach (i) folgt aus $\text{MSE}_\theta[S^{(n)}] \rightarrow 0 \quad \forall \theta \in \Theta$, dass $\text{Var}_\theta[S^{(n)}] \rightarrow 0$ und ebenfalls $\text{Bias}_\theta[S^{(n)}] \rightarrow 0 \quad \forall \theta \in \Theta$. Also gilt für alle $\theta \in \Theta$

$$\begin{aligned} P_\theta(|S^{(n)} - \theta_j| > \epsilon) &\leq P_\theta\left(|S^{(n)} - \mathbb{E}_\theta[S^{(n)}]| > \frac{\epsilon}{2}\right) + P_\theta\left(|\text{Bias}_\theta[S^{(n)}]| > \frac{\epsilon}{2}\right) \\ &\leq \frac{4}{\epsilon^2} \text{Var}_\theta[S^{(n)}] + 1_{\{|\text{Bias}_\theta[S^{(n)}]| > \frac{\epsilon}{2}\}}(\theta). \end{aligned} \quad (16.19)$$

Da $\text{Bias}_\theta[S^{(n)}] \rightarrow 0$ existiert ein n_0 so dass der zweite Summand auf der rechten Seite für alle $n > n_0$ identisch verschwindet. Dann konvergiert aber wegen $\text{Var}_\theta[S^{(n)}] \rightarrow 0$ die ganze rechte Seite gegen 0.

(iii) Asymptotische Erwartungstreue bedeutet gerade $\text{Bias}_\theta[S^{(n)}] \rightarrow 0$ für alle $\theta \in \Theta$. Gilt zusätzlich $\text{Var}_\theta[S^{(n)}] \rightarrow 0 \quad \forall \theta \in \Theta$, dann verschwindet nach nach (i) der MSE und nach (ii) folgt daraus die Konsistenz. \square

Zum Abschluss dieses Abschnitts soll noch die **Methode der Momente** beschrieben werden, mit der parametrische Modelle mittels der Daten 'geeicht' werden können. Hierbei berechnet man in einem parametrischen Modell mit ausschließlich kontinuierlichen Parametern $\theta \in \Theta \subseteq \mathbb{R}^q$ offen die ersten q Momente empirisch und als Funktionen von θ im parametrischen Modell

$$\mathcal{M}_{\theta,p}[X] = \mathbb{E}_\theta[X^p]. \quad (16.20)$$

16. Statistische Modelle

Hier sind die Ausdrücke links und in der Mitte durch die rechte Seite definiert.

Im nächsten Schritt setzen wir die ersten q empirischen und theoretischen Momente gleich

$$\begin{aligned}\bar{X}_n &= \mathcal{M}_{\theta,1}[X] \\ \overline{X^2}_n &= \mathcal{M}_{\theta,2}[X] \\ &\vdots \\ \overline{X^q}_n &= \mathcal{M}_{\theta,q}[X] .\end{aligned}\tag{16.21}$$

Die Ausdrücke auf der linken Seite können wir jeweils aus den Stichprobendaten errechnen. Indem wir diese q -Gleichungen nach $\theta \in \Theta \in \mathbb{R}^q$ auflösen (wenn das möglich ist), erhalten wir einen Schätzwert $\hat{\theta}$ als Lösung dieses Gleichungssystems.

Wenn die Gleichungen, die $\mathcal{M}_{\theta,p}[X]$ als Funktionen von θ ausdrücken stark nichtlinear sind oder Momente hoher Ordnung > 4 auftreten ist die Methode der Momente jedoch nicht zu empfehlen, da solche Momente sehr stark von einzelnen (großen) Werten abhängen. Damit ist die Methode der Momente etwas, das zwar im Prinzip funktioniert (nach dem starken Gesetz der großen Zahlen konvergieren auch die empirischen Momente, denn dies sind ja auch arithmetische Mittelwerte), jedoch nicht sonderlich praxistauglich ist. Für die Normalverteilung ist das Verfahren jedoch durchaus o.k., obwohl es auch hier besser geht, da der Varianzschätzer aus der Methode der Momente nicht erwartungstreu ist.

Beispiel 16.9. Wir erproben die Methode der Momente anhand von zehn normalverteilten Zufallszahlen, die mittels eines Zufallszahlengenerators zu den Parametern $\mu = 2$ und $\sigma = 2$ ($\sigma^2 = 4$) erzeugt wurden. Außerdem runden wir auf die 2. Nachkommastelle. Die Zahlen sind wie folgt:

	1	2	3	4	5	6	7	8	9	10
x	1.60	-0.17	2.33	1.97	2.44	-1.40	-0.45	1.20	4.32	-1.80

Im nächsten Schritt berechnen wir die Werte für \bar{x} und $\bar{x^2}$:

	1	2	3	4	5	6	7	8	9	10	mean
x	1.60	-0.17	2.33	1.97	2.44	-1.40	-0.45	1.20	4.32	-1.80	1.004
x^2	2.56	0.03	5.43	3.88	5.95	1.96	0.20	1.44	18.66	3.24	4.336

Nun stellen wir die Momentengleichungen auf

$$\begin{aligned}\mu &= \mathbb{E}_{(\mu,\sigma^2)}[X] = \bar{X}_n = 1.004 , \\ \mu^2 + \sigma^2 &= \mathbb{E}_{(\mu,\sigma^2)}[X^2] = \overline{X^2}_n = 4.336 ,\end{aligned}\tag{16.22}$$

und lösen $\hat{\mu} = 1.004$ und $\hat{\sigma}^2 = 4.336 - 1.004^2 = 3.277$. Wie bei einer so kleinen Stichprobe zu erwarten, sind die geschätzten Werte weder richtig passend noch völlig daneben.

Satz 16.10. *Es sei $(\Omega, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \{X_j\}_{j \in \mathbb{N}})$ ein statistisches Produktmodell mit Ergebnisraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Es gelte $\Theta \subseteq \mathbb{R}^q$, die Menge $\mathcal{M} = \{(\mathcal{M}_{\theta,1}[X], \dots, \mathcal{M}_{\theta,q}[X]) : \theta \in \Theta\} \subseteq \mathbb{R}^q$ sei offen und es existiere nur eine eindeutig bestimmte, stetige Lösungsabbildung zu (16.21), $\mathbb{M}^{-1} : \mathcal{M} \rightarrow \Theta$. Des Weiteren gelte $X_j \in L^{2q}(\Omega, \mathcal{A}, P_\theta; \mathbb{R})$.*

Dann ist der Multiparameterschätzer nach der Methode der Momente

$$\hat{\theta}_{\text{MM}}(X_1, \dots, X_n) = \mathbb{M}^{-1}(\overline{X}, \overline{X^2}, \dots, \overline{X^q}) \quad (16.23)$$

konsistent.

Beweis. Nach dem starken Gesetz der großen Zahlen, angewendet auf $Y_j = X_j^l$, $l = 1, \dots, q$, gilt für $\theta \in \Theta$ beliebig

$$\overline{X_n^l} \longrightarrow \mathcal{M}_{\theta,l}[X] \quad \text{für } n \rightarrow \infty \quad \text{fast sicher.} \quad (16.24)$$

Da \mathcal{M} offen ist, gilt fast sicher, dass, für hinreichend große n , $(\overline{X_n}, \dots, \overline{X_n^q}) \in \mathcal{M}$. Daher macht (16.23) für solche n Sinn und wegen der vorausgesetzten Stetigkeit von \mathbb{M}^{-1} gilt

$$\mathbb{M}^{-1}(\overline{X_n}, \overline{X_n^2}, \dots, \overline{X_n^q}) \longrightarrow \mathbb{M}^{-1}(\mathcal{M}_{\theta,1}[X], \dots, \mathcal{M}_{\theta,q}[X]) = \theta \quad (16.25)$$

fast sicher für $n \rightarrow \infty$. □

16. Statistische Modelle

1. Übungsaufgaben

2. Literaturliste

17. Konfidenzintervalle

Wenn man einen Parameter mit einem konsistenten Schätzer annäherungsweise bestimmen möchte, muß man zu hinreichend großen Stichproben greifen. Da der Grenzübergang für den Stichprobenumfang gegen unendlich in der Praxis nicht durchführbar ist, stellt sich die Frage, wie groß eine Stichprobe sein muss, damit eine hinreichende Genauigkeit in der Schätzung erzielt werden kann. Umgekehrt ist von Interesse, mit welchem Schätzfehler man bei gegebenem Stichprobenumfang rechnen muss.

Bei der Beantwortung dieser Fragen muß man berücksichtigen, dass zumindest theoretisch auch bei einem konsistenten Schätzer bei endlichem Stichprobenumfang erhebliche Fehlschätzungen nicht ausgeschlossen werden können – sie sind nur eben sehr selten. Aus diesem Grund gibt man eine niedrige Wahrscheinlichkeit $\alpha \in (0, 1)$ vor und konstruiert dann einen Bereich um den Schätzwert herum, so dass der wahre Wert zu $1 - \alpha$ im Inneren dieses Bereiches liegt und nur mit W.-keit α außerhalb.

Definition 17.1. Es sei $(\Omega, \mathcal{A}, \{P_\theta\}_{\theta \in \Theta}, \{X_j\}_{j \in \mathbb{N}})$ ein statistisches Modell über dem Ergebnisraum (E, \mathcal{B}_E) , vgl. Definition 16.1, und θ_j die j -te Komponente von $\theta = (\theta_1, \dots, \theta_q)$.

- (i) Es seien S_-, S_+ zwei Schätzer im Sinne von Definition 16.3 (i) und $\alpha \in (0, 1)$ eine (kleine) Irrtumswahrscheinlichkeit. Falls $S_- \leq S_+$ und

$$\begin{aligned} P_\theta(\theta_j \in [S_-, S_+)) &= P_\theta(\{\theta_j \in [S_-(X_1, \dots, X_n), S_+(X_1, \dots, X_n))\}) \\ &= 1 - \alpha \quad \forall \theta \in \Theta, \end{aligned} \quad (17.1)$$

dann heißt das Paar S_-, S_+ ein **Intervallschätzer** für θ_j zum Konfidenzniveau $1 - \alpha$.

- (ii) Das zufallsabhängige Intervall

$$[S_-(X_1, \dots, X_n), S_+(X_1, \dots, X_n)) \quad (17.2)$$

nennen wir auch **Konfidenzintervall** für θ_j zum Konfidenzniveau $(1 - \alpha)$.

Nun sollen zunächst Konfidenzintervalle für den Erwartungswert μ konstruiert werden. Wir beginnen mit einem besonders einfachen Fall:

Satz 17.2. Es seien X_j normalverteilte i.i.d. Zufallsvariablen $X_j \sim N(\mu, \sigma^2)$, wobei die Varianz σ^2 als bekannt vorausgesetzt wird und $\theta = \mu \in \mathbb{R}$ den im Produktmodell zu schätzenden Parameter darstellt. Mit z_p , dem p -Quantil der Standardnormalverteilung, erhält man

17. Konfidenzintervalle

- (i) Das symmetrische, beidseitige Konfidenzintervall für den Erwartungswert μ mit gegebener Varianz ist definiert durch

$$S_{\pm}(X_1, \dots, X_n) = \bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (17.3)$$

und es ist ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha$ im Sinne von Definition 17.1.

- (ii) Das linksoffene, einseitige Konfidenzintervall für den Erwartungswert μ mit gegebener Varianz ist definiert durch

$$S_- = -\infty, \quad S_+(X_1, \dots, X_n) = \bar{X} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \quad (17.4)$$

und es ist ebenfalls ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha$. Gleiches gilt für das rechtsoffene, einseitige Konfidenzintervall

$$S_+ = \infty, \quad S_-(X_1, \dots, X_n) = \bar{X} - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}. \quad (17.5)$$

Beweis. (i) Für unabhängige Z.V. $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ gilt $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. Indem wir dies induktiv anwenden, sehen wir $n\bar{X} \sim N(n\mu, n\sigma^2)$ und daher $\bar{X} \sim N(\mu, \frac{1}{n}\sigma^2)$. Wir beweisen nun (i) wie folgt

$$P\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = P\left(Z - z_{1-\alpha/2} \leq 0 < Z + z_{1-\alpha/2}\right) \quad (17.6)$$

mit $\mu = \mathbb{E}[X_1]$ dem wahren Erwartungswert. Hier haben wir \bar{X} standardisiert zu $Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$. Wegen $-z_{1-\alpha/2} = z_{\alpha/2}$ erhalten wir mit F der Verteilungsfunktion der Standardnormalverteilung

$$P\left(Z - z_{1-\alpha/2} \leq 0 < Z + z_{1-\alpha/2}\right) = F(z_{1-\alpha/2}) - F(-z_{1-\alpha/2}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha. \quad (17.7)$$

- (ii) Wir rechnen wiederum mittels Standardisieren für das linksoffene Konfidenzintervall

$$\begin{aligned} P\left(\mu < \bar{X} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}\right) &= P(0 < Z + z_{1-\alpha}) = P(-z_{1-\alpha} < Z) \\ &= P(z_{\alpha} < Z) = 1 - P(Z \leq z_{\alpha}) = 1 - \alpha. \end{aligned} \quad (17.8)$$

Die Rechnung für das rechtsoffene Konfidenzintervall verläuft analog. \square

Beidseitige Konfidenzintervalle geben einen Vertrauensbereich, indem man mit sehr großer W.-keit den wahren Wert findet. Man kann ihre (halbierte) Breite als typischen Schätzfehler auffassen.

Anders liegt der Anwendungsfall bei den halboffenen Konfidenzintervallen. Hier möchte man eine obere bzw. untere Schranke angeben, die von dem unbekannten Erwartungswert μ nur mit der Fehler-W.-keit α über- bzw. unterschritten wird.

Typische Werte für α sind $\alpha = 10\%, 5\%, 1\%$ oder 0.1% , was zu Konfidenzniveaus von $90\%, 95\%, 99\%$ bzw. 99.9% korrespondiert.

Beispiel 17.3. Eine Abfüllanlage füllt Limonade in Flaschen, dabei kommt es zu einer Ungenauigkeit (Standardabweichung) von 0.01l. Die normale Füllmenge soll eigentlich 0.75l betragen, hängt aber natürlich von der Maschineneinstellung ab. Der Produzent will wissen, ob die tatsächliche durchschnittliche Füllmenge unter 0.75l liegt, denn er will ja schließlich nichts verschenken. Dazu lässt er eine Stichprobe von 10 Flaschen entnehmen. Der Durchschnittswert der Füllmengen beträgt $\bar{x} = 0.749$. Kann sich der Produzent seiner Sache nun zu 99% sicher sein?

Lösung. Wir setzen wieder eine Normalverteilung der Füllmenge X mit bekanntem $\sigma_X = 0.01$ und unbekanntem μ_X voraus. Des Weiteren berechnen wir hier ein linksoffenes Konfidenzintervall, denn der Produzent will eine abgesicherte Aussage über eine obere Schranke für die durchschnittliche Füllmenge. Mit $z_{0.99} = 2.33$ haben wir, dass zu 99% der wahre Wert μ_X kleiner ist als

$$\bar{x} + z_{0.99} \frac{\sigma_X}{\sqrt{n}} = 0.749 + 2.33 \frac{0.01}{\sqrt{10}} = 0.756 . \quad (17.9)$$

Das reicht zur Beantwortung der Frage mit der geforderten Genauigkeit nicht aus. Zwar ist der gemessene Wert unterhalb der 0.75l und sogar unterhalb der 0.75l, doch ist der Term für die Berücksichtigung der statistische Streuung noch so groß, dass das Konfidenzintervall die 0.75l mit einschließt. Hier sollte eine größere Stichprobe gezogen werden, um Klarheit zu schaffen. \diamond

Wir betrachten als nächstes den Fall, dass im Gauss-Produktmodell σ^2 ebenfalls unbekannt ist und aus den empirischen Werten mittels der empirischen Varianz $\hat{\sigma}^2$ geschätzt wird. Man sollte vermuten, dass die statistische Fluktuation des Varianzschätzers zur Vergrößerung der Konfidenzbereiche beiträgt und dies ist auch tatsächlich der Fall. Im ersten Schritt beschreiben wir diese Fluktuationen wahrscheinlichkeitstheoretisch.

Definition 17.4. Es seien X_1, \dots, X_n i.i.d. und standardnormalverteilt, $X_j \sim N(0, 1)$. Definiere die Zufallsvariable

$$K^2 = K^2(X_1, \dots, X_n) = \sum_{j=1}^n X_j^2 . \quad (17.10)$$

Dann ist die χ^2 -Verteilung mit n Freiheitsgraden definiert als $\text{Vert}(K^2)$.

In Abbildung 17.1 (a) sieht man die Dichten einiger χ^2 -Verteilungen. Im Prinzip sind die χ^2 -Verteilungen alte Bekannte aus Definition 8.14.

Lemma 17.5. Die χ^2 -Verteilung mit n Freiheitsgraden stimmt mit der Gammaverteilung zu den Parametern $b = \frac{1}{2}$ und $\beta = \frac{n}{2}$ überein.

Beweis. Unter den Voraussetzungen von Definition 17.4 ist der Zufallsvektor $X = (X_1, \dots, X_n)$ multivariat normalverteilt mit Erwartungswert 0 und Kovarianzmatrix die Identitätsmatrix in \mathbb{R}^n , vergleiche Übungsblatt 10, Aufgabe 2. Wir haben in Vektornotation

17. Konfidenzintervalle

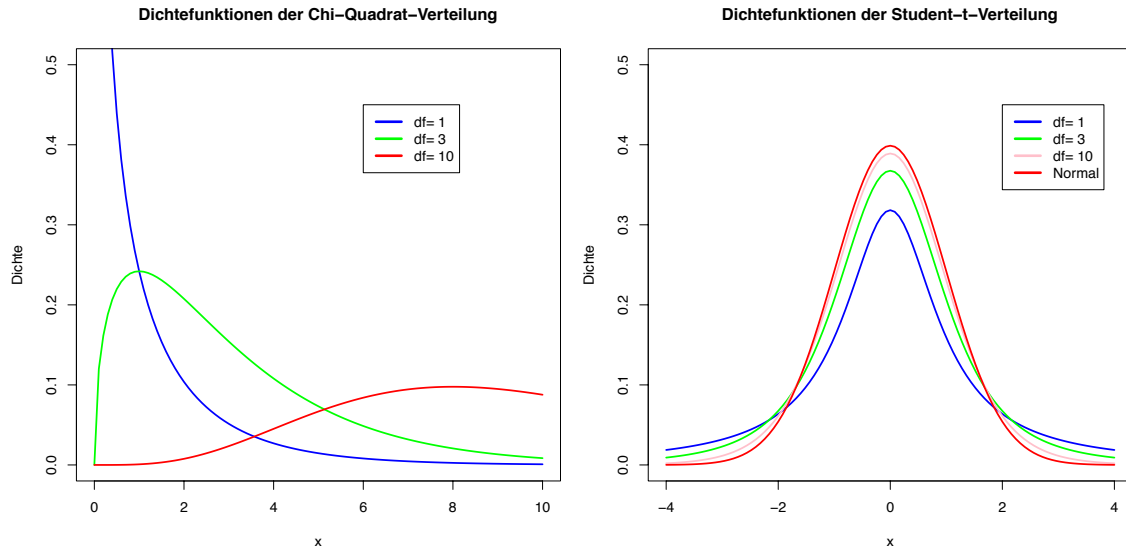


Abbildung 17.1: Dichten der χ^2 und der t -Verteilung.

$K^2 = X'X = \|X\|^2$ mit $\|\cdot\|$ der Euklidischen Norm auf \mathbb{R}^n . Wir berechnen daher, unter Verwendung von n -dimensionalen Kugelkoordinaten und der bekannten Formel für die Oberfläche der $(n-1)$ dimensionalen Sphäre $S^{(n-1)}$, den folgenden Ausdruck für die Verteilungsfunktion der $\chi^2(n)$ -Verteilung

$$\begin{aligned}
 F_{\chi^2(n)}(\lambda) &= \int_{\{x \in \mathbb{R}^n : \|x\|^2 \leq \lambda\}} f_X(x) dx \\
 &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\{x \in \mathbb{R}^n : \|x\|^2 \leq \lambda\}} e^{-\frac{1}{2}\|x\|^2} dx \\
 &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{S^{(n-1)}} dA \int_0^{\sqrt{\lambda}} e^{-\frac{1}{2}r^2} r^{n-1} dr \\
 &= \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{1}{2} \int_0^\lambda e^{-\frac{1}{2}\rho} \rho^{\frac{n}{2}-1} d\rho. \tag{17.11}
 \end{aligned}$$

Im letzten Schritt haben wir zudem den Koordinatenwechsel $r \rightarrow \rho = r^2$ verwendet. Durch geeignetes Kürzen und Differentiation nach λ ergibt sich nun die Behauptung, da die so gewonnene Dichte der $\chi^2(n)$ -Verteilung mit der Dichtefunktion der $(\frac{1}{2}, \frac{n}{2})$ -Gammaverteilung übereinstimmt. \square

Satz 17.6. Im Gauß-Produktmodell $P_\theta^{(n)} = N(\mu, \sigma^2)^{\otimes n}$ mit zu schätzenden Parametern $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ = \Theta$ sind das arithmetische Mittel \bar{X} und die empirische Varianz $\hat{\sigma}^2$ unabhängig. Es gilt

$$Z = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1) \quad \text{und} \quad (n-1) \frac{\hat{\sigma}^2}{\sigma^2} = (n-1) \frac{\hat{\sigma}^2(X_1, \dots, X_n)}{\sigma^2} \sim \chi(n-1). \tag{17.12}$$

Beweis. Wir definieren $D = \{\lambda \mathbf{1}_n : \lambda \in \mathbb{R}\}$, $\mathbf{1}_n = (1, 1, \dots, 1)'$ (n Mal), den linearen Unterraum der totalen Diagonalen in \mathbb{R}^n . Sei $x \in \mathbb{R}^n$, dann ist $\bar{x}(1, \dots, 1) = \Pi x$ die

Projektion von x auf D . Mit $\Pi^\perp x = x - \Pi x$ bezeichnen wir die Projektion auf das orthogonale Komplement von D in \mathbb{R}^n .

Wir möchten nun zeigen, dass für $X = (X_1, \dots, X_n)$, $X_j \sim N(\mu, \sigma^2)$ i.i.d. die D bzw. D^\perp -wertigen Zufallsvariablen ΠX und $\Pi^\perp X$ unabhängig voneinander sind. Dazu wähle Borelmeßbare $A \subseteq D$ und $B \subseteq D^\perp$ und rechne nach

$$\begin{aligned} P(\{\Pi X \in A\} \cap \{\Pi^\perp X \in B\}) &= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \int_{\{\Pi x \in A\} \cap \{\Pi^\perp x \in B\}} \exp\left\{-\frac{1}{2} \frac{\|x - \mu \mathbf{1}_n\|^2}{\sigma^2}\right\} dx \\ &= \frac{1}{(2\pi)^{\frac{1}{2}} \sigma} \int_A \exp\left\{-\frac{1}{2} \frac{\|\Pi x - \mu \mathbf{1}_n\|^2}{\sigma^2}\right\} d\Pi x \\ &\times \frac{1}{(2\pi)^{\frac{n-1}{2}} \sigma^{n-1}} \int_B \exp\left\{-\frac{1}{2} \frac{\|\Pi^\perp x\|^2}{\sigma^2}\right\} d\Pi^\perp x . \end{aligned} \quad (17.13)$$

Hier haben wir ausgenützt, dass $\|x - \mu \mathbf{1}_n\|^2 = \|\Pi x - \mu \mathbf{1}_n\|^2 + \|\Pi^\perp x\|^2$ und einen Koordinatenwechsel von x zu $(\Pi x, \Pi^\perp x)$ vorgenommen. Das erste Integral auf der rechten Seite nimmt mitsamt seinem Vorfaktor für $A = D$ den Wert eins an und das zweite für $B = D^\perp$, denn es handelt sich um multivariate Gaußdichten auf diesen Unterräumen, siehe Definition 11.8. Also ist die rechte Seite identisch mit $P(\Pi X \in A)P(\Pi^\perp X \in B)$, was zu beweisen war.

Da nun $\bar{X} = \frac{1}{n} \mathbf{1}_n' X = \frac{1}{n} \mathbf{1}_n' \Pi X$ und $\hat{\sigma}^2 = \frac{1}{n-1} \|\Pi^\perp X\|^2$ sind auch die Schätzer für Erwartungswert und Varianz unabhängig.

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ wurde bereits im Beweis von Satz 17.2 diskutiert. Indem wir D^\perp mit \mathbb{R}^{n-1} identifizieren, sehen wir, dass $\Pi^\perp X$ multivariat normalverteilt ist mit Erwartungswert $0 \in \mathbb{R}^{n-1}$ und Kovarianzmatrix $\sigma^2 \text{Id}_{n-1}$. Nach den Varianzrechenregeln ist deshalb $\frac{\Pi^\perp X}{\sigma} \sim N(0, \text{Id}_{n-1})$. Also

$$(n-1) \frac{\hat{\sigma}^2}{\sigma^2} = \left\| \frac{\Pi^\perp X}{\sigma} \right\|^2 \sim \chi^2(n-1) . \quad (17.14)$$

□

Wir können nun die Verteilung definieren, welche bei der Konstruktion von Konfidenzintervallen für den Fall der geschätzten Varianz die Rolle der Standardnormalverteilung übernimmt. Wesentlich ist hierbei – wie schon bei der Verteilung von Z – die Unabhängigkeit von den Parametern μ und σ^2 .

Definition 17.7. Gegeben seien zwei unabhängige Zufallsvariablen $X \sim N(0, 1)$ und $Y \sim \chi^2(n)$. Dann definiert die Verteilung der Zufallsvariable

$$T = \sqrt{n} \frac{X}{\sqrt{Y}} \sim t(n) \quad (17.15)$$

die **Student- t -Verteilung** zu n **Freiheitsgraden**.

17. Konfidenzintervalle

In Abbildung 17.1 (b) sind die Dichten einiger t -Verteilungen gezeigt. Wegen der Achsensymmetrie der Dichten gilt auch hier für die p -Quantile $t_p(n)$ der $t(n)$ Verteilung, dass für $p < 0.5$ gilt $-t_p(n) = t_{1-p}(n)$.

Im Folgenden zeigen wir, dass die Standardisierung mit der empirischen Standardabweichung im Gauß-Produktmodell zu t -verteilten Z.V. führt.

Satz 17.8. *Es seien wie oben X_1, \dots, X_n i.i.d. Zufallsvariablen mit $X_j \sim N(\mu, \sigma^2)$. Dann ist $T = T(X_1, \dots, X_n)$*

$$T = \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} \quad (17.16)$$

Student- t -verteilt mit $n - 1$ Freiheitsgraden, $T \sim t(n - 1)$.

Beweis. Wir rechnen nach

$$T = \sqrt{n} \frac{\bar{X} - \mu}{\hat{\sigma}} = \sqrt{n-1} \frac{\sqrt{n} \frac{\bar{X} - \mu}{\sigma}}{\sqrt{(n-1) \frac{\hat{\sigma}^2}{\sigma^2}}} = \sqrt{n-1} \frac{Z}{\sqrt{K^2}}. \quad (17.17)$$

Wende nun Satz 17.6 und Definition 17.7 an, um den Beweis zu beenden. \square

Nun können wir die Resultate von Satz 17.2 für den Fall des Gauss-Produktmodells mit unbekanntem Erwartungswert und unbekannter Varianz reproduzieren.

Satz 17.9. *Es seien, für $j = 1, \dots, n$, X_j normalverteilte i.i.d. Zufallsvariablen $X_j \sim N(\mu, \sigma^2)$ mit $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ = \Theta$ den im Produktmodell zu schätzenden Parametern. Mit $t_p(n-1)$, dem p -Quantil der Student- t -Verteilung mit $n-1$ Parametern, erhält man*

- (i) *Das symmetrische, beidseitige Konfidenzintervall für den Erwartungswert μ mit geschätzter Varianz ist definiert durch*

$$S_{\pm}(X_1, \dots, X_n) = \bar{X} \pm t_{1-\alpha/2}(n-1) \frac{\hat{\sigma}}{\sqrt{n}} \quad (17.18)$$

und es ist ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha$ im Sinne von Definition 17.1.

- (ii) *Das linksoffene, einseitige Konfidenzintervall mit geschätzter Varianz ist definiert durch*

$$S_- = -\infty, \quad S_+(X_1, \dots, X_n) = \bar{X} + t_{1-\alpha}(n-1) \frac{\hat{\sigma}}{\sqrt{n}} \quad (17.19)$$

und es ist ebenfalls ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha$. Gleiches gilt für das rechtssoffene, einseitige Konfidenzintervall

$$S_+ = \infty, \quad S_-(X_1, \dots, X_n) = \bar{X} - t_{1-\alpha}(n-1) \frac{\hat{\sigma}}{\sqrt{n}}. \quad (17.20)$$

Beweis. Der Beweis folgt genau den Rechenschritten des Beweises von Satz 17.2, wobei die Standardabweichung σ durch die empirische Standardabweichung $\hat{\sigma}$ zu ersetzen ist, die Z.V. $Z \sim N(0, 1)$ durch $T \sim t(n-1)$ und die Quantile der Standardnormalverteilung entsprechend durch die Quantile der t -Verteilung mit $n-1$ Freiheitsgraden. \square

Beispiel 17.10. Bei einer Untersuchung einer Stichprobe von $n = 30$ im Jahre 2011 Maisanbauflächen à 1ha wird ein mittlerer Ertrag von 25t/ha mit einer (empirischen) Standardabweichung von 5t vorgefunden (fiktive Zahlen). Können wir einen Bereich (zweiseitiges Konfidenzintervall) angeben, so dass man zu 95% davon ausgehen kann, dass der wahre Wert μ für den zu erwartenden Maisertrag darin liegt? Die Verteilung des 1ha-Flächenertrags X ist hier als normalverteilt vorausgesetzt.

Lösung. Wir haben aus der Aufgabenstellung $n = 30$, $\bar{x} = 25$, $\hat{\sigma} = 5$ und $\alpha = 0.05$.

Aus der Tabelle der Quantile (siehe Anhang) der $t(n-1)$ -Verteilung lesen wir das $t_{1-\alpha/2}(n-1) = t_{0.975}(n-1)$ Quantil der $t(n-1) = t(29)$ -Verteilung ab. Man erhält $t_{0.975}(29) = 2.045$. Für das Konfi-Intervall ergibt sich (mit Rundung der Zahlen)

$$\left[\bar{x} - t_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \right] = \left[25 - 2.045 \frac{5}{\sqrt{30}}, 25 + 2.045 \frac{5}{\sqrt{30}} \right] = [23.1, 26.9] . \quad (17.21)$$

Wir können demnach zu 95% Konfidenz davon ausgehen, dass der tatsächliche Durchschnittsertrag eines ha Mais in der BRD im Jahre 2011 zwischen 23.1 und 26.9 t liegt. \diamond

17. Konfidenzintervalle

1. Übungsaufgaben

2. Literaturliste

18. Asymptotische Konfidenzintervalle

Nun möchten wir uns von den Einschränkungen auf das Gauß-Produktmodell im vorangegangenen Kapitel befreien. Durch den zentralen Grenzwertsatz 15.1 wissen wir ja, dass in beliebigen Produktmodellen $Z = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$ schwach nach Verteilung gegen $N(0, 1)$ konvergiert. Ersetzen wir hier σ durch die empirische Standardabweichung $\hat{\sigma}$ (oder durch einen anderen konsistenten Schätzer) sollte sich intuitiv hieran nicht viel ändern. Dass dies auch den mathematischen Tatsachen entspricht wird, unter geeigneten Voraussetzungen, durch die Grenzwertsätze von Slutsky abgesichert, siehe z.B. T. Ferguson, A Course in Large Sample Theory, Chapman & Hall 1996.

Definition 18.1. Es seien X_j , $j = 1, \dots, n$, i.i.d. Zufallsvariablen aus einem beliebigen Produktmodell, so dass die 2. Momente von X_j existieren. Dann bezeichnen wir die in Satz 17.2 definierten (beidseitigen, links- und rechtsoffenen) Intervallschätzer – mit σ ersetzt durch die empirische Standardabweichung $\hat{\sigma}$ – als **asymptotische Konfidenzintervalle** für den wahren Erwartungswert μ . D.h. die W.-keit, dass der wahre Wert μ in dem Zufallsintervall liegt, konvergiert für Stichprobengröße $n \rightarrow \infty$ gegen $1 - \alpha$.

Die Konvergenzgeschwindigkeit in der obigen Definition hängt natürlich von der Verteilung der X_j und damit von den Details des Produktmodells ab. Es ist jedoch üblich für moderat kleine Fehler-W.-keiten $\alpha \approx 5\%$ ab $n \approx 30$ von einer für praktische Zwecke ausreichenden Konvergenz auszugehen und asymptotische Konfidenzintervalle zu verwenden.

Beispiel 18.2. In einer Versicherung möchte man aus den Daten der letzten 3 Jahre schätzen, wieviele Fälle von Berufsunfähigkeit auf 10.000 Versicherte in einem Monat eintreten. Über die 36 Monate ergebe sich ein Schnitt von 3 Versicherungsfällen pro Monat mit einer Standardabweichung von 0.086. Man gebe eine obere Schranke an, so dass der Erwartungswert von X =Versicherungsfälle pro Monat zu 99.9% darunter liegt.

Lösung. Obwohl X vermutlich eher poissonverteilt ist, können wir wegen des großen Stichprobenumfangs >30 den Intervall-Schätzer für ein normalverteiltes Merkmal zum Konfi-Niveau 99.9% verwenden.

Gefragt ist nach einem linksoffenen Konfi-Intervall. Sowohl μ_X als auch σ^2 sind unbekannt. In approximativ 99.9% der Fälle solcher Stichprobenerhebungen haben wir

$$\mu_X < \bar{x} + z_{1-\alpha} \frac{\hat{\sigma}}{\sqrt{n}} = 3 + 3.09 \times \frac{0.086}{\sqrt{36}} = 3.44 . \quad (18.1)$$

Hier ist zu beachten, dass für $n - 1 = 35$ die t -Verteilung einigermaßen durch die Standardnormalverteilung approximiert wird. Daher ist es üblich für $n - 1 \geq 30$ z -Werte für

18. Asymptotische Konfidenzintervalle

die Konstruktion der Konfidenzintervalle zu verwenden. Im vorliegenden Fall macht diese Abweichung allerdings c.a. 8% der Breite des Konfidenzintervalls aus. \diamond

Oft treten Merkmale bzw. Zufallsvariablen auf, die nur zwei Werte annehmen können: Ein/e Patient/in hat eine bestimmte Krankheit, oder nicht. Ein produziertes Teil ist Ausschuß oder kann verkauft werden. Die ganze Grundgesamtheit unterscheidet sich also in statistische Einheiten, welche diese Eigenschaft haben, und andere, die sie nicht haben. Dies modelliert man durch eine Z.V. X mit $X(\omega) = 1$ falls ω die Eigenschaft hat und $X(\omega) = 0$ sonst. Die Verteilung von solch einem X ist gegeben durch das Bernoulli-Modell $X \sim B(1, p)$ mit $P(X = 1) = p$ und $P(X = 0) = 1 - p$.

Gefragt ist nun ein Konfidenzintervall für den Wert p , welches auf n -facher unabhängiger Beobachtung von X beruht, d.h. wir befinden uns im Bernoulli-Produktmodell. Wir beantworten diese Frage für Stichprobengrößen n , so dass $pn > 20$ mittels asymptotischer Konfidenzintervalle.

Dazu seien X_1, \dots, X_n die Stichprobenvariablen. Als konsistenten, erwartungstreuen Schätzer für p verwenden wir \bar{X} , da $\mathbb{E}[X] = P(X = 1) = p$. Wir erinnern uns an die Formel für die Standardabweichung von $B(1, p)$, $\sigma_X = \sqrt{p(1-p)}$. Dadurch erhalten wir einen Schätzer für die Standardabweichung

$$\hat{\sigma}_B = \sqrt{\bar{X}(1 - \bar{X})}. \quad (18.2)$$

Approximativ gelten deshalb folgende Konfidenzintervalle:

Definition 18.3. Im Bernoulli-Produktmodell mit Stichprobengröße n ist

$$S_{\pm} = \bar{X} \pm z_{1-\alpha/2} \frac{\hat{\sigma}_B}{\sqrt{n}} = \bar{X} \pm z_{1-\alpha/2} \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}} \quad (18.3)$$

ein **asymptotisches, beidseitiges Konfidenzintervall für den wahren Anteilswert** p . Hier ist $z_{1-\alpha/2}$ das Quantil der Standardnormalverteilung. Für die links- bzw. rechtsseitigen Konfidenzintervalle zum Niveau $1 - \alpha$ erhält man

$$\left[0, \bar{X} + z_{1-\alpha} \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}} \right) \text{ bzw. } \left[\bar{X} - z_{1-\alpha} \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}, 1 \right]. \quad (18.4)$$

Beispiel 18.4. Ein/e Händler/in möchte eine Lieferung von 100.000 Glühbirnen kaufen. Um die Qualität zu prüfen, werden 500 Glühbirnen entnommen (mit Zurücklegen) und getestet. Es sind 39 Birnen kaputt. Gefragt ist ein 95%-Konfi-Intervall für den Anteil der defekten Glühbirnen.

Lösung. Gefragt ist ein zweiseitiges Konfi-Intervall zum Niveau 0.95, also zur Fehler-W-keit $\alpha = 0.05$. Wir haben $n = 500$, $\bar{x} = 39/500$. Also

$$p = \bar{x} \pm z_{1-\alpha/2} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}} = \frac{39}{500} \pm z_{0.975} \sqrt{\frac{(39/500)(1 - 39/500)}{500}}. \quad (18.5)$$

Mit $z_{0.975} = 1.96$ ergibt sich

$$p = \frac{39}{500} \pm 1.96 \sqrt{\frac{(39/500)(1 - 39/500)}{500}} = 0.078 \pm 0.023 \quad (18.6)$$

zu 95% Konfidenz. Mit 95% Sicherheit liegt der Anteil der defekten Glühlampen zwischen 5.5% und 9.1%. \diamond

18. *Asymptotische Konfidenzintervalle*

1. Übungsaufgaben

2. Literaturliste

19. Maximum Likelihood Schätzer

Im letzten Kapitel wurden wünschenswerte Eigenschaften von Schätzern diskutiert und einige spezielle Beispiele gegeben. Hier soll nun diskutiert werden, wie man geeignete Schätzer finden kann.

Definition 19.1. $X : (G, \mathcal{B}) \rightarrow (E, \mathcal{B}_E)$ sei ein Merkmal und $\chi = (g_1, \dots, g_n) \in G^{\times n}$ eine Stichprobe aus der Grundgesamtheit G . Gegeben seien die gemessenen Daten $x_1, \dots, x_n \in E$ des Merkmals X , $x_i = X(g_i)$. Des Weiteren sei $\{P_\theta^{(n)}\}_{\theta \in \Theta, n \in \mathbb{N}}$ ein parametrisches statistisches Modell über (E, \mathcal{B}_E) .

- (i) Ist (E, \mathcal{B}_E) abzählbar, dann ist die Maximum-Likelihood Funktion $\mathcal{L} : E^{\times n} \times \Theta \rightarrow \mathbb{R}$ definiert durch

$$\mathcal{L}(x_1, \dots, x_n | \theta) = P_\theta^{(n)}(\{(x_1, \dots, x_n)\}) \quad (19.1)$$

- (ii) Ist $(E, \mathcal{B}_E) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ (multivariat) kontinuierlich und ist $f_n(x_1, \dots, x_n | \theta)$ die kontinuierliche W.-keitsdichte zu $P_\theta^{(n)}$ auf $(\mathbb{R}^{d \times n}, \mathcal{B}(\mathbb{R}^{d \times n}))$, dann definieren wir die Maximum Likelihood Funktion $\mathcal{L} : E^{\times n} \times \Theta \rightarrow \mathbb{R}$ als

$$\mathcal{L}(x_1, \dots, x_n | \theta) = f_n(x_1, \dots, x_n | \theta) \quad (19.2)$$

Die Likelihood Funktion misst, wie 'wahrscheinlich' die tatsächlich beobachteten Werte unter der Verteilungshypothese $P_\theta^{(n)}$ wären. Wörtlich zu nehmen ist dies allerdings nur im Falle von Definition 19.1 (i). Liegt ein statistisches Modell mit kontinuierlicher Dichte vor, so müsste man die Dichte noch mit (infinitesimalen) Volumina $dx_1 \cdots dx_n$ Multiplizieren, um auf eine (infinitesimale) W.-keit zu kommen. Da diese Volumina aber nicht θ -abhängig sind, wird das nun folgende Maximum Likelihood Prinzip nicht von diesen Faktoren beeinflusst.

Maximum Likelihood Prinzip: Unter allen zur Verfügung stehenden W.-Maßen eines parametrischen statistischen Modells wähle dasjenige aus, für das die gemessenen Daten die größte Wahrscheinlichkeit (bzw. Wahrscheinlichkeitsdichte) aufweisen.

Diesen klaren Gedanken versuchen wir nun im Rahmen unserer Theorie der Schätzer mathematisch umzusetzen. Zunächst eine kleine Wiederholung: Für eine Funktion $f : \Theta \rightarrow \mathbb{R}$ definieren wir

$$\arg \max_{\theta \in \Theta} f(\theta) = \{\theta \in \Theta : f(\theta) \geq f(\theta') \forall \theta' \in \Theta\}. \quad (19.3)$$

Die Menge $\arg \max_{\theta \in \Theta} f(\theta)$ kann leer sein, kann ein oder mehrere Elemente aus Θ enthalten. Falls $\arg \max_{\theta \in \Theta} f(\theta)$ genau ein Element $\hat{\theta}$ enthält, schreiben wir vereinfachend auch $\hat{\theta} = \arg \max_{\theta \in \Theta} f(\theta)$.

19. Maximum Likelihood Schätzer

Definition 19.2. Unter denselben Voraussetzungen wie in Definition 19.1 (i) bzw. (ii) heißt $\hat{\theta} = \{\hat{\theta}^{(n)}\}_{n \in \mathbb{N}}$ ein Maximum Likelihood Schätzer, (bzw. genauer ein Maximum Likelihood Multiparameterschätzer) falls

$$\hat{\theta}(x_1, \dots, x_n) \in \arg \max_{\theta \in \Theta} \mathcal{L}(x_1, \dots, x_n | \theta) \quad \forall x_1, \dots, x_n \in E^{\times n}. \quad (19.4)$$

In der praktischen Arbeit mit dem Maximum Likelihood Schätzverfahren nimmt man an, dass Gleichung (19.4) als definierende Gleichung für den Maximum Likelihood Schätzer verstanden werden kann, dass also

$$\hat{\theta}(X_1, \dots, X_n) = \arg \max_{\theta \in \Theta} \mathcal{L}(X_1, \dots, X_n | \theta) \quad (19.5)$$

für $P_\theta^{(n)}$ -fast sicher den Schätzer eindeutig definiert $\forall \theta \in \Theta$ und $(X_1, \dots, X_n) \sim P_\theta^{(n)}$. Diese Bedingung müßte streng genommen Modell für Modell bewiesen werden, bzw. allgemeine hinreichende Bedingungen sollten aufgestellt werden. Auch müßte man Messbarkeitseigenschaften der $\arg \max$ -Definition genauer unter die Lupe nehmen. Es ist bedauerlich, dass dies in der Statistikliteratur kaum betrachtet wird. Mit einem gewissen Gefühl der Unzufriedenheit folge ich deshalb an dieser Stelle der statistischen Folklore und sehe (19.5) als Definition an.

Lemma 19.3. Zu denselben Voraussetzungen wie in Definition 19.1 (i) oder (ii) nehmen wir zusätzlich an, dass $P_\theta^{(n)} = P_\theta^{\otimes n}$ mit P_θ einer Familie von W .-Maßen auf (E, \mathcal{B}_E) , d.h. dass ein Produktmodell vorliegt. Dann gilt

$$\arg \max_{\theta \in \Theta} \mathcal{L}(x_1, \dots, x_n | \theta) = \arg \max_{\theta \in \Theta} \log \mathcal{L}(x_1, \dots, x_n | \theta) = \arg \max_{\theta \in \Theta} \sum_{j=1}^n l(x_j | \theta) \quad (19.6)$$

wobei $l(x_j | \theta) = \log P_\theta(\{x_j\})$ falls E abzählbar ist und $\log f(x_j | \theta)$ falls $(E, \mathcal{B}_E) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ und $f(\cdot | \theta)$ die kontinuierliche Dichte von P_θ ist.

Beweis: Siehe Satz 12.6 und Satz 12.7 für

$$\mathcal{L}(x_1, \dots, x_n | \theta) = \prod_{j=1}^n P_\theta(\{x_j\})$$

im diskreten Fall bzw.

$$\mathcal{L}(x_1, \dots, x_n | \theta) = \prod_{j=1}^n f(x_j | \theta)$$

im kontinuierlichen Fall. Wenn wir nun das $\arg \max$ für $\mathcal{L}(x_1, \dots, x_n | \theta)$ bestimmen, so können wir das äquivalent auch für $\log \mathcal{L}(x_1, \dots, x_n | \theta)$ tun, da $\log(t)$ streng monoton in t steigt. Nun wendet man die Rechenregeln für \log an, um den Beweis zu beenden. ■

In manchen Fällen kann man die Form der Schätzer auch analytisch bestimmen. Dazu nimmt man an, dass $\Theta \subseteq \mathbb{R}^q$ offen ist und man stellt die Maximum Likelihood Gleichungen auf

$$\nabla_\theta \log \mathcal{L}(x_1, \dots, x_n | \theta) = \sum_{j=1}^n \nabla_\theta l(x_j | \theta) \stackrel{!}{=} 0. \quad (19.7)$$

Lösungen zu diesen Gleichungen sind unter geeigneten Annahmen an die Hessematrix von $\log \mathcal{L}$ Kandidaten für den Maximum Likelihood Schätzer.

Beispiel 19.4. Wir betrachten die Zahlen N_1, \dots, N_n , die wir mit einem $B(\tilde{n}, p)$ Binomial-Produktmodell mit fest gegebenem \tilde{n} modellieren. Dann ist

$$\begin{aligned} \log \mathcal{L}(N_1, \dots, N_n | p) &= \log \left(\prod_{j=1}^n \binom{\tilde{n}}{N_j} p^{N_j} (1-p)^{\tilde{n}-N_j} \right) \\ &= \sum_{j=1}^n \left[\log \binom{\tilde{n}}{N_j} + N_j \log p + (\tilde{n} - N_j) \log(1-p) \right]. \end{aligned}$$

Ableiten nach p liefert

$$0 \stackrel{!}{=} \sum_{j=1}^n \left[\frac{N_j}{\hat{p}} - \frac{(\tilde{n} - N_j)}{(1 - \hat{p})} \right] = \frac{\sum_{j=1}^n N_j}{p} - \frac{n\tilde{n} - \sum_{j=1}^n N_j}{1 - p}.$$

Dies ist äquivalent zu (für $p \notin \{0, 1\}$)

$$\frac{p}{\sum_{j=1}^n N_j} = \frac{1 - p}{n\tilde{n} - \sum_{j=1}^n N_j} \Leftrightarrow \hat{p} = \frac{1}{\frac{n\tilde{n} - \sum_{j=1}^n N_j}{\sum_{j=1}^n N_j} + 1} = \frac{\sum_{j=1}^n N_j}{n\tilde{n}}$$

Der Maximum Likelihood Schätzer läuft also wieder auf den Anteilswert heraus.

Maximum Likelihood Schätzer sind im Allgemeinen nicht erwartungstreu. Von Interesse ist deshalb vor Allem die Konsistenz der Maximum Likelihood Schätzer. Hierbei spielt die folgende Begriffsbildung eine große Rolle:

Definition 19.5. Es sei für $\theta \in \Theta$, $p_j(\theta)$ eine Zähldichte auf $(\mathbb{Z}^d, \mathbb{P}(\mathbb{Z}^d))$ bzw. $f(x|\theta)$ eine kontinuierliche Dichte auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ so dass sich diese Dichten für je zwei unterschiedliche $\theta, \theta' \in \Theta$ unterscheiden. Dann definieren wir die Kullback-Leibler Information als

$$K(\theta|\theta_0) = \mathbb{E} \left[-\log \left(\frac{p_X(\theta)}{p_X(\theta_0)} \right) \middle| \theta_0 \right] = \sum_{j \in \mathbb{Z}^d} -\log \left(\frac{p_j(\theta)}{p_j(\theta_0)} \right) p_j(\theta_0) \quad (19.8)$$

bzw.

$$K(\theta|\theta_0) = \mathbb{E} \left[-\log \left(\frac{f(X|\theta)}{f(X|\theta_0)} \right) \middle| \theta_0 \right] = \int_{\mathbb{R}^d} -\log \left(\frac{f(t|\theta)}{f(t|\theta_0)} \right) f(t|\theta_0) dt \quad (19.9)$$

Hier arbeiten wir mit dem natürlichen Logarithmus. In der Literatur wird teils auch \log_2 verwendet.

Satz 19.6. Es sei $K(\theta|\theta_0)$ definiert wie in Definition 19.5. Dann gilt

$$(i) \quad K(\theta|\theta_0) \geq 0;$$

19. Maximum Likelihood Schätzer

$$(ii) \quad K(\theta|\theta_0) = 0 \Leftrightarrow \theta = \theta_0$$

Beweis: Wir führen den Beweis nur im kontinuierlichen Fall. Der diskrete Fall wird analog bewiesen.

(i) Mit der Ungleichung von Jensen (siehe das nachfolgende Lemma) gilt

$$\begin{aligned} K(\theta|\theta_0) &= \mathbb{E} \left[-\log \left(\frac{f(X|\theta)}{f(X|\theta_0)} \right) \middle| \theta_0 \right] \\ &\geq -\log \left(\mathbb{E} \left[\frac{f(X|\theta)}{f(X|\theta_0)} \middle| \theta_0 \right] \right) \\ &= -\log \left(\int_{\mathbb{R}^d} \frac{f(t|\theta)}{f(t|\theta_0)} f(t|\theta_0) dt \right) \\ &= -\log \left(\int_{\mathbb{R}^d} f(t|\theta) dt \right) = -\log(1) = 0 \end{aligned}$$

(ii) Da \log strikt konkav ist, gilt nach der Jensenschen Ungleichung die Gleichheit nur dann, wenn $Y = \frac{f(X|\theta)}{f(X|\theta_0)}$ P -fast sicher konstant ist. Dann ist $f(t|\theta) = cf(t|\theta_0)$ und die Normiertheit der beiden Dichten führt zu $c = 1$. ■

Lemma 19.7. (*Jensensche Ungleichung*) Es sei $I \subseteq \mathbb{R}$ ein offenes Intervall in \mathbb{R} , $\phi : I \rightarrow \mathbb{R}$ eine konvexe Funktion und $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Werten in I eine Zufallsvariable.

(i) Sind X und $\phi(X)$ integrierbar, so gilt

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X]) \quad (19.10)$$

(ii) Ist ϕ strikt konvex, d.h. $\phi''(t) > 0$ auf I und X nicht konstant, dann gilt

$$\mathbb{E}[\phi(X)] > \phi(\mathbb{E}[X]) \quad (19.11)$$

Beweis: Wir nehmen zur Erleichterung des Beweises an, dass ϕ zweifach differenzierbar ist, $\phi''(t) \geq 0$. Dann gilt

$$\phi(t) \geq \phi(s) + \phi'(s)(t - s)$$

und

$$\phi(t) = \sup_{s \in I} [\phi(s) + \phi'(s)(t - s)].$$

Wir haben also $\phi(X) \geq \phi(s) + \phi'(s)(X - s)$ und ist X nicht konstant, so gilt $>$ für eine Menge mit Maßgrößer 0, denn X kann nicht P -fast sicher mit s übereinstimmen.

Mit der Monotonie des Integrals haben wir für alle $s \in I$

$$\mathbb{E}[\phi(X)] \geq \phi(s) + \phi'(s)\mathbb{E}[X - s] = \phi(s) + \phi'(s)(\mathbb{E}[X] - s)$$

Also

$$\mathbb{E}[\phi(X)] \geq \sup_{s \in I} [\phi(s) + \phi'(s)(\mathbb{E}[X] - s)] = \phi(\mathbb{E}[X]).$$

Um auch den Beweis von (ii) zu vereinfachen, nehmen wir zusätzlich an, dass $\phi''(t) > \epsilon > 0$ gilt. Dann haben wir nach der Taylor Formel mit Lagrange-Restglied

$$\phi(t) \geq \phi(s) + \phi'(s)(t-s) + \frac{1}{2}\epsilon(s-t)^2.$$

Und daher für alle $s \in I$

$$\begin{aligned} \mathbb{E}[\phi(X)] &\geq \phi(s) + \phi'(s)(\mathbb{E}[X] - s) + \frac{1}{2}\epsilon\mathbb{E}[(X-s)^2] \\ &\geq \phi(s) + \phi'(s)(\mathbb{E}[X] - s) + \frac{1}{2}\epsilon\text{Var}[X] \end{aligned}$$

und

$$\begin{aligned} \mathbb{E}[\phi(X)] &\geq \sup_{s \in I} [\phi(s) + \phi'(s)(\mathbb{E}[X] - s)] + \frac{1}{2}\epsilon\text{Var}[X] \\ &= \phi(\mathbb{E}[X]) + \frac{1}{2}\epsilon\text{Var}[X] > \phi(\mathbb{E}[X]). \end{aligned}$$

■

Satz 19.6 können wir so verstehen, dass $\theta_0 = \arg \max_{\theta \in \Theta} [-K(\theta|\theta_0)]$ gilt. Wir sehen nun, dass die Maximum Likelihood Funktionen modifiziert werden können, ohne die arg max-Mengen zu verändern, dass sie P -fast sicher gegen $-K(\theta|\theta_0)$ konvergieren.

Satz 19.8. *Unter den Bedingungen von Lemma 19.3 und der zusätzlichen Annahme, dass $l(X|\theta) \in L^2((E, \mathcal{B}_E, P_{\theta_0}); \mathbb{R}) \forall \theta, \theta_0 \in \Theta$ gilt für $X_j \sim P_{\theta_0}$ i.i.d., dass*

(i) *Die Maximum Likelihood Schätzer können als Maximierer der wie folgt modifizierten Likelihood Funktion verstanden werden*

$$\log \mathcal{L}'(X_1, \dots, X_n|\theta, \theta_0) = \left[\frac{1}{n} \sum_{j=1}^n (l(X_j|\theta) - l(X_j|\theta_0)) \right] \quad (19.12)$$

(ii)

$$\log \mathcal{L}'(X_1, \dots, X_n|\theta, \theta_0) \longrightarrow -K(\theta|\theta_0) \text{ } P\text{-fast sicher } \forall \theta, \theta_0 \in \Theta \quad (19.13)$$

Beweis: (i) Offensichtlich gilt

$$\arg \max_{\theta \in \Theta} \log \mathcal{L}(X_1, \dots, X_n|\theta) = \arg \max_{\theta \in \Theta} \left[\frac{1}{n} \sum_{j=1}^n (l(X_j|\theta) - l(X_j|\theta_0)) \right].$$

(ii) Nach dem starken Gesetz der großen Zahlen gilt

$$\begin{aligned} \log \mathcal{L}'(X_1, \dots, X_n|\theta, \theta_0) &= \left[\frac{1}{n} \sum_{j=1}^n (l(X_j|\theta) - l(X_j|\theta_0)) \right] \\ &\rightarrow \mathbb{E} [l(X_1|\theta) - l(X_1|\theta_0) | \theta_0] \\ &= \mathbb{E} [\log f(X_j|\theta) - \log f(X_j|\theta_0) | \theta_0] \\ &= \mathbb{E} \left[\log \left(\frac{f(X_j|\theta)}{f(X_j|\theta_0)} \right) \middle| \theta_0 \right] = -K(\theta|\theta_0) \end{aligned}$$

19. Maximum Likelihood Schätzer

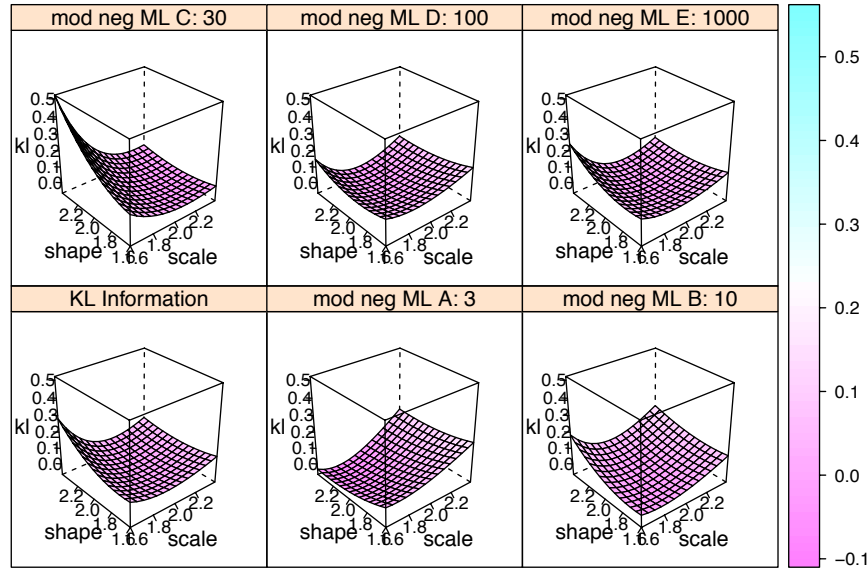


Abbildung 19.1: Negative modifizierte log-Likelihood und Kullback-Leibner Information für $Wei(2, 2)$ und verschiedene Anzahlen n von Pseudozufallszahlen über $\theta = (\eta, \beta) = (\text{scale}, \text{shape})$.

■

Satz 19.8 besagt, dass die modifizierte Likelihood Funktion fast sicher punktweise gegen die Kullback-Leibner Information konvergiert, die ein eindeutiges $\arg \max$ aufweist, und zwar den wahren Parameter θ_0 . Die Vermutung liegt nun nahe, dass dann auch die $\arg \max$ der Likelihood Funktion gegen θ_0 konvergieren, was gleichbedeutend mit der Konsistenz ist. Leider sind einige technische Voraussetzungen nötig, um dieses Resultat auch zu beweisen.

Proposition 19.9. *Es sei $\Theta \subseteq \mathbb{R}^d$ kompakt und $u : E \times \Theta \rightarrow \mathbb{R}$ eine Funktion. Für alle $x \in E$ sei $u(x|\theta)$ eine stetige Funktion in θ und für alle $\theta \in \Theta$ eine messbare Funktion in x . Falls $\exists K : (E, \mathcal{B}_E) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ so dass $K(x) \geq |u(x, \theta)|$ und $\mathbb{E}[K(X)|\theta] < \infty \quad \forall \theta \in \Theta$, dann gilt für X_j i.i.d. mit $X_j \sim P_{\theta_0}$*

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{j=1}^n u(X_j|\theta) - \mathbb{E}[u(X_1|\theta)|\theta_0] \right| \rightarrow 0 \quad P\text{-fast sicher} \quad (19.14)$$

Beweis: Siehe T. S. Ferguson, A Course in Large Sample Theory, Chapman & Hall 1996 Kap. 16. ■

Korollar 19.10. *Erfüllt $u(x, \theta) = l(x, \theta) - l(x, \theta_0)$ die Bedingungen aus Proposition 19.9, so ist die Konvergenz in (19.13) uniform, d.h.*

$$\sup_{\theta \in \Theta} |\log \mathcal{L}'(X_1, \dots, X_n|\theta, \theta_0) + K(\theta|\theta_0)| \rightarrow 0 \quad P\text{-fast sicher} \quad \forall \theta_0 \in \Theta \quad (19.15)$$

Abbildung 19.1 zeigt die Konvergenz der modifizierten Maximum-Likelihood Funktionen gegen die Kullback-Leibner Information für die Weibullverteilung mit simulierten Daten aus $Wei(2, 2)$ für $n = 3, 10, 30, 100$ und 1000 .

Satz 19.11. Es sei $\Theta \subseteq \mathbb{R}^d$ ein kompakter Parameterraum und P_θ eine Familie von paarweise verschiedenen W .-Maßen auf (E, \mathcal{B}_E) , so dass $u(x, \theta) = l(x|\theta) - l(x|\theta_0)$ die Bedingungen von Proposition 19.9 erfüllt $\forall \theta_0 \in \Theta$, mit $l(x|\theta) = \log f(x|\theta)$ für P_θ kontinuierlich und $l(x|\theta) = \log p_x(\theta)$ für P_θ diskret. Ist $X_j : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{B}_E)$, $X_j \sim P_\theta$ i.i.d., dann gilt

$$\hat{\theta}(X_1, \dots, X_n) \longrightarrow \theta_0 \quad P - \text{fast sicher } n \rightarrow \infty \quad (19.16)$$

also ist der Maximum Likelihood Schätzer $\hat{\theta}$ konsistent.

Beweis: Der Beweis orientiert sich wieder an Ferguson (1996). Aufgrund des Satz von Lebesgue mit L^1 -Dominante $K(X)$ für $u(X, \theta)$ und mit der Stetigkeit von $u(x, \theta)$ in θ und deshalb der punktweise Konvergenz von $u(X, \theta_n) \rightarrow u(x, \theta)$ für $\theta_n \rightarrow \theta$ folgt dass $K(\theta|\theta_0)$ stetig ist in θ .

Wir betrachten nun die kompakte Menge $Q = \Theta \setminus B_{\theta_0}(\epsilon)$ mit $B_{\theta_0}(\epsilon) \subseteq \mathbb{R}^d$ der offenen Kugel mit Radius ϵ und Mittelpunkt θ_0 . Dann gilt, dass $-K(\theta|\theta_0)$ sein Maximum $\delta = \sup_{\theta \in Q} -K(\theta|\theta_0)$ auf Q annimmt. Deshalb gilt mit Satz 19.6 $\delta < 0$.

Wir nehmen nun an, dass $\hat{\theta}(X_1(\omega), \dots, X_n(\omega))$ unendlich oft in Q liegt und zeigen, dass ω dann in einer Ausnahmемenge vom Maß 0 liegt. Wegen Korollar 19.10 gibt es fast sicher ein $N = N(\omega)$ so dass die linke Seite in (19.15) kleiner $|\delta|$ ist. Für unendlich viele $n \in \mathbb{N}$ gilt deshalb

$$\hat{\theta}(X_1, \dots, X_n) = \arg \max_{\theta \in Q} \log \mathcal{L}'(X_1, \dots, X_n|\theta, \theta_0) < 0.$$

Dies ist jedoch im Widerspruch dazu, dass $\hat{\theta}(X_1, \dots, X_n)$ der Maximum-Likelihood Schätzer ist, denn

$$\log \mathcal{L}'(X_1, \dots, X_n|\theta_0, \theta_0) = \frac{1}{n} \sum_{j=1}^n [l(X_j|\theta_0) - l(X_j|\theta_0)] = 0.$$

Folglich ist

$$P(\limsup \{\hat{\theta}(X_1, \dots, X_n) \in Q\}) = 0.$$

Wähle nun $Q = Q(k)$ mit $\epsilon = 1/k$, $k \in \mathbb{N}$. Dann gilt auch

$$P\left(\bigcup_{k \in \mathbb{N}} \limsup \{\hat{\theta}(X_1, \dots, X_n) \in Q(k)\}\right) = 0.$$

Dies bedeutet aber gerade, dass $\hat{\theta}(X_1, \dots, X_n)$ P -fast sicher gegen θ_0 konvergiert. ■

Verschiedene Bedingungen können hier noch abgeschwächt werden. Beispielsweise erscheint Θ kompakt als ein Ärgernis, weil dies viele unserer Standardmodelle ausschließt. In der Praxis kann man jedoch oft einen kompakten Wertebereich angeben in dem man suchen möchte, da außerhalb liegende Werte zu offensichtlich absurden Verhältnissen führen würden.

In der Praxis werden die Maximum-Likelihood Gleichungen oft numerisch gelöst. Auch hier demonstrieren wir dies wieder für die Weibullverteilung mit Pseudozufallszahlen aus Wei(2, 2).

19. Maximum Likelihood Schätzer

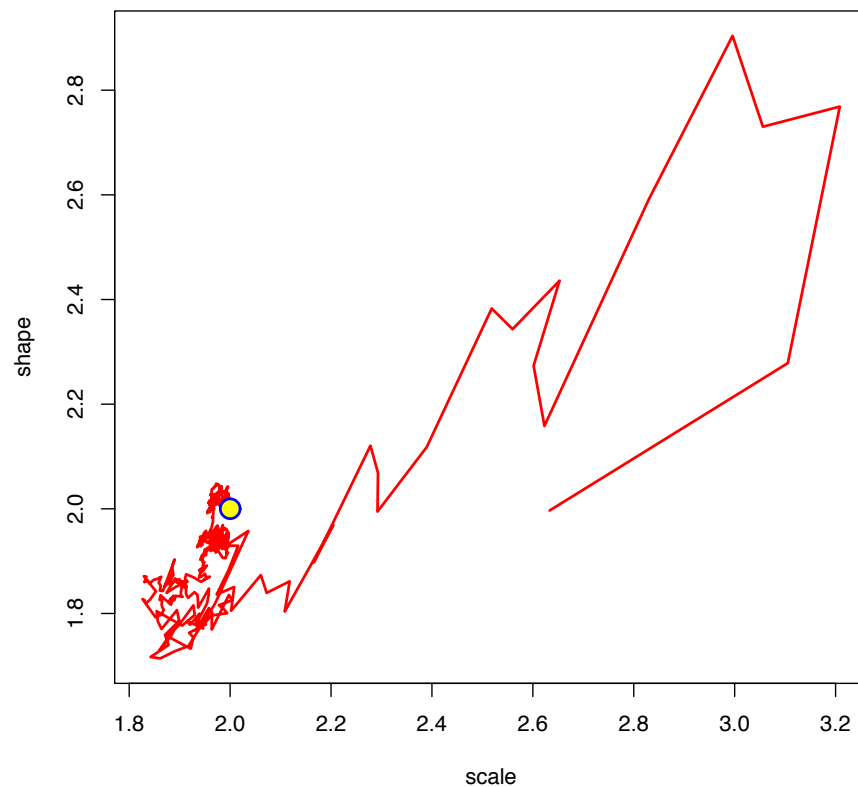


Abbildung 19.2: Konvergenzgeschichte von Maximum Likelihood Schätzungen von $n = 3$ bis $n = 1000$, $\theta_0 = (2, 2)$ für eine sukzessiv erweiterte Stichprobe von $\text{Wei}(\theta_0)$ -Pseudozufallszahlen.

```
n=100                                # set number of
                                     # Wei(2,2) Pseudos
theta0=c(2,2)                         # set  $\theta_0$ 
nlL=function(theta,X) -sum(dweibull(X, # def.
  scale=theta[1],shape=theta[2],log=TRUE)) # negLogLikelihood
X=rweibull(n,scale=theta0[1],shape=theta0[2]) # generate sample
thetaS=c(1,1)                         # set start as Exp(1)
optim(thetaS,nlL,X=X)                 # calculate estimate
                                     # by optimization of  $\theta$ 
```

Abbildung 19.2 zeigt die Konvergenzgeschichte für $n = 3$ bis $n = 1000$, wobei sukzessiv jeweils das *sample* um eine $\text{Wei}(2, 2)$ - Zufallszahl erhöht wird.

1. Übungsaufgaben

19. *Maximum Likelihood Schätzer*

2. Literaturliste

A. Tabellen

A.1. Verteilungsfunktion der Standardnormalverteilung

Verteilungsfunktion der Standardnormalverteilung $F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$										
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

A.2. Quantile der Student- t -Verteilung

\tilde{x}_p -Quantile der Student- t_n -Verteilung							
n	p						
	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
1	3,078	6,314	12,706	31,821	63,656	318,289	636,578
2	1,886	2,920	4,303	6,965	9,925	22,328	31,600
3	1,638	2,353	3,182	4,541	5,841	10,214	12,924
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,894	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,421	3,689
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,396	3,660
∞	1,282	1,645	1,960	2,327	2,577	3,092	3,293
Die letzte Zeile " ∞ " enthält die Quantile der Standard-Normalverteilung und gilt in guter Näherung für die t_n -Verteilung mit $n \geq 30$							

A.3. Quantile der χ^2 -Verteilung

\tilde{x}_p -Quantile der χ_n^2 -Verteilung									
n	p								
	0,005	0,01	0,025	0,05	0,9	0,95	0,975	0,99	0,995
1	0,000	0,000	0,001	0,004	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	9,236	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,180	2,733	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	15,987	18,307	20,483	23,209	25,188
11	2,603	3,054	3,816	4,575	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	33,196	36,415	39,364	42,980	45,558

Bibliography

- [Arb00] Louis Francois Antoine Arbogast. *Calcul des derivations*. Levrault (Strassburg), 1800.
- [Bau90] Heinz Bauer. *Maß- und Integrationstheorie*. Berlin etc.: Walter de Gruyter. xviii, 259 S. , 1990.
- [Bau91] Heinz Bauer. *Wahrscheinlichkeitstheorie, 4., völlig überarb. u. neugestaltete Aufl. des Werkes: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. De Gruyter Lehrbuch. Berlin etc.: Walter de Gruyter. xvii, 520 S. , 1991.
- [Bay70] F.R.S. Bayes. *An essay towards solving a problem in the doctrine of chances*. 1970.
- [Ber99] Jakob Bernoulli. *Wahrscheinlichkeitsrechnung (Ars conjectandi)*. Uebersetzt u. herausg. von R. Haussner. Leipzig: W. Engelmann. 2 Bde. 162 S. resp. 172 S. kl. 8°. (Ostwald's Klassiker d. exact. Wissensch. No. 107 u. 108.) , 1899.
- [Dev09] Keith Devlin. *The unfinished game. Pascal, Fermat, and the seventeenth-century letter that made the world modern. A tale of how mathematics is really done. Transl from the English by Enrico Heinemann. (Pascal, Fermat und die Berechnung des Glücks. Eine Reise in die Geschichte der Mathematik.)*. München: Verlag C. H. Beck. 205 p. , 2009.
- [dM67] A. de Moivre. *The doctrine of chances: or, A method of calculating the probabilities of events in play*. Reprint of the 3rd edition. London 1756; New York, N.Y.: Chelsea Publishing Company 1967. XII, 368 p. , 1967.
- [FHT96] Ludwig (ed.) Fahrmeir, Alfred (ed.) Hamerle, and Gerhard (ed.) Tutz. *Multivariate statistical methods. (Multivariate statistische Verfahren. Unter Mitarbeit von Wolfgang Brachinger, Walter Häußler, Heinz Kaufmann, Peter Kemény, Christian Kredler, Willi Nagl, Friedemann Ost, Heinz Pape.)* 2., überarb. Aufl. Berlin: de Gruyter. xvi, 902 S., 1996.
- [Fra55] Francesco Faà di Bruno. Sullo sviluppo delle funzioni. *Annali di Scienze Matematiche e Fisiche*, 6:479–480, 1855.
- [Geo09] Hans-Otto Georgii. *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Berlin: de Gruyter, 4th revised and extended ed. edition, 2009.

Bibliography

- [Geo11] Hans-Otto Georgii. *Gibbs measures and phase transitions. 2nd extended ed.* De Gruyter Studies in Mathematics 9. Berlin: de Gruyter. xix, 545 p. EUR 119.95; \$ 180.00/hbk , 2011.
- [Hac07] Ian Hacking. *The emergence of probability. A philosophical study of early ideas about probability, induction and statistical inference. 2nd ed.* Cambridge: Cambridge University Press. 209 p., 2007.
- [Hal74] Paul R. Halmos. *Measure theory. 2nd printing.* Graduate Texts in Mathematics. 18. New York - Heidelberg - Berlin: Springer-Verlag. XI, 304 p. , 1974.
- [JJ04] Konrad Jacobs and Dieter Jungnickel. *Introduction to combinatorics. (Einführung in die Kombinatorik.) 2., völlig neu bearbeitete und erweiterte Aufl.* de Gruyter Lehrbuch. Berlin: Walter de Gruyter. xii, 406 S. , 2004.
- [Kle06] Achim Klenke. *Probability theory. (Wahrscheinlichkeitstheorie.).* Berlin: Springer. xii, 602 p. , 2006.
- [Kol33] Andreij N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitstheorie.* Springer Berlin, 1933.
- [KRY09] Joseph P. S. Kung, Gian-Carlo Rota, and Catherine H. Yan. *Combinatorics. The Rota way.* Cambridge Mathematical Library. Cambridge: Cambridge University Press. xii, 396 p. , 2009.
- [Mar06] A.A. Markov. Extension of the law of large numbers to quantities, depending on each other (1906). Reprint. 2006.
- [Pau02] Djura Paunić. History of measure theory. *Handbook of Measure Theory*, 2002.
- [Pri13] Nicolas Privault. *Understanding Markov chains. Examples and applications.* Springer Undergraduate Mathematics Series. Singapore: Springer. ix, 354 p. EUR 37.44; \$ 49.99/net , 2013.
- [Rob07] Christian P. Robert. *The Bayesian choice. From decision-theoretic foundations to computational implementation. 2nd ed., 1st paperback ed.* Springer Texts in Statistics. New York, NY: Springer. xxv, 602 p., 2007.
- [Sch09] Klaus D. Schmidt. *Actuarial mathematics. (Versicherungsmathematik.) 3rd revised and extended ed.* Springer-Lehrbuch. Berlin: Springer. xi, 366 p. EUR 34.95 , 2009.
- [vM72] Richard von Mises. *Wahrscheinlichkeit, Statistik und Wahrheit. 4. Aufl., durchgesehen von Hilda Geiringer.* Library of Exact Philosophy. 7. Wien-New York: Springer-Verlag. XX,293 S. , 1972.